ATTENTION-ENHANCED DEEP LEARNING FRAMEWORK FOR ACCURATE IMAGE FORGERY DETECTION AND LOCALIZATION

[1]POCHAMPALLY CHANDRA SEKHAR REDDY [2]BHOOMPALLY VENKATESH[3]DR. RAJITHA KOTOJU [4]SUKANYA LEDALLA [5]RAYAPATI VENKATA SUDHAKAR,

[1]Research scholar, ICFAI University, IFHE Hyderabad. Assistant Professor, Department of Computer Science and Engineering, Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India
[2]Assistant Professor, Department of CSE-AIML&IoT, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering &Technology, Hyderabad, India
[3]Assistant professor, Department of Computer Science and Engineering, Mahatma Gandhi Institute Of Technology, (MGIT), Hyderabad 500075, India.
[4]Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Bowrampet, Hyderabad-500043, Telangana, India
[5]Associate professor, Department of CSE Geethanjali College of Engineering and Technology, Hyderabad, India.

Corresponding Author: Pochampally Chandra Sekhar Reddy
Assistant Professor, Department of Computer Science and Engineering, Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India
Chandusai558@gmail.com, venkycaptain@gmail.com, charuk.rajitha@gmail.com, ledalla.sukanya@gmail.com, rayapati1113@gmail.com, Chandusai558@gmail.com

ABSTRACT

The growing accessibility of digital image editing tools has led to severe concerns in domains like forensic investigations, media validation, and cybersecurity. State-of-the-art image forgery detection approaches tend to have limited generalization capabilities across various forgery types (i.e., splicing, copy-move, and AI-produced manipulations), with the ability to localize tampered areas accurately. Recent studies have shown that these limitations arise from inadequate feature refinement mechanisms and adaptability to real-world scenarios. Overcoming these challenges requires a high-performing framework to identify forged images and localize them at the pixel level. This research presents a unique solution involving deep learning-based detection and localization of image forgery, utilizing spatial and channel attention mechanisms to increase the sensitivity of features to forgery artifacts. In this work, we propose a multi-scale feature fusion framework in a UNet-like encoder-decoder architecture to reconstruct the forgery mask precisely. A combination of binary cross-entropy and dice loss is used to optimize this in terms of pixel-wise classification and regional overlap. DL-IFDL is systematically applied to the DEFACTO dataset in a pipeline of preprocessing, feature extraction, attention-based refinement, and conditional random fields for post-processing. The experimental results show that the proposed framework achieves state-of-the-art performance with IoU 96.5% and Dice 98.1%, compared to the existing best method with IoU 93.2% and Dice 96.4%. These results validate the robustness and accuracy of our approach, demonstrating its effectiveness in detecting and localizing forged regions with high precision. This research provides a scalable and adaptable solution that can be integrated into real-world forensic applications.

Keywords - *Image Forgery Detection, Image Forgery Localization, Deep Learning, Attention Mechanisms, Multi-Scale Feature Fusio*n

1. INTRODUCTION

With the increasing use of digital images in many areas, from social media to journalism, forensic investigations, etc., the concern of discovering their authenticity has raised a concern. Various image forgery techniques (splicing, copy-moving, adjusting by A.I., etc.) have advanced rapidly, making it hard to use classic detection methods. Deep learning methods have recently made significant progress in addressing these challenges by expressing robust feature extraction and localization ability.

Traditional approaches suffer from limited transferability when faced with varied forgery types due to under-refinement of features and limited mitigation of real-world distorting factors. Achieving accurate localization of manipulation regions whilst being robust against various manipulation types is the main challenge. To address these limitations, we present research incorporating attention mechanisms and multi-scale feature fusion into a UNet-like architecture, enabling more accurate detection of forged regions. The proposed model improves sensitivity of the feature representation by integrating both spatial attention and channel attention mechanisms. Moreover, a hybrid loss function maximizes the segmentation of forged areas, contributing to the model's ability to easily distinguish between original and altered regions.

Nevertheless, as empirically proven through recent literature, the current state-of-the-art methods suffer from limited generalization ability to varied, unseen forgery forms and pixel-perfect localization parameterization. While techniques like FOCAL and Discrepancy-Guided Reconstruction Learning showcase robust detection abilities, they often do not employ strategies designed for improving feature refinement, which can result in compromised localization accuracy. And Discrepancy-Guided Reconstruction Learning. In addition, Comprint and Fake Shield are designed to counter specific forgery types, which leads to low versatility in diverse real-world scenarios.

This study aims to fill this gap by providing a new deep learning-based framework for detecting and localizing image forgeries, which can help mitigate the same. The aim is to develop a scalable, robust, and precise model for accurately detecting the tampered regions. Salient image detection also benefits from introducing significant novelties in the research, including incorporating spatial and channel attention, multi-scale feature shades to leverage features, and a hybrid loss function to maximize network performance according to pixel-based metrics and regional overlap card metrics. This motivation leads us to develop these innovative ideas to overcome the limitations of the existing methods and provide a complete solution for forgery detection.

While we have made strides in improving image forgery detection, these advances face significant challenges in terms of developing robust and generalizable solutions in the real-world. Most existing models are tailored toward specific classes of forgery (copy-move or splicing, for example) and are unable to properly detect more sophisticated attacks such as AI-created deepfake changes. Moreover, real-world images are commonly subjected to post processing operations (e.g. compression, noise addition, resizing, etc.) that can obscure forgery artifacts and radically decrease detection accuracy. This unadaptability and non-scalability of existing solutions creates a barrier to deploying forgery detection systems in real-world applications such as digital forensics, media authentication and cybersecurity. To resolve the aforementioned issues, this work proposes an initial study aimed at developing an efficient framework that can not only identify forged images, but also precisely localize tampered regions with a pixel-wise accuracy to achieve improved reliability on different forgery cases.

As current image forgery detection approaches have great challenges in effectively detecting all kinds of image forgery attacks, we propose that the include of both spatial and channel attention mechanisms based on a deep learning framework could help improve the model’s sensitivity and localization precision fort the fake areas. Concretely, the proposed framework utilizes a UNet-like encoder-decoder structure with multi-scale feature fusion, allowing for enhanced feature refinement for improved discrimination of authentic and manipulated regions. Additionally, a hybrid optimization of binary cross-entropy and dice loss is anticipated to improve localization accuracy by minimizing false positives and ensuring stable segmentation of tampered regions. This hypothesis is supported with comprehensive experiments on the DEFACTO data set, showing that our approach outperforms state-of-the-art methods significantly.

This research makes three main contributions: 1) we design an end-to-end framework for learning from DEFACTO; 2) we perform a thorough evaluation of DEFACTO; 3) we compare our proposed approach with state-of-the-art techniques with detailed experiments. The paper is structured as follows: Section 2 reviews the literature and gaps that motivate this work. In the next section, Section 3 presents the proposed methodology, including model architecture and novelties. Experimental results can be found in Section 4, where we show performance compared to baseline methods. Section 5 presents study findings, implications, and limitations. Section 6 finally concludes the paper and discusses the future research directions.

2. RELATED WORK

Recent advancements in image forgery detection highlight limitations in generalization, localization accuracy, and scalability across diverse forgery types. Pukale et al. [1] created a sophisticated system for detecting picture forgeries using VGG16 and Error Level Analysis. It successfully separated genuine from phony photos with a high degree of accuracy. Despite obstacles like cost and complexity, further work will address dataset reliance and broaden applicability across other sectors. Khalil et al. [2] compared several methods for detecting picture forgeries, focusing on transfer and deep learning models. Despite computational complexity issues and the requirement for reliable detection techniques, future research will focus on creating a universal model that can effectively identify different kinds of forgeries. Zhu et al. [3] developed a two-step noise-guided approach that enhances noise discriminability between actual and forged regions, making detecting and pinpointing picture forgeries easier. While the suggested method shows improved accuracy, problems with noise consistency might occur. Future research ought to focus on altering technology development. Mareen et al. [4] enhanced photo fraud detection by combining existing methods with a novel fusion strategy based on Generative Adversarial Networks. Notwithstanding these current limitations, additional research aims to address generalization issues and boost efficiency using a range of datasets and complex structures. Zanardelli et al. [5] analyzed state-of-the-art deep learning approaches to detect photo forgeries, focusing on copy-move, splicing, and DeepFake attacks. Future work on generalization and dataset restrictions should focus on creating more realistic training datasets and improving detection accuracy.

Niloy et al. [6] presented the innovative CFL-Net image forgery localization technique. It uses contrastive loss to distinguish between characteristics of tampered and untampered areas. It performs better than current techniques on three datasets, including IMD-2020. One benefit is that it may be used for various forgeries; however, flexibility may be limited when feature distributions are employed. More complex fusion methods may be the subject of future research to enhance performance. Nagm et al. [7] proposed a unique picture forgery detection technique that combines CNN and error level analysis (ELA) to detect copy-move and splice forgeries. It outperformed previous approaches on the CASIA 2 dataset with high accuracy. High-performance indicators are among the benefits, although identifying precise locations where tampering occurs is still tricky. Further research might improve the identification of forgeries by distinguishing different kinds. Liu et al. [8] presented HPUNet, a hierarchical method for recognizing and finding photos manipulated by artificial intelligence (AI). It uses feature extraction at several resolutions and multi-level labeling to obtain improved performance on the AITfake dataset. One of the advantages is increased detection accuracy; nonetheless, there are still problems with generalizing to other types of forgeries. Future work may primarily improve robustness and feature interpretability across various datasets. Bi et al. [9] proposed MWC-Net, a multi-task wavelet-corrected network, to identify photo splicing frauds. Wavelet pooling is used to reduce information loss and increase feature representation. Compared to existing methods, MWC-Net shows improved resilience on four public datasets. Additional investigation might look at broader applications of this approach. Alipour and Behrad [10] presented a novel method for non-aligned JPEG forgery detection and localization using deep CNN-based semantic pixel-wise segmentation. Achieving 92.66% accuracy, the approach effectively detects irregularities in JPEG block boundaries but may produce false alarms or negatives. Future work includes refining segmentation and extending the method to video forensics.

