Copy Move Image Forgery Detection

V. Sree Teja Pradeep (22BCE1737)
School of Computer Science of Engineering (SCOPE)
Vellore Institute of Technology, Chennai Campus
Chennai, India

Pragya Bose (22BCE1680)
School of Computer Science of Engineering (SCOPE)
Vellore Institute of Technology, Chennai Campus
Chennai, India

***Abstract*—Ever since the era of digital images came into existence, the authenticity of them has been constantly questioned due to the emergence of various manipulation techniques and attacks to tamper with a digital image, causing fake news and paranoia, these manipulations go unnoticed by human eye making it necessary for automated machines to help detect the manipulations made in the image. In this work we focus on detecting the manipulations in an image using a U-Net model with a ResNet-50 encoder to predict the binary masks which show the manipulated regions in an image. We have also utilized keypoint detection method along with our CNN model to visually identify the important regions of the image and find out the authenticity of the image.**

***Index Terms*—digital images, authenticity, manipulation, tampering, binary masks**

I. Introduction

Due to the rapid development of multimedia technology, image forgery has become extremely easy [1]. Accessing a digital camera has become effortless and affordable, making it easy to capture more than any number of images one could imagine a couple of decades ago. In addition to that, accessing images on the Internet has also become an effortless task with sources like Google Images, Instagram, Facebook, WhatsApp, X (formerly known as Twitter), etc. This provides individuals with a wide range of digital images. Furthermore, numerous user-friendly image editing softwares are introduced that can be used by any individual to craft the necessary modifications. Although these technological advancements are fascinating, several kinds of forgeries have been emerged due to the widespread of these softwares.

There are multiple types of digital image forgery attacks, one prevalent form of attack involves, copying a portion of the image and pasting it somewhere else in the same image, it can be detected by finding out unnatural noise patterns and change in the texture of image, this is the simplest form of attacks as this can be done using obvious and basic softwares such as Google Snapseed, etc.

Another attack which is worth mentioning is **splicing**, in which parts of two images are combined without leaving any overlapping regions unlike copy-move forgery, various adjustment methods are then applied to blend the images seamlessly

There are several other attacks involving geometric transformations of images such as rotation, scaling, skewing which often leave traces of interpolation. Object removal, smoothing and increasing contrast can also be considered as some of the **active image** forgery attacks which are often used.

Some common methods used to detect copy-move forgery are: Keypoint based methods and block based methods. Keypoints are basically the important parts of an image which are distinct and stand out in the image, these keypoints are based on their uniqueness and indicate areas where copy-move forgery might have taken place. Block based methods divide the image into multiple blocks and extract features from each block and compare them with the features of other blocks to find out if there has been a copy move forgery attack. Recently, Deep-Learning based methods became known for detecting copy-forgery as the CNN based methods learn efficiently and are effective for image segmentation.

A. Motivation

With the ever growing study in this field, there has been a lot of progress in detecting forged images, but our study wanted to explore the idea of using an encoder-decoder model, since the architecture of this model enables it to locate pixel-wise spatial information. Therefore we've decided to utilize U-Net model as it has an encoding-decoding architecture and is suitable for segmenting images and build binary masks based on the tampered and authentic regions of an image.

B. Contribution

The key contributions of this paper are:

a) **Implementation of pixel-wise forgery location using U-Net:** We trained a U-Net model on CASIA 2.0 dataset to accurately predict the manipulated regions in an image. This model demonstrated a strong performance compared to the traditional CNN methods. The U-Net model is created with ResNet-50 as the backbone feature extractor.

b) **Standalone usage of Keypoint detection:** We are using Good Features to Track (GFTT) algorithm from the opencv library which is a classic tecnique to detect keypoints followed by DBSCAN Clustering, which clusters the keypoints to give us a rough idea of where the manipulation has taken place. We have also tried to implement another keypoint based copy move forgery detection using CenSure keypoint detectors and FREAK descriptors, which were used for reference.

II. Related Work

In [2], the authors of the paper "Detecting Facial Image Manipulations with Multi-Layer CNN Models" (2024) proposed and analyzed the use of convolutional neural networks (CNNs) for the detection of manipulated facial images. The authors of the paper conducted a c analysis comparing the three complex CNN architectures, assessing their ability to classify and localize manipulations that may have occurred during modifications, that achieved up to 76% accuracy in distinguishing manipulated images from genuine images.

The authors of "ObjectFormer for Image Manipulation Detection and Localization" introduced a method that detects and localizes image manipulations and also combines high-frequency features as well as RGB features which is done by utilizing learnable object prototypes, which are then refined to patch embeddings to capture patch-level consistencies, that performed better than existing tampering detection and localization methods [3].

