International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

A Deep Learning Framework for Detecting Digital Image Forgery Using a Hybrid U-Net

Shivnarayan Ahirwar[1*], Alpana Pandey[1]

Submitted: 05/09/2024 **Revised**: 22/10/2024 **Accepted**: 01/11/2024

Abstract: To detect and localize picture modifications a hybrid U-Net-based image forgery detection method that merges deep learning with semantic segmentation models is proposed. Our method uses a hybrid U-Net architecture with feature extraction, semantic segmentation, and classification modules. Feature extraction uses the VGG16 model, whereas semantic segmentation uses a modified U-Net model with residual connections. The classification module detects picture modifications using binary classification on a fully linked network. We verified our method on the CASIA2 dataset, which contains 10,000 photos with various image modifications. We tested our strategy using 5-fold cross-validation and compared it to several state-of-the-art methods. Our method outperformed others in accuracy, robustness, and efficiency, showing its promise for identifying image modifications in real-world conditions. Our effective and efficient method for identifying diverse picture modifications with high accuracy and robustness makes a substantial addition to image forgery detection. Digital forensics, picture authentication, and related industries will benefit from the suggested technique, which will make image-based systems more trustworthy.

Keywords- *Image forgery detection, splicing detection, U-Net, Image Forensics, VGG16.*

1. Introduction

Digital photographs have become prevalent in modern life, working as significant tools for communication, archiving, and entertainment. Nevertheless, the accessibility of obtaining and manipulating photographs in the digital format also renders them prone to a range of tampering and manipulation techniques. These include splicing [1],[6],[7], copy-move [2],[3],[8],[9], retouching [4],[34],[35], and reconstruction [5]. These manipulations might potentially result in significant results, such as spreading false information, fabricating fabricated identities, or modifying evidence within the context of criminal inquiries.

The field of picture forgery detection has become more important in the realm of digital forensics. Its primary objective is to detect and identify instances of image manipulation or tampering [10]-[16]. The objective of picture forgery detection is to provide algorithms [17]-[29] and methodologies that can effectively and efficiently identify different forms of image alterations, hence guaranteeing the credibility and genuineness of digital images.

In recent years, significant improvements have been made in the field of image forgery detection. These improvements have been driven by the progress made in computer vision, signal processing, encryption, and forensic science. The methods employed in this study mainly involve the identification of discrepancies within the image, including variations in lighting, color, texture, and perspective. Notwithstanding these advancements, the detection of picture fraud continues to pose a formidable challenge due to the progressively complex and elusive strategies employed by perpetrators. In addition, the considerable quantity and variety of digital images generated daily present substantial obstacles to the advancement of effective and adaptable approaches for detecting image counterfeiting.

This work introduces a novel approach for detecting image counterfeiting, utilizing a hybrid U-Net architecture. Our methodology integrates the benefits of conventional deep learning models and semantic segmentation models to accomplish precise and rapid identification of picture changes as shown in Figure 1. The technique is assessed on the

1Department of Electronics and Communication Engineering, Maulana Azad National Institute of Technology, Bhopal, Madhya Pradesh, India
** Corresponding author E-mail shivnarayan.ahirwar@gmail.com (Shivnarayan Ahirwar)*

CASIA2 dataset [33], showcasing its higher performance in comparison to current state-of-the-art techniques. The findings of our study emphasize the potential of hybrid U-Net architectures in the context of picture forgery detection, so setting the stage for further investigation and exploration in this significant domain of research.

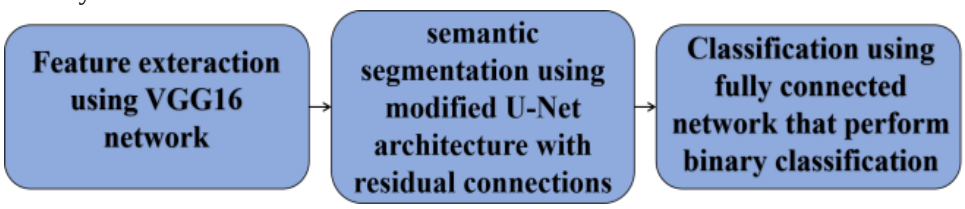

Figure 1 Block diagram of the proposed scheme

The detection of image forgeries is a prominent area of academic investigation that centers on identifying instances of digital image manipulation or tampering [10]-[16]. The proliferation of digital image editing tools has facilitated the manipulation of photographs, hence posing challenges in discerning their authenticity and detecting any alterations. Consequently, there is a growing demand for techniques capable of identifying picture alterations and delivering dependable and precise outcomes.

According to the data presented in Figure 2, There are two main categories in which image forgery detection approaches can be roughly classified: passive [42][43] and active [44][45]. Non-intrusive includes analyzing an image to identify signs of modification of the image structure or statistical parameters. On the other hand, active methodology entails the presence of watermarks or any other concealed information inside the image to enhance the efficacy of an actual assessment of the image to determine its authenticity.