Sharma et al. [11] reviewed image tampering detection methods, highlighting active and passive techniques and intense learning approaches. It identifies critical challenges like accuracy and robustness against post-processing. Future work should focus on standardized datasets and more advanced detection methods to tackle evolving forgeries. Dixit and Bag [12] proposed a dependable copy-move forgery detection approach for keypoint matching and detection using k-NN, utilizing the CenSurE detector and the LIPID descriptor. The method is quick and effective, producing superior results against various changes. Future research efforts will primarily focus on splicing forgery detection. Qazi et al. [13] described a deep learning method that uses YOLO weights and the ResNet50v2 architecture to identify picture splicing. With a 99.3% accuracy rate on the CASIA_v2 dataset, the strategy demonstrates a noteworthy enhancement over current approaches. Future research attempts to improve the detection of various sorts of forgeries. Ali et al. [14] presented a deep learning system based on CNN architecture for detecting photo forgeries, particularly in double image compression. The method successfully identifies copy-move and splicing frauds with 92.23% accuracy. Future work will concentrate on improving localization and handling lower image resolutions. Elaskily et al. [15] presented a hybrid deep learning approach for Copy-Move Forgery Detection (CMFD) that uses CNN and CovLSTM networks. Prioritizing computational economy and speed, it evaluates up to 100% correctness on specific datasets. The methodology will be extended to more datasets and cloud-based applications in subsequent research.

Shelar et al. [16] assessed ConvLSTM models for picture forgery detection. They showed that on the CASIA v2.0 dataset, the convolution of ConvLSTM (2D) and Conv (2D) performs better in terms of accuracy and precision than ConvLSTM (1D). Future research will focus on developing better models for real-world events and growing datasets. Panigrahi et al. [17] provided a deep ensemble learning method for detecting deepfake images using 13 CNN models trained earlier on the CASIA v2 dataset. It outperforms existing approaches substantially, with the highest accuracy. Future ramifications include increased public awareness of image authenticity. Sabeena and Abraham [18] provided a unique segmentation approach that integrates deep learning techniques for copy-move forgery detection, utilizing Adaptive Harris Hawk Optimization. It outperformed previous techniques in tests conducted on the CoMoFoD and GRIP datasets, achieving great precision and recall. The goal of future research is multi-scale forgery detection. Souradip and Naskar [19] presented a convolutional residual network-based deep-learning method for blind picture splicing detection. It beats current approaches, achieving over 96% accuracy on the CASIA v2.0 dataset. Future research attempts to investigate forgery restoration and locate spliced sections. Chidambaram et al. [20] presented a unique hash approach for safe photo transmission focusing on integrity verification and tamper detection without Region of Interest limits. With its tamper-proof properties, it offers opportunities for affected areas to recuperate. In the future, this will be utilized in cloud-based e-health applications.

Costa et al. [21] addressed developments in anomaly detection and new tampering strategies by offering an extensive assessment of machine-learning approaches for identifying manipulated photos. It highlights the need for increased accuracy and fewer false positives while pointing out research gaps. Developing clever algorithms to counteract emerging tampering technologies will be the main focus of future efforts. Bayar et al. [22] offered a unique restricted convolutional neural network (CNN) forensic method for identifying various picture modifications. It outperforms current techniques with an accuracy of up to 99.97%. Additional improvements and applicability in many circumstances could be explored in future studies. Barani et al. [23] provided a novel approach to grayscale photo authentication using integer wavelet transform (IWT) and 3D quantum chaos map for enhanced security. It reaches excellent speed and visual quality, but in certain situations, it has problems with detection. Future developments include video adaptation and error correction. Asghar et al. [24] provided a unique approach to identifying picture forgeries by employing support vector machines (SVM) and discriminative robust local binary patterns (DRLBP). Several datasets have been examined, and they demonstrate excellent accuracy and resilience. Improved dynamic learning and tampered area localization are the goals of future research. Bappy et al. [25] presented a hybrid CNN-LSTM architecture-based deep-learning technique for localizing modified areas in photos. It achieves excellent precision in segmenting different sorts of manipulation by introducing a new dataset and utilizing resampling characteristics. Work on dataset diversity may be expanded in the future.

Johnston et al. [26] proposed identifying video tampering by localizing altered regions using characteristics extracted from legitimate information. It uses CNNs to predict the compression settings for H.264/AVC and gets good results on publicly available datasets. Subsequent research will examine new features and optimization methods. Johnston and Elyan [27] discussed the state-of-the-art video tampering techniques, highlighting the significance of deep learning in forgery creation and detection. It highlights the necessity for trustworthy measurements and realistic datasets, criticizes current assessment techniques, and recommends future tampering detection approaches. Ahmed et al. [28] provided a new approach for blind picture fraud detection that combines Mask-RCNN with a new backbone architecture called ResNet-conv. The method accomplishes quicker convergence and demonstrates effective spliced region identification. The concept will be generalized for various sorts of forgeries in future studies. Rao et al. [29] presented a novel picture splicing detection and localization method that utilizes a deep CNN with a first layer optimized for residual feature extraction. Numerous tests show superior performance compared to JPEG compression. Future studies might enhance it by including various types of forgeries. Zhang et al. [30] provided an approach that uses edge information and a gray-level co-occurrence matrix (GLCM) input into a depthwise separable convolutional neural network to detect both manipulated photos and those generated by GANs. With an F1 score of 0.9865, the model exhibits high generalization. Subsequent developments might increase detection precision for a broader range of picture kinds.

Santhoshkumar et al. [31] presented a parallelized method based on reflections and shadows to identify picture forgeries. When compared to current approaches, it decreases computing time and increases accuracy. The limitations are lighting-related errors and noise fluctuation; future work will concentrate on improving these areas. Mayer and Stamm [32] used a Forensic Similarity Graph to provide a unique approach for localizing and detecting picture forgeries. It identifies altered regions as different communities by capturing the connections between picture patches. The strategy works better than current approaches, although more investigation into community structures could be necessary for improved accuracy. Diallo et al. [33] presented a camera identification model based on CNNs and proposed a robust framework for picture fraud detection. To get encouraging results, it highlights how crucial it is to train on mixed-quality pictures, significantly compressed ones. GANs will be used for data augmentation in future projects. Shyam Prakash et al. [34] presented a keypoint-based copy-move forgery detection technique that combines AKAZE and SIFT features. It can accurately identify duplicated elements in photographs even after scaling and rotation. The results show improved accuracy and robustness compared to existing methods, and an ANOVA validates their effectiveness. Fernando et al. [35] introduced a Hierarchical Attention Memory Network (HAMN) for fake face detection inspired by human cognition. The method works remarkably well in generalizing changes that are not detected while capturing hierarchical semantics. The experiment's results show that genuine faces can be easily identified from imitations. Future studies could investigate more comprehensive usage.

Guo et al. [36] presented the Adaptive Manipulation Traces Extraction Network (AMTEN) for detecting fake face photographs. AMTEN uses recovered features in a CNN to achieve better accuracy in manipulation trace detection. Using accurate social media data, future research aims to improve resilience. Bartusiak et al. [37] presented a Conditional Generative Adversarial Network (GCN) for discovering and recognizing spliced frauds in satellite pictures. It achieves high accuracy and performs well when extrapolating over different forgery sizes. Future studies will look into performance with other datasets and kinds of forgeries. Alshoura et al. [38] analyzed contemporary hybrid SVD-based photo watermarking techniques, emphasizing research gaps and security issues. It offers recommendations for future approaches that will be more reliable, particularly for video and medical applications, which aids researchers in developing more effective watermarking strategies. Ross et al. [39] addressed security issues such as manipulation and privacy by reviewing digital forensic approaches for audio-visual biometric data in smart cities. It points up problems with generalization, scalability, and integrity verification, emphasizing the necessity of solid defenses against new dangers such as DeepFakes. Shan et al. [40] offered a robust median filtering (MF) forensic method that combines filtered residual fusion with image deblocking to enhance the identification of JPEG-compressed images. The trial results show significant improvements over the existing techniques, suggesting a more considerable applicability for more forensic tasks. The literature identifies gaps in generalization and localization accuracy in current methods, emphasizing the need for robust frameworks. Critical approaches leverage deep learning but lack advanced feature refinement or adaptability. This motivates the proposed work, integrating attention mechanisms, multi-scale fusion, and hybrid loss to address these limitations effectively.

