Document Forgery Detection using CNN

Kajal Awate AnkitaKafare Prof. S. N. Patankar

Abstract- Nowadays document forgery detection is becoming increasingly important as forgery techniques are becoming available even to untrained users. The use of tools like Adobe Photoshop, GNU Gimp to create manipulated fraud documents is a major concern for the government in this digital era. It is extremely crucial to detect image forgery done by computer in these documents. This proposed system focusses on the use of image processing techniques and algorithm to detect the image is forged or not. A Convolutional Neural Network (CNN) is being adopted to extract the features from image of the document to analyze it further and classify it. The image of the document extract the features like brightness of the pixel, font format and resolution of image. These features provide us with details to analyze the similarity between the digital forged images and enable us to develop an algorithm to detect and manipulation in digital images and tampering in documents.

I.INTRODUCTION

Document forgery is a common problem affecting manyareas of our daily lives. For example, customers may presentfake documents to banks in order to obtain a loan or evenpresent tampered documents to insurance companies in orderto obtain the insurance amount. Several approaches may beused to falsify a document. The organization once receiving the image verifies the details with the naked eye and just confirms if there is no physical tampering done to the document whose image has been sent, they do not verify whether the image has been digitally manipulated or forged using any digital manipulation tools easily available to the common users on the Internet. Due to this there is an increase in the field of image forgery and digitally manipulating the document or the image of the document which has been scanned or picture of the document clicked. The forgery or image manipulation that might be done on the document is not visible to the normal naked human eye. This leads to lot of loss of the organization as well as increases the criminal activity. By using image processing along with machine learning and deep learning it is possible to detect these forgeries or digital manipulation done on these documents and images.

II. PROBLEM STATEMENT

The recent advance in the use of image processing applications has benefitted many areas including the forensic and digital verification techniques in cybercrime detection. At the same time the features of image processing techniques are used for producing digital evidence in criminal activities. Image processing tools have been associated with a variety of crimes, including counterfeiting of currency notes, cheques, as well as manipulation of important government documents, wills, financial deeds or educational certificates. The proposed system to provide a solution to this problem. The proposed system focusses on the use of image processing techniques and algorithms along with machine learning to detect the forgery and image manipulation done in a government document which is used nationally on a widescale.

III. LITERATURE SURVEY

Jing Hong Duan, et. al. has proposed the technology which has been widely used in many fields, especially in printing industry for security documents. Though there are many anticounterfeiting methods, Data hiding method has become the mainstream method. This paper presents a computational inexpensive method. It can hide the message while the image is halftoned, and the message can be extracted easily. Halftoning is a traditional printing technique which converts the continuous toned image into a binary image so that the image can be displayed or printed with bi level devices such as digital inkjet printer and printing machine.

Dong Hyun Kim, et. al. has proposed that an image manipulation might be misused by the criminals of counterfeiters for the purpose of counterfeiting. The filter that is used for acquiring the hidden features is High pass. And, it becomes easily available to apply it to various multimedia as well as image. Digital Forensics will be needed to detect such illegal purposes. This paper can be used to check whether or not the image is manipulated or not and can be applied for detection of the manipulation techniques.

S.PraylaShyry, et. al. describes multiple types of picture forgery and detecting them techniques and methods have been elaborately explained. The picture and image manipulation by different means and methods have been discussed. At the beginning multiple types of attacks are categorized and the passive approach is been explained and discussed.

ZhuangXiong, et. al. describes a non-local scheme which is related and based on the 3D convolutional neural network for the image and super resolution has been proposed. The method and techniques used is been built to sharpen the non-local patches. The analysis of the method indicate in results the higher reconstruction accuracy.

IV.PROPOSED MODEL

From the literature survey we have studied that there are n number of algorithms that match the requirements of our project.

The best suited algorithm for our project that we are going to use is the CNN algorithm.

The system uses Convolution Neural Network (CNN), which is an image classification algorithm. CNN is made up of neurons, each having an independent weight assigned to it.CNN is a class of deep neural network specially used for image recognition and image processing. CNN takes the input as an image, identify and assign priority to various features of the image and it differentiates the features from one another. The preprocessing required for CNN isless and have ability to learn image characteristics. CNN consists ofseveral set of convolution layer, pooling layer, flatten and dense. The sets of convolution and pooling layer are used for feature extraction and the number of such set may vary. Convolution layer is the basic building block of the CNN and is used for extracting features from an input image.

The following figure represents the system flow of the proposed system model.