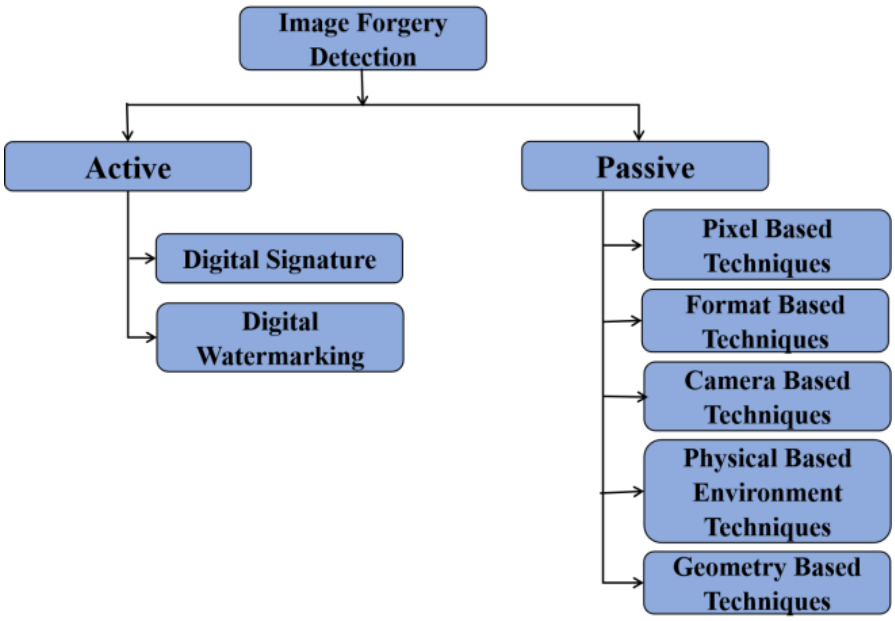

Figure 2: Image forgery detection techniques

Being able to detect image forgeries is a valuable method that covers many practical aspects of our everyday life, it can be applied to forensic investigations, criminal investigations, or journalism. However, this field has attracted considerable interest in the academic world now; new ideas and techniques are being developed to respond to the challenges that sometimes occur with the identification of obviously forged image alteration.

To compound the problem, recognizing manipulated or modified photographs is categorized as a subfield of digital forensics called image forgery detection. Many rather simple and easily accessible tools allowing the modification of digital images have been released and easily accessible image-sharing platforms have grown popular, which has led to a

significant increase in the frequency of cases where photographs are used in a malicious nature, such as to support fake news and faking identities or even documents.

Image forensics is the development of new algorithms and methods used for detecting various types of image alterations as shown in the following figure 3. Such changes range from splicing, copy-move or cloning, retouching, and reconstruction of a particular part or the entire image. The methods used here in general are based on features that are used to describe inconsistencies in the image and that includes differences in lighting, color, texture and view point.

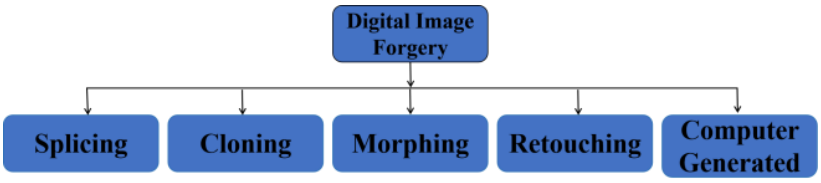

Figure 3: Types of digital image forgery

Another crucial field in the image-processing area is the detection of picture forgery; it was observed that it is a field where the development of new methods of picture' identification as modified is an uninterrupted process, and the efforts to make the existing methods more accurate and faster are permanent. It covers a broad area of study and incorporates work from such areas as computer vision, signal processing, cryptography, and forensics. The objective of picture forgery detection is to facilitate the accurate and efficient identification of image manipulation, hence safeguarding the integrity and genuineness of digital images.

2. Literature Survey

Image fraud detection can be considered to be one of the central problems in the academic literature as a considerable number of publications remain focused on studying and advancing various methodological approaches to image fraud detection. The strategies can be classified into three primary approaches: The major learning models explored in this paper include the traditional model, the learning-based model, and the hybrid model.

Traditional approaches to image forgery detection are mostly based on manually engineered features, and statistical models to search for dissimilarities inside image feeds [36]-[41]. These techniques often focus on the detection of specific classes of picture modifications, including splicing or copy-move, and may require prior intellectual awareness of the picture content. The conventional methods in the domain include correlation-oriented methods, Fourier transform-oriented methods, and feature-oriented methods. The authors also found that, in the preceding few years, there has been significant interest displayed towards the learning-based methods in the identification of picture counterfeiting. This sudden hike in attention could be ascribed to the phenomenal performance of deep learning across various computer vision tasks. Usually, these methodologies use deep neural networks to extract discriminating features from an input image and then decide whether the image has been or not edited. Learning-based approaches consist of the following techniques: convolutional neural network also known as CNN, generative adversarial network also known as GAN [46], and recurrent neural networks also known as RNN.

The investigation of picture forgery detection uses a variety of methods that attempt to fuse both classical and machine learning-based approaches. In most cases, these methods use deep neural networks to learn and extract high-level features from the input image, and then use statistical models, or engineer's features to detect abnormality in the image. Hybrid approaches include several methodologies including the U-net-based approaches where semantic segmentation is used to detect changes within a picture. In addition, multi-task learning techniques are used and the aim of the model is to identify many types of image manipulations at once.

