JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

VISION LEARNER BASED IMAGE TAMPERING DETECTION AND LOCALIZATION

[1]Anil Kumar MG , [2]Divya M, [3]Nagaratna Naik, [4]Nandana CL, [5]Sujatha R Upadhyaya

[1]Student, [2]Student, [3]Student, [4]Student, [5]Professor
[1]Department of Computer Science and Engineering,
[1]PES University, Bengaluru, India

Abstract : Image tampering is a prevalent issue in the digital age, where images and visual content are manipulated for malicious purposes. In the past, there have been multiple attempts to detect tampered images. Most of these methods focus on copy-move or image splicing, which are recognized as two major types of tampering techniques. While most of such efforts use machine learning as their primary technique for tamper detection, this paper presents a comprehensive approach to address both types of image tampering detection using Vision Learner, a deep learning technique. The system classifies images as authentic or tampered and also localizes the tampered regions with a high degree of accuracy. This research uses UNET, a well-known CNN architecture, for image splicing localization and DCT for copy-move localization. A passive image tampering detection technique is used to classify and localize image tampering based on the inconsistencies within the image. The CASIA dataset, consisting of more than 12600 generic images, has been used to demonstrate the experiments and results.

IndexTerms **- Image Tampering, CNN, Vision Learner, Localization, Image Splicing, Copy-Move Image, UNet, DCT.**

I. INTRODUCTION

With the growing number of social media and other online platforms that encourage people to share images, the issue of image tampering has become very common. The ease with which digital photographs can be altered has simply fuelled such cases. Initially, basic tools like Adobe Photoshop were extensively used for tampering images. However, as digital forensics capabilities improved more complex tampering techniques have emerged.

Image tampering detection systems are crucial in various fields like content verification, forensics, journalism, and legal proceedings, as they ensure the integrity and validity of visual information. There are different types of tampering, like image splicing, copy-move, and removal. Image splicing is a type of tampering where some regions of an image are copied and then pasted onto another image. Copy-move tampering is a type where some regions of an image are copied and then pasted within the same image only. Removal is a type of tampering where some regions of an image are removed. In addition, localization of image splicing and copy-move tampering are also detected. Some progress has been made in this field, such as the use of Ycbcr for preprocessing, the use of a Steganalysis Rich Model (SRM) filter to produce noise maps, the use of Local Binary Patterns (LBPs) on individual streams of Ycbcr, Region-based Convolutional Neural Networks (RCNN) for detection and localization of tampering, and Support Vector Machine (SVM) for the classification of the image as authentic or tampered.

There are two types of tampering detection techniques: active and passive. Active image tampering detection involves adding additional information or features during image capturing and the image creation to help image tampering detection. Some examples of active image tampering detection techniques are watermarking, digital signatures, and fingerprinting. Watermarking means adding a unique identifier or signature to the image. In digital signatures, various cryptographic techniques are employed to generate a unique signature for an image based on its content. Then, if any changes are made to the image, it will result in a different signature, using which one can detect if tampering has occurred or not. Fingerprinting is similar to digital signatures, but it involves embedding a unique identifier into the image. Then it can be used to trace the origin of an image and detect an unauthorized alteration. Active methods are more proactive but require the cooperation of image creators to embed additional information.

Passive image tampering detection involves analyzing the image content without additional embedded information like watermarking, digital signatures, or fingerprinting. Passive methods mainly depend on detecting inconsistencies, anomalies, or statistical irregularities in the image data. Forensic analysis techniques, such as examining the noise patterns, color inconsistencies, and compression artifacts, can be used to identify tampered regions. Image forensics algorithms may analyze the image metadata, such as EXIF data, to identify inconsistencies or timestamps that do not match the expected properties of an authentic image. Some examples of passive techniques used for detecting image tampering are noise analysis, JPEG compression analysis, color inconsistency detection, blind image forensics, edge inconsistency detection, texture analysis, metadata examination, etc. Passive methods are more widely applicable but may be less robust compared to sophisticated tampering techniques. These passive

techniques are often used in combination with or complemented by machine learning algorithms to enhance the accuracy of image tampering detection. Advanced methods, including deep learning approaches, can automate the process of identifying subtle inconsistencies in large datasets. In this paper, a passive approach to detect and localize tampered regions has been used.

