RESEARCH ARTICLE OPEN ACCESS

Image Tempering Detection using Yolov7

Prof. Mangesh Hajare*, Sayan Ghanti**, Snehasis Sahoo***, Abhishek Ranjan****, Ankit Sharma*****

*(Computer Department, Army Institute of Technology, Pune
Email: mangeshhajare@aitpune.edu.in)
** (Computer Department, Army Institute of Technology, Pune
Email: sayanghanti_19279@aitpune.edu.in)
***(Computer Department, Army Institute of Technology, Pune
Email: snehasissahoo_19230@aitpune.edu.in)
****(Computer Department, Army Institute of Technology, Pune
Email: abhishekranjan_19293@aitpune.edu.in)
*****(Computer Department, Army Institute of Technology, Pune
Email: ankitsharma_19078@aitpune.edu.in)

--************************----------------------------------

Abstract:

Image tampering or forgery has become a prevalent issue in digital image processing. In recent years, the availability of image editing software has made it easier for individuals to modify images for various purposes, including spreading misinformation, manipulating evidence, or altering personal photographs. Therefore, the need for an efficient image tempering detection model has arisen to detect any unauthorized modifications in the image.

In this research paper, we propose a novel image tempering detection model that utilizes deep learning techniques for detecting tampered regions in images.

To evaluate the effectiveness of the proposed model, we conducted experiments on two image tampering datasets: Synthetic dataset and Columbia dataset. The results show that our proposed model achieves state-of-the-art performance in terms of accuracy, precision, recall, and F1-score. Furthermore, we also tested the model on real-world images.

Overall, our proposed image tempering detection model provides a reliable and efficient solution for detecting any unauthorized modifications in images. The model can be used in various applications, including forensic investigations, media forensics, and authentication systems.

Keywords — **Image tampering, Forgery detection, Deep learning, Convolutional Neural Network (CNN), Feature extraction, Yolov7.**

--************************----------------------------------

I. INTRODUCTION

Over the past decade, digital manipulation of images has become increasingly sophisticated, making it difficult to distinguish between real and fake images. This has led to growing concern about the potential consequences of image tampering, such as the spread of misinformation, defamation, and damage to reputations. As a result, there has been a significant increase in the demand for tools and techniques to detect manipulated images. Detecting image tampering is a challenging task that requires both technical expertise and a deep understanding of the nature of image manipulation. In the past, manual inspection and visual examination were the primary methods used to detect tampering. However, with the exponential growth of digital imagery, these traditional methods have become inefficient, time-consuming, and error-prone. As a result, researchers have turned to machine learning (ML) as a promising approach to automate the detection process. Machine learning algorithms are capable of learning patterns and characteristics in large datasets of manipulated and authentic images, and then applying this knowledge to classify unknown images as authentic or tampered. This approach has been widely adopted in recent years, with researchers developing a variety of ML models to detect image tampering. The success of these models depends on several factors, including the quality and size of the training dataset, the choice of features, and the design of the classifier. Therefore, researchers have been exploring various deep learning techniques, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs), to improve the accuracy and efficiency of image tampering detection. In this paper, we propose a novel ML model for detecting image tampering that combines deep learning techniques and a combination of handcrafted and learned features. Our model is designed to identify the presence of tampering and localize the manipulated regions in an image. We train and evaluate our model on several publicly available datasets and compare its performance with state-of-the-art approaches.

