SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

A Multidomain Deep Learning Approach for Accurate Detection and Localization of Copy-Move Forgeries using CLMD-BFA

Anita Mudgal[1*], Vijander Singh[2], Asha Rani[3]

[1]Instrumentation and Control Engineering Department Netaji Subhas University of Technology, Dwarka, New Delhi.
E-mail: anita.mudagal.ic20@nsut.ac.in
[2]Instrumentation and Control Engineering Department Netaji Subhas University of Technology, Dwarka, New Delhi.
E-mail: vijaydee@nsut.ac.in
[3]Instrumentation and Control Engineering Department Netaji Subhas University of Technology, Dwarka, New Delhi.
E-mail: asha.rani@nsut.ac.in
*Corresponding Author: Anita Mudgal, E-mail: anita.mudagal.ic20@nsut.ac.in

Abstract
The problem of image forgery is defined in terms of detecting, classifying, and localizing manipulations. Existing forgery localization techniques have been found to be limited either in terms of particular types of forgery they can handle or the need for increased computational cost when dealing with multiple manipulations. In order to solve such problems, this paper presents an efficient cascaded learning-based approach to multimodal image forgery localization through differential bio-inspired feature analysis (CLMD-BFA). In the proposed approach, the initial segmentation of forged/real image regions is carried out using the U-Net architecture and subsequent classification of forgery types is done through cascaded learning. The multi-modal features like frequency, entropy, Gabor, wavelet and convolutional features are obtained and optimized through Grey Wolf Optimizer (GWO). As compared to the latest techniques in image forgery detection and localization such as CW-HPF, DCNN, MSTA-Net, and transformer-based architectures, the proposed CLMD-BFA framework offers a more advanced multimodal feature representation and increased localization ability while keeping computational efficiency intact for varied forgery types. The optimized common spatial and temporal patterns (CSTPs) are used to train a 1D-CNN for patch-level forgery localization. Experimental results on several datasets with multimodal image forgery show the effectiveness of the proposed technique to detect and localize forgery accurately. Through comparative analysis, it is observed that the proposed approach performs better than traditional hand-crafted feature based techniques and modern deep learning based approaches.

Keywords: Image Forgery Detection; Image Forgery Localization; Digital Image Forensics; U-Net Segmentation; Grey Wolf Optimizer; Cascaded CNN; Bio-inspired Feature Analysis; Deep Learning.

1. Introduction

Digital images are widely used in social media, news reporting, surveillance systems, medical applications, and forensic investigations. The increasing availability of advanced image editing software and artificial intelligence-based content generation techniques has made image manipulation easier and more accessible. Consequently, forged images can be created with high visual quality and may be difficult to distinguish from authentic images. Such manipulations can be used to spread misinformation, conceal evidence, influence public opinion, and compromise the credibility of digital content [1], [2].

Image forgery refers to the deliberate modification of image content to alter its original meaning or context. Some of the common techniques used for creating image forgery are copy-move forgery, image splicing, object removal, and image inpainting. The use of modern technology makes these techniques very hard to be detected since it is possible to achieve almost undetectable manipulation. Hence, the development of forgery detection techniques is an area of active research [3]–[6].

The existing approaches to forgery detection focus on handcrafting the forensic features of the images using artifacts that occur during the process of acquisition and image compression. Methods based on the JPEG compression artifacts, demosaicing, prediction residuals, and illumination have been commonly applied for detecting regions of the manipulated images [9], [17], [24], [40]. However, such approaches suffer when dealing with images that were subjected to several processing operations. However, with the recent advances in deep

learning techniques, image forgery detection algorithms have achieved better results. The CNN is capable of automatically extracting discriminative feature representation for images and has shown promising results in copy-move forgery detection, image splicing detection, seam carving forgery detection, and forgery localization [7], [8], [21], [27]. Also, attention-based models like AR-Net and MSTA-Net enhance the localization performance by utilizing manipulation-related features and multi-scale information [11], [37].
On the other hand, transformer-based models have received much attention lately in image forensics owing to the fact that they are capable of modeling long-term dependencies and global information within the images. Algorithms like Edge-Enhanced Transformer, Forgery-Aware Adaptive Transformer, Fusion Transformer, Vision Transformer-based localization models, Detective SAM, and Forensics SAM have shown promising results in forgery detection and localization tasks [49]–[57]. However, they normally rely on large datasets and computing power.

1.1 Research Gap

Despite the successes that the existing image forgery detection algorithms have made, there still exist some challenges in this domain. Traditional handcrafted forgery detection techniques are usually devised for detecting specific manipulations and can be less successful when dealing with diverse forgery tasks [9], [17], [24], [40]. The deep learning-based algorithms enhance forgery detection due to their ability to learn features automatically, but the majority of CNN-based frameworks use only one feature domain and are specialized for certain forgery categories [7], [8], [21], [27]. Likewise, transformer-based techniques ensure the better localization results, but at the same time, they demand large amounts of training data and high computational costs [49]–[57].
Moreover, it is important to highlight that little attention was paid to the combined usage of multidomain feature representations and feature optimization algorithms in a single forgery detection system. Thus, the creation of such a system becomes especially relevant since it is required to develop an efficient multimodal approach that would utilize additional information from multiple domains (spatial, frequency, texture, deep).
To address this gap, this paper proposes a Cascaded Learning-Based Multimodal Image Forgery Localization through Differential Bio-Inspired Feature Analysis (CLMD-BFA) framework that integrates segmentation, multidomain feature extraction, feature optimization, forgery classification, and localization within a unified architecture.

1.2 Contributions

The major contributions of this work are summarized as follows:
1. An approach for combined detection, classification and localization of forged images using a cascaded multimodal structure referred to as CLMD-BFA is proposed.
2. The extraction of features from multiple domains is carried out in this research work through the combination of spatial, frequency, entropy, Gabor, wavelet and convolutional feature descriptors.
3. The Grey Wolf Optimizer (GWO) technique is used to extract optimal Common Spatial and Temporal Patterns (CSTPs) to increase the discrimination of features and eliminate redundancy among features.
4. A patch localization scheme that uses the 1D-CNN network is proposed in this work for accurate localization of forged image patches.
5. The performance of the proposed system is evaluated experimentally and compared with other existing schemes such as CW-HPF, DCNN and MSTA.

1.3 Related Work

Image forgery detection has been gaining lots of attention among researchers because of the rapid increase in sophistication of image editing software. The earlier methods used to detect image forgeries were mainly dependent on forensics features that were manually crafted through image acquisition and compression processes. Many forensic methods that were based on the detection of JPEG artifacts, demosaicing, prediction residue, and illumination inconsistencies have been extensively used for detecting the manipulated parts of images [9], [17], [24], [40]. While these methods are valuable in terms of forensics evidence, they are not highly robust to more complex image manipulations.
Deep learning has brought a significant improvement in image forgery analysis. Marra et al. have developed a deep end-to-end CNN-based approach to detect image forgery [7] whereas Hosny et al. have developed a deep CNN-based model for detecting copy-move forgery in images [8]. Other methods have also used deep CNN models for detecting splicing, seam carving forgery, and localization of forgeries [21], [27].

In an effort to boost the capability of feature discrimination and localization, many researchers have tried attention-based learning algorithms. For instance, AR-Net has used adaptive attention and residual refinement techniques to boost forgery detection of the copy-move method [11]. In the same way, MSTA-Net applies multi-scale self-texture attention to detect manipulation traces and improves localization [37]. Despite promising performances, the efficiency of these frameworks largely relies on complicated model architecture.

However, transformer-based methods can be effective tools to detect and localize image forgery thanks to global context exploitation and long-distance relations. Examples include Edge-Enhanced Transformers [49], Forgery-Aware Adaptive Transformers [50], Fusion Transformers with Object-Mask Guidance [51], Vision Transformer-Based localization framework [52], Detective SAM [56], and ForensicsSAM [57]. Though they give high-performing forgery detection and localization, their implementation is costly since they need much training data and computational costs.

Nevertheless, despite many studies made about the subject, most existing works concentrate only on hand-crafted features or convolution-based feature learning or transformer-based representation individually. The combination of multiple-domain features, bio-inspired features optimization, and cascade learning in unified forgery detection, classification, and localization still lacks consideration. The aforementioned issue has motivated the proposed framework, CLMD-BFA.

The rest of this paper is organized as follows. Methodology and model architecture are introduced in section 2. In section 3, experimental results and comparative performance analysis are described. Section 4 concludes the paper.

2. Proposed methodology

The existing image forgery analysis methods either support a limited number of manipulation categories or exhibit increased computational complexity when extended to multiple forgery types. In addition, the majority of these models cannot localize multiple forgery types, limiting their scalability. This section discusses the design of an efficient cascaded-learning model for the localization of multimodal image forgeries using differential bioinspired feature analysis in order to address these issues. Figure 1 illustrates the overall architecture of the proposed CLMD-BFA framework, which shows that the model initially collects authentic and forged images for training a U-Net Model, which aids in the segmentation of authentic and forged images. The segmented images are processed by a collection of Binary Cascaded Convolutional Neural Networks (BCCNNs), which aid in the identification of multiple types of forgery. The considered forgery categories include copy-move, splicing, and inpainting manipulations. The model can be extended to other forgeries by gradually adding BCCNNs with similar configurations. After identifying forgeries, a module for patch-level feature extraction is used to estimate multidomain feature sets. These features include frequency-domain, entropy-based, convolutional, Gabor-based spatial, and wavelet-domain descriptors.