The paper "Holistic Image Manipulation Detection using Pixel Co-occurrence Matrices" (2021) introduces a novel approach that holistically detects tampered images by combining pixel co-occurrence matrices with deep learning to detect tampered or untampered images, achieving high accuracy, demonstrating robustness across various manipulation types [4].

In [5] A new block based copy-move forgery method was introduced in the paper "A new block-based copy-move forgery detection method in digital images" (2016) where DCT transform is used to extract feature vectors from images which are not overlapping and are sorted lexicographically and the blocks which are copied are selected based on a criteria from similar feature vectors.

III. Dataset Used

The dataset used for this project is CASIA 2.0, which is publicly available on kaggle, github, etc. It consists of 12,472 images out of which 7,491 are authentic images which have not been tampered whereas the rest 4981 are tampered images along with their groundtruth binary mask. The white regions in the mask indicate tampered regions whereas the black regions indicate untampered regions. The image forgery attacks used to tamper the images in this dataset include copy-move, splicing, geometric transformations. All the images are resized to (256,256) and normalized using ImageNet mean and standard values to ensure uniformity and easier model training, if the images are not normalized the feature distributions of the inputs will be very different from what the encoder expects which results in slower convergence and less stable training leading to poor performance. A custom built PyTorch class handles the pre-processing of the dataset which includes loading the images and their ground truth masks, which are binarized into 0 and 1 tensors for tampered images and 0 tensors for authentic images, since the ground truth of an authentic image should suggest no manipulated regions in the image. The training and testing datasets were split in the ratio of 80-20 to ensure fair evaluation during training.

IV. Proposed Approach

The proposed method aims to produce the binary mask of a potentially tampered image using a U-Net based architecture with a ResNet-50 encoder.

The U-Net architecture was a major discovering in the field of Image Segmentation, which basically refers to finding objects and boundaries in a given image, which is very useful in our case, since we have to identify the tampered objects in our image.

The U-Net architecture involves 2 main components: **a contracting path** and **an expanding path**. The contracting path constantly decreases the spatial dimensions of the image while capturing relevant information about the image, whereas the expanding path upsamples the feature map and produce a segmentation map using the patterns learnt in the contracting path. The contracting path uses a combination of convolution and pooling layers which are used to extract the features in an image and reduce the spatial dimensions at the same time. The expanding path uses convolution and up-convolution until it builds the segmentation map from the information obtained from the contacting path.

All the input images and the groundtruth masks are resized to (256,256) which are compatible with the encoder's expectations. Masks are binarized using a threshold value 0.5 after converting them to tensors. Binary Cross Entropy Loss (BCE Loss) is used to train the model, which is appropriate in this case since binary segmentation is being performed and Adam optimizer is used as the optimizer to update network weights with a carefully chosen learning rate to ensure stable convergence.

Along with this approach, we've implemented 2 methods to detect copy move forgery based on keypoint detection methods. One is using GFTT(Good Features to Track) algorithm from opencv which is a popular method used to detect high quality keypoints such as corners, blobs and edges of image that has been through various transformations. The keypoints are then clustered using DBSCAN algorithm to create a mask of all the important parts in an image.

The second method which we've implemented based on the keypoint detection approach is using CenSure keypoint detector with FREAK descriptors to extract feature vectors from the descriptor and apply brute force matcher with hamming distance to find knn (K-nearest-neighbours) for each descriptor and cluster the matched points using Agglomerative clustering which indicate copy move forged regions in the tampered image.

V. Proposed Algorithm

This project mainly consists of 3 independent sections which are used to detect copy-move forgery namely: U-Net model with ResNet-50 as encoder, keypoint detection using GFTT and keypoint detection and clustering using CenSure Keypoint detectors with FREAK descriptors. This project is implemented in Python using PyTorch framework and was implemented and executed in a Kaggle notebook using T4

GPU. The algorithms for each section of this project are listed as follows:

Algorithm 1 U-Net Based Copy-Move Forgery Detection

1: **Input:** Load training data using a DataLoader with batch size 8.
2: Initialize a U-Net model with a ResNet50 encoder.
3: Load pre-trained weights into the model.
4: Use Binary Cross Entropy (BCE) Loss and Adam optimizer .
5: Train the model for a set number of epochs on the training data.
6: For each epoch, compute predictions, calculate loss, and update weights.
7: Load a sample from the validation set and visualize its prediction using the trained model.
8: **Output:** Binary forgery mask

Algorithm 2 GFTT based keypoint detection

1: **Input:** Image from the dataset.
2: Convert the input RGB image to grayscale.
3: Use OpenCV's GFTT detector to identify the keypoints.
4: Apply DBSCAN clustering from scikit-learn library to the detected keypoints.
5: Create a binary mask from the clustered points.
6: **Optional:** Apply dilation using a 5×5 kernel to fill small holes and gaps in the mask.
7: **Output:** Binary mask of keypoints.