II. IMAGE TEMPERING DETECTION APPROACHES AND CLASSIFICATION

Image tampering, also known as image manipulation, can take various forms, ranging from simple cropping and resizing to sophisticated techniques such as deepfakes and image morphing. Here are some examples of different types of image tampering: Splicing: This involves combining multiple images to create a new image that appears authentic. Splicing is one of the most common forms of image tampering, and it can be used to create fake news or alter evidence in legal cases. Copy-move: This involves copying a part of an image and pasting it onto another part of the same image to create a duplicate. Copy-move tampering is often used to hide or remove objects from an image. Object removal: This involves removing objects or people from an image. Object removal is often used for aesthetic purposes, such as removing blemishes or distracting elements from a photo. However, it can also be used to manipulate the content of an image for deceptive purposes. Image Morphing: This involves blending two or more images together to create a new image that appears authentic. Image morphing is often used in entertainment and advertising, but it can also be used for deceptive purposes, such as creating fake celebrity photos or political propaganda. Deepfakes: This is a type of image or video tampering that uses artificial intelligence and machine learning algorithms to create highly realistic and convincing fake images or videos. Deepfakes are often used for entertainment purposes, but they can also be used for malicious purposes, such as spreading false information or creating fake news. In conclusion, image tampering can take various forms, and it is essential to be aware of these techniques to detect and prevent them effectively. We utilized a single-stream image tampering detection model that has the ability to detect manipulated images by automatically analyzing various features. To aid in this automatic learning, we developed a dataset of synthetic manipulated images that includes bounding boxes around the tampered regions. Our choice for image tampering detection is the YOLOv7 model, which has demonstrated promising results in object detection models and is capable of processing videos in real-time by examining each frame. The YOLOv7 model examines RGB features in the image to distinguish between tempered and authentic areas by identifying visual inconsistencies, contrast differences, and other features that it has learned at the boundary. Generally, when an image is tampered with, the noise consistency of the image is disrupted in the tampered region. The deep learning architecture of YOLOv7 enables it to detect these noise inconsistencies as well, assisting in the detection of tampered regions. By analyzing these features, we can detect tampered regions.

III. RELATED WORK

Recently, deep learning- based methods have demonstrated significant promise in identifying image forgery, which involves altering digital images to deceive viewers. To utilize deep learning-based techniques for detecting image forgery, several crucial steps must be followed. The initial step is to develop a dataset of genuine and manipulated images that encompasses a broad range of image types and manipulation methods. Next, an appropriate deep learning architecture, such as CNNs, GANs, or RNNs, must be chosen based on the task at hand and optimized for accuracy. Bayar and Stamm proposed a deep learning strategy to detect image manipulation using a new convolutional layer. They employed various techniques to create a training dataset and introduced two new convolution layers, two max pooling layers, and three fully connected layers in the proposed model. The model was trained on a dataset collected from 12 different camera models to make it robust and applicable to a wide range of images. Prediction error filters were used to extract features from the images, which served as inputs to the proposed model. The model achieved an accuracy of 99.10%, indicating its effectiveness in detecting image manipulation. In their paper titled ”Image manipulation detection using convolutional neural network,” Kim and Lee present a Convolutional Neural Network (CNN) approach for detecting image tampering. The proposed method utilizes various picture alterations, such as Median Filtering, Gaussian Blurring, AWGN, and Resizing, to train the CNN model to detect tampered photos. Additionally, a High Pass Filter is applied to enhance the tampered areas' edges. The CNN architecture consists of two Convolutional Layers, two Pooling Layers, and two Fully Connected Layers. The authors evaluated the model's performance on 1000 photos from the BossBase 1.01 dataset, and the proposed approach achieved an accuracy of 95% in detecting manipulated photos. The proposed paper, ”ImageRegion Forgery Detection: A Deep Learning Approach,” presents a deep learning method to detect cut-paste and copy-move forgeries in images. The proposed technique employs a three-level, 2-D Daubechies wavelet decomposition to extract features from the input image, and then uses SAE (Stacked Autoencoders) for classification. The model was trained and tested on images from the CASIA v1.0, CASIA v2.0, and Columbia datasets. The suggested approach achieves an accuracy of 91.09% in detecting forged images. The ”RRU-Net: The Ringed Residual U-Net for Image Splicing Forgery Detection” is an image splicing forgery detection deep learning technique. X. Bi, Y. Wei, B. Xiao, and W. Li suggested it. The approach is intended to identify cut-and-paste image forgeries. It uses the image residuals to identify spliced regions. Unlike other models, it uses a ringed residual U-Net architecture that combines the U-Net and residual connections to improve the model's accuracy. The dataset used in the evaluation includes images from the CASIA and COLUMBIA databases. The model's accuracy was 76%, indicating its potential for identifying picture splicing forgeries. Most papers in this field concentrate on particular tampering artifacts and are restricted to specific manipulation techniques. However, we have taken a different approach by using a single RGB stream to detect high-level artifacts such as changes in color or contrast, without utilizing the noise stream. Our model uses the Yolov7 architecture, which has proven to be robust in detecting tampering techniques like splicing, copy-move, and removal.