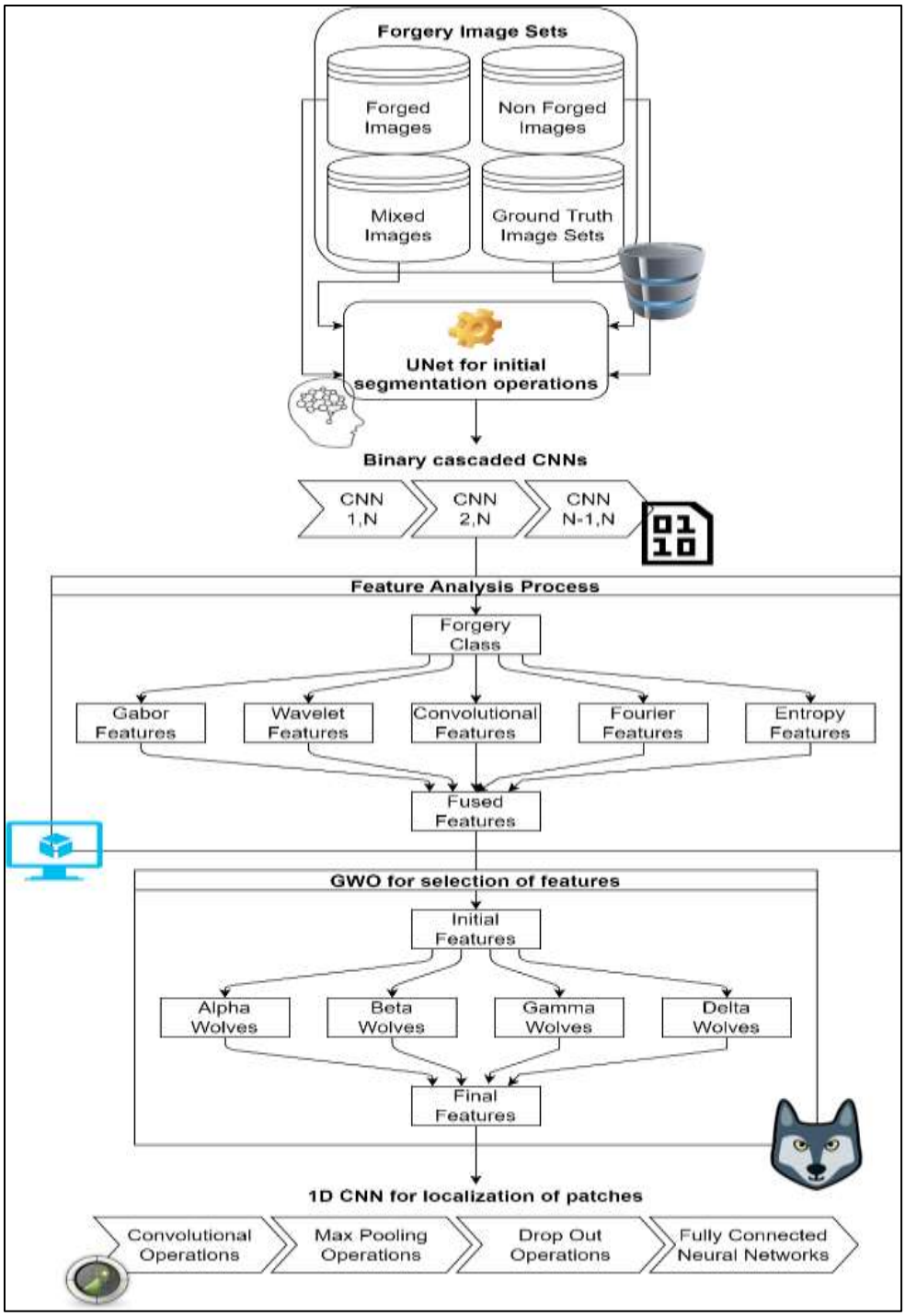

Figure 1. Design of the proposed U-Net based model for segmented classification & localization of different forgery types

The extracted patterns are processed by a Grey Wolf Optimizer (GWO), which identifies class-level feature sets with high variance. These feature sets are differentially correlated across forgeries in order to identify common spatial and temporal patterns (CSTPs). The CSTPs are employed to train a 1D CNN that aids in patch-level marking for various types of forgeries.

As illustrated in figure 1, the input image is initially segmented using a U-Net architecture to extract relevant Regions of Interest (RoIs). The model initially segments the input image, and extracts different Regions of Interest (RoIs) via U-Net based segmentation process. The flow of this process is depicted in figure 2, where

different Max Pooling, Up Sampling, Convolution, Rectilinear Unit (ReLU), and Cropping layers are used for identification of high-entropy regions.

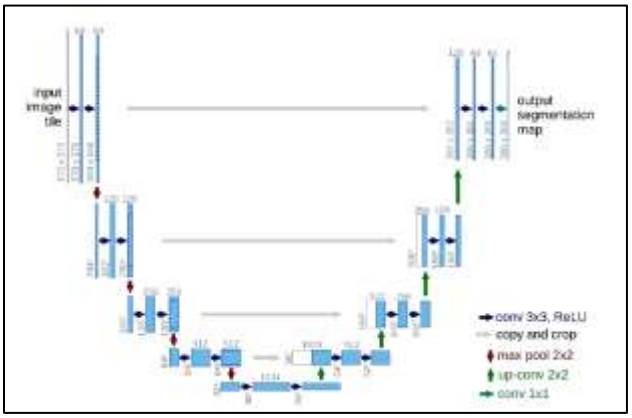

Figure 2. Design of the U-Net model for initial segmentation operations

The proposed model initially extracts multiple bit-planes from the input image using equation (1).

$$S_i = \bigcup_{r,c}^{N,M} \left(P_{r,c} \bigoplus 2^i \right) \quad \ldots (1)$$

Where, N & M are dimensions of the input image, while P represents pixel intensity level. The value of i is modified between (1, 64) to get augmented image samples. Each of these samples are evaluated for RGB images, and converted into intensity (Y), chrominance (Cr), and luminance (Cb) levels using equations (2), (3) and (4) as follows:

$$Y = 16 + (65.481 * R + 128.553 * G + 24.966 * B) \quad \ldots (2)$$

$$C_b = 128 + (-37.797 * R - 74.203 * G + 112 * B) \quad \ldots (3)$$

$$C_r = 128 + (112 * R - 93.786 * G - 18.214 * B) \quad \ldots (4)$$

The extracted intensity values are fused using equation (5) to determine the average pixel intensity of the corresponding image slice.

$$I_{avg_i} = \frac{1}{N * M} * \sum_{r,c}^{N,M} C_i \quad \ldots (5)$$

Where, $C \in (Y, C_b, C_r)$, and average level is evaluated for all colour levels. Based on this evaluation, the U-Net model extracts distance between different pixels using equation 6,

$$d(s_i, s_j) = \frac{1}{N * M} * \sum_{i=1}^{N_S} I_{avg_i} * \sqrt{\sum_{r,c}^{N,M} \frac{(P_{i_r} - P_{j_r})^2 + (P_{i_c} - P_{j_c})^2}{Var(s_i, s_j)}} \quad \ldots (6)$$

Where, s_i & s_j are different slices that are obtained from equation 1, while P represents their pixel levels, and $var(s_i, s_j)$ is the variance between these pixels, which is estimated using equation (7),

$$var(x, y) = \frac{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}}{N} \quad \ldots (7)$$

Where, $\bar{x}$ represents average of all elements in the list, while N is the number of elements present in the list that is being evaluated by this process. These distance metrics are fused using equation (8) in order to estimate U-Net entropy levels,

$$E_f = \frac{1}{N * M} \sum_{i=1}^{N} \sum_{j=1}^{M} d(i,j) * \log\big(d(i,j)\big) \quad \ldots (8)$$

This process is repeated for each U-Net layer, and a series of E_f levels are generated, which assist in iterative segmentations. Pixels with $P > E_f$ are passed to the next layer, while others are removed from the image, which assists in incremental segmentation operations. The results of this initial segmentation process can be observed from figure 3(a), 3(b), and 3(c), where different input images are segmented to identify respective RoIs.