Algorithm 3 Keypoint based method using CenSure keypoint detector and FREAK descriptor

1: **Input:** Image from the dataset.
2: Convert the input RGB image to grayscale.
3: Detect Keypoints using CenSurE.
4: Extract Descriptors using FREAK.
5: Perform k-NN Matching.
6: Apply Agglomerative Clustering to group matched points based on their spatial proximity.
7: **Output:** Visualization of the clusters in the original image.

VI. Evaluation and metrics

The U-net based model with ResNet-50 as the encoder was trained on CASIA 2.0 dataset, which consists of images which are tampered using multiple forgery methods and variety of authentic images. The dataset was split into 80-20 ratio for training and testing. The model was trained for 40 epochs with a learning rate of 1e-4 using Adam optimizer and Binary cross entropy loss as the loss function. To evaluate the performance, several metrics such as accuracy, precision, recall, F1 Score, IoU (Intersection over Union) and dice coefficient are used. The metrics evaluated on the testing dataset are summarized below:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

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

$$\text{Dice Coefficient} = 1 - \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

TABLE I
Evaluated Metrics for Digital Image Forgery Detection Model

Metric	Value
Accuracy	0.9927
Precision	0.7801
Recall	0.8612
F1 Score	0.8028
IoU	0.7199
Dice Coefficient	0.8028

The reason for the unusually high accuracy despite the other metrics being relatively less is because, the ground truth masks and the predicted masks have more common black regions or authentic regions than the manipulated regions, as the model succeeds in locating the tampered region which means it classified the authentic region correctly and since the authentic regions in the image dominate, the accuracy is high as the model is able to classify the authentic and tampered regions.

Our model achieved a decent precision with 78% score which means that out of all the positives predicted by the model 78% of the predictions are correct demonstrating a considerably low false positive rate.

Although the model shows a moderate precision, 86% recall score indicates that our model correctly predicted 86% of the outputs, which means that 86% of the predictions were correct.

F1 score is the harmonic mean of precision and recall showing a balance of the two metrics. F1-score being 80% indicates the robustness of our model. For segmentation tasks, it's the overlap between the predicted region and the actual ground truth region, divided by their union.

Since IoU is measured by pixel-to-pixel overlapping, there might not be a perfect overlap since our model succeeds in locating the tampered region but does not identify the tampered region pixel-to-pixel, causing the reduce in IoU score.

The Dice coefficient measured is 0.8028, it confirms the concurrence between predicted masks and the ground truth binary masks. This metric indicates the model's effectiveness in identifying both the extent and the shape of tampered regions.

Fig. 1. illustrates the outputs the model has predicted from the testing dataset, showing the input image and the ground truth mask from the CASIA 2.0 dataset and the predicted mask

Fig. 1. Example of input image, ground truth mask, and predicted mask

and another image highlighting the forgery/manipulation in the input image.

Fig. 2. indicates the outputs produced from one of the keypoint based methods where GFFT(Good Features to Track) Algorithm to detect keypoints. Fig. 2. shows the input image, the ground truth mask, the image with keypoints highlighted, the output mask produced by our U-Net model and the mask of the keypoints.

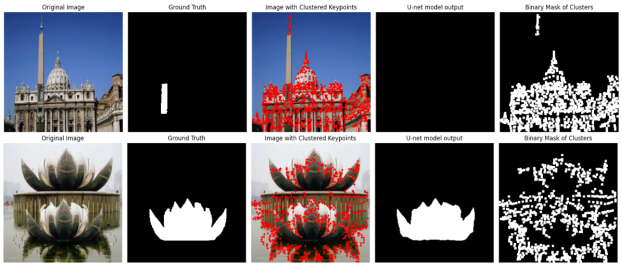

Fig. 2. Example of input image, ground truth mask, and predicted mask

Fig. 3. illustrates the outputs of the second keypoint based Copy-Move forgery detection method, which involves detecting of keypoints using CenSure keypoint detectors and using FREAK descriptors to extract feature vectors and clustering the matched points using Agglomerative Clustering.