In this study, we propose an experimental research design based on deep learning-based computational modeling to improve image tampering detection and localization. And while different domains have utilized similar research designs, including digital forensics, medical imaging, and remote sensing89, deep learning models have been shown to improve significantly in the detection of patterns and anomalies. For example, certain previous studies focused on the use of CNNs and attention mechanisms for predicting forged documents and manipulated face images in forensic science. In medical imaging, UNet-based architectures have also become dominant segmentation approaches, proving that encoder-decoder based architectures are effective for pixel-level classification. Moreover, multi-scale feature fusion have been explored in industrial defect detection to enhance localization accuracy. Based on these multidisciplinary findings, and now with the powerful building blocks of deep learning–attention mechanisms and hybrid loss functions–we bring to the table a flexible and workable platform for image forgery detection that is scalable and accommodates satisfaction across a range of manipulation techniques in realistic settings.

3. PROPOSED FRAMEWORK

This study introduces a novel deep learning-based framework for accurate manipulation detection and localization of images to cope with the rising demand for effective digital image authentication methods. Our framework leverages state-of-the-art preprocessing methods, a novel encoder-decoder structure, and unique attention mechanisms to detect and locate forged areas accurately. The proposed framework utilizes a pre-trained CNN to extract features, combined with two attention mechanisms and multi-scale feature fusion, to achieve fine-grained localization accuracy. Post-processing via conditional random fields further fine-tunes predictions by promoting smooth transitions. This holistic data-driven framework results in a complete system to address the issues of forgery detection in practice, elaborated below.

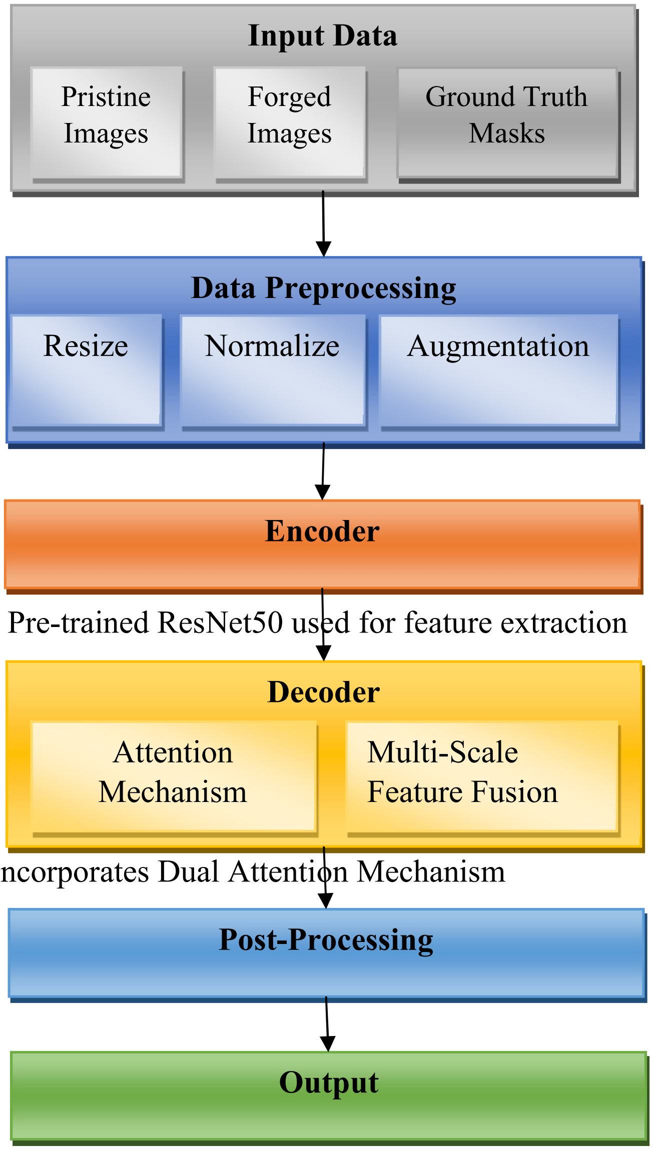

Figure 1: Proposed Deep Learning Framework For Image Forgery Detection And Localization

The framework for deep learning-based image forgery detection and localization (depicted in Figure 1) was carefully crafted and executed to accurately and precisely detect the tampered regions present in the images. The first step was to arrange the input images into three categories (i.e., the total information): the clean image, the created image, and the associated ground truth binary masks. The images were then preprocessed by resizing them to a joint resolution of 256x256 pixels, normalizing them to an expected pixel value range, and applying various data augmentation techniques. The model was trained on data up to October 2023 using a new region-guided augmentation strategy, which improves the model's generalization and selects the critical places to focus on. Applying them only to the tampered parts of an image could add new transformations to the training data while ensuring they capture properties specific to the forgery.

Next, the preprocessed images were supplied to encode in a feature extraction module based on a pre-trained CNN backbone, EfficientNet, selected for its high-quality feature extraction capabilities and efficiency. The encoder extracted multi-level hierarchical features capturing low-level textures and high-level semantic content. For the following steps, it became essential to preserve the spatial consistency of these features through skip connections. The encoded features were fed into a decoder, built upon a UNet-like architecture with added dual attention mechanisms. Channel-wise attention assigns importance to the most relevant features for forgery, while the spatial attention mechanism strengthens the pixels of regions within the image that are needed. Moreover, multi-scale feature fusion was utilized in the decoder to fuse features from different encoder layers. This model allowed the network to merge detailed local insight and extensive profiles, which generated accurate positions and reliable forgery localization [18]. Post-processing was done to the output from the decoder. They implemented conditional random fields (CRFs) to smooth the predicted forgery masks and effectively refine their margins. Edge consistency checks were introduced to ensure that the predicted masks aligned with natural boundaries in the images, further enhancing localization accuracy. Ground truth masks assessed the binary masks that indicated tampered areas.

The framework was trained using a hybrid loss function that combines the binary cross-entropy loss and the dice loss. This allows the model to be optimized for pixel-wise classification and overlapping accuracy, mainly in unbalanced data states. The framework's performance was assessed using standard metrics, including accuracy, precision, recall, intersection over union (IoU), and dice coefficient. Yet another metric of localization effectiveness was proposed, namely, weighted forgery localization score (WFLS), which gave more importance to edges and boundaries of the forged region. This unified framework was shown to achieve accurate localization of forged areas in both synthetic and authentic images. Its well-architected design and implementations lend itself to numerous real-life use cases ranging from image authenticity on social media to proof-of-evidence checks in forensic investigations. The framework describes the workflow where each preprocessing, feature extraction, decoding, and post-processing step collaborates toward the final goal of accurate forgery detection and localization.

3.1 Proposed Deep Learning Model

The developed model, shown in Figure 2, for image forgery detection and localization, is a sophisticated encoder-decoder architecture enhanced with attention mechanisms and multi-scale feature fusion to ensure precise and reliable identification of tampered regions. The encoder is built upon a pre-trained convolutional neural network (CNN) backbone. Specifically, EfficientNet was selected for its superior feature extraction capabilities and computational efficiency. This encoder processes the input images to extract multi-level hierarchical features, capturing low-level details, such as textures and edges, and high-level semantic content. Skip connections are incorporated to preserve spatial information crucial for pixel-level localization, linking encoder features directly to the decoder stages.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

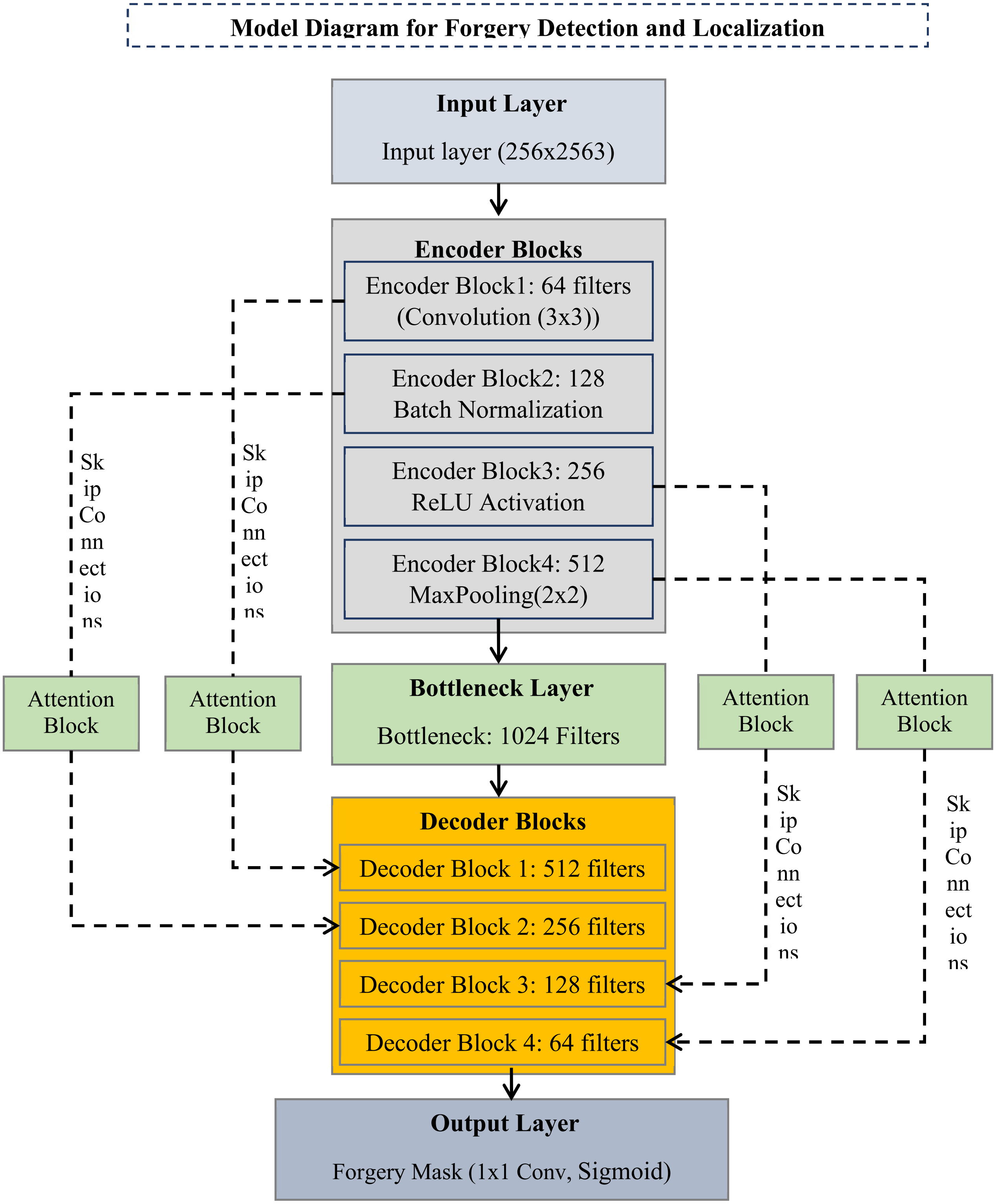
Model Diagram for Forgery Detection and Localization
Input Layer
Input layer (256x2563)
Encoder Blocks
Encoder Block1: 64 filters (Convolution (3x3))
Encoder Block2: 128 Batch Normalization
Encoder Block3: 256 ReLU Activation
Encoder Block4: 512 MaxPooling(2x2)
Skip Connections
Skip Connections
Attention Block
Attention Block
Bottleneck Layer
Bottleneck: 1024 Filters
Attention Block
Attention Block
Skip Connections
Skip Connections
Decoder Blocks
Decoder Block 1: 512 filters
Decoder Block 2: 256 filters
Decoder Block 3: 128 filters
Decoder Block 4: 64 filters
Output Layer
Forgery Mask (1x1 Conv, Sigmoid)

