Digital signature Forgery Detection using CNN

Lakkoju Chandra Kiran[1], Gorantla Akhil Chowdary[2], Manchala Shalem Raju[3], Kondaveeti Gopi Krishna[4]

[5]***Guided by: Asst. Prof. Siva Shankar,*** *Dept. of Computer Science, KKR & KSR Institute of Technology and Sciences,Vinjanampadu, Guntur, Andhra Pradesh.*

---***---

ABSTRACT: Representing information in visual ways has become very important in the digital computing environment. In recent years, we have seen a major increase in the accessibility and transmission of digital images using imaging technologies such as digital cameras and scanners, due to advancements in computer and network technologies. Any type of image manipulation operation is regarded as a picture forgery when manipulating digital images involving copying one part of the image into another part of the image. It is an important field of research to verify the integrity of images and detect traces of tampering without requiring additional prior knowledge of the image content or any embedded signatures. An attempt is made to survey the recent developments in the field of digital image forgery detection, and blind methods for forgery detection are presented in full bibliography. First, different methods of detection of image forgery are classified and then its generalised structure is developed. Together with a suggestion for future studies, the current status of the image forgery detection technique is addressed. There is therefore an urgent need for automated tools that are capable of detecting false multimedia content and preventing dangerous false information from spreading.

KEYWORDS: Digital Signatures; Forgery Detection;Active Authentication;Convolutional Neural Networks; Machine Learning ;Deep Learning.

INTRODUCTION:

Fig (a) Forged Image (b)Real Image

The aim of this paper is to present an analysis of the methods for verification of visual media integrity, i.e. the detection of manipulated images. In this paper, we studied active approaches to digital image forgery detection.

Digital images are the primary source of information in a Digital World Today. Because of their ease of procurement and Storage is the fastest means of transmitting information.Photos may be used as evidence in the court of any case.The photos seen on every TV news are acknowledged as being the truthfulness certificate of that news. Digital pictures In many uses, from military to military, they are used. Medical diagnosis and from artwork to photography by consumers. Therefore, digital image forensics emerges as an increasingly rising need for Society, Thus the photographs must be real. can transfer their The development of computers and computers in today's scenario, the availability of low-cost hardware

and software resources is very critical. The digital images can be easily manipulated without leaving the Clear manipulation traces. It has become hard to trace the Manipulations of these. As a result, the dignity and integrity of Digital photographs are losing their validity.

For certain malicious purposes, such as hiding some essential traces from an image, this manipulation of images can be used. Adjusted photos in this way Used