Localization refers to the process of identifying and highlighting the specific regions within an image that has been tampered. UNet is a deep learning technique used for image splicing localization, employing a U-shaped architecture to segment and localize spliced regions within an image. Discrete Cosine Transformation (DCT) is a technique used in copy-move localization that converts image data into frequency components, enabling the identification of duplicated regions. By highlighting the location where tampering has occurred, the Image Tampering Detection System prevents the misleading of image tampering and ensures the safety and reliability of digital images.

The objective of this research is to contribute to the field of image forensics by developing a robust image tampering detection system. The system goes beyond binary classification and provides detailed localization of tampered regions, leveraging the power of deep learning.

The contributions of this research are:

1. A methodology to detect 'splice' and 'copy-move' types of tampering using Vision Learner.
2. A methodology to highlight the tampered regions within the image.
3. Experimental validation for the above approaches and performance testing.

In Section II, a background of the research work is presented with a discussion on contemporary research work. In Section III, the approach adopted in this research to achieve the objective of detecting and localizing tampered images is presented. Section IV explains the experimental setup used. In Section V, a discussion of the results of the experimentation is elaborated. Lastly, section VI presents the conclusions and future directions of the research.

II. Related Work

There has been research on various methods for the detection of image tampering. One common approach involves the extraction of features from RGB images and the use of a Convolutional Neural Network (CNN) for classification. Zhou et al. [1] propose a 2-stream faster RCNN network, extracting features from the RGB stream. Then the SRM filter is used to derive noise feature maps from RGB images followed by the selection of features from both streams through a Region of Interest (RoI) pooling layer. Spatial co-occurrence features are then combined to determine whether a predicted area has been tampered with. When an image is subjected to hybrid post-processing transformations, detecting tampered regions, localizing them, and performing segmentation become challenging tasks. Therefore, Shivanandappa et al. [2] present an Improved Convolutional Neural Network (ICNN) model. The goal of this model is to attain a strong correlation between neighboring pixels, which is something that previous models frequently fall short of, and influences segmentation results. In addition to the vertical and horizontal layers, the TDS-ICNN adds a third layer known as the correlation layer. As a result, even in the face of small-smooth post-processing tampering efforts, it successfully localizes and segments tampered locations.

The methods proposed by Thakur et al.[3] and Manu et al.[4] focus only on image splicing tampering detection. Thakur et al.[3] proposed an approach which initiates image decomposition using Discrete Wavelet Transform (DWT) that will employ an input image to minimize the image size representation. This helps identify the inconsistency between two edge coefficients. After DWT, Speed-Up Robust Features (SURF) is applied on calculated coefficients which extracts specific image features that help to determine the location as well as orientation and scale parameters of the spliced region of an image. Classification of images as Authentic or Forged is done by SVM with the help of extracted features. If it is classified as forged then the spliced region is also detected. Until now all the approaches used RGB images but Manu et al.[4] suggest converting the RGB images to YCbCr to reveal image statistics. After conversion, two methods are proposed for image tampering detection. In the first method, texture descriptors of the image are used along with pattern histograms. In the second method, texture descriptors, entropy histograms, and measurement of image quality artifacts are combined. Subsequently, both methods use SVM to classify the image as tampered with or not and to identify tampered boundaries for image splicing. Using the traditional UNET as a backbone, Aminu et al. [5] present the Deep Residual UNET with Stacked Dilated Convolution, a network designed for the detection and localization of tampered images. For the encoder path and the decoder path, two different kinds of residual units are used. Dilated convolutions are employed in the residual units to increase the convolutional kernels' receptive field size. The use of residual units speeds up training and facilitates information propagation between lower and higher layers. Then the trained model is used for prediction.