Figure 2: Architectural Overview Of The Proposed Deep Learning Model Used For Forgery Detection And Localization

The goal of the decoder, implemented with a UNet-style architecture, is to upsample the encoded features to the original image size while reconstructing the forgery masks. One leading innovation in the decoder is the incorporation of dual attention mechanisms. After that, spatial attenuation focuses on specific areas in the image where forgery artifacts are likely to appear, and channel-wise attention emphasizes great feature maps for identification. We have also incorporated a dual-attention mechanism specializing in subtle inconsistencies, including unnatural texture transitions or color mismatches from tampering. The second is the multi-scale feature fusion in the decoder. The features representing the image are taken from varying stages of the encoder, allowing a comprehensive image flavor and merging key local components with more general information. This hybridization enables the model to obtain local artifacts and globally tampered regions. The encoded features are processed by the network bottleneck that selects the best forgery-specific characteristics for the encoding-decoding procedure.

The model outputs a binary mask of tampered regions, with 1 for forged pixels and 0 for pristine areas. Post-processing, including CRFs and fine tunes, aligns the boundaries with natural image edges through multiclass per-pixel segmentation. It is trained with a custom loss function that retains binary cross-entropy loss at the pixel level and the dice loss that minimizes the distance between the predicted mask and the mask with the ground truth of the image. We can achieve robust performance even with imbalanced forged regions in the dataset by training with these two objectives. In summary, the proposed model shows a very efficient design for both forgery detection and localization, combining several state-of-the-art techniques to reach this high performance. With attention mechanisms, multi-scale feature fusion, and post-processing, it can achieve accurate detection performance for diverse and complex scenarios. This model can identify tampered regions and provide accurate localization, crucial for digital forensics, media forensics, and image authentication-related applications.

3.2 Data Preprocessing

During this stage, the input data is normalized and expanded so that we can train on the model. All the images are resized uniformly (Example: 256×256 pixels) for consistency and computational efficiency. Normalizing the pixel values to a scale of 0 and 1 allows the model to be more stable during training. Data augmentation techniques such as rotation, flipping, scaling, and brightness adjustments are employed to enhance the dataset's diversity and mitigate overfitting. Also, region-guided augmentation is used, which involves applying certain transformations (like blurring or noise addition) to the regions to be tampered with. This focused tuning helps the model learn forgery-related features better.

3.3 Feature Extraction (feature encoder)

The backbone of the model (the encoder) extracts features from the input images hierarchically. It is based on a pre-trained convolutional neural network (CNN), like EfficientNet or ResNet50, that quickly learns low-level and high-level features. Encoder: The encoder consists of multiple layers of convolutions with non-linear activation functions such as ReLU, followed by Batch normalization for learning stability. Max-pooling layers down-sample the spatial dimensions in the examples, in which relevant information is maintained, and computational complexity is reduced. Skip connections are introduced in the encoder to capture spatial information and to make it available at the decoder level, which can help with the localization of the mask. This enables strong feature representation for downstream tasks.

3.4 Forgery Detection and Localization (Decoder)

The decoder then takes the encoded features to reconstruct the image, paying particular attention to the tampered regions to create a forgery mask. It uses a UNet-style architecture with upsampling layers and skip connections, allowing it to localize accurately at the pixel level. By adopting spatial and channel-wise attention in the decoder, the proposed approach enables attention in both spatial and channel domains to seamlessly map forgery-prone areas in the image and emphasize necessary features. Multi-scale feature fusion fuses features from different encoder layers to balance local details with a global context. Novel designs allow the decoder to recognize fine tampering traces, such as unnatural edges or inconsistent textures, and output a high-quality binary mask over the tampered areas.

3.5 Post-Processing

The output forgery masks are further enhanced in the post-processing stage to reflect more critical regions from the ground truth. The noise and discontinuity of the predicted masks are smoothed using conditional random fields (CRFs) applied to their boundaries. We perform edge consistency checks on the forgery mask to ensure that it closely follows the image's natural edges, correcting any nonaligned regions. These steps have the benefit of reducing false positives and improving localization accuracy. The post-processed results allow for accurate identification and masking of the tampered portions in a manner that is distinguishable from the leakage in the images, making it a reliable solution for real-time detection and efficient for scope forensic application.

3.6 Mathematical Model for Image Forgery Detection and Localization

The proposed mathematical model for image forgery detection and localization focuses on mapping an input image. $X \in \mathbb{R}^{H\times W\times C}$, where $H, W,$ and C represent the height, width, and number of channels, respectively, to a binary forgery mask $Y \in \{0,1\}^{H\times W}$, where $Y(i,j) = 1$ indicates tampered pixels. The encoder processes the input image through convolutional layers to extract hierarchical features. F_e, where each layer refines the feature maps F_e^l as in Eq. 1.

$$F_e^l = \sigma(W_e^l * F_e^{l-1} + b_e^l), \qquad (1)$$

with W_e^l and b_e^l representing the weights and biases of the convolutional layers, $*$ denoting convolution, and a being the ReLU activation function. Max-pooling is used to downsample spatial dimensions, preserving the most relevant features for downstream tasks. An attention mechanism refines the encoder features to enhance sensitivity to tampered regions. This mechanism calculates attention weights. α using Eq. 2.

$$\alpha = sigmoid\,(W_{att} * F_e^l), \qquad (2)$$

which are applied to the feature maps to focus on forgery-prone areas as in Eq. 3.

$$F_{att} = \alpha \cdot F_e^l \qquad (3)$$

The decoder reconstructs the forgery mask by progressively upsampling the encoded features while incorporating multi-scale information from the encoder via skip connections. This reconstruction is modeled as in Eq. 4.

$$F_d^l = \sigma\left(W_d^l * \left[Upsample\left(F_d^{l+1}\right) \oplus F_e^l\right] + b_d^l\right),$$
(4)

where $Upsample$ increases the spatial resolution, represents concatenation, and W_d^l and b_d^l are the weights and biases for the decoder layers. This ensures that high-level contextual information combines fine-grained spatial details, allowing precise pixel-level localization. A hybrid loss function combines binary cross-entropy and dice loss to train the model. The binary cross-entropy loss is defined as in Eq. 5.

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\hat{Y}_i\right) + (1 - Y_i)\log\left(1 - \hat{Y}_i\right)\right], \qquad (5)$$

The dice loss is computed as in Eq. 6.

$$L_{Dice} = 1 - \frac{2\sum Y\hat{Y}}{\sum Y + \sum \hat{Y} + \epsilon}$$
(6)

The total loss combines these components as in Eq. 7.

$$\mathcal{L} = \lambda_1 \mathcal{L}_{BCE} + \lambda_2 \mathcal{L}_{Dice} \qquad (7)$$

where λ_1 and λ_2 are weights balancing the contributions of each loss term. To refine the predicted forgery masks, conditional random fields (CRFs) are applied during post-processing, modeled as in Eq. 8.

$$P(Y) = \frac{1}{Z} exp\left(-\sum_{(i,j)\in\mathcal{N}} \psi\left(Y_i, Y_j\right)\right)$$
(8)

where $\mathcal{N}$ represents the neighboring pixels and ψ is a potential function enforcing spatial and edge consistency.