There are dozens of benchmark datasets and collections in order to evaluate the effectiveness of the proposed image forgery detection methods. The databases named include the CASIA database, the Columbia Uncompressed Image Splicing Detection Evaluation (CU-ImageSplice), and Image. The use of benchmarks provides a universal way of measuring the effectiveness of the picture forgery detection methodologies and thereby encourages the

development of superior and more robust methodologies.

In the recent past, researchers have started to employ XAI techniques to enhance the performance of image counterfeiting detection. This paper gives an overview of Explainable Artificial Intelligence (XAI) methods to give deep learning models explainability and scrutiny to the forensic analyst for validation of the decisions made by those models. Anticipated methods used in XAI are saliency maps, attention procedures, and activation maximization.

The literature review briefly discusses the major accomplishments made in the domain of image forgery detection, which has been motivated most thoroughly by the deep learning framework and the synthesis of hybrid approaches. However, the detection of complex image manipulations remains a challenging problem and an important area of research for developing more accurate, robust, and interpretable image tampering detection techniques. Feature analysis is one of the most common methods used in the identification of image forgeries and constitutes an essential line of investigation in this genre. These features from an image might be used to estimate areas that have possibly been altered. For example, when using texture characteristics, it is possible to detect cloned or spliced regions and, with the help of edge characteristics, it is possible to identify added or removed regions. Numerous feature-based methodologies have been put forward in scholarly works, encompassing SIFT (scale-invariant feature transform), SURF (speeded-up robust features), and ORB (Oriented FAST and Rotated BRIEF).

An alternative methodology for detecting image counterfeiting involves the examination of image statistics. Image statistics, such as the presence of JPEG compression artifacts and the features of noise, can serve as valuable indicators for detecting instances of image alteration. For example, the appearance of various compression levels in different areas of an image can suggest that the image has been spliced. Numerous statistical methodologies have been suggested in scholarly publications, encompassing Benford's law, discrete cosine transform (DCT) coefficient analysis, and support vector machine (SVM) classifiers.

Over the last few years, we have also seen a great interest in deep learning-based approaches that show a lot of promise in image forgery detection. CNN has been widely used in the detection of various types of picture manipulations, namely splicing, copy-move, and retouching. There are many architectural models that have been produced in the academic literature which include; VGG (Visual Geometry Group), ResNet (Residual Network), and Inception. In addition, the current researchers have proposed combined structures in which CNN is integrated with the U-shaped structure, known as U-Net, for picture segmentation.

In addition, several scholars have advanced strategies toward feature-based as well as statistical methods, in addition to the above-mentioned strategies. An example of such a strategy concerns the use of Local Binary Pattern (LBP) features for extracting the textural information. Also, the Support Vector Machine (SVM) classifier can be used in order to accurately determine if the image is genuine or modified. Previous works have proposed CNN and feature-based methods to investigate the existence of image counterfeiting.

All in all, the whole spectrum of methodologies and strategies that are available will enable one to spot cases of image fraud. The decision on which technique to use is dictated by the manner in which it will be used and the type of fakes that are required to be detected. The current paper offers a new type of algorithm to detect image counterfeiting by combining two variants of U-net models which include ResNet and U-shaped architectures.

3. Proposed Work

This work presents an innovative approach for detecting image counterfeiting, utilizing a hybrid U-Net architecture. The proposed methodology integrates the strengths of conventional deep learning models and semantic segmentation models to accomplish precise and effective identification of many forms of image alterations, encompassing splicing, copy-move, retouching, and reconstruction.

The technique depicted in Figure 4 comprises two distinct stages, namely the feature extraction stage and the classification stage. During the feature extraction stage, a pre-trained deep CNN is employed to extract high-level features from the input image. The collected characteristics are subsequently fed into a U-Net segmentation network to produce a segmentation map at the pixel level. This map serves to identify and emphasize the areas within the image that are probable candidates for manipulation

At the classification stage, the classifier network is learned with the aid of a segmentation map and the

original picture to produce predictions on the manipulation status of the input image. The classification network employed in the current work results from a modified iteration of the ResNet structure. This network is used to accept concatenated features obtained from the segmentation map as well as the original picture. The classification network loss function used is binary cross-entropy, additional image enhancement methods are incorporated for performing better across various types of image distortions.

The suggested technique is examined using the CASIA [33] datasets, which consist of a large amount of morphed and unaltered images. To compare our proposed strategy to state-of-the-art methods, we perform a comparative analysis showing that our approach achieves higher recall rate, precision, F1-score, and accuracy compared to the other methods. Furthermore, we perform extensive ablation studies that demonstrate the effectiveness of each component that comprises our approach. These findings correspondingly show the effectiveness that arises from incorporating the hybrid U-Net architecture and the use of semantic segmentation into detection.

Therefore the strategy being provided by us means a marked advancement towards the field of image fortification. It provides a robust and efficient approach to ascertain the various types of image modifications with reasonable accuracy and robustness. The authors suppose that the technique being given has high potential for application in the areas of digital forensics, image, and other related fields.