Instead of converting RGB images to YCbCr, some methods use a different approach by converting them to Error Level Analysis (ELA) images. Chakraborty et al.[6], Madake et al.[7] propose the conversion of RGB images to ELA images. Chakraborty et al.[6] further proposed the use of noise residual images which can be extracted using the SRM filter. This technique calculates the statistics required to extract specific features from the noise residuals surrounding a pixel. Changes made to an image's details will have an impact on the related residuals because the residuals are tied to those details. Subsequently, a dual-branch convolutional neural network is employed to differentiate between authentic and forged. Madake et al.[7] further proposes the use of metadata analysis along with ELA. Then CNN is used for classification.

A technique is proposed by Wang et al.[8] to locate altered regions in a losslessly compressed altered image when the JPEG decompressor's output contains its unaltered region. The important realization is that the way the altered and unaltered regions react to JPEG compression is different. In particular, the altered area exhibits more intense high-frequency quantization noise in contrast to the unaltered area. For altered region localization, the suggested method uses Principal Component Analysis (PCA) to extract high-frequency quantization noise and segregate various spatial frequency quantization noises. In Post-processing, high-frequency noise is normalized using morphology operations to obtain the final localization result .

Some approaches only detect copy-move tampering. One of them is proposed by Elaskily et al.[9]. The authors suggest an algorithm that operates on the entire image and is not block-based. The images are preprocessed, and then feature extraction and classification are done by CNN. Next, utilizing DCT, four square mean features, and stationary wavelet transform (SWT), Pugar et al.[10] presented a copy-move detection algorithm. Before using the chosen channel to break down into four subbands using the SWT, the input image is first transformed into the YCbCr color space. After choosing the approximate subband, it is split up into overlapping blocks. Then, DCT is employed for all overlapping blocks. Subsequently, it marks the regions deemed duplicated

regions to produce the required output image. Further Singh et al.[11] and Warbhe et al.[12] provide an active method to spot copy-move tampering in the photos. A technique proposed by Singh et al.[11] divides the image into overlapped patches to detect tampering in BMP images. The forged area's correlation coefficients are then compared to the original image's correlation coefficients. Next, the algorithm's efficiency is calculated for a range of different mask sizes. Warbhe et al.[12] proposed a Normalized Cross-Correlation. It has three primary processes: determining the rotation angle and scaling detection; detecting copy-move tampering using coarse-scaled, rotated tamper detection (CSRTD); and detecting it through fine-scaled, rotated tamper detection (FSRTD).

III. Research Methodology

The proposed architecture is shown in Figure 1. There are four stages of processing, namely Pre-processing, Classification, Localization, and Output Generation, each of these stages is explained below. Each of these stages was experimented with multiple techniques and one technique was finalized. Accordingly, ELA was finalized as the best Pre-processing technique and Vision Learner was the best Classification method. UNet was chosen to be the best technique for Localization of Image Splicing tampering and DCT was the best technique for Localization of Copy Move tampering. Further explanation of various techniques employed is given under respective subsections.

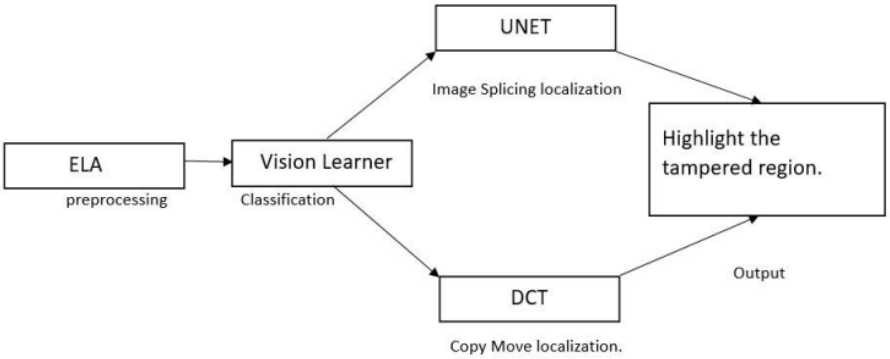

Fig. 1. Proposed Methodology

3.1 Preprocessing

Preprocessing refers to all the transformations on the raw data before it is fed to the deep learning algorithm or machine learning. Preprocessing of data to enhance desired features or reduce artifacts that can bias the network. Because raw data frequently contains noise, errors, or irrelevant information that might adversely affect the performance of the models. For example, resizing the image input to match the size of an image input layer of the model.