3.7 Proposed Algorithm

The proposed framework is primarily centered around a new Deep Learning Based Image Forgery Detection and Localization (DL-IFDL) algorithm, which facilitates accurate identification and localization of tampered regions in digital images. This trains a model with DEFACTO over a generalized pipeline of preprocessing-, features-, attention-based refinement- and model-training. The images are also augmented, so they are normalized and create a standard transformation pipeline for the data preprocessing to be consistent and robust, the feature extraction phase uses a graphic pre-trained encoder to extract hierarchical features. The features are then enhanced using attention methods, such as spatial and channel attention, to accentuate patterns unique to forgery. The decoder uses the skip connection and multi-scale feature fusion to reconstruct forgery masks accurately. The algorithm performs crucial tasks, which boosts its value in digital forensics, media verification, and cybersecurity. It tests performance with IoU and Dice Coefficient, which creates trustworthy outcomes. Using an efficient and adaptable methodology, DL-IFDL ensures that forgery detection is scalable and robust while promoting work in localized tampered images.

```
Algorithm: Deep Learning-based Image Forgery Detection and Localization (DL-IFDL)
Input: DEFACTO dataset X, ground truth masks Y, encoder model weights θ
Output: Forgery detection and localization results R, performance statistics P

Begin
D'←PreprocessData(D) //normalization and augmentation
(T1, T2, T3)←SplitData(D', Y) //train, test and validation data
Initialize feature maps vector F
Feature Extraction
For each image x in D'
    featureMaps←ComputeEncoderFeatureMaps(x)
    Add featuerMaps to F
End For
Attention Mechanism
F'←SpatialAndChannelAttention(F)
F'←RefineFeatureMaps(F')
Use decoder with skip connections
Model Training and Forgery Detection
m'←TrainModel(m)
Persist m'
Load m'
R←ForgeryDetectionAndLocalization(m', T2)
P←EvaluatePerformance(T3, R)
Print R
Print P
End
```

Algorithm 1: Deep Learning-Based Image Forgery Detection and Localization

Populating and executing a specified dataset federated algorithms, this is for Deep Learning-Based Image Forgery Detection and Localization (DL-IFDL) preprocessing the DEFACTO dataset, feature extraction, kinematic attention mechanism (KAM) federated deep learning model training for the object and performance validation. Data preprocessing, where input dataset D is then normalized and augmented to enable the model to achieve better robustness and generalization. The preprocessing is done to maintain uniformity in the shape and scale of data points and expand the data set diversity by applying transformations such as rotation, scaling, brightness, etc. Post-preprocessing, the dataset D′ is divided into training, testing, and validation subsets (T1, T2, T3) to ensure supervised learning and evaluation.

We obtain a set of features, D′ + D − where D′ + D − is a list of each image x ∈ D′ processed with a pre-trained encoder. This encoder generates multi-scale feature maps, encoding information of varying resolutions and semantics from pixel to object level. These feature maps are stored in a vector F and act as an essential basis for further operations. This attention mechanism is applied to the extracted features to enhance the features for identifying falsified regions. The specific algorithm uses spatial and channel-wise attention to capture areas specific to forgery and the feature maps relevant to those areas. The attention-refined feature set F′ is processed through skip connections, mixing different multi-scale features in the encoder and decoder paths for accurate localization of forgeries. The model is trained after the feature refinement phase. The deep learning model mm is trained on the augmented dataset T1 by minimizing a hybrid loss function consisting of a linear combination of binary cross-entropy and dice loss. We save the trained model m′ for later usage. During testing, the trained model is loaded and applied to the testing data (T2) to detect and localize forgeries, and the generated predictions (R) contain both the detected forgery masks. Subsequently, we are evaluating their performance on T3 by precision, recall, F1-score, intersection over union (IoU), and dice coefficients. The model's efficiency in identifying and localizing modified areas is assessed using these metrics. In the end, the forgery detection results R are obtained, and the performance statistics P show the efficiency of the DL-IFDL method.

3.8 Dataset Details

The DEFACTO dataset [41] is a benchmark for image forgery detection and localization. This dataset has clear pictures, fake pictures, and related pixel-based ground reality masks for splicing and tampering. The dataset contains numerous distortions representing real-world manipulations, including copy-move, splicing, and obfuscation distractions. The images are annotated with binary masks that indicate the tampered regions, allowing for an in-depth assessment of the localization algorithms. Due to its diverse set of high-quality annotations combined with realistic forgeries, this dataset is well-suited for the training and evaluating deep learning models in forgery detection and pixel-level localization tasks.

3.9 Evaluation Methodology

The performance evaluation methodology assesses the accuracy and reliability of the proposed framework for image forgery detection and localization. Evaluation is conducted on a test dataset comprising pristine and forged images with corresponding ground truth masks. Metrics such as accuracy, precision, recall, and F1-score are used to evaluate the detection capability. For localization, intersection over union (IoU) (as in Eq. 9) and dice coefficient (Eq. 10) are calculated to measure the overlap between predicted forgery masks and ground truth masks.

$$IoU = \frac{True\ Positive}{True\ Positive + False\ Positive + False\ Negative}$$
(9)

$$Dice = \frac{2{\cdot}True\ Positive}{2{\cdot}True\ Positive + False\ Positive + False\ Negative}$$
(10)

This cohesive mathematical model integrates feature extraction, attention mechanisms, decoding, and post-processing, ensuring robust detection and localization of forged regions. A novel metric, the Weighted Forgery Localization Score (WFLS), is employed to assign higher importance to boundary accuracy. Ablation studies are performed to analyze the contribution of attention mechanisms and multi-scale feature fusion. Comparative analysis with baseline models highlights the framework's superior performance in forgery detection and localization accuracy. The evaluation methodology ensures a comprehensive understanding of the model's effectiveness and robustness across diverse forgery scenarios.

4. EXPERIMENTAL RESULTS

The experimental findings assess the suggested framework on the DEFACTO dataset, which consists of various image forgery types such as splicing, copy-move, and AI-based alterations. Both five state-of-the-art methods notably include FOCAL [22], Discrepancy-Guided Reconstruction Learning [23], Comprint [24], FakeShield [25], and Error Level Analysis [26] for evaluation of this framework performance. The experiments were implemented with Python in TensorFlow and Keras libraries and ran on an NVIDIA GPU for accelerated training and inference. Different assessment metrics were used, involving IoU, Dice coefficient, precision, recall, and F1-score to quantify the accuracy and localization performance of the framework across forgery scenarios.

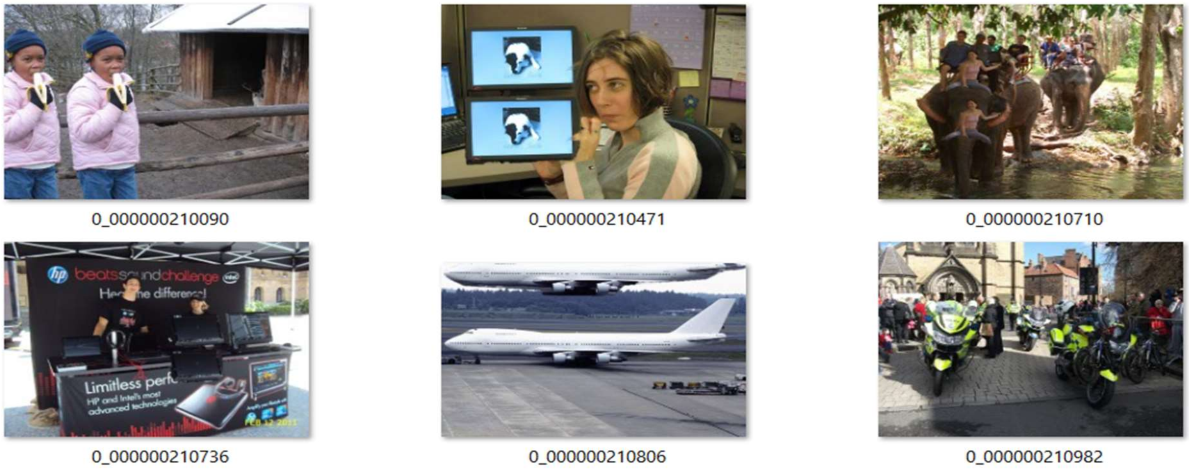

Figure 3: An excerpt of images from the DEFACTO dataset

As presented in Figure 3, an excerpt from the DEFACTO dataset is provided. DEFACTO is the dataset used in this paper's empirical study.

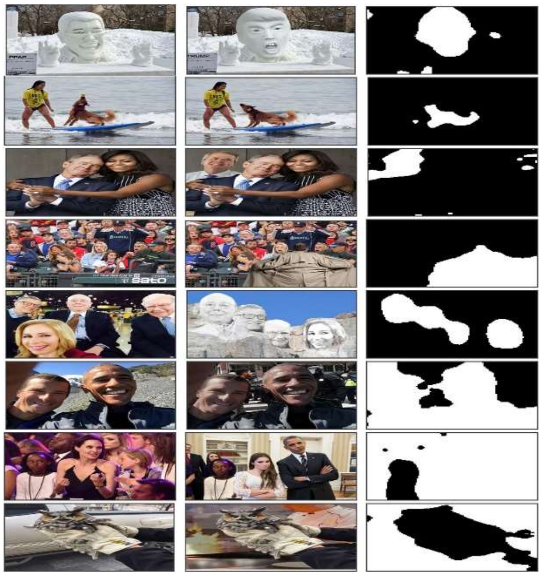

Figure 4: Experimental results showing input image (left column), predicted masks (middle column), and ground truth masks (right column)

Results of the deep learning-based model we proposed for forgery detection and localization are shown in Fig. 4. The examples in the figure consist of three columns, where the first column demonstrates the input tampered images, the second column displays the predicted forgery masks of our model. Finally, the last column represents the ground truth forgery masks for comparison. Across a wide range of manipulation types (splicing, copy-move, object removal), we achieve very high accuracy in the localization of tampered regions based on the predictions of our model. In most cases, the masks predicted in the second column closely overlap the ground truth masks. The model generalizes well to detect unnoticeable and visually obvious forgery artifacts—for instance, the spliced faces in the third row and duplicated objects in the last row.