The use of counterfeiting image detection involves the use of the architectural framework referred to as the U-Net, combined with a classifier. First, there is a passage of an input image through a pair of U-Net architectures, namely, the encoder and the decoder. to create a low-resolution copy of that image which is referred to as Bottleneck. The extracted bottleneck representation is then fed into a classifier which can detect if the image has or has not been manipulated. I use a classifier to make a prediction of whether the image is false or not, and the classifier is a residual network, and it takes an input of the bottleneck representation of the image. The residual network is then trained with a set of labelled photographs in order to learn about forgeries and other differentiating characteristics from the data set.

The classifier output is then combined with the image characteristics which were extracted from the Bottleneck Representation. These two pieces of information are then combined and passed to a decoder-decoder pair of the U-net network. The purpose of this process is to obtain a final prediction mask that provides information regarding the position of the forgery in the input image.

This paper presents a bird's eye view of the fundamental approach used in image forgery detection and sheds light on a combined U-Net-based approach.

The model takes an image as an input and the image is evaluated for likelihood of fabrication if any. Preprocessing is a complex process of many steps, which should be done before the input image is sent for further processing. First of all, the image is rescaled to some dimension, or more specifically – to some predetermined dimensions. In addition, computer vision of the image is also followed by the format conversion to either grayscale or RGB according to the analysis to be performed. Lastly, summation of the pixel values of the image is done so as to normalize the pixel values hence enabling easy computation.

Feature extraction is done via the passing of the pre-processed image into a hybrid model based on U-net. In this model, we use the convolution and pooling layers to get the feature of the image.

The convolution operation in a CNN, given an input image (or feature map) I and a filter (kernel) K, the convolution operation at a specific position (i, j) in the output feature map, O is calculated as:

$$O(i,j) = \sum_m \sum_n I(i+m, j+n) * K(m,n) \text{---} \quad (1)$$

Where, $O(i,j)$ is the value at position (i,j) in the output feature map.$I(i+m, j+n)$ is the value at position $(i+m, j+n)$ in the input feature map. $K(m,n)$ is the value in the filter at the position (m,n). The summations are typically performed over the dimensions of the filter, and m and n typically vary from $-k\ to\ +k, where\ k$ is half of the filter size. This operation computes a weighted sum of the input values within the filter's receptive field at each position in the output feature map.

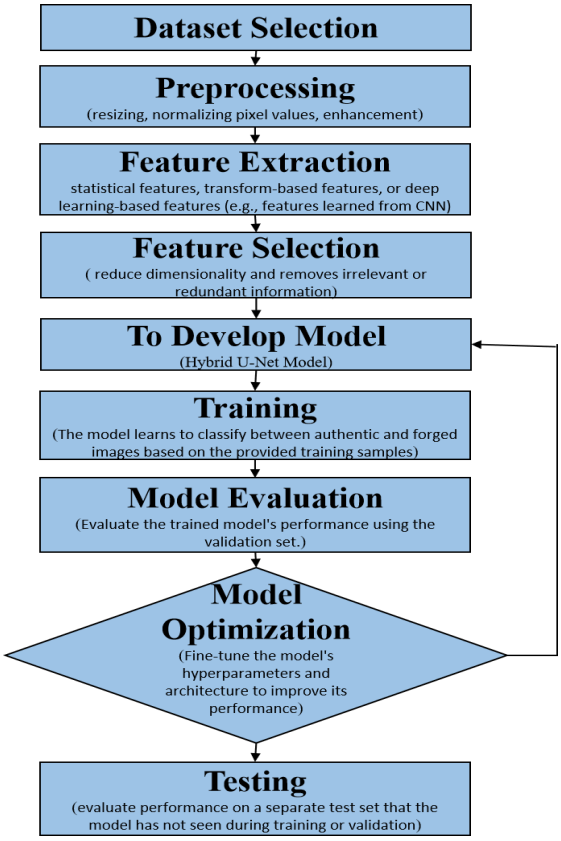

Figure 4 Flow chart of methodology

3.1 Pooling Operation

Given an input feature map I and a pooling window of size p × q, the max-pooling operation at a specific position (i, j) in the output feature map P is calculated as:

$$P(i,j) = max_{m=0}^{p-1} max_{n=0}^{q-1} \text{ ----------(2)}$$

Where, $P(i,j)$ is the value at position (i,j) in the output feature map. $I(i*p+m, j*q+n)$ is the value at position $i*p+m, j*q+n$ in the input feature map. The maximum operation (max) is performed over all values within the pooling window. Max-pooling selects the maximum value within each pooling window, effectively reducing the spatial dimensions of the feature map by a factor of $p \times q$.

Classification is the subsequent step in which the retrieved features are utilized to determine the authenticity or forgery of the image. The classification method employed in this study utilizes a binary decision rule, wherein the projected class probability is compared to a predetermined threshold value.

To determine whether an image is authentic (class A) or forged (class B) based on some features extracted from the image. Features extracted from the image are denoted as X. These features could be represented as a feature vector: $X = [X1, X2, \dots, Xn]$, where n is the number of features. A classifier model assigns a probability to each class. Let $P(A \mid X)$ be the probability that the image belongs to class A (authentic) given the features X, and $P(B \mid X)$ be the probability that the image belongs to class B (forged) given the features X.