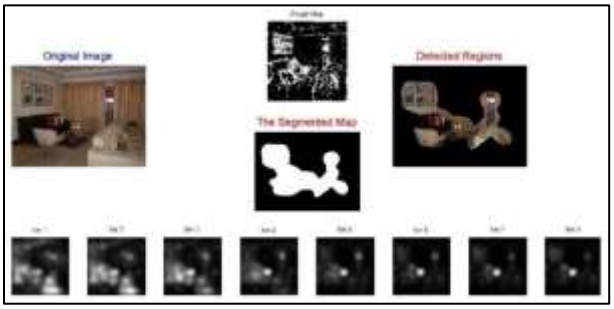

Figure 3 (a). RoIs for complex backgrounds

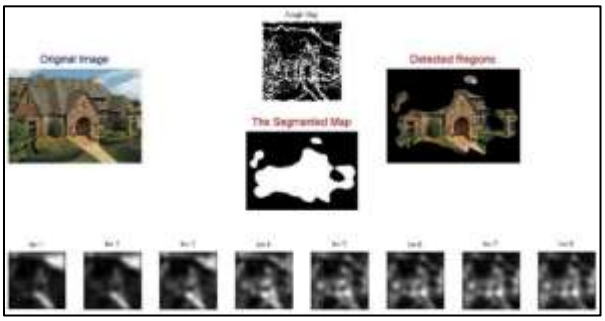

Figure 3 (b). RoIs for simple backgrounds

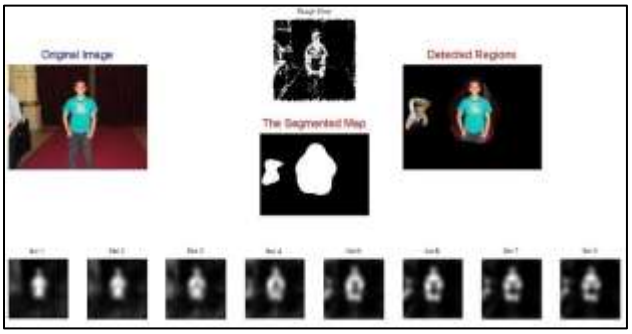

Figure 3 (c). RoIs for single entity analysis

The segmentation process is applied to both authentic and forged images. The segmented images are processed via a Binary Cascaded CNN (BCCNN), which assists in identification of different forgery types. Design of this BCCNN can be observed from figure 4, where multiple Convolution, Max Pooling, & Drop Out layers are combined for extraction of augmented feature sets. These feature sets are categorized into binary classes via the Fully Connected Neural Network (FCNN) layer, that performs binary classifications.

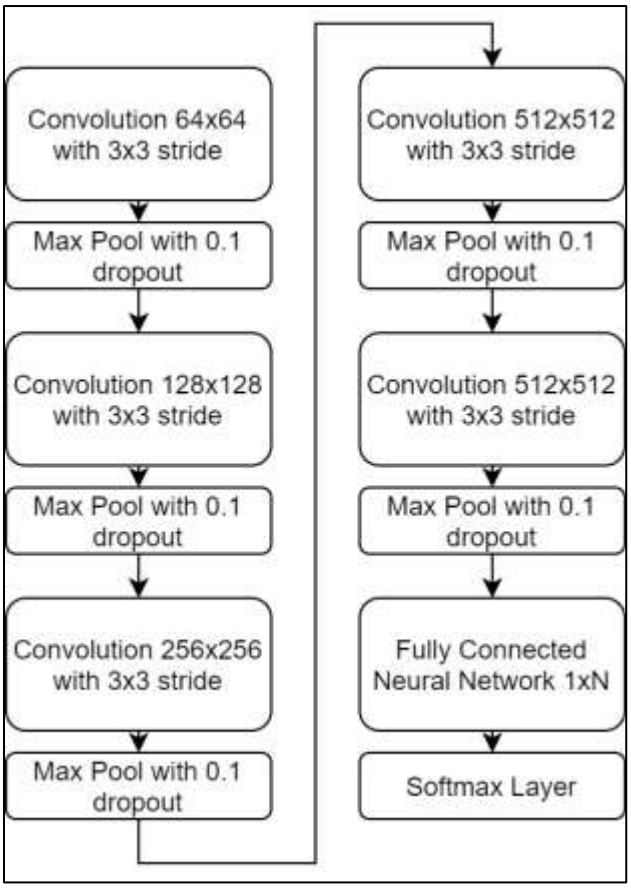

Figure 4. Design of a single binary cascaded CNN for classification of images into 'authentic', and 'forged' categories

The model initially extracts 2D convolutional features from these images via equation 9,

$$Conv_{s_{i,j}} = \sum_{a=-\frac{m}{2}}^{\frac{m}{2}} \sum_{b=-\frac{n}{2}}^{\frac{n}{2}} I_s(i-a, j-b) * ReLU\left(\frac{m}{2}+a, \frac{n}{2}+b\right) \quad \ldots (9)$$

Where, n, m represents 2D window sizes, I_s represents the segmented image, and a, b represents different stride sizes. These convolutional features are extracted by shifting the feature window by stride dimensions, which allows the model to extract high-density feature sets. Due to the high-density of these features, redundancies are prevalent, which are removed by a set of Max Pooling operations. These operations include,

- Calculation of a Max Pooling threshold using equation (10),

$$f_{th} = \left(\frac{1}{X_k} * \sum_{x \in X_k} x^{p_k}\right)^{\frac{1}{p_k}} \quad \ldots (10)$$

Where, x & p are the feature values, and their variance probability levels.

- Features with $x > f_{th}$ are passed to next iteration, while others are removed from current iteration due to low variance levels.

This process is repeated for each set of Convolutional & Max Pooling layers. The drop out layers remove all features that have less than 10% variance, which assists in identification of highly variant feature sets. These feature sets are categorized into 'authentic' or 'forged' classes via a SoftMax based Fully Connected Neural Network (FCNN) that uses equation (11.1) for fusion of extracted features with their weights & biases.

$$c_{out} = SoftMax\left(\sum_{i=1}^{N_f} f_i * w_i + b_i\right) \quad \ldots (11.1)$$

Where, N_f are total number of extracted features from the Convolution, Max Pooling & Drop Out layers. The weights & biases are auto-tuned by the CNN model, in order to obtain higher accuracy for the binary classification process. For this purpose, the images are categorized into 'copy move', 'inpainting', 'splicing', and 'authentic' classes. The classification process is repeated for each set of binary classes (keeping 'authentic' class as constant), in order to identify the final forgery class using equation (11.2),

$$c_{final} = Normal, if\ converge$$
$$else, \bigvee_{i=1}^{N_d} C_i \quad \ldots (11.2)$$

Where, N_d represents total number of forgery classes. The 'forgery' class images are further processed via multidomain feature analysis, wherein a window size of 5x5 is used to obtain a set of 25 pixels. A series of such pixel-sets are extracted for each image, and represented via multidomain analysis. This analysis includes the following operations,

- Extraction of frequency features using Fourier transform from equation (12),

$$DFT_i = \sum_{j=1}^{N_f} x_j * \left[cos\left(\frac{2 * \pi * i * j}{N_f}\right) - \sqrt{-1} * sin\left(\frac{2 * \pi * i * j}{N_f}\right)\right] \quad \ldots (12)$$

Where, x is the set of pixels, and N_f are total number of feature points ($N_f = 25$), that can be tuned for different window sizes.

- Entropy components are estimated using the Discrete Cosine Transform (DCT), as defined in equation (13),

$$DCT_i = \frac{1}{\sqrt{2 * N_f}} * x_i \sum_{j=1}^{N_f} x_j * \cos\left[\frac{\sqrt{-1} * (2 * i + 1) * \pi}{2 * N_f}\right] \quad \ldots (13)$$

- Window based 1D Convolutional components are extracted using equation (14),

$$Conv_{out_i} = \sum_{a=-\frac{m}{2}}^{\frac{m}{2}} x(i-a) * LReLU\left(\frac{m + 2a}{2}\right) \quad \ldots (14)$$

Where, $LReLU$ is a leaky Rectilinear Unit that is used for quantization of features using equation (15),

$$LReLU(x) = l_a * x, when\ x < 0\ , else\ LReLU(x) = x \quad \ldots (15)$$

Where, l_a is a constant used for quantization of negative features.

- Spatial features are extracted using Gabor analysis using equation (16),

$$G(x,y)_s = e^{\frac{-x`^2+\partial^2*y'^2}{2*\emptyset^2}} * \cos\left(2*\frac{pi}{\lambda}*x'\right) \quad \ldots(16)$$

Where, x & y are sample number, and pixel levels, while ∂ & ϕ are angular constants and ranges between $(0, 2\pi)$ for augmentation of pixel sets. The values of x & y are further modified using equations (17) & (18) as follows,

$$x' = x * \cos(\phi) + y * \sin(\phi) \quad \ldots(17)$$
$$y' = -x * \sin(\phi) + y * \cos(\phi) \quad \ldots(18)$$

- Approximate & detailed wavelet features are extracted using equations (19) & (20) as follows,

$$W_a = \frac{x_i + x_{i+1}}{2} \quad \ldots(19)$$

$$W_d = \frac{x_i - x_{i+1}}{2} \quad \ldots(20)$$

These feature sets are fused for each set of windows to form a Forgery Fusion Vector (FFV) that contains inherent class-level feature redundancies. These redundancies are reduced via use of a Grey Wolf Optimization model that assists in increasing feature-level variance via the following process

- The feature selection GWO process starts by setting up the following constants,
  - A constant set of Wolves that will be used for reconfiguration process (N_w)
  - A constant set of Iterations that will reconfigure the Wolves (N_i)
  - A Wolf-level learning rate (L_w), that assists each Wolf to learn from other Wolf configurations
- The optimization process starts via generation of N_w Wolf configurations as follows,
  - Select N features from the set of multidomain features via equation (21),

$$N = STOCH(L_w * N(FFV), N(FFV)) \quad \ldots(21)$$

Where, $STOCH$ represents a Stochastic Markovian process for generation of number sets, while $N(FFV)$ are the total number of features present in the FFV sets.