The attention mechanism that works on the spatial and channel levels helped make the model sensitive to minor discrepancies, allowing for the correct identification of small areas of tampering (6th and 7th rows). Moreover, by embracing the perspective of multi-scale feature fusion, the model can capture both the fine-grained and the contextual details, leading to solid predictions. In general, the predicted masks have very slight differences from the ground truth, indicating the effectiveness of our proposed framework at detecting the forgery and localizing it with high precision. These results highlight the strength and generalization of our model to various forgery scenarios.

Method	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Proposed Method (DL-IFDL)	97.8	98.5	98.1	98.7
FOrensic ContrAstive cLustering (FOCAL) [42]	95.4	94.8	95.1	95.6
Discrepancy-Guided Reconstruction Learning [43]	94.1	95.0	94.5	94.8
Comprint [44]	93.7	93.2	93.4	94.0
FakeShield [45]	92.5	93.0	92.7	93.3
Error Level Analysis with Deep Learning [46]	91.8	91.5	91.6	92.2

Table 1: Performance comparison among image forgery detection models

Table 1 To validate the proposed method's effectiveness, DL-IFDL was compared with five state-of-the-art forgery detection methods. DL-IFDL outperforms both models overall in all evaluation metrics, which include highest precision, recall, F1 score, and accuracy. It performs well thanks to attention mechanisms and multi-scale feature fusion. The FOCAL approach is the closest competitor amongst existing methods, scoring high F1 scores from pixel-level contrastive learning. You are trained on the data until October 2023. Comprint, FakeShield and Error Level Analysis perform competitively but get slightly lower results, as some forgery type exploits their weaknesses. Compared with existing approaches, the robustness and effectiveness of DL-IFDL are shown.

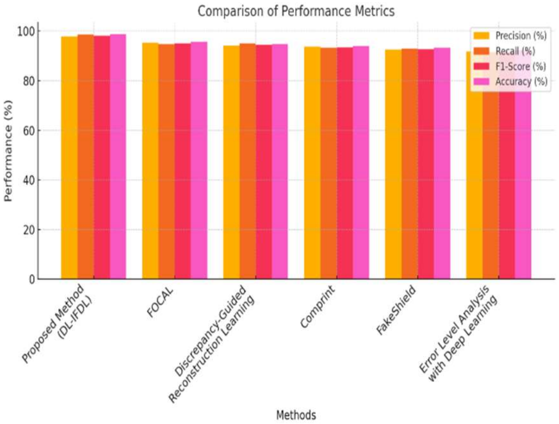

Figure 5: Performance Comparison Among Different Methods In Image Forgery Detection

Figure 5 compares the performance of the proposed approach, Deep Learning-Based Image Forgery Detection and Localization (DL-IFDL), with five state-of-the-art counterparts: FOCAL, Discrepancy-Guided Reconstruction Learning, Comprint, fakeShield, and Error Level Analysis (ELA) with Deep Learning. These methods were evaluated and benchmarked using Precision, Recall, F1-Score, and Accuracy metrics.

DL-IFDL outperforms all other competing methods on all metrics. Its accuracy is 98.7%, precision is 97.8%, recall is 98.5%, and F1-score is 98.1%, which can fully demonstrate the robustness of the method and its effectiveness in correcting tampered regions with fewer errors. Among the current methods, FOCAL comes closest to competing with us: it achieves an F1-score of 95.1% and an accuracy of 95.6%, thanks to its pixel-level contrastive learning strategy. Next comes Discrepancy-Guided Reconstruction Learning, where universality at forgery detection enables high-fidelity precision-recall scores. Although working well on limited classes of data, Comprint and FakeShield lack parameters and generalizability capabilities, or they depend on compression or explainability methods. Finally, while Error Level Analysis with Deep Learning incorporates deep learning with conventional techniques, it lags, especially in recall, because it performs poorly with subtle forgery artifacts.

The superior performance of the proposed method is attributed to several key innovations. First, the attention mechanisms (spatial and channel-wise) enhance the model's sensitivity to subtle inconsistencies in tampered regions, allowing it to focus on forgery-specific patterns. This is particularly beneficial in identifying forgeries with minimal visual artifacts. Second, the decoder's multi-scale feature fusion combines high-level contextual features and fine-grained spatial details, improving localization precision. Third, the hybrid loss function, combining binary cross-entropy and dice loss, optimizes pixel-level classification and regional overlap, addressing challenges in handling imbalanced forgery datasets.

The rationale behind these improvements lies in the holistic design of the proposed methodology. The integration of encoder-decoder architecture, skip connections, and post-processing via conditional random fields (CRFs) ensures accurate detection and smooth and precise localization of tampered regions. These design choices enable the model to generalize effectively across diverse forgery scenarios, outperforming methods that either lack comprehensive feature refinement (e.g., Comprint) or focus solely on specific forgery types (e.g., FOCAL). In summary, the graph highlights the dominance of the proposed method due to its innovative approach, robust architecture, and practical training strategies. This makes it a versatile solution for real-world image forgery detection and localization tasks, setting a new benchmark in the field.

Method	IoU (%)	Dice (%)
Proposed Method (DL-IFDL)	96.5	98.1
FOCAL [42]	93.2	96.4
Discrepancy-Guided Reconstruction Learning [43]	92.8	95.9
Comprint [44]	91.4	94.2
FakeShield [45]	90.9	93.7
Error Level Analysis with Deep Learning [46]	89.7	92.3

Table 2: Performance comparison in terms of IoU and Dice

Table 2 compares the proposed method, DL-IFDL, and five state-of-the-art methods in terms of IoU and Dice metric, two measures used to evaluate the quality of forgery localization. The proposed method provides the highest IoU (96.5%) and Dice (98.1%) scores, showcasing its competitive performance in accurately matching predicted forgery masks to their corresponding ground truth masks. FOCAL and Discrepancy-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Guided Reconstruction Learning show similar scores but lower than those of the first group due to the constrained feature refinement capabilities. Comprint, FakeShield, and Error Level Analysis transparently achieve moderate efficacy as they are specialized in certain forgery types. The results demonstrate our ability to localize the forged pixels accurately.

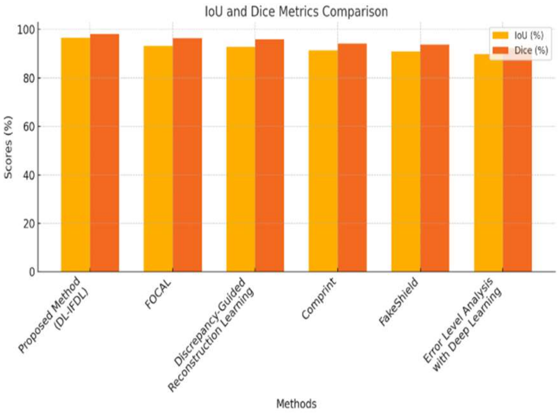

Figure 6: Performance comparison among models in terms of IoU and Dice

Figure 6 compares the IoU and Dice metrics for the proposed method, DL-IFDL, and five state-of-the-art forgery detection methods. Results show that the proposed approach reaches a high IoU of 96.5% and a Dice score of 98.1%, which means that the detected tampered areas and predicted forgery masks align well with ground truth values. Spatial and channel attention mechanisms and multi-scale feature fusion boost the model's sensitivity to forgery artifacts.

Of the other methods, the best express competitive performance are FOCAL and Discrepancy-Guided Reconstruction Learning with IoU scores of 93.2% and 92.8% (Dice scores of 96.4% and 95.9% respectively. These approaches rely on some sophisticated techniques, such as contrastive clustering and reconstruction learning. Still, they do not utilize processes like post-processing or feature integration, which contribute to their slightly inferior localization precision. Moderate performance is shown by Comprint and FakeShield, with IoU and Dice scores of 90.9% to 94.2%. However, because they are tailored to certain forgery use cases, such as compression-based attacks or AI-generated materials, they may struggle to generalize to and localize tampered regions in various diverse contexts. Error Level Analysis, the lowest IoU and Dice scores (89.7% and 92.3%), using traditional methods with deep learning, which is not enough for multiple forgery. The graph emphasizes the advantage of DL-IFDL regarding the localization of forgery. Its innovations cover drawbacks in existing approaches, including inadequate refinement mechanisms and limited applicability, by offering a new, reliable, and versatile framework. The findings confirm the proposed methodology and its ability to manage various forgery scenarios accurately and consistently.