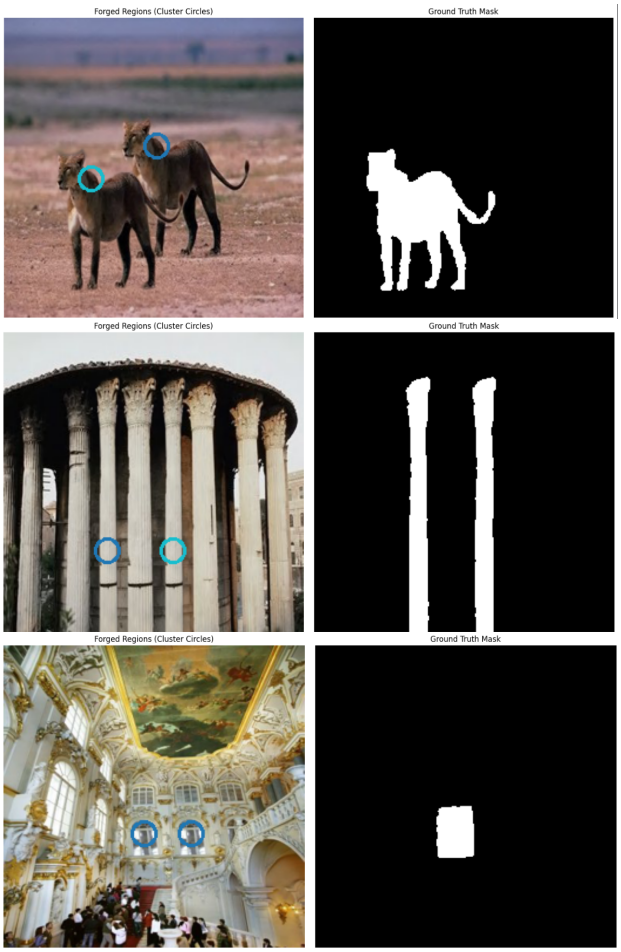

Fig. 3. Example of input image, ground truth mask, and predicted mask

VII. Conclusion and Future Work

In this work, we present to you a deep learning based approach to detect tampered regions in a digital image with the help of U-Net convolutional neural network (CNN) architecture with ResNet-50 as the encoder and two keypoint based methods using GFFT for detecting the keypoints and applying DBSCAN Clustering, and using CenSure keypoint detector and FREAK descriptors and finding the matches using KNN with Brute Force matcher and clustering with Agglomerative clustering to detect possible manipulated regions in the image. The GFFT method used to detect the keypoints performed poorly in terms of detecting the manipulated regions, but identified a lot of keypoints concerning the corners, edges and blobs. The CenSure keypoint detector and FREAK descriptor method's performance was moderate as it seemed to identify the clusters which are similar and therefore seemed to locate a part of the manipulated object but it failed to identify the

spatial information of the manipulated region and generate the binary mask for the manipulated regions and moreover the keypoint locator failed to identify a lot of keypoints and couldn't find a good number of matches to identify the clusters, the manipulated clusters in very few of the images in the testing dataset are located successfully but failed to spatially identify the manipulated region like the U-Net model does. As part of future work, we aim to explore diverse loss functions which are compatible with the U-Net architecture and improve fine-grained localization and overall performance of the model. An algorithm to fuse all three of the methods would create an ideal model for detecting copy-move forgery in digital images as all three methods contribute individually, a fusion would make use of the robust localization of the DFFT keypoint detector and the effective detection of CenSure+FREAK keypoint based method, along with the strong output produced by the U-Net Model to detect the copy-move forged areas in a digital image.

REFERENCES

[1] J. Charpe and A. Bhattacharya, "Revealing image forgery through image manipulation detection," 2015 Global Conference on Communication Technologies (GCCT), Thuckalay, India, 2015, pp. 723-727, doi: 10.1109/GCCT.2015.7342759.

[2] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Niessner, "FaceForensics++: Learning to Detect Manipulated Facial Images," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 1-11, doi: 10.1109/ICCV.2019.00009.

[3] J. Wang et al., "ObjectFormer for Image Manipulation Detection and Localization," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 2354-2363, doi: 10.1109/CVPR52688.2022.00240.

[4] [4] Lakshmanan Nataraj, Michael Goebel, Tajuddin Manhar Mohammed, Shivkumar Chandrasekaran, B. S. Manjunath "Holistic Image Manipulation Detection using Pixel Cooccurrence Matrices"

[5] H. Moradi-Gharghani and M. Nasri, "A new block-based copy-move forgery detection method in digital images," 2016 International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 2016, pp. 1208-1212, doi: 10.1109/ICCSP.2016.7754344.