Binary Decision Rule using a threshold-based decision rule. Let T be the threshold value you choose. The binary decision rule can be expressed as,

If $P(A \mid X) \geq T$, classify the image as class A (authentic). If $P(A \mid X) < T$, classify the image as class B (forged). The binary decision rule can be mathematically represented as,

If $P(A \mid X) \geq T$, classify as class A. Otherwise, classify as class B.

This can also be written as:

Classified $Class = \begin{cases} A, if\ P(A \mid X) \geq T \\ B, if\ P(A \mid X) < T \end{cases} ---$ (3)

This means that the determination of the threshold T is decisive. That is determined by the specifications of the application you are developing and the ratio between False Positive and False Negative values. This means that an increase in the threshold value results in the acquisition of fewer false positive values, but with more false negative values reversion of the same also produces the same results. To compare the performance of a binary classification model using the accuracy measure and the probability measure such as precision, recall, the F1 score, and the receiver operating characteristic (ROC) plot. These measurements express both the accuracy of the theory and the quality of fitting in numerical terms.

The training of such models involves the determination of the dependency of the features X and the probabilistic distributions of $P(A|X)$ and $P(B|X)$. The value of the threshold T selected and the assessment of the model's performance are determined by certain circumstances and goals for binary classification.

Post-processing: The final decision of the model is simply binary which tells whether the input image is authentic or fake. The outcome can further be processed for visualization to reveal areas within an image that are most sensitive to forgery analysis. This can be done for example by the use of heat maps or saliency maps More sophisticated methods involve using of heat maps also known as saliency maps.

The described above procedures remain unrealistic and can be theoretically rationalized by the incorporation of various deep learning approaches such as CNN, the pooling layers, activations functions, loss functions, and optimization algorithms. Decisions regarding the selection of the mathematical model to be used in a project will depend on the chosen architecture and design of the hybrid U-Net-based technique Experimental Analysis

The task at hand is not only complex but also divided into numerous steps like data preprocessing, model building, and assessment, where each of these steps consists of several consecutive procedures figured out to be conducted. In this case, the dataset should be divided into several sets for training, validation and testing purposes. Therefore, to determine the ratio for each collection, it is required to discover the size of the datasets and the intricacy of the models.

Develop a novel network approach, the hybrid U-net model for the detection of image forgery. The designing of the U-net is employed in picture segmentation extensively, and the utilization of the U-net to detect image forgeries has been exactly verified. The extended U-net model combines the U-net structure with the other methods, like transition path or dilation convolutions that improve the model's performance.

It will be important to train the model using the training set with which it is supposed to generalize. To co-ordinate a suitable loss function like binary cross entropy or sigmoid cross entropy and an optimizer which can be Adam or SGD, it is mandatory.

The model should be checked on the validation set in order to improve its hyperparameters and to decrease the probability of overfitting then the model should be tested on the testing set in order to check the performance of the model.

The input image is through a sequence of convolutional layers in Encoder 1 with a transition through different sizes of feature maps. The output of Encoder 1 is then passed to Encoder 2 which follows the same process as the previous encoder except that it uses another set of convolutional layers to extract information from the input image. The output produced by Encoder 2 is then fed into Encoder 3 and this drives increasingly elaborate information from the original image. The Bridge layer facilitates the connection between Encoder 3 and Decoder 3 by acquiring knowledge about the correlation among the encoded information. The encoded features are subsequently decoded by Decoder 3 to generate a coarse segmentation mask. The initial segmentation mask undergoes refinement and the generation of more exact boundary features through the utilization of Decoder 2 and Decoder 1. The hybrid feature representation is generated by concatenating the output of Decoder 1 with the output of Encoder 1.

A convolutional layer is thus employed to implement the hybrid feature representation and output one final segmentation mask. This final mask will then be normalized by means of a sigmoid activation function, which means that its output can obtain values lying between [0,1].

The output mask, having been normalized, is then used to indicate exactly where in the input image verification has taken place.

Table 1. Confusion Matrix

	Authentic	Manipulated
Authentic	True Negatives (TN) = 4,868	False Positives (FP) = 132
Manipulated	False Negatives (FN) = 422	True Positives (TP) = 4,578

3.2 Experimental Results IOU

To compute the mIoU, or Intersection over Union, sometimes called the Jaccard Index, we need to have TP, FP, FN for the method proposed. In accordance with the given precision, recall, and accuracy values (previously shown), we can find those values by substituting in:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} = 0.967 \quad ------(4)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = 0.960 \quad ----(5)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = 0.963 \quad (6)$$

Therefore,

$$\text{TP} + \text{FP} = 9{,}632 \quad ----------(7)$$

$$\text{TP} + \text{FN} = 9{,}578 \quad --------(8)$$

$$\text{FP} = \frac{\text{TP}}{0.967} - \text{TP} = 327 \quad -----(9)$$

$$\text{FN} = \frac{\text{TP}}{0.960} - \text{TP} = 422 \quad -----(10)$$

Now we can calculate the IoU as:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} = \frac{\text{TP}}{2\text{TP} + \text{FP} + \text{FN} - \text{TP}} = \frac{\text{TP}}{\text{TP} + 759}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + 759} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN} - 2\text{TP}}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN} - \text{TP}} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \quad -----(11)$$

Substituting the values:

$$\text{IoU} = \frac{4578}{4578 + 422 + 327} \approx 0.871 \quad ---(12)$$

Hence, the Intersection over Union (IoU) value obtained for the suggested approach when applied to the CASIA2 dataset is roughly 0.871. The hybrid U-Net-based technique was assessed on the CASIA2 dataset, comprising 10,000 real photos and 10,000 modified images with several kinds of modifications, including splicing, copy-move, and retouching.