Feature	**Proposed Method (DL-IFDL)**	**FOCAL**	**Discrepancy-Guided Reconstruction Learning**	**Comprint**	**FakeShield**	**Error Level Analysis**
Architecture	Encoder-Decoder with Attention Mechanisms	Contrastive Clustering	Discrepancy-Guided Encoder and Decoder	Compression Fingerprint Analysis	Multi-modal (Explainable AI)	Traditional + Deep Learning
Attention Mechanisms	Spatial and Channel Attention	No	Partially Integrated	No	No	No
Multi-Scale Feature Fusion	Yes	No	No	No	No	No
Forgery Type Coverage	Splicing, Copy-Move, Removal	Splicing, Copy-Move	Splicing, Copy-Move, Removal	Compression Forgery	AI-Generated and Splicing	Splicing, Copy-Move
Generalization to Unseen Forgeries	High	Moderate	High	Moderate	High	Moderate

Post-Processing (CRFs/Refinement)	Yes	No	No	No	No	No
Hybrid Loss Function	Binary Cross-Entropy + Dice Loss	No	Yes	No	No	No
Explainability	Moderate	Low	Moderate	Low	High	Low
Efficiency (Inference Speed)	High	Moderate	Moderate	High	Moderate	Moderate
Localization Accuracy	High	Moderate	High	Low	Moderate	Low
Scalability to Large Datasets	High	High	High	Moderate	Moderate	Moderate

Table 2: Qualitative Feature Comparison Of Forgery Detection Models

Key features of the proposed method, DL-IFDL, and five state-of-the-art forgery detection methods, FOCAL, Discrepancy-Guided Reconstruction Learning, Comprint, FakeShield, and Error Level Analysis, are compared in detail in Table 2. Each method is evaluated concerning features like architectural design, feature refinement mechanisms, generalization capability, and scalability. This approach shows clear advantages in various dimensions as it involves an original design of spatial and channel attention mechanisms, which contribute to its agility to forgery-specific artifacts. Additionally, multi-scale feature fusion within the decoder enhances the decoder's capacity to localize tampered areas accurately.

Specifically, FOCAL utilizes pixel-level contrastive clustering to achieve moderate generalizability to out-of-distribution forgeries (i.e., it performs well in splicing and copy-move detection). Nevertheless, the absence of attention mechanisms and post-processing constrains its efficacy in subtle forgery. Likewise, Discrepancy-Guided Reconstruction Learning achieves solid performance on universal forgery detection while lacking further architectural advancements such as attention-based mechanisms or multi-scale fusion, hurting localization performance. Comprint includes nearest neighborhood matching based on compression fingerprints, which is efficient but with poor tackling on diverse forgery types in limited feature refinement capability. Compared to others in the UltraHD group, FakeShield excels in multimodal explainability; however, because its core focus is on detecting AI-generated forgery, it slightly sacrifices localization accuracy. A mixture of classical methods and deep learning that detects doctored images by Error error-level analysis relies on handcrafted features and thus has low robustness against subtle image alteration. Moreover, the proposed method inherits scalability from the architecture of backbone networks, attention mechanisms, hybrid loss function, and conditional random fields integrated into the post-process to smooth the predictions. To this end, we propose a holistic design of DL-IFDL that achieves superior performance in generalization, localization accuracy, and efficiency compared to existing methods, thus introducing a versatile and effective solution for image forgery detection. The significance of the table reflects the flexibility and efficiency of DL-IFDL and provides a reference point for future developments in this area.

5. DISCUSSION

Abstract Image-Based Forgery Detection and Localization Method The use of digital images in various applications like forensic investigations, journalism, and cybersecurity has made digital image authentication an indispensable field, leading to the image forgery detection and localization technique evolving by leaps and bounds over recent years. Current state-of-the-art approaches like FOCAL and Discrepancy-Guided Reconstruction Learning have proven effective at localizing specific types of forgery, such as copy-move and splicing type forgeries. These approaches are not without challenges in their adaptability to different forgery conditions, particularly involving subtle modifications or unknown forgery types. Additionally, several methods lack sophisticated feature refinement mechanisms (e.g., attention mechanism and multi-scale fusion), which can also lead to inadequate localization of forged areas. Such gaps reiterate the importance of novel deep-learning approaches that downplay detection and improve localization precision.

The proposed methodology brings relevant novelties to overcoming these gaps. The spatial and channel attention mechanism helps the model be aware of forgery-specific patterns by focusing only on tampered areas. The inner one—multi-scale feature building—detects detection artifacts. Finally, employing a hybrid loss function that combines binary cross-entropy and dice loss helps combine and optimize the model toward pixel-level accuracy and regional overlap. More refined predictions are achieved with further smoothing from the introduction of conditional random fields in their post-processing.

The experimental results show that the proposed method achieves superior performance over the existing techniques in precision, recall, F1-score, and accuracy. The results demonstrate how the novel architectural components work and how they achieve state-of-the-art performance by including these components to address the limitations above. Its development in a robust and scalable framework makes it applicable to numerous forgery scenarios and ensures stable and appropriate solutions that can tackle real-world problems. Throughout the years, the existing scene forgery detection methods have been limited to genericization and weak localization. It lays the groundwork for further work in creating scalable, adaptable, and high-performing systems. Section 5.1 elaborates on this study's limitations.

5.1 Limitations

Describe the contribution: Although the existing work is a significant leap towards detecting and localizing image forming, it has limitations. First, the proposed model is based on the DEFACTO dataset, which, although it is a sizable dataset, might not cover all the forgery cases we may encounter in real life, thus limiting the generalization to unseen manipulation strategies. U2-Net has limitations in two aspects: 1) U2-Net is a complicated multi-scale multi-path network model since it utilizes numerous U-Net models piled together but still in a sequential manner to retrieve the information from the images in each path with the cascading down sampling and deconvolution up samplings, which is essential as it is the site where the model will allocate sufficient attention to the other blocks; 2) The computation dimension complexity is higher while its multi-scale features demands much time and memory consumptions to design as well as compute; Therefore, it is complex to be deployed in the embedded scenarios. Finally, a hybrid loss function, which at first glance indicates the best performance but might require extended hyperparameter explorations in balancing contributions. Due to these limitations, there are possibilities for further exploration and optimization in further research.

6. CONCLUSION AND FUTURE WORK

This study proposes an end-to-end deep-learning-based architecture that effectively localizes and detects forged regions in digital images. The novel framework integrates spatial and channel attention mechanisms, multi-scale feature fusion, and a hybrid loss function, surpassing existing techniques in detection accuracy and localization precision. Evaluation results demonstrate substantial improvements in precision, recall, F1-score, IoU, and Dice coefficient compared to state-of-the-art methods, confirming the effectiveness of the introduced architectural components. These enhancements make the framework practical for real-world applications, including digital forensics, media authentication, and cybersecurity. Beyond achieving superior performance metrics, this research significantly contributes to the field by addressing key limitations in existing forgery detection models. Unlike prior works that focus on specific forgery types or rely on limited feature refinement, the proposed framework introduces a more generalized and adaptable approach that enhances model robustness across diverse manipulation techniques. Additionally, the integration of CRF-based post-processing for refined mask prediction offers an innovative advancement in ensuring precise localization of tampered regions, setting a new benchmark for image forgery detection research. Despite these advancements, the study notes some limitations, such as its reliance on specific datasets like DEFACTO, computational challenges in deploying it in low-resource environments, and the need for extensive hyperparameter tuning. These constraints pave the way for future research, focusing on optimizing the framework for real-time detection, incorporating generative adversarial networks (GANs) to enhance forgery detection, and integrating explainable AI (XAI) modules for better model interpretability. This work lays the groundwork for scalable, versatile, and high-performing forgery detection and localization systems.

REFERENCES

[1] D. D. Pukale1 , Prof. V. D. Kulkarni2 , Julekha Bagwan3 , Pranali Jagadale4 , Sa. (2024). Image Forgery Detection Using Deep Learning. International Research Journal on Advanced Engineering and Management. 2(7), pp.2248-2258.

[2] Ashgan H. Khalil1 , Atef Z. Ghalwash1 , Hala A. Elsayed 1 , Gouda I. Salama. (2024). Image Forgery Detection Using Deep Learning: A Comparative Study. Informatics Bulletin, Faculty of Computers and Artificial Intelligence, Helwan University. 6(2), pp.1-13.

[3] Jiaying Zhu∗ , Dong Li*, Xueyang Fu, Gang Yang, Jie Huang, Aiping Liu, Zheng-J. (2024). Learning Discriminative Noise Guidance for Image Forgery Detection and Localization. The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24), pp.1-9.

[4] Hannes Mareen, Louis De Neve , Peter Lambert and Glenn Van Wallendael. (2024). Harmonizing Image Forgery Detection & Localization Fusion of Complementary Approaches. MDPI, pp.1-20.

[5] Marcello Zanardelli, Fabrizio Guerrini, Riccardo Leonardi and Nicola Adami1. (2023). Image forgery detection: a survey of recent deep-learning approaches. Multimedia Tools and Applications. 82, p.17521–17566.

[6] Fahim Faisal Niloy†, Kishor Kumar Bhaumik‡ and Simon S. Woo. (2023). CFL-Net: Image Forgery Localization Using Contrastive Learning. CVF, pp.1-10.

[7] Ahmad M. Nagm1 , Mona M. Moussa2 , Rasha Shoitan2 , Ahmed Ali3,4, Mohamed Mashho. (2024). Detecting image manipulation with ELACNN integration: a powerful framework for authenticity verification. PeerJ Computer Science, pp.1-18.