  - As per this selection, a Wolf fitness is evaluated by equation (22),

$$f_w = \sqrt{\frac{\left(\sum_{i=1}^{N}\left(x_i - \sum_{j=1}^{N}\frac{x_j}{N}\right)^2\right)}{N+1}} \quad \ldots(22)$$

  - Using the same operations, generate N_w Wolf configurations.
- Once a set of N_w Wolves are generated, calculate fitness threshold using equation (23),

$$f_{th} = \sum_{i=1}^{N_w} f_{w_i} * \frac{L_{w_i}}{N_w} \quad \ldots(23)$$

- Based on this threshold, mark each of the Wolves as follows,

Mark Wolf as $'Alpha'$ if $f \geq 2 * f_{th}$ ... (24)

Mark Wolf as $'Beta'$ if $f \geq f_{th}$ ... (25)

Mark Wolf as $'Gamma'$ if $f \geq \frac{f_{th}}{2}$ ... (26)

Else, Mark Wolf as $'Delta'$

- For all 'Beta' Wolves, stochastically replace its features using equation (27),

$$F(Beta) = STOCH\big(F(Alpha)\big) \quad \text{... (27)}$$

- Repeat the same operation for 'Gamma' Wolves, and 'Delta' Wolves, by using features from 'Beta' Wolves and 'Gamma' Wolves respectively, which assists in iterative learning operations.

- This process is repeated for N_i set of iterations, and Wolf configurations are updated to optimize the feature selection process.

Once all iterations are completed, then the final features are selected via equation (28),

$$F(Final) = \bigcup_{i=1}^{f_w \geq 2 * f_{th}} F_i \quad \text{... (28)}$$

These features are used to train a 1D CNN Model that assists in localization of forged patches. Design of this 1D CNN Model is depicted in figure 5 as follows,

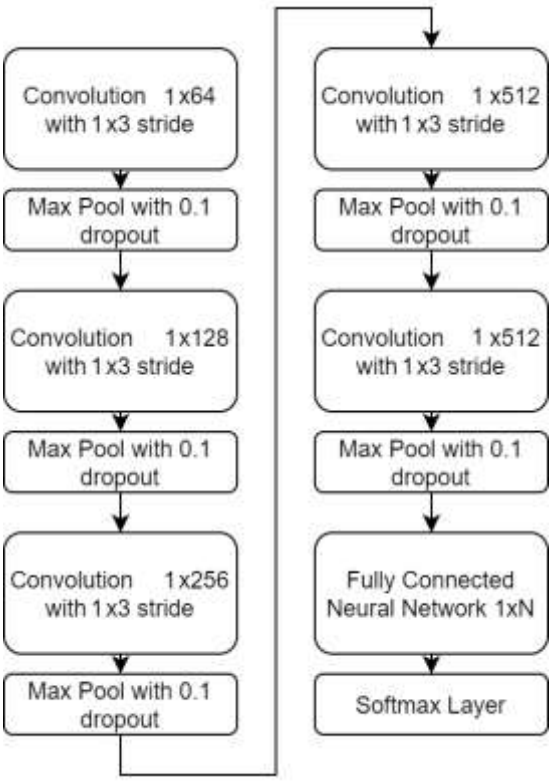

Figure 5. Design of the 1D CNN Model for localization of forged regions

The model uses a combination of 1D Convolutional Layers with 1D Max Pooling & Dropout layers for further augmentation of multidomain features. These features are classified into 'forged' & 'non-forged' regions via a Purely Linear classifier as per equation (29) as follows,

$$c_{out} = PureLin\left(\sum_{i=1}^{N_f} f_i * w_i + b\right) \quad \ldots (29)$$

Where, N_f are number of features selected by the GWO process, while w & b are their respective weights and biases. This model is trained for different forgery types, and its accuracy, precision, recall, and inference time are evaluated in the next section.

3. Results and Discussion

3.1. Dataset Construction and Augmentation

The dataset used in this study is constructed by integrating three benchmark image-forensics repositories: the Forgery Image Data Samples [46], the IEEE IFS-TC Image Forensics Challenge Corpus[47], and the CoMoFoD[48] dataset. These databases are chosen since they contain authentic photos as well as different manipulations, namely, copy-move, splicing, and inpainting/object removal forgeries.

In order to provide efficient learning of the suggested cascaded CNN and 1D CNN localization models, two problems – data unavailability and class imbalance – are tackled through careful use of augmentation methods. Apart from simple scaling, normalization, and segmentation with the help of U-Net, the data is augmented through patch extraction, bit-plane manipulation, and ROI-based sampling. As a result, a balanced dataset consisting of 80,000 images, equally divided into four classes of 20,000 images each, is obtained. Table 1 outlines the distribution of the final dataset across its original sources and the experimental splits (75% training, 15% validation, and 10% testing).

Table1: Dataset details (training, validation and testing)

Original dataset	Total images	Training Set (75%)	Validation Set (15%)	Testing Set (10%)
IEEE IFS-TC	20,000	15,000	3,000	2,000
Forgery Image Data Samples	30,000	22,500	4,500	3,000
CoMoFoD	30,000	22,500	4,500	3,000
Total images	80,000	60,000	12,000	8,000

Proportional partitioning of the data set is meticulously followed at every step in order to avoid data set bias and to provide equal representation of all forgery types during model evaluation. Incorporation of these diverse data sets enables the learning model to detect global and localized manipulation artefacts. This is because the IEEE IFS-TC corpus offers challenging forensic test cases. The Forgery Image Data Samples add diversity to the forgery categories; and CoMoFoD increases the representation of copy move and geometrically transformed forgeries.

Finally, the large volume of the augmented data set containing 80,000 images offers the necessary statistical basis in order to maintain the learning consistency through the different aspects of the proposed architecture, including segmentation, binary cascaded classification, multi-domain feature extraction, GWO-based feature selection, and patch level localization. This balanced setup enables the evaluation of the model performance in terms of its detection accuracy, localization abilities, precision, and recall.

(a). Results of forgery detection process. (b). Copy Move Forgery Detection

(c). Inpainting Forgery Detection. (d). Splicing Forgery Detection

Figure 6. Qualitative results of the proposed framework.

3.2. Evaluation metrics

The performance of the proposed CLMD-BFA model was evaluated and compared with the state-of-the-art methods, namely CW-HPF [3], DCNN [28], and MSTA [37]. The comparative analysis was conducted using four widely adopted evaluation metrics: forgery detection accuracy (A), forgery localization accuracy (AL), localization precision (P), and localization recall (R). Detection accuracy measures the overall capability of the model to correctly classify forged and authentic images, while localization accuracy evaluates the effectiveness of identifying manipulated regions. Precision and recall are further employed to assess the quality and completeness of forgery localization. The accuracy, localization accuracy, precision and recall metrics are computed using equations (30)-(33), respectively.

$$A = \frac{1}{N_t}\sum_{i=1}^{N_t}\frac{t_{p_i} + t_{n_i}}{t_{p_i} + t_{n_i} + f_{p_i} + f_{n_i}} \quad \ldots(30)$$

$$AL = \frac{1}{N_t}\sum_{i=1}^{N_t}\frac{R_{d_i}}{TR_i} \quad \ldots(31)$$

$$P = \frac{1}{N_t}\sum_{i=1}^{N_t}\frac{t_{p_i}}{t_{p_i} + t_{n_i}} \quad \ldots(32)$$

$$R = \frac{1}{N_t}\sum_{i=1}^{N_t}\frac{t_{p_i} + f_{n_i}}{t_{p_i} + t_{n_i} + f_{p_i} + f_{n_i}} \quad \ldots(33)$$

where N_t denotes the total number of test images, t_{p_i}, t_{n_i}, f_{p_i}, and f_{n_i} represent the true positive, true negative, false positive, and false negative outcomes respectively. Furthermore, R_{d_i} denotes the number of regions correctly detected as forged, while TR_i represents the total number of forged regions present in the corresponding test image.

3.3. Experimental setup

The experimentation is conducted using Python 3.9 and TensorFlow/Keras on a workstation which had Intel Core i7 CPU, 16 GB RAM and NVIDIA RTX 3060 GPU (12 GB VRAM). The time taken to infer is calculated based on the entire test data set.

3.4. Performance analysis of the proposed methodology

The forgery detection performance of the proposed model and the rival techniques are evaluated with varying sizes of testing samples. Comparative results of forgery detection accuracy are shown in table 2, while the graph representing the same in figure 7.

Table 2. Accuracy for detection of forgeries under different sample sizes

N	CW-HPF [3]	DCNN [28]	MSTA [37]	Proposed (CLMD-BFA)
10k	82.5	87.95	90	96.78
20k	82.62	88.82	90.58	96.89
40k	82.9	90.5	91.65	97.18
60k	83.1	91.65	92.38	97.36
80k	83.26	92.64	93.01	97.52

Table 2 shows the comparative forgery detection accuracy of the proposed CLMD-BFA approach and existing approaches, such as CW-HPF [3], DCNN [28], and MSTA [37], for various testing sample sizes. It is worth noting that all approaches demonstrate a gradual increase in accuracy with an increase in sample size; however, the proposed approach demonstrates better and consistent results as compared to the other techniques. The forgery detection accuracy of the proposed model is increased from 96.78% at 10k samples to 97.52% at 80k samples, which means stable learning and high level of generalization ability. Meanwhile, the maximum results of CW-HPF, DCNN, and MSTA are 83.26%, 92.64%, and 93.01%, correspondingly. Improved results of the proposed approach are based on the use of the cascade learning strategy, which involves the combination of U-Net-based segmentation, binary CNN classification, multi-domain feature extraction, and GWO-based feature optimization. As a result, the proposed approach is able to detect several types of forgeries with high accuracy.