Before finalizing ELA as the preprocessing strategy, many different strategies were used. The first of them was to directly load RGB images and resize them based on different pre-trained models. Then, Ycbcr was used as the preprocessing method. In this method, the input RGB image is divided into three streams. Y is luma (brightness), Cb is blue minus luma (B-Y), and Cr is red minus luma (R-Y). Using YCbCr gave better results compared to directly loading RGB images, but it was not good enough.

Finally, ELA was used for preprocessing which gave the best results. An ELA image is an image which is obtained when the image is resaved in the system with a predetermined compression rate. In this ELA image, the regions that are tampered with will have variations in error levels, whereas the regions that are not tampered with will have the same error levels. Then images are resized to (224, 224), and augmentations such as rotation are performed on the image.

3.2 Classification

Classification is a deep learning or machine learning method that uses a model to predict the correct label for the given input image. In the Image Tampering Detection system, the labels the model predicts are either Authentic or Tampered.

Before finalizing Vision Learner - Resnet34 for Classification, many different approaches and pre-trained models were used. In the First approach models with basic CNN layers were used. However, these models did not perform well. Then CNN pre-trained models were used. These models gave better results compared to basic CNN models but it was still not good enough. But then the use of Vision Learner gave the best performance.

The Vision Learner is a component of Fastai, a deep learning library built on top of PyTorch, designed to make it easier to train powerful and accurate neural networks. The fastai.vision module specifically focuses on tasks related to computer vision, such as image classification, object detection, and image segmentation. Within the fastai.vision module, the learner class is a key component. The learner is responsible for handling the training loop, optimization, and various other aspects of the deep learning training process.

Here the preprocessed images were loaded into the Vision Learner model, where feature extraction is done. Later, authentic images are labeled as Au and tampered images as Tp, and the image is either classified as Au or Tp. After training, the performance metrics were collected for each of the models.

3.3 Localization for Image Splicing and Copy-Move

After the image is classified as tampered, the tampered regions in the image are localized. So for localizing tampered regions in spliced images, UNet is used. UNet is a widely used CNN architecture for image segmentation. Here additional preprocessing is done, which involves resizing and augmentation (rotation, horizontal flips). Then UNet performs feature extraction, training, validation, and testing phases. The performance metrics are also given. According to the performance metrics, UNet is the best-suited approach for image splicing localization.

Now to localize copy-move tampered images, a method called DCT, which is discrete cosine transformation. First, the RGB image is converted to a grayscale image because a single colour stream is suitable for applying DCT. The grayscale image is divided into non-overlapping blocks, and DCT is applied to each block. DCT coefficients are quantized and organized into a matrix. Then similar blocks are identified by comparing DCT coefficients and pixel distances. A threshold is applied to determine similarity. If two blocks are similar enough, their coordinates and shift vectors are recorded. Similar shift vectors are eliminated based on a limit. If there are too many similar vectors, the image has likely been tampered with. Based on the identified similar blocks and their shift vectors, a predicted mask is generated.

3.4 Highlighting the Tampered Region

Now that the localization phase of both spliced images and copy-move images are done, the next step is to highlight the tampered region on the input image. Here, red is used to highlight image-splicing tampered regions, and blue is used to identify copy-move tampered regions.

IV. EXPERIMENTS

To validate the efficiency of the proposed methodology, extensive experiments were conducted on diverse datasets. Experiments are primarily conducted to establish the best pre-processing, classification, and localization strategies.

4.1 Selection of the Best Classifier

In the first set of experiments, the performance of a basic CNN model is compared with three pre-trained CNN models namely Exception, FB Net, and MobileNet V2. The experiments used the entire dataset (CASIA 2) for training the models, leaving out 1200 images testing. Before that, a set of experiments with a basic CNN model alone was conducted to ensure the right train: validate proportions. This set of experiments also compared the performance of three different pre-processing techniques, namely RGB, YCbCr, and ELA. Figure 2 shows the performance of the Basic CNN model and Figure 3 shows the performance of CNN pre-trained models.