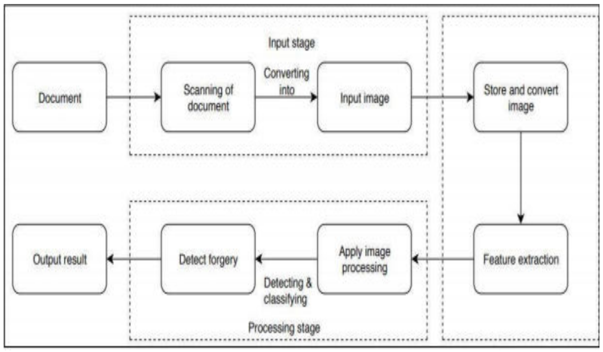


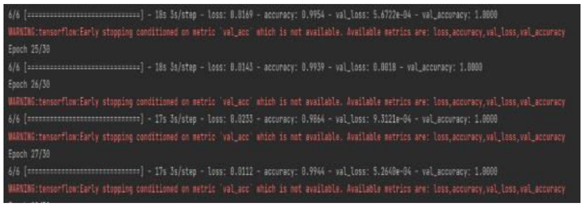
Figure shows the System Flow diagram of Forgery Detection in document process. The diagram is explained as follows: The proposed system uses Convolution model which consists of multiple layers for the purpose of feature extraction from the image. Training data is provided to the model for better prediction of whether the document is forged or not. The classification of document with taking into consideration of various features in the document. The input image is processed, and multiple features are extracted and analyzed. Multiple convolutional layers used to provide better prediction. The testing of the model done using images of document will show its accuracy. Hence, the model classifying the document and the document which has been forged.

For experimental analysis 80% training data is used and 20% data is used for validation.

While CNN training involves a larger portion of data, there are no large public datasets that contain numerous image pairs marked with their copy-move manipulations and ground truth. Therefore, we generated our own dataset.
The training data were designed to present two datasets categories: non-forged and forged.
The forged images are generated as follows:

- Portions of images are cropped and pasted arbitrarily
- The portions of copied image regions undergo processing under various functions such as resizing, rotation or other distortions and then being pasted to produce the tampered image.
- Once the portions are cropped and pasted, postprocessing operation will takes place to complete the process of generating tampered images.

The following figure shows the training validation accuracy of the algorithm ran on the dataset and it gives a validation accuracy of 0.98.



```
6/6 [==============================] - 18s 3s/step - loss: 0.0169 - accuracy: 0.9954 - val_loss: 5.6722e-04 - val_accuracy: 1.0000
WARNING:tensorflow:Early stopping conditioned on metric `val_acc` which is not available. Available metrics are: loss,accuracy,val_loss,val_accuracy
Epoch 25/30
6/6 [==============================] - 18s 3s/step - loss: 0.0143 - accuracy: 0.9939 - val_loss: 0.0018 - val_accuracy: 1.0000
Epoch 26/30
WARNING:tensorflow:Early stopping conditioned on metric `val_acc` which is not available. Available metrics are: loss,accuracy,val_loss,val_accuracy
6/6 [==============================] - 17s 3s/step - loss: 0.0233 - accuracy: 0.9864 - val_loss: 9.3121e-04 - val_accuracy: 1.0000
WARNING:tensorflow:Early stopping conditioned on metric `val_acc` which is not available. Available metrics are: loss,accuracy,val_loss,val_accuracy
Epoch 27/30
6/6 [==============================] - 17s 3s/step - loss: 0.0112 - accuracy: 0.9944 - val_loss: 5.2640e-04 - val_accuracy: 1.0000
WARNING:tensorflow:Early stopping conditioned on metric `val_acc` which is not available. Available metrics are: loss,accuracy,val_loss,val_accuracy
```

V.CONCLUSION

Document forgery detection is done using CNN was proposed in this work. The algorithm uses Convolution model which consists of multiple layers for the purpose of feature extraction from the image.These features are used by the convolutional layers to process the image and give a binary classified output as whether the input image is forged or not. The system will use the features to process and will detect and classify whether the image is forged or not.As a part of future scope, the proposed model will be implemented on a difficult dataset to validate the betterment of the projected model.

REFERENCES

[1] JinghongDuan, “An Anti-Counterfeiting Method for Printed Image by Digital Halftoning Method”, IEEE International Congress on Image and Signal Processing, 2012.

[2] Dong-Hyun Kim,Hae-Yeoun Lee, “Image Manipulation Detection using Convolutional Neural Network”, International Journal of Applied Engineering Research, 2017, pp. 11640-11646.

[3] S.PraylaShyry, SaranyaMeka , “Digital Image Forgery Detection”, International Journal of Recent Technology and Engineering(IJRTE),2019,pp 658-661.

[4] ZhuangXiong, Xiaoming Tao, Nan Zhao, Baihong Lin, “SINGLE IMAGE SUPER-RESOLUTION USING A NON-LOCAL 3D CONVOLUTIONAL NEURAL NETWORK”, GlobalSIP,2018.

[5] DamirDemirovic, Emir Skejic, “Performance of some image processing algorithms in TensorFlow”, IEEE,2018

[6] TianmeiGuo, JiwenDong ,HenjianLi'YunxingGao, “Simple Convolutional Neural Network on Image Classification”, IEEE, 2017

[7] Daniel Oliveira Dantas, Helton DaniloPassos Leal, “FAST 2D AND 3D IMAGE PROCESSING WITH OPENCL”, IEEE, 2015