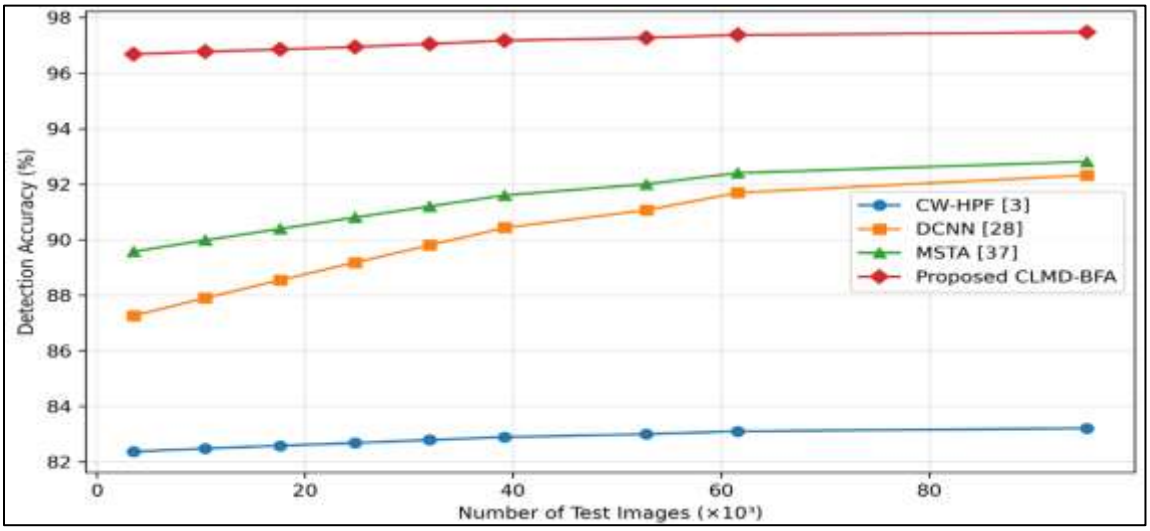

Figure 7. Accuracy for detection of forgeries under different sample sizes

With the largest test sample, the developed CLMD-BFA model obtains better performance than CW-HPF, DCNN, and MSTA by about 14.26%, 4.88%, and 4.51%, respectively. From the above analysis, it is evident that the proposed CLMD-BFA framework is capable of dealing with such tasks due to its robustness and scalability.

Table 3. Accuracy for localization of forgeries under different sample sizes

N	CW-HPF [3]	DCNN [28]	MSTA [37]	Proposed (CLMD-BFA)
10k	80.54	84.18	87.01	95.63
20k	80.7	85.08	87.61	95.77
40k	80.95	86.57	88.56	96.04
60k	81.15	87.77	89.34	96.23
80k	81.3	88.67	89.92	96.38

Table 3 presents the forgery localization accuracy achieved by the proposed CLMD-BFA model and the benchmark methods for different test sample sizes. It is observed from the results obtained that localization accuracy increases gradually with the increase in the number of test samples for all methods. However, the proposed CLMD-BFA model consistently outperforms the competing approaches across all experimental settings. At 80k samples, the proposed model achieves a localization accuracy of **96.38%**, compared with **89.92%**, **88.67%**, and **81.30%** obtained by MSTA, DCNN, and CW-HPF, respectively. The superior performance can be attributed to the integration of multi-domain feature extraction, GWO-based feature selection, and cascaded learning, which collectively enhance the discrimination and localization of forged regions. It is observed from evaluation analysis and figure 8 that the proposed model is able to achieve 12.4% higher localization accuracy than CW HPF [3], 8.3% higher localization accuracy than DCNN [28], and 6.5% higher localization accuracy than MSTA [37] for different forgeries. This is because the model uses multidomain feature extraction that fuses frequency, entropy, Gabor and Wavelets with convolutional features. Due to fusion and use of simplified 1D CNN process, the model is applicable for a wide variety of real-time forgery localization scenarios.

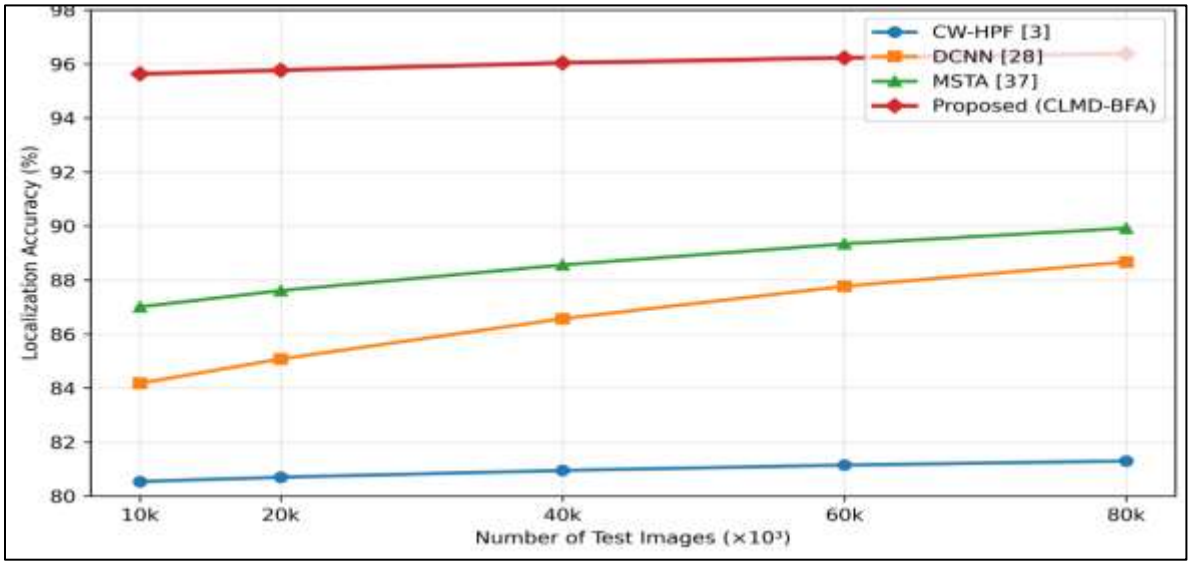

Figure 8. Accuracy for localization of forgeries under different sample sizes

Table 4 shows that the precision of all methods increases with the number of test samples. However, the proposed CLMD-BFA framework consistently achieves the highest precision across all experimental sizes. At 80k test images, the proposed model attains a localization precision of 96.05%, compared with 90.21%, 89.87%, and 80.75% for MSTA, DCNN, and CW-HPF, respectively.

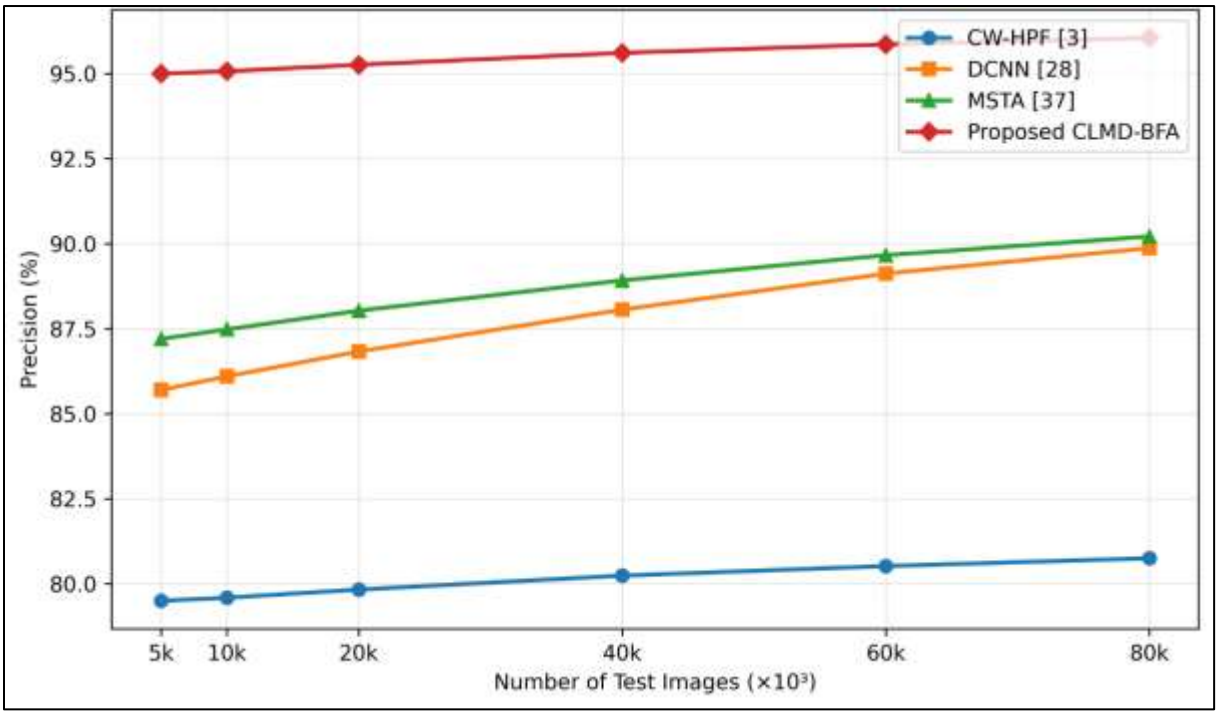

Figure 9. Precision for localization of forgeries under different sample sizes

The improvement is attributed to the integration of multidomain feature extraction, Grey Wolf Optimizer-based feature selection, and cascaded learning, which collectively enhance the discriminative capability of the localization framework while maintaining lower computational complexity.