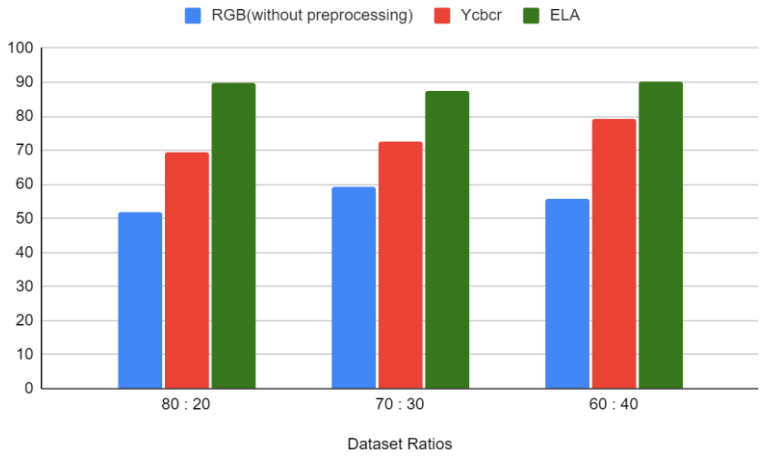

Fig. 2. Performance of Basic CNN model

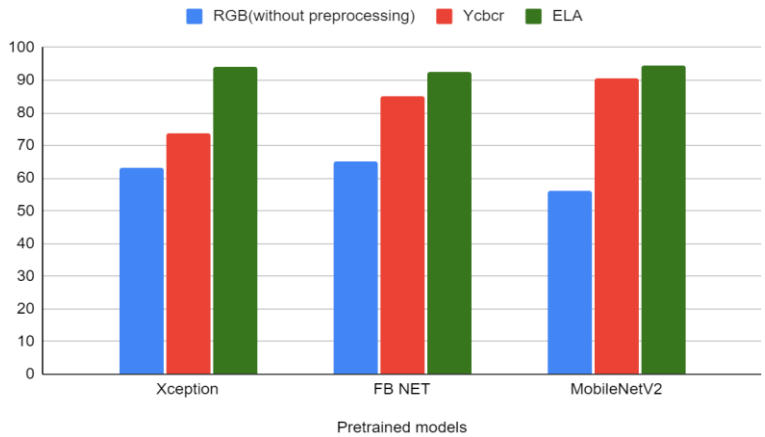

Fig. 3. Accuracy Comparison of different CNN Pre-trained models

4.2 Vision Learner for Improved Performance

Different classification techniques are used to improve the performance of the model. First, the model with only basic CNN layers was used, which didn't give good results. Then CNN layers plus pre-trained models are used. This increased the performance compared to only the basic CNN layers, but it still wasn't good enough. So then Vision Learner was used, which gave an excellent performance compared to other methods. Figure 4 shows the performance of different vision models.