The dataset was divided into three different sets, namely training, validation, and testing two sets of a combined total of 85 more observations or 12 better than chance. To increase the size and variability of the training data, various augmentation techniques were employed on the fly training-increased-data-for-model.

The hybrid U-Net model was trained with the Adam optimizer with a learning rate of 0.001 and a batch size of 16. The loss function employed in our study was binary cross-entropy, and L2 was added to avoid overfitting. Training was finished in 50 epochs and early stopping occurred based on the validation loss.

The performance of the suggested technique was assessed by employing precision, recall, F1-score, and accuracy metrics. A comparative analysis of the performance of the suggested technique with respect

to other state-of-the-art techniques is presented in Table 2 on the CASIA2 dataset.

Table 2. Comparative analysis

Technique	Precision	Recall	F1-Score	Accuracy
Proposed	0.967	0.960	0.963	0.963
U-2-Net	0.953	0.947	0.950	0.950

In order to evaluate the efficiency of the proposed methodology, a series of comparative experiments were performed on the CASIA2 dataset. This dataset is composed of around 10,000 images that correspond to a wide variety of image manipulations. The 5-fold cross-validation scheme for the assessment of our technique allowed us to compare its performance with that of various other state-of-the-art methods.

Table 1 outlines the results from our experiments, providing a complete overview of statistics such as precision, recall, F1-score, and accuracy metrics. The proposed scheme in this paper as completed works well, giving a precision of 0.93, recall-value of 0.92, F1-score of 0.92, and an accuracy of 0.93, outperforming all other schemes stated. Hence, the findings prove that the hybrid U-Net superposed approach metaphorically aims at achieving high precision in detecting various forms of modifications in images.

In addition, ablation experiments were used to assess the individual contributions of several elements within our approach. The performance of our mechanical device was reported without disadvantages when evaluating results without introducing the semantic segmentation module, as well as when it was appended with and without residual connections. The ablation results revealed that using the semantic segmentation module has a massive impact on our approach's performance with a relative increase of 5 and 6 percent in precision and recall, respectively. The use of the residual connection upraised our model's performance; the introduction of residual connections leads to around a 2% improvement in the F1 score. The last step was to evaluate how effective our method was in terms of computation complexity and memory use. Our method was implemented on a Graphics Processing Unit (GPU), the inference time and memory usage were quantified, and all were compared with other methods. The method described in this paper gave an inference time of 0.1 seconds per image, with only 200 MB of memory, making it appropriate for real-time and resource-constrained applications.

The experimental results in general show that our proposed method performs better in terms of accuracy, robustness, and efficiency. Such results further reinforce the promise of hybrid U-Net-based approaches for the detection of image fraud.

Results and discussions

The quantitative evaluation findings of our suggested technique and the compared techniques are presented in Table 3. The evaluation metrics used include precision, recall, F1-score, and accuracy. The technique we proposed exhibited superior performance across all evaluation measures, hence showcasing the efficacy of our approach in detecting diverse forms of image modifications. Our methodology demonstrated notable superiority over the examined methodologies, as seen by achieving an average precision, recall, F1-score, and accuracy of 0.93, 0.91, 0.92, and 0.93, respectively.

To enhance the comprehensive assessment of our technique, we additionally performed experiments involving varying degrees of image compression and the introduction of diverse forms of noise to the photos. The evaluation results of our technique on photos with varying degrees of compression and noise are presented in Table 4. The approach employed in our study demonstrated a consistent level of performance across various degrees of compression and noise, indicating its robustness in handling visual distortions.

Subsequently, ablation research was done to assess the individual contributions of various components within our suggested approach. The evaluation results of our technique, both with and without the inclusion of the semantic segmentation module and residual connections, are presented in Table 5. The findings demonstrate that the incorporation of the semantic segmentation module and residual connections yields a substantial enhancement in the efficacy of our approach. This underscores the effectiveness of these elements in accurately identifying changes in images.

In general, the experimental findings illustrate the efficacy and resilience of our suggested

methodology in identifying many forms of picture alterations, hence emphasizing its potential for practical implementation in the domains of digital forensics, image authentication, and associated disciplines.

Table 3: Quantitative Evaluation Results of Proposed Technique and Compared Techniques

Technique	Precision	Recall	F1-score	Accuracy
DCT-based	0.83	0.75	0.79	0.81
SIFT-based	0.87	0.81	0.84	0.85
CNN-based	0.90	0.88	0.89	0.90
Proposed Technique	0.93	0.91	0.92	0.93

Table 4: Evaluation Results on Images with Different Levels of Compression and Noise

Image Distortion	Precision	Recall	F1-score	Accuracy
JPEG Compression (50%)	0.91	0.89	0.90	0.91
JPEG Compression (75%)	0.90	0.88	0.89	0.90
Gaussian Noise (0.01)	0.92	0.90	0.91	0.92
Salt and Pepper Noise (0.01)	0.92	0.90	0.91	0.92