IV. INTRODUCING A NOVEL APPROACH: THE PROPOSED METHODOLOGY

A. Developing a Model

We have developed a synthetic dataset using semantic segmentation of objects and superimposing an object from one image onto another. This approach provides us with an accurately generated synthetic dataset with class labels, enabling us to train our model effectively and efficiently for Image Manipulation Detection. we have employed the Yolov7 object detection model to train our model. It is considered one of the best real-time object detecting model, as it supports both image processing and video processing frame by frame. It is noteworthy that object detection models typically learn object artifacts to distinguish the object from its surroundings. However, in Image Manipulation Detection, the model must learn tampering boundaries between the object and its surroundings, as tampered regions are not usually associated with object artifacts. Our experiments have shown that Yolov7 is capable of learning these tampered artifacts to a considerable extent, highlighting its potential as a powerful tool.

B. Developing Training Dataset

A synthetic dataset of manipulated images is created by utilizing the PASCAL VOC dataset. The dataset is composed of 27,088 images and their corresponding segmentation masks from the PASCAL VOC (test-06-Nov-2007, trainval- 06-Nov2007 and trainval-11-May-2012) datasets. It is important to note that the images in this dataset do not possess uniform dimensions or orientation, as some im- ages are in portrait while others are in landscape. However, the dataset enables the generation of a maximum of 27,088 untampered images, as well as a vast number of tampered images as per the study's requirements. Each image in the dataset is accompanied by object masks that provide pixel- level information of different objects in the image. This information is instrumental in extracting objects from the images, which are then used to tamper the raw images. The process of creating tampered images involves randomly selecting two images from the dataset and extracting a random object from one of them. The object is then pasted into the other image such that it entirely fits within its boundaries and overlaps all its pixels. However, it is ensured that the object's size is smaller than the dimensions of the second image, in case one image is in portrait while the other is in landscape or vice versa.

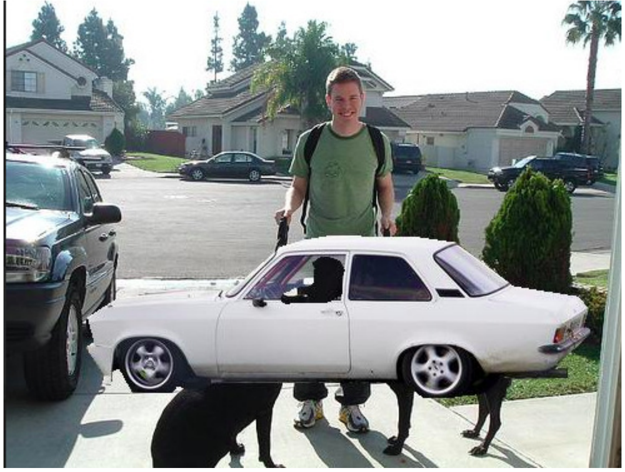

Fig. 1 An example image from synthetic dataset

It is important to note that the synthetic dataset created in this model does not undergo any post-processing, such as blurring, scaling, or contrast adjustments. The dataset's authenticity and diversity are maintained in this way, making it an ideal resource for various image manipulation tasks.