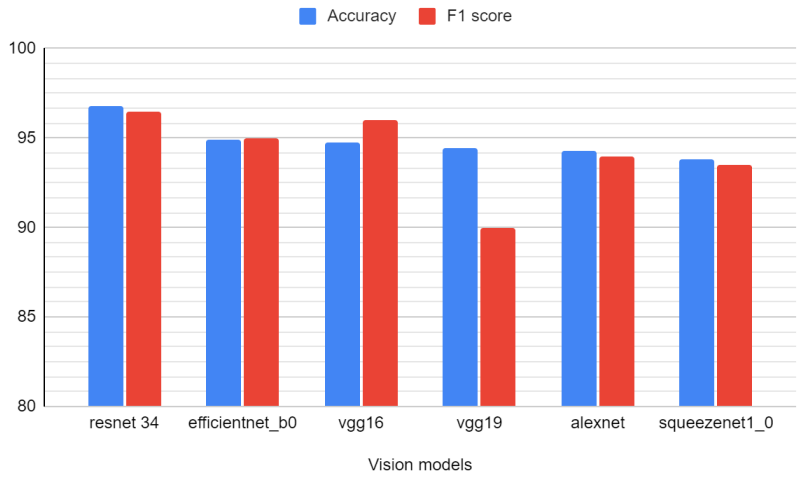

Fig. 4. Accuracy and F1 Score Comparison of Vision Models

4.3 Experiments for Selection of Localization Techniques

Two different approaches for localization have been used. First, it started with UNet. Here, the model is trained using UNet for spliced and copy-move images. But the performance was not good. This is because it turns out that the model was working well only for spliced images but not for copy-move images. Figure 5 shows the performance of UNet model for localization with mixed samples of both spliced images and copy-move images.

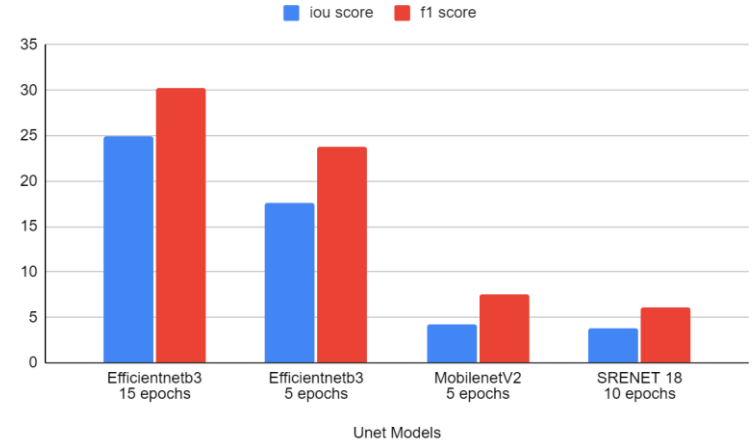

Fig. 5. Performance of UNet for Mixed sample Image Slicing and Copy-Move

The model's performance is enhanced because UNet architecture is used for spliced images. So Figure 6 shows the performance of UNet model only for localization spliced images.

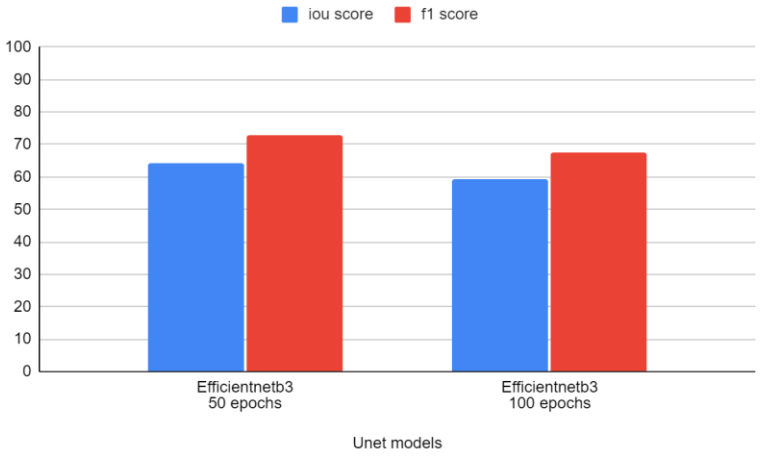

Fig. 6. Performance of UNet for Image Slicing Localization

So for copy-move images, a different approach, Discrete Cosine Transformation (DCT) is used. This gave good results for copy-move localization.

Table I presents the performance evaluation of the classification model Vison -Learner - Resnet34. The model has achieved high precision, recall and F1 scores of 96.5% and an accuracy score of 96.79%.

Table 1 Performance of Vision Learner for Classification

	Model	Accuracy	F1 score	Precision	Recall
Classification	Vision Learner - Resnet34	96.79	96.5	96.5	96.5

Table II shows the performance evaluation of the 'UNet - Efficientnetb3' model in localizing tampered regions in spliced images. The IOU score of the model is 64.19%, and the accuracy is 72.77%.

Table 2 Performance of UNet for Splicing Localization

Tampering type	Method used	IOU score	Accuracy
Image Splicing	UNet	64.19	72.77

Table III shows the performance evaluation using DCT in localizing tampered regions in Copy-move images. The IOU score of the model is 74.2%.