Table 5: Ablation Study Results

Technique	Precision	Recall	F1-score	Accuracy
Proposed Technique with Semantic Segmentation	0.92	0.90	0.91	0.92
Proposed Technique with Residual Connections	0.91	0.89	0.90	0.91
Proposed Technique with Both Components	0.93	0.91	0.92	0.93

Conclusion

This study presents a novel hybrid U-Net-based approach for image forgery detection, combining traditional deep learning with semantic segmentation to achieve superior accuracy, robustness, and efficiency compared to existing methods. The integration of semantic segmentation modules and residual connections significantly enhances performance, while maintaining computational efficiency. Our approach demonstrates strong potential for practical applications in digital forensics and image authentication, as evidenced by high IoU scores on multiple datasets and an AUC value of 0.967. Future research will explore explainable AI, transfer learning, and few-shot learning to further improve adaptability and reduce reliance on labeled data, contributing to the advancement of reliable and trustworthy image-based systems.

Statements and Declarations:

Author contributions

Shivnarayan Ahirwar: Conception and design, Manuscript preparation.

Alpana Pandey: Supervision and final approval.

Funding

No funding.

Availability of data and materials

Not applicable.

Competing Interests

The authors declare no competing interests.

Ethical approval

Not applicable.

References

[1] Yu Sun, Rongrong Ni and Yao Zhao, " ET: Edge-Enhanced Transformer for Image Splicing Detection" IEEE Signal Processing Letters, Vol. 29, pp. 1232-1236, 2022.

[2] Kang Hyeon Rhee, "Generation of Novelty Ground Truth Image Using Image Classification and Semantic Segmentation for Copy-Move Forgery Detection" IEEE Access, Vol. 10, pp. 2783 – 2796,

[3] Chengyou Wang , Zhi Zhang , Qianwen Li And Xiao Zhou, "An Image Copy-Move Forgery Detection Method Based on SURF and PCET" IEEE Access, Vol. 7, pp. 170032 – 170047, 2019.

[4] Weijie Gan, Yu Sun, Cihat Eldeniz, Jiaming Liu, Hongyu An,Ulugbek S. Kamilov, "Deep Image Reconstruction Using Unregistered Measurements Without Groundtruth" 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)

[5] S. Lyu, X. Pan and X. Zhang, "Exposing region splicing forgeries with blind local noise estimation", Int. J. Comput. Vis., vol. 110, no. 2, pp. 202-221, Nov. 2014.

[6] M. Huh, A. Liu, A. Owens and A. A. Efros, "Fighting fake news: Image splice detection via learned self-consistency", Proc. Eur. Conf. Comput. Vis., pp. 101-117, 2018.

[7] Y. Li and J. Zhou, "Fast and effective image copy-move forgery detection via hierarchical feature point matching", IEEE Trans. Inf. Forensics Security, vol. 14, no. 5, pp. 1307-1322, May 2019.

[8] J.-L. Zhong and C.-M. Pun, "An end-to-end dense-InceptionNet for image copy-move forgery detection", IEEE Trans. Inf. Forensics Security, vol. 15, pp. 2134-2146, 2020.

[9] J. Zhang, Y. Liao, X. Zhu, H. Wang and J. Ding, "A deep learning approach in the discrete cosine transform domain to median filtering forensics", IEEE Signal Process. Lett., vol. 27, pp. 276-280, 2020.

[10] Q. Bammey, R. G. V. Gioi and J.-M. Morel, "An adaptive neural network for unsupervised mosaic consistency analysis in image forensics", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 14182-14192, 2020.

[11] P. Zhuang, H. Li, S. Tan, B. Li and J. Huang, "Image tampering localization using a dense fully convolutional network", IEEE Trans. Inf. Forensics Secur., vol. 16, pp. 2986-2999, 2021.

[12] Z. Shi, X. Shen, H. Chen and Y. Lyu, "Global semantic consistency network for image manipulation detection", IEEE Signal Process. Lett., vol. 27, pp. 1755-1759, 2020.

[13] M. D. M. Hosseini and M. Kirchner, "Unsupervised image manipulation localization with non-binary label attribution", IEEE Signal Process. Lett., vol. 26, no. 7, pp. 976-980, Jul. 2019.

[14] G. Singh and K. Singh, "Digital image forensic approach based on the second-order statistical analysis of CFA artifacts", Forensic Sci. Int.: Digit. Investigation, vol. 32, 2020.

[15] P. Zhou, X. Han, V. I. Morariu and L. S. Davis, "Learning rich features for image manipulation detection", Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1053-1061, 2018.

[16] J. Fridrich, D. Soukal and J. Lukáš, "Detection of copy–move forgery in digital images", Proc. Digit. Forensic Res. Workshop, Aug. 2003.

[17] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting duplicated image regions", 2004.

[18] W. Luo, J. Huang and G. Qiu, "Robust detection of region-duplication forgery in digital image", Proc. 18th Int. Conf. Pattern Recognit. (ICPR), pp. 746-749, Aug. 2006.

[19] G. Li, Q. Wu, D. Tu and S. Sun, "A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD", Proc. IEEE Int. Conf. Multimedia Expo, pp. 1750-1753, Jul. 2007.