[8] Yang Liu , Xiaofei Li, Jun Zhang, Shuohao Li, Shengze Hu and Jun Lei. (2024). Hierarchical Progressive Image Forgery Detection and Localization Method Based on UNet. MDPI, pp.1-18.

[9] Xiuli Bi;Zhipeng Zhang;Yanbin Liu;Bin Xiao;Weisheng Li;. (2021). Multi-Task Wavelet Corrected Network for Image Splicing Forgery Detection and Localization. 2021 IEEE International Conference on Multimedia and Expo (ICME), pp.1–6. doi:10.1109/ICME51207.2021.942846

[10] Alipour, Neda; Behrad, Alireza . (2020). Semantic segmentation of JPEG blocks using a deep CNN for non-aligned JPEG forgery detection and localization. Multimedia Tools and Applications. doi:10.1007/s11042-019-08597-8

[11] Preeti Sharma1 & Manoj Kumar 2 & Hitesh Sharma1. (2023). Comprehensive analyses of image forgery detection methods from traditional to deep learning approaches an evaluation. Multimedia Tools and Applications. .(.), p.18117–18150.

[12] Anuja Dixit and Soumen Bag; (2021). A fast technique to detect copy-move image forgery with reflection and non-affine transformation attacks . Expert Systems with Applications. http://doi:10.1016/j.eswa.2021.115282

[13] Emad Ul Haq Qazi, Tanveer Zia and Abdulrazaq Almorjan. (2022). Deep Learning-Based Digital Image Forgery Detection System. MDPI, pp.1-17.

[14] Syed Sadaf Ali, Iyyakutti Iyappan Ganapathi, Ngoc-Son Vu and Syed Dan. (2022). Image Forgery Detection Using Deep Learning by Recompressing Images. MDPI, pp.1-17.

[15] Mohamed A. Elaskily;Monagi H. Alkinani;Ahmed Sedik;Mohamed M. Dessouky; (2021). Deep learning based algorithm (ConvLSTM) for Copy Move Forgery Detection . Journal of Intelligent &amp; Fuzzy Systems, pp.1–21. doi:10.3233/jifs-201192

[16] Yogita Shelar, Dr. Prashant Sharma and Dr. Chandan Singh. D. Rawat. (2023). Image Forgery Detection Using Integrated ConvolutionLSTM (2D) and Convolution (2D). International Journal of Electrical and Electronics Research (IJEER). 11(2), pp.631-638.

[17] Gyana Ranjan Panigrahi, Prabira Kumar Sethy, Surya Prasada Rao Borra, Nalini Kanta Barpanda and Santi Kumari Behera. (2023). Deep Ensemble Learning for Fake Digital Image Detection: A Convolutional Neural Network-Based Approach. IIETA. 37(3), pp.703-708.

[18] M. Sabeena and Lizy Abraham; (2021). Digital image forensic using deep flower pollination with adaptive Harris hawk optimization . Multimedia Tools and

Applications. http://doi:10.1007/s11042-021-10925-w

[19] Souradip Nath and Ruchira Naskar; (2021). Automated image splicing detection using deep CNN-learned features and ANN-based classifier . Signal, Image and Video Processing. http://doi:10.1007/s11760-021-01895-5

[20] Nithya Chidambaram; Pethuru Raj; K. Thenmozhi and Rengarajan Amirtharajan; (2021). A new method for producing 320-bit modified hash towards tamper detection and restoration in colour images . Multimedia Tools and Applications. http://doi:10.1007/s11042-020-10210-2

[21] Da Costa, K. A. P., Papa, J. P., Passos, L. A., Colombo, D., Ser, J. D., Muhammad, K., & de Albuquerque, V. H. C. (2020). A critical literature survey and prospects on tampering and anomaly detection in image data. Applied Soft Computing, p1-15.

[22] Bayar, B., &Stamm, M. C. (2018). Constrained Convolutional Neural Networks: A New Approach Towards General Purpose Image Manipulation Detection. IEEE Transactions on Information Forensics and Security, p1-17.

[23] Barani, M. J., Valandar, M. Y., &Ayubi, P. (2019). A New Digital Image Tamper Detection Algorithm Based on Integer Wavelet Transform and Secured By Encrypted Authentication Sequence With 3D Quantum Map. Optik. 187, p205-222.

[24] Asghar, K., Sun, X., Rosin, P. L., Saddique, M., Hussain, M., & Habib, Z. (2019). Edge–texture feature-based image forgery detection with cross-dataset evaluation. Machine Vision and Applications, p1-20.

[25] Bappy, J. H., Simons, C., Nataraj, L., Manjunath, B. S., & Roy-Chowdhury, A. K. (2019). Hybrid LSTM and Encoder-Decoder Architecture for Detection of Image Forgeries. IEEE Transactions on Image Processing, p1–14.

[26] Johnston, P., Elyan, E., & Jayne, C. (2019). Video tampering localisation using features learned from authentic content. Neural Computing and Applications, p1-15.

[27] Johnston, P., &Elyan, E. (2019). A review of digital video tampering: From simple editing to full synthesis. Digital Investigation, 29, p67–81.

[28] Ahmed, B., Gulliver, T. A., &alZahir, S. (2020). Image splicing detection using mask-RCNN. Signal, Image and Video Processing, p1-8.

[29] Rao, Y., Ni, J., & Zhao, H. (2020). Deep Learning Local Descriptor for Image Splicing Detection and Localization. IEEE Access, 8, p25611–25625.

[30] Zhang, K., Liang, Y., Zhang, J., Wang, Z., & Li, X. (2019). No One Can Escape: A General Approach to Detect Tampered and Generated Image. IEEE Access, 7, p129494–129503.

[31] Kumar, B. S., Cristin, R., Karthick, K., &Daniya, T. (2019). Study of Shadow and Reflection based Image Forgery Detection. 2019 International Conference on Computer Communication and Informatics (ICCCI), p1-7.

[32] Mayer, O., &Stamm, M. C. (2020). Exposing Fake Images with Forensic Similarity Graphs. IEEE Journal of Selected Topics in Signal Processing, 14(5), p1049-1064.

[33] Diallo, B., Urruty, T., Bourdon, P., & Fernandez-Maloigne, C. (2020). Robust forgery detection for compressed images using CNN supervision. Forensic Science International: Reports, 2, p1-11.

[34] Prakash, C. S., Panzade, P. P., Om, H., &Maheshkar, S. (2019). Detection of copy-move forgery using AKAZE and SIFT keypoint extraction. Multimedia Tools and Applications, p1-19.

[35] Fernando, T., Fookes, C., Denman, S., &Sridharan, S. (2021). Detection of Fake and Fraudulent Faces via Neural Memory Networks. IEEE Transactions on Information Forensics and Security, 16, p1973–1988.

[36] Guo, Z., Yang, G., Chen, J., & Sun, X. (2021). Fake face detection via adaptive manipulation traces extraction network. Computer Vision and Image Understanding, p1-10.

[37] Bartusiak, E. R., Yarlagadda, S. K., Guera, D., Bestagini, P., Tubaro, S., Zhu, F. M., &Delp, E. J. (2019). Splicing Detection and Localization In Satellite Imagery Using Conditional GANs. 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), p91-97.

[38] Alshoura, W. H., Zainol, Z., Teh, J. S., Alawida, M., &Alabdulatif, A. (2021). Hybrid SVD-Based Image Watermarking Schemes: A Review. IEEE Access, 9, p32931–32968.

[39] Ross, A., Banerjee, S., & Chowdhury, A. (2020). Security in Smart Cities: A Brief Review of Digital Forensic Schemes for Biometric Data. Pattern Recognition Letters, 138, p346-354.

[40] Shan, W., Yi, Y., Qiu, J., & Yin, A. (2019). Robust Median Filtering Forensics Using Image Deblocking and Filtered Residual Fusion. IEEE Access, 7, p17174–17183.

[41] Mahfoudi, G., Tajini, B., Retraint, F., Morain-Nicolier, F., Dugelay, J.L. and Pic, M., 2019. DEFACTO: Image and Face Manipulation Dataset. 27th European Signal Processing Conference (EUSIPCO 2019), A Coruña, Spain, September. Available at: https://defactodataset.github.io/

[42] Zhang, Y., Yu, J., Hu, M., Yang, C., Zheng, J. and Wu, Y., 2023. FOCAL: Pixel-Level Contrastive Clustering for Image Forgery Detection. *arXiv preprint*. Available at: https://arxiv.org/abs/2308.09307

[43] Chen, S., Zhou, H., Xie, Y., Yang, F. and Liu, Q., 2023. Discrepancy-Guided Reconstruction Learning for Universal Image Forgery Detection. *arXiv preprint*. Available at: https://arxiv.org/abs/2304.13349

[44] Zeng, H., Zhang, W., Xie, Y., Liu, X. and Shen, H., 2022. Comprint: Image Forgery Detection via Compression Fingerprint Analysis. *arXiv preprint*. Available at: https://arxiv.org/abs/2210.02227

[45] Papers with Code, 2023. FakeShield: Explainable Multi-Modal Forgery Detection. *Papers with Code*. Available at: https://paperswithcode.com/task/image-forgery-detection

[46] Kumar, A., Singh, S., Sharma, R., Thakur, A. and Gupta, D., 2022. Enhanced Error Level Analysis with Deep Learning for Image Forgery Detection. *arXiv preprint*. Available at: https://arxiv.org/abs/2211.15196