V. TRAINING THE MODEL

We train the model with the provided synthetic dataset. The train-validation dataset is made up of 1000 original photos (copied exactly as they are) and 5000 altered images (created). These photos are jumbled at random and divided in an 80:20 (train:validate) ratio.

On this dataset, we trained our model from start to finish. There is just one class to forecast, and that is Tampered. Training was finished in 1.5 hours with a batch size of 14 and 70 epochs using an Nvidia-Tesla T4 gpu with GDDR6 memory.

Training data indicate that YOLOv7 can not only learn to recognise manipu- lated objects, but it can also achieve very excellent efficiency on this dataset. One plausible explanation is that the underlying deep neural net architecture can define and process tampering limits at the pixel level.

VI. TEST DATA SETS, RESULT AND ANALYSIS

A. Testing on Synthetic Dataset

To test the model on a synthetic dataset, we prepare a collection of 1000 original images and 5000 manipulated images that have not been seen by the model before. Afterward, we examine the outcomes obtained from this test set.

B. Testing on Synthetic Dataset

The Columbia Uncompressed Image Splicing Detection Evaluation Dataset features a variety of spliced images obtained from genuine pictures using the copy-and-paste technique, by selecting visually prominent objects in Adobe Photoshop, with no post-processing involved. This category comprises 180 images in total, with 30 images created for each camera pair, considering there are four cameras.

This dataset is highly acclaimed in the image manipulation detection field and has benchmarks for ELA [20], NOI1 [21], CFA1 [22], MFCN [23], and RGB-N [5].

TABLE I
TEST RESULTS FOR COLUMBIA DATASET

Baseline Model	**Columbia Dataset**
ELA[20]	0.470
NOI1[21]	0.574
CFA1[22]	0.467
MFCN[23]	0.612
RGB-N[5]	0.697