Table 3 Performance of DCT for Copy-Move Localization

Tampering type	Method used	IOU score
Copy - move	DCT	74.2

Figure 7 shows the system output when the given input image is spliced. It first displays the input image, then the predicted ground-truth mask, followed by the image where the tampered region is highlighted on the input image. Finally, it displays a text indicating that the input image is tampered.

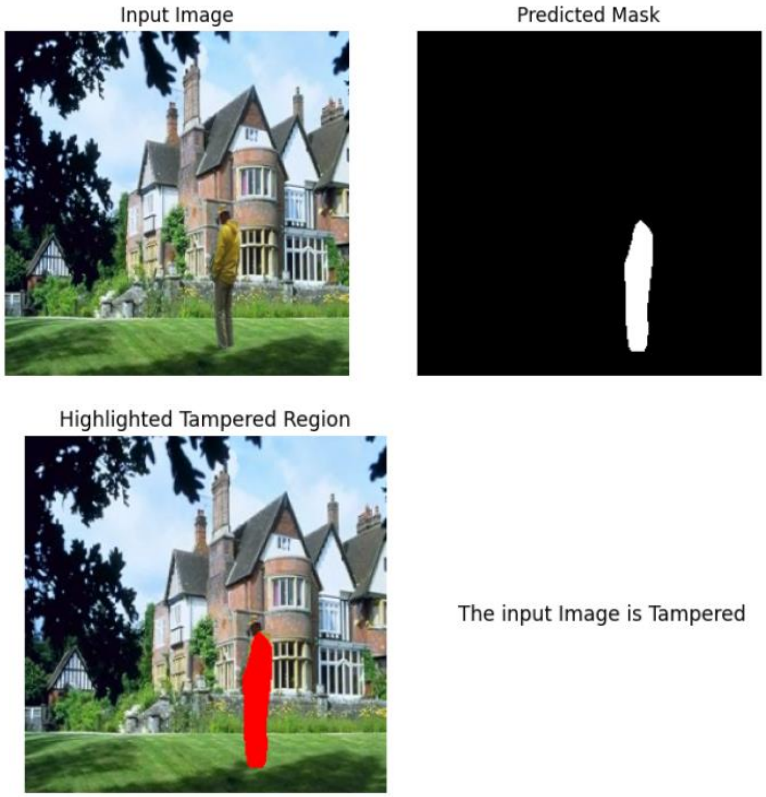

Fig. 7. Output of Image Tampering Detection System when the input has Image splicing tampering

Figure 8 shows the system output when the given input image has Copy-Move tampering.