Table 4. Precision for localization of forgeries under different sample sizes

N	CW-HPF [3]	DCNN [28]	MSTA [37]	Proposed CLMD-BFA
10k	79.59	86.1	87.48	95.07
20k	79.83	86.83	88.03	95.26
40k	80.24	88.06	88.92	95.61
60k	80.52	89.12	89.66	95.86
80k	80.75	89.87	90.21	96.05

The forgery localization approaches are evaluated based on recall rates and the same is shown in table 5. In all the test cases, the newly developed CLMD-BFA approach has produced the best recall rate. For instance, with 80k test images, the proposed model has produced a recall rate of 94.83%, compared to 89.52%, 88.04%, and 78.72% that were obtained by MSTA, DCNN, and CW-HPF approaches, respectively. The improved recall rate has resulted from the combination of feature extraction and GWO based feature optimization techniques.

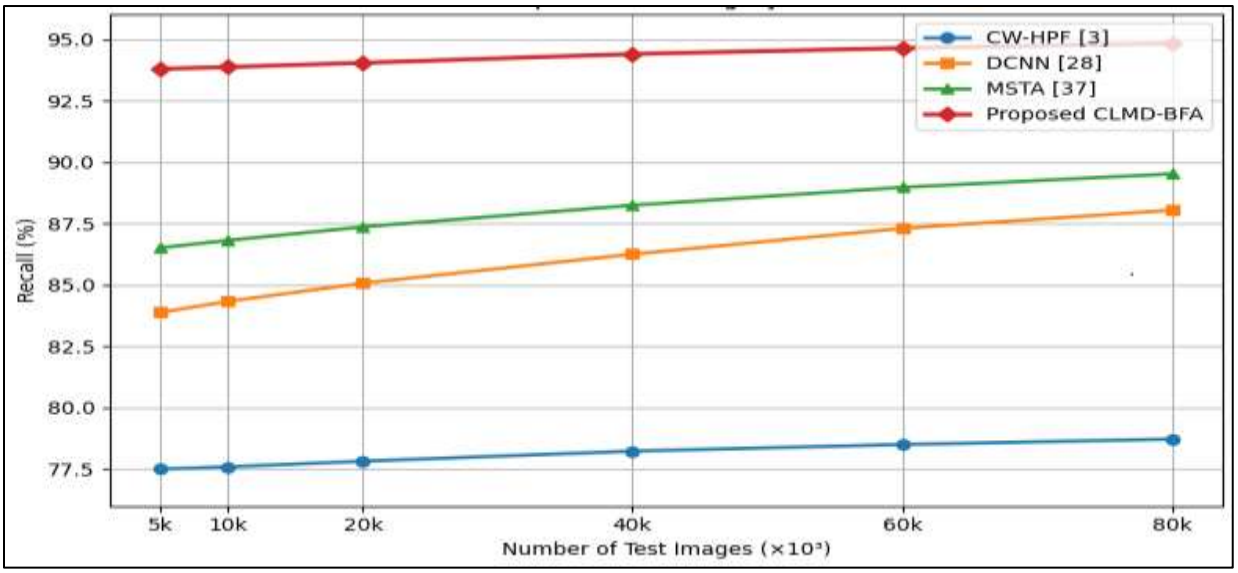

Figure 10. Recall for localization of forgeries under different sample sizes

Table 5. Recall for localization of forgeries under different sample sizes

N	CW-HPF [3] R (%)	DCNN [28] R (%)	MSTA [37] R (%)	Proposed CLMD-BFA R (%)
10k	77.59	84.33	86.81	93.87
20k	77.82	85.07	87.36	94.04
40k	78.23	86.25	88.24	94.4
60k	78.5	87.3	88.97	94.64
80k	78.72	88.04	89.52	94.83

3.5. Classification performance assessment using ROC curves

The ROC curves for CLMD-BFA algorithm along with three other classifiers are shown in figure 11. The ROC curve shows the correlation between the True Positive Rate (TPR) and the False Positive Rate (FPR) at different decision levels. The closer the ROC curve of the classifier to the top left region of the graph, the better discriminating power it has, since it detects more true positives while keeping the false positives low.

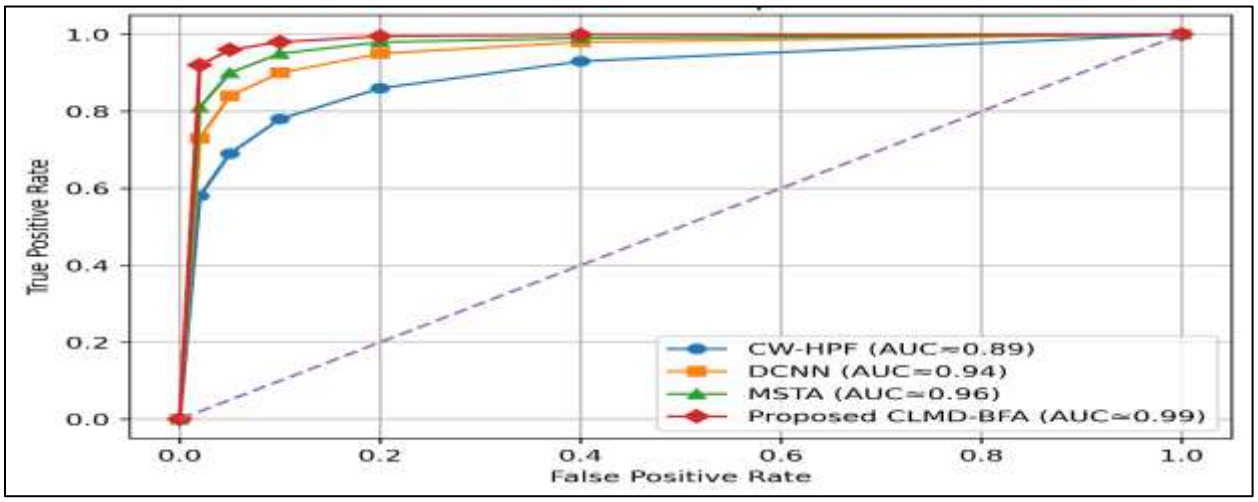

Figure 11. ROC curves and AUC comparison of CW-HPF, DCNN, MSTA, and the proposed CLMD-BFA model for image forgery detection.

Figure 11 reveals that the proposed CLMD-BFA outperforms other competing models at any given operating point. The ROC curve tends to reach the ideal coordinate (TPR=1, FPR=0), suggesting that the model is capable of detecting forged images with very little possibility of generating wrong detections. The Area under the Curve (AUC) of the proposed model is around 0.99, which is much greater than that of CW-HPF (0.89), DCNN (0.94), and MSTA (0.96).

The high value of the AUC indicates that the proposed model has higher discrimination power in detecting forged vs. genuine images. The improvement of the performance is due to the following features of the model:

- U-Net-based forgery region segmentation,
- Binary Cascaded CNN (BCCNN) classification,
- Multi-domain feature extraction (frequency, entropy, Gabor, wavelet, and convolutional features),
- Grey Wolf Optimizer (GWO)-based feature selection, and
- 1D CNN-based localization.

The overall performance of forgery detection and localization in an image is significantly enhance through the combination of these factors. Table 6 shows the Area under the Curve (AUC) values obtained through the use of the CLMD-BFA approach and the other benchmark models. The AUC value serves as a comprehensive indicator of the discrimination capability of a model by assessing its discriminative capability at various classification thresholds. It is noted from table 6 that the CLMD-BFA model has recorded the best AUC value among all the competing models, demonstrating the excellent discrimination capability of the approach in identifying the forgery images. The superior performance of the AUC value results from the use of the combination of features such as the U-Net based segmentation, cascaded CNN classification, multidomain feature extraction, feature optimization through GWO, and 1D-CNN localization. The use of these features enables the CLMD-BFA approach to identify multiple forgery features in the image while avoiding false identifications.

These components collectively enhance feature representation and reduce false classifications, thereby improving the robustness of forgery detection and localization.

Table 6. AUC performance comparison

Method	AUC
CW-HPF [3]	0.89
DCNN [28]	0.94
MSTA [37]	0.96
Proposed CLMD-BFA	**0.99**

3.6. Computational complexity

The proposed CLMD-BFA framework demonstrates a favorable trade-off between detection performance and computational efficiency. Although the model incorporates multidomain feature extraction and GWO-based feature selection, the use of a lightweight 1D CNN localization stage reduces computational overhead. Consequently, the proposed model achieves an average inference time of approximately 0.19 s per image, which is significantly lower than that of CW-HPF, DCNN, and MSTA while simultaneously providing higher detection and localization accuracy.