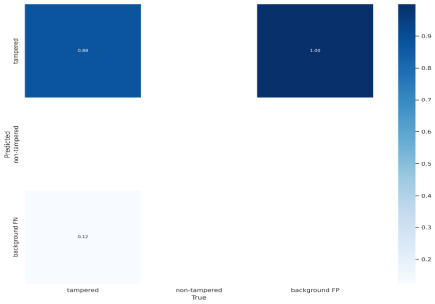

Fig. 2 Testing on synthetic dataset: Confusion matrix

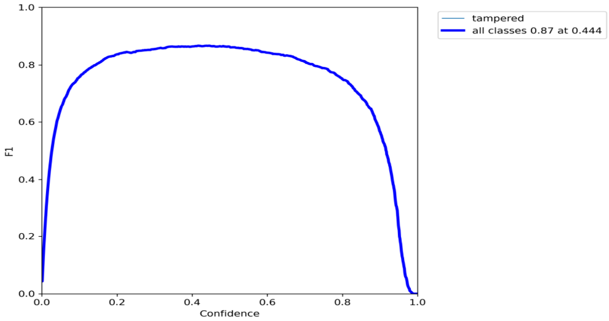

Fig. 3 Testing on synthetic dataset: F1 vs Confidence

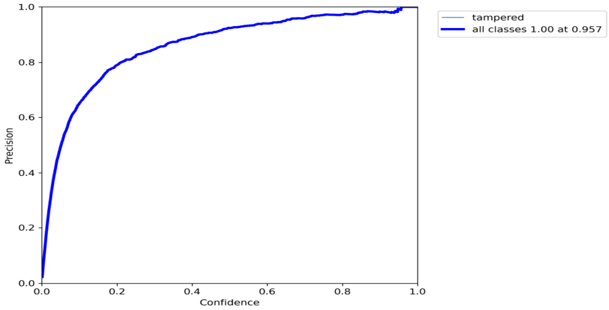

Fig. 4 Testing on synthetic dataset: Precision vs Confidence

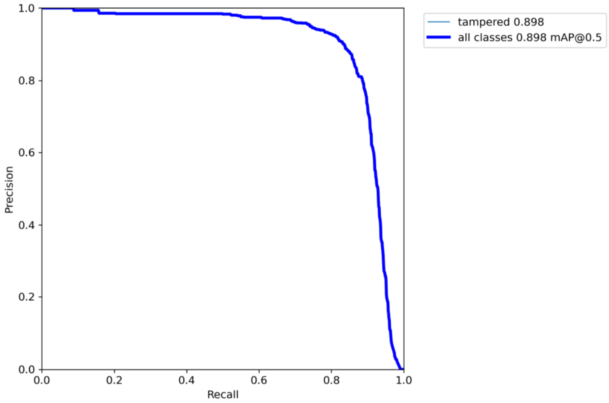

Fig. 5 Testing on synthetic dataset: Precision vs Recall

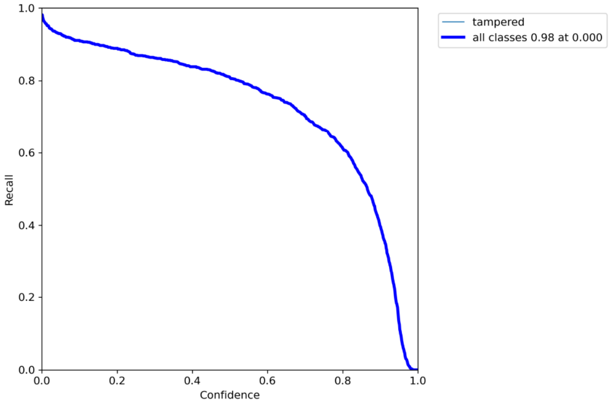

Fig. 6 Testing on synthetic dataset: Recall vs confidence

TABLE II
TEST RESULTS FOR COLUMBIA DATASET

Metric	**Value**
No. Of Images	6000
Precision	0.962
Recall	0.909
mAP@.5	0.949
mAP@.5:.95	0.848

VII. CONCLUSIONS

Our proposed approach involves utilizing the YOLOv7 architecture and training it on a synthetic dataset that is composed of images with overlaid objects from the same or different images. We were able to demonstrate that the YOLOv7 model is capable of effectively adapting to such manipulations, treating them in the same manner as it would with any other object. With this trained model, we were able to successfully detect manipulations in the COLUMBIA Uncompressed Image Splicing Detection Evaluation Dataset, producing fair and unbiased results.

This deep learning model based on the RGB stream takes into consideration both the global and local features of the image to detect tampering. To achieve this, we analyze the inconsistencies between the tampered region and the authentic region, by extracting image features through multiple neural networks. The model has proven effective in detecting various types of tampering operations, such as splicing, copy-move, and removal. However, we leave tampering detection in compressed images as a subject of future study.

VIII. FUTURE SCOPE

To further improve the accuracy of the models in video forensics, future work could explore the potential of utilizing information from sequential frames to detect deepfakes and any other manipulations. This approach would greatly enhance the models' ability to identify tampering in video content.

Moreover, other machine learning models can be developed in further research, which could detect tampering operations beyond the existing ones, such as blending, blurring, compressing, and more. This expansion of the models' capabilities would enable them to identify a broader range of tampering techniques, contributing to more robust and accurate tampering detection.

REFERENCES

[1] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, pp. 5–10, ACM, 2016.

[2] Kim, D. H., Lee, H. Y. (2017). Image manipulation detection using convolutional neural network. International Journal of Applied Engineering Research, 12(21), 11640–11646.

[3] Y. Zhang, J. Goh, L. L. Win, and V. L. Thing, "Image region forgery detection: A deep learning approach.," in SG-CRC, pp. 1–11, 2016.

[4] X. Bi, Y. Wei, B. Xiao, and W. Li, "RRU-Net: The Ringed Residual U- Net for Image Splicing Forgery Detection," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019, pp. 30-39, DOI: 10.1109/CVPRW.2019.00010.