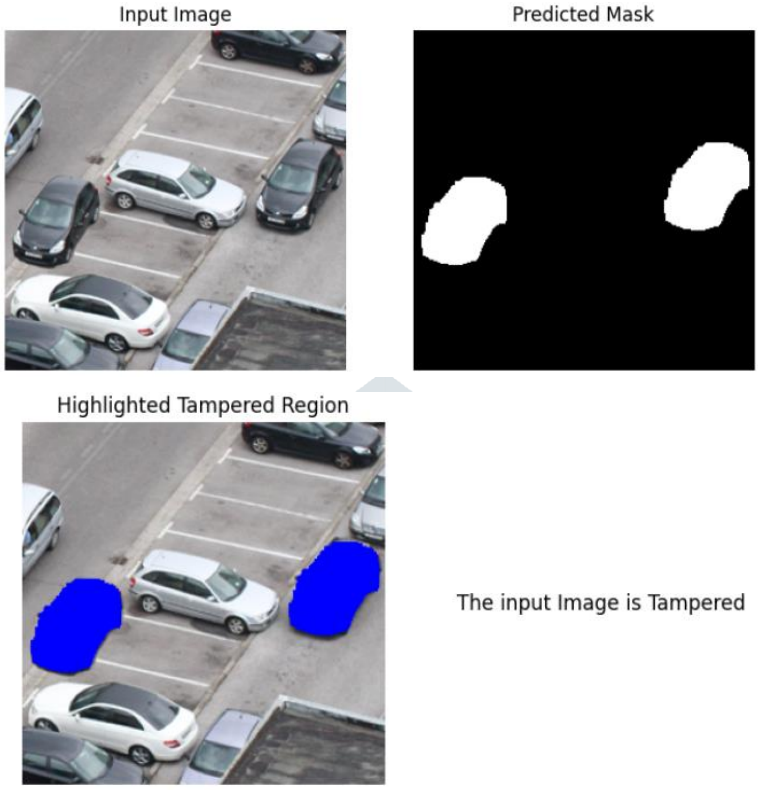

Fig. 8. Output of Image Tampering Detection System when the input image has Copy-Move tampering

Figure 9 shows the system output when the given input image is authentic. It first displays the input image and then presents a text indicating that the input image is authentic.

Fig. 9. Output of Image Tampering Detection System when the input is Authentic

V. CONCLUSION

The objective of the research, of establishing the best and most optimal strategy for image tampering detection was achieved through thorough experimentation on a standard dataset. It was shown experimentally that the Vision Learner model performed the best for tampering detection, while ELA was the best strategy for pre-processing images. UNet architecture performed very well for the localization of spliced images, while DCT was the best strategy for the localization of 'Copy-Move' images. These results indicate that the proposed system can effectively detect and localize tampered regions in images. One can also deduce that deep learning techniques can be effectively used for image tampering detection compared to machine learning techniques. The focus of this research has been to find the best strategies for all stages of experimentation from pre-processing to addressing localization. One of the obvious future work would be to expand the experimentation to cover more types of image tampering, such as removal, and re-touching.

REFERENCES

[1] Zhou, P., Han, X., Morariu, V. I., and Davis, L. S. 2018. Learning rich features for image manipulation detection. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[2] Shivanandappa, M., Patil, M. M., Prabhu. V. S. S., and Swetha. M. D. 2023. Tampering detection and segmentation model for multimedia forensic. International Journal of Advanced Computer Science and Applications, 14(9), 878-887.

[3] Thakur, T., Singh, K., and Yadav, A. 2018. Blind approach for digital image forgery detection. International Journal of Computer Applications, 179(10).

[4] Manu, V. T., and Mehtre, B. M. 2019. Tamper detection of social media images using quality artifacts and texture features. Forensic Science International, 295, 100-112.

[5] Aminu, A. A., Agwu, N. N., and Steve, A. 2021. Detection and localization of image tampering using deep residual UNET with stacked dilated convolution. IJCSNS International Journal of Computer Science and Network Security, 21(9).

[6] Chakraborty, S., Chatterjee, K., & Paramita, P. 2022. Detection of image tampering using deep learning, error levels noise residuals. Research Square.

[7] Madake, J., Meshram, J., Mondhe, A., and Mashalkar, P. 2023. Image tampering detection using error level analysis and metadata analysis. 4th International Conference for Emerging Technology (INCET), Belgaum, India, 1-7.

[8] Wang, W., Dong, J., and Tan, T. 2011. Tampered region localization of digital color images based on JPEG compression noise. Digital Watermarking, Lecture Notes in Computer Science, Springer Berlin Heidelberg Volume 6526, 120.

[9] Elaskily, M. A., et al. 2020. A novel deep learning framework for copy-move forgery detection in images. Multimedia Tools and Applications, 79, 19167-19192.

[10] Pugar, F. H., Muzahidin, S., and Arymurthy, A. M. 2019. Copy-move forgery detection using SWT-DCT and four square mean features. Proceedings of the International Conference on Electrical Engineering and Informatics (ICEEI), Bandung, Indonesia, 63-68.

[11] Singh, P. B., Shalini, M., and Goel, S. 2016. Correlation based image tampering detection. IJCSIT, 7, 990-995.

[12] Warbhe, A. D., Dharaskar, R. V., and Thakare, V. M. 2016. Digital image forensics: An affine transform robust copy-paste tampering detection. Proc. 10th Int. Conf. Intell. Syst. Control (ISCO), Coimbatore, India, 1-5.