Table 6. Computational complexity and average inference time

Method	Computational Complexity	Average Inference Time/Image (s)
CW-HPF [3]	$O(n^2)$	0.41
DCNN [28]	$O(n^2k)$	0.34
MSTA [37]	$O(n^2k + m)$	0.27
Proposed CLMD-BFA	O(n logn + kd)	0.19

Where:

- n = number of image pixels,
- k = convolution kernel operations,
- m = attention-map computations,

- d = optimized feature dimensions selected by GWO.

3.7. Limitation

The performance of the proposed CLMD-BFA approach has been studied in the context of copy-move, splicing, and inpainting forgeries. However, there is a need to assess its performance against novel forms of forgeries using artificial intelligence techniques. Further, the current approach lacks low-complexity features including color, texture, and edge maps.

4. Conclusion

The proposed model is a fusion of U-Net, Binary Cascaded Convolutional Neural Networks (BCCNNs) and Grey Wolf Optimizer (GWO). The U-Net and BCCNNs help in segmentation of both authentic and forged images and process the segmented images to identify various forgery types respectively. Grey Wolf Optimizer (GWO) assists in identifying class-level feature sets with high variance and used to process the extracted patterns. This makes it possible to spot frequent spatial and temporal patterns (CSTPs). In contrast to CW-HPF, which is based on handcrafted forensic features, DCNN that mainly employs convoluted features, and MSTA that makes use of multi-scale self-texture attention for forgery detection, the proposed CLMD-BFA uses U-Net segmentation, multidomain features extraction, GWO-based features optimization, and 1D-CNN localization. It is noticed that the proposed framework is 15.4% more efficient than CW HPF [3], 4.9% more efficient than DCNN [28] and 3.5% more efficient than MSTA [37] in terms of forgery detection. This is because of the highly efficient cascaded process of CNN, which integrates multiple classification models for the identification of heterogeneities in forgery detection. Therefore, it is possible to extend the application of this model to other types of forgeries simply through addition of new layers in the framework of the binary cascaded CNN. It is also observed that the proposed model achieved localization accuracy 12.4% better than CW HPF [3], 8.3% better than DCNN [28], and 6.5% better than MSTA [37] for diverse types of forgeries. This is because of the ability of the model to extract multidomain features and incorporate convolutional features with frequencies, entropy, Gabor and Wavelet Indices. It is also concluded from the results that the proposed model has the ability to deliver 14.5% better localization accuracy as compared to CW HPF [3], 5.9% better localization accuracy as compared to DCNN [28], and 4.5% better localization accuracy as compared to MSTA [37], respectively. The proposed model is able to achieve localization recall levels 12.9% higher than CW HPF [3], 5.5% higher than DCNN [28], and 3.9% higher than MSTA [37] for various forgeries when evaluated in terms of scalability measures. Therefore, it is concluded that proposed model is outperforming the other models in terms of forgery detection, localization accuracy and computational efficiency.

In future, the model will be validated on higher number of forgeries and can be extended via integration of low-complexity feature analysis techniques like color maps, texture maps, edge maps, etc. which will assist in speeding up the localization process. This performance can also be improved via integration of Generative Adversarial Networks (GANs), Transformer Networks, and other incremental learning methods, which can be applied on large-scale forgery data sample sets.

Conflict of Interest Statement

The authors declare that they have no known financial interests, personal relationships, commercial affiliations, or competing interests that could have appeared to influence the work reported in this manuscript. All authors have contributed to the research and preparation of this manuscript, have reviewed and approved the final version, and agree with its submission for publication. The authors further confirm that there are no conflicts of interest related to the research, authorship, or publication of this article.

References

[1] K. H. Rhee, "Generation of Novelty Ground Truth Image Using Image Classification and Semantic Segmentation for Copy-Move Forgery Detection," in IEEE Access, vol. 10, pp. 2783-2796, 2022, doi: 10.1109/ACCESS.2021.3136781.

[2] X. Lin and C. -T. Li, "PRNU-Based Content Forgery Localization Augmented With Image Segmentation," in IEEE Access, vol. 8, pp. 222645-222659, 2020, doi: 10.1109/ACCESS.2020.3042780.

[3] L. Zhuo, S. Tan, B. Li and J. Huang, "Self-Adversarial Training Incorporating Forgery Attention for Image Forgery Localization," in IEEE Transactions on Information Forensics and Security, vol. 17, pp. 819-834, 2022, doi: 10.1109/TIFS.2022.3152362.
[4] A. -T. Phan-Ho and F. Retraint, "A Comparative Study of Bayesian and Dempster-Shafer Fusion on Image Forgery Detection," in IEEE Access, vol. 10, pp. 99268-99281, 2022, doi: 10.1109/ACCESS.2022.3206543.
[5] H. Wu, J. Zhou, J. Tian, J. Liu and Y. Qiao, "Robust Image Forgery Detection Against Transmission Over Online Social Networks," in IEEE Transactions on Information Forensics and Security, vol. 17, pp. 443-456, 2022, doi: 10.1109/TIFS.2022.3144878.
[6] A. -R. Gu, J. -H. Nam and S. -C. Lee, "FBI-Net: Frequency-Based Image Forgery Localization via Multitask Learning with Self-Attention," in IEEE Access, vol. 10, pp. 62751-62762, 2022, doi: 10.1109/ACCESS.2022.3182024.
[7] F. Marra, D. Gragnaniello, L. Verdoliva and G. Poggi, "A Full-Image Full-Resolution End-to-End-Trainable CNN Framework for Image Forgery Detection," in IEEE Access, vol. 8, pp. 133488-133502, 2020, '= doi: 10.1109/ACCESS.2020.3009877.
[8] K. M. Hosny, A. M. Mortda, M. M. Fouda and N. A. Lashin, "An Efficient CNN Model to Detect Copy-Move Image Forgery," in IEEE Access, vol. 10, pp. 48622-48632, 2022, doi: 10.1109/ACCESS.2022.3172273.
[9] E. A. Armas Vega, E. González Fernández, A. L. Sandoval Orozco and L. J. García Villalba, "Passive Image Forgery Detection Based on the Demosaicing Algorithm and JPEG Compression," in IEEE Access, vol. 8, pp. 11815-11823, 2020, doi: 10.1109/ACCESS.2020.2964516.
[10] I. -J. Yu, S. -H. Nam, W. Ahn, M. -J. Kwon and H. -K. Lee, "Manipulation Classification for JPEG Images Using Multi-Domain Features," in IEEE Access, vol. 8, pp. 210837-210854, 2020, doi: 10.1109/ACCESS.2020.3037735.
[11] Y. Zhu, C. Chen, G. Yan, Y. Guo and Y. Dong, "AR-Net: Adaptive Attention and Residual Refinement Network for Copy-Move Forgery Detection," in IEEE Transactions on Industrial Informatics, vol. 16, no. 10, pp. 6714-6723, Oct. 2020, doi: 10.1109/TII.2020.2982705.
[12] Y. Sun, R. Ni and Y. Zhao, "ET: Edge-Enhanced Transformer for Image Splicing Detection," in IEEE Signal Processing Letters, vol. 29, pp. 1232-1236, 2022, doi: 10.1109/LSP.2022.3172617.
[13] J. Yang, A. Li, S. Xiao, W. Lu and X. Gao, "MTD-Net: Learning to Detect Deepfakes Images by Multi-Scale Texture Difference," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 4234-4245, 2021, doi: 10.1109/TIFS.2021.3102487.
[14] Y. -X. Luo and J. -L. Chen, "Dual Attention Network Approaches to Face Forgery Video Detection," in IEEE Access, vol. 10, pp. 110754-110760, 2022, doi: 10.1109/ACCESS.2022.3215963.
[15] M. Aloraini, M. Sharifzadeh and D. Schonfeld, "Sequential and Patch Analyses for Object Removal Video Forgery Detection and Localization," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 3, pp. 917-930, March 2021, doi: 10.1109/TCSVT.2020.2993004.
[16] P. Yu, J. Fei, Z. Xia, Z. Zhou and J. Weng, "Improving Generalization by Commonality Learning in Face Forgery Detection," in IEEE Transactions on Information Forensics and Security, vol. 17, pp. 547-558, 2022, doi: 10.1109/TIFS.2022.3146781.
[17] C. W. Park, Y. H. Moon and I. K. Eom, "Image Tampering Localization Using Demosaicing Patterns and Singular Value Based Prediction Residue," in IEEE Access, vol. 9, pp. 91921-91933, 2021, doi: 10.1109/ACCESS.2021.3091161.
[18] S. Lai, L. Jin, Y. Zhu, Z. Li and L. Lin, "SynSig2Vec: Forgery-Free Learning of Dynamic Signature Representations by Sigma Lognormal-Based Synthesis and 1D CNN," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 10, pp. 6472-6485, 1 Oct. 2022, doi: 10.1109/TPAMI.2021.3087619.
[19] S. I. Lee, J. Y. Park and I. K. Eom, "CNN-Based Copy-Move Forgery Detection Using Rotation-Invariant Wavelet Feature," in IEEE Access, vol. 10, pp. 106217-106229, 2022, doi: 10.1109/ACCESS.2022.3212069.
[20] H. Wu and J. Zhou, "IID-Net: Image Inpainting Detection Network via Neural Architecture Search and Attention," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 3, pp. 1172-1185, March 2022, doi: 10.1109/TCSVT.2021.3075039.
[21] Y. Rao, J. Ni and H. Zhao, "Deep Learning Local Descriptor for Image Splicing Detection and Localization," in IEEE Access, vol. 8, pp. 25611-25625, 2020, doi: 10.1109/ACCESS.2020.2970735.
[22] C. Miao et al., "Learning Forgery Region-Aware and ID-Independent Features for Face Manipulation Detection," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 4, no. 1, pp. 71-84, Jan. 2022, doi: 10.1109/TBIOM.2021.3119403.