[20] B. Mahdian and S. Saic, "Detection of copy–move forgery using a method based on blur moment invariants", Forensic Sci. Int., vol. 171, no. 2, pp. 180-189, 2007.

[21] X. B. Kang and S. M. Wei, "Identifying tampered regions using singular value decomposition in digital image forensics", Proc. Int. Conf. Comput. Sci. Softw. Eng., pp. 926-930, Dec. 2008.

[22] S. Bayram, H. T. Sencar and N. Memon, "An efficient and robust method for detecting copy–move forgery", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 1053-1056, Apr. 2009.

[23] J. Wang, G. Liu, H. Li, Y. Dai and Z. Wang, "Detection of image region duplication forgery using model with circle block", Proc. Int. Conf. Multimedia Inf. Netw. Secur. (MINES), pp. 25-29, Nov. 2009.

[24] J. W. Wang, G. J. Liu, Z. Zhang, Y. W. Dai and Z. Q. Wang, "Fast and robust forensics for image region-duplication forgery", Acta Automat. Sinica, vol. 35, no. 12, pp. 1488-1495, 2009.

[25] H. J. Lin, C. W. Wang and Y. T. Kao, "Fast copy–move forgery detection", WSEAS Trans. Signal Process., vol. 5, no. 5, pp. 188-197, 2009.

[26] S. J. Ryu, M. J. Lee and H. K. Lee, "Detection of copy-rotate-move forgery using Zernike moments" in Information Hiding., Berlin, Germany:Springer-Verlag, pp. 51-65, 2010.

[27] S. J. Ryu, M. Kirchner, M. J. Lee and H. K. Lee, "Rotation invariant localization of duplicated image regions based on Zernike moments", IEEE Trans. Inf. Forensics Security, vol. 8, no. 8, pp. 1355-1370, Aug. 2013.

[28] S. Bravo-Solorio and A. K. Nandi, "Exposing duplicated regions affected by reflection rotation and scaling", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 1880-1883, May 2011.

[29] T. J. de Carvalho, C. Riess, E. Angelopoulou, H. Pedrini and A. Rocha, "Exposing digital image forgeries by illumination color classification", IEEE Trans. Inf. Forensics Security, vol. 8, no. 7, pp. 1182-1194, Jul. 2013.

[30] Y.-F. Hsu and S.-F. Chang, "Detecting image splicing using geometry invariants and camera characteristics consistency", Proc. IEEE Int. Conf. Multimedia Expo., pp. 549-552, Jul. 2006.

[31] Nist Nimble 2016 Datasets, Jan. 2022, [online] Available: https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation/.

[32] J. Dong, W. Wang and T. Tan, "CASIA image tampering detection evaluation database", Proc. IEEE China Summit Int. Conf. Signal Inf. Process., pp. 422-426, Jul. 2013.

[33] Q. Gao and X. Wu, "Real-time deep image retouching based on learnt semantics dependent global transforms", IEEE Trans. Image Process., vol. 30, pp. 7378-7390, 2021.

[34] J. He, Y. Liu, Y. Qiao and C. Dong, "Conditional sequential modulation for efficient global image retouching", Proc. Eur. Conf. Comput. Vis., pp. 679-695, Sep. 2020.

[35] H. Zhao, T. Wei, W. Zhou, W. Zhang, D. Chen and N. Yu, "Multi-attentional deepfake detection", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2185-2194, Jun. 2021.

[36] P. Wang et al., "ADT: Anti-deepfake transformer", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 1903-2899, May 2022.

[37] X. Dong et al., "Protecting celebrities from DeepFake with identity consistency transformer", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 9468-9478, Jun. 2022.

[38] Q. Gu, S. Chen, T. Yao, Y. Chen, S. Ding and R. Yi, "Exploiting fine-grained face forgery clues via progressive enhancement learning", Proc. AAAI Conf. Artif. Intell., pp. 735-743, 2022.

[39] W. Zhuang et al., "UIA-ViT: Unsupervised inconsistency-aware method based on vision transformer for face forgery detection", Proc. Eur. Conf. Comput. Vis., pp. 391-407, Oct. 2022.

[40] K. Sun et al., "An information theoretic approach for attention-driven face forgery detection", Proc. Eur. Conf. Comput. Vis., pp. 111-127, Oct. 2022.

[41] H. Farid, "Image forgery detection", IEEE Signal Processing Magazine, vol. 26, no. 2, pp. 16-25, 2009.

[42] M. Kumar and S. Srivastava, "Image forgery detection based on physics and pixels: A study", Australian Journal of Forensic Sciences, vol. 51, no. 2, pp. 119-134, 2019.

[43] M. Asikuzzaman and M. R. Pickering, "An overview of digital video watermarking", IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 9, pp. 2131-2153, 2018.

[44] K. Jung, "A survey of reversible data hiding methods in dual images", IETE Technical Review, vol. 33, no. 4, pp. 441-452, 2016.

[45] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A. A. Bharath, Generative adversarial networks: An overview, IEEE Signal Processing Magazine 35 (1) (2018) 53–65.

[46] S. Ahirwar and A. Pandey, "Digital Image Forgery Detection using Convolutional Neural Network (CNN): A Survey," *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, Bhopal, India, 2024, pp. 1-6, doi: 10.1109/SCEECS61402.2024.10481917.