[23] C. Kumawat and V. Pankajakshan, "A JPEG Forensic Detector for Color Bitmap Images," in IEEE Open Journal of Signal Processing, vol. 2, pp. 280-294, 2021, doi: 10.1109/OJSP.2021.3075917.
[24] Y. Niu, B. Tondi, Y. Zhao, R. Ni and M. Barni, "Image Splicing Detection, Localization and Attribution via JPEG Primary Quantization Matrix Estimation and Clustering," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 5397-5412, 2021, doi: 10.1109/TIFS.2021.3129654.
[25] N. Le and F. Retraint, "A Statistical Modeling Framework for DCT Coefficients of Tampered JPEG Images and Forgery Localization," in IEEE Access, vol. 10, pp. 71143-71164, 2022, doi: 10.1109/ACCESS.2022.3188299.
[26] L. Zhao, C. Chen and J. Huang, "Deep Learning-Based Forgery Attack on Document Images," in IEEE Transactions on Image Processing, vol. 30, pp. 7964-7979, 2021, doi: 10.1109/TIP.2021.3112048.
[27] S. -H. Nam, W. Ahn, I. -J. Yu, M. -J. Kwon, M. Son and H. -K. Lee, "Deep Convolutional Neural Network for Identifying Seam-Carving Forgery," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 8, pp. 3308-3326, Aug. 2021, doi: 10.1109/TCSVT.2020.3037662.
[28] H. Chen, X. Yang and Y. Lyu, "Copy-Move Forgery Detection Based on Keypoint Clustering and Similar Neighborhood Search Algorithm," in IEEE Access, vol. 8, pp. 36863-36875, 2020, doi: 10.1109/ACCESS.2020.2974804.
[29] K. Kadam, S. Ahirrao, K. Kotecha and S. Sahu, "Detection and Localization of Multiple Image Splicing Using MobileNet V1," in IEEE Access, vol. 9, pp. 162499-162519, 2021, doi: 10.1109/ACCESS.2021.3130342.
[30] B. Chen, W. Tan, G. Coatrieux, Y. Zheng and Y. -Q. Shi, "A Serial Image Copy-Move Forgery Localization Scheme With Source/Target Distinguishment," in IEEE Transactions on Multimedia, vol. 23, pp. 3506-3517, 2021, doi: 10.1109/TMM.2020.3026868.
[31] O. Mayer and M. C. Stamm, "Exposing Fake Images With Forensic Similarity Graphs," in IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 1049-1064, Aug. 2020, doi: 10.1109/JSTSP.2020.3001516.
[32] G. Jia et al., "Inconsistency-Aware Wavelet Dual-Branch Network for Face Forgery Detection," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, no. 3, pp. 308-319, July 2021, doi: 10.1109/TBIOM.2021.3086109.
[33] R. Sinhal, I. A. Ansari and C. W. Ahn, "Blind Image Watermarking for Localization and Restoration of Color Images," in IEEE Access, vol. 8, pp. 200157-200169, 2020, doi: 10.1109/ACCESS.2020.3035428.
[34] Y. Quan and C. -T. Li, "On Addressing the Impact of ISO Speed Upon PRNU and Forgery Detection," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 190-202, 2021, doi: 10.1109/TIFS.2020.3009583.
[35] K. H. Rhee, "Composition of Visual Feature Vector Pattern for Deep Learning in Image Forensics," in IEEE Access, vol. 8, pp. 188970-188980, 2020, doi: 10.1109/ACCESS.2020.3029087.
[36] S. Walia, K. Kumar, M. Kumar and X. -Z. Gao, "Fusion of Handcrafted and Deep Features for Forgery Detection in Digital Images," in IEEE Access, vol. 9, pp. 99742-99755, 2021, doi: 10.1109/ACCESS.2021.3096240.
[37] J. Yang, S. Xiao, A. Li, W. Lu, X. Gao and Y. Li, "MSTA-Net: Forgery Detection by Generating Manipulation Trace Based on Multi-Scale Self-Texture Attention," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 7, pp. 4854-4866, July 2022, doi: 10.1109/TCSVT.2021.3133859.
[38] J. -L. Zhong and C. -M. Pun, "An End-to-End Dense-InceptionNet for Image Copy-Move Forgery Detection," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2134-2146, 2020, doi: 10.1109/TIFS.2019.2957693.
[39] Y. Wang, C. Peng, D. Liu, N. Wang and X. Gao, "ForgeryNIR: Deep Face Forgery and Detection in Near-Infrared Scenario," in IEEE Transactions on Information Forensics and Security, vol. 17, pp. 500-515, 2022, doi: 10.1109/TIFS.2022.3146766.
[40] F. Matern, C. Riess and M. Stamminger, "Gradient-Based Illumination Description for Image Forgery Detection," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1303-1317, 2020, doi: 10.1109/TIFS.2019.2935913.
[41] Y. Liu, C. Xia, X. Zhu and S. Xu, "Two-Stage Copy-Move Forgery Detection With Self Deep Matching and Proposal SuperGlue," in IEEE Transactions on Image Processing, vol. 31, pp. 541-555, 2022, doi: 10.1109/TIP.2021.3132828.
[42] Y. Zheng, Y. Cao and C. -H. Chang, "A PUF-Based Data-Device Hash for Tampered Image Detection and Source Camera Identification," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 620-634, 2020, doi: 10.1109/TIFS.2019.2926777.

[43] J. Brogan et al., "Fast Local Spatial Verification for Feature-Agnostic Large-Scale Image Retrieval," in IEEE Transactions on Image Processing, vol. 30, pp. 6892-6905, 2021, doi: 10.1109/TIP.2021.3097175.
[44] H. Malik, R. Gjomemo, V. N. Venkatakrishnan, R. Ansari and A. Irtaza, "Remote Check Truncation Systems: Vulnerability Analysis and Countermeasures," in IEEE Access, vol. 8, pp. 59485-59510, 2020, doi: 10.1109/ACCESS.2020.2982620.
[45] K. H. Rhee, "Detection of Spliced Image Forensics Using Texture Analysis of Median Filter Residual," in IEEE Access, vol. 8, pp. 103374-103384, 2020, doi: 10.1109/ACCESS.2020.2999308.
[46]http://web.archive.org/web/20171013200331/http://ifc.recod.ic.unicamp.br/fc.website/index.py (IEEE IFS-TC Image Forensics Challenge Corpus)
[47] https://ieee-dataport.org/documents/forgery-image-dataset(Forgery Image Data Samples)
[48] https://www.vcl.fer.hr/comofod/ (CoMoFoD Dataset Samples)
[49] H. Liu, Z. Tan, C. Tan, Y. Wei, J. Wang, and Y. Zhao, "Forgery-aware adaptive transformer for generalizable synthetic image detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024, pp. 10770–10780.
[50] D. Karageorgiou, G. Kordopatis-Zilos, and S. Papadopoulos, "Fusion transformer with object mask guidance for image forgery analysis," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024, pp. 4345–4355.
[51] I. G. Atak and A. Yasar, "Image forgery detection by combining visual transformer with variational autoencoder network," Appl. Soft Comput., vol. 165, p. 112068, 2024.
[52] D. Pawar, R. Gowda, and K. Chandra, "Image forgery classification and localization through vision transformers," Int. J. Multimedia Inf. Retrieval, vol. 14, no. 1, p. 8, 2025.
[53] M. Abdelmaksoud, B. Youssef, K. Wassif, and R. A. El-Khoribi, "Hybrid framework for image forgery detection and robustness against adversarial attacks using vision transformer and SVM," Sci. Rep., vol. 15, no. 1, p. 40371, 2025.
[54] Z. Chen, H. Li, and J. Hu, "Multi-scale high-frequency vision transformer for face forgery detection," IEEE Trans. Biometrics, Behavior, Identity Sci., 2025.
[55] Q. Man and Y.-I. Cho, "Transformer Based on Multi-Domain Feature Fusion for AI-Generated Image Detection," Electronics, vol. 15, no. 3, p. 716, 2026, doi: 10.3390/electronics15030716.
[56] G. Lek, C. Zhu, P.-Y. Chen, R. Birke, and L.-Y. Chen, "Detective SAM: Adapting SAM to Localize Diffusion-based Forgeries via Embedding Artifacts," in Data in Generative Models—The Bad, the Ugly, and the Greats, 2025.
[57] R. Peng, S. Tan, C. Kong, A. Luo, A. C. Kot, and J. Huang, "ForensicsSAM: Toward Robust and Unified Image Forgery Detection and Localization Resisting to Adversarial Attack," arXiv preprint arXiv:2508.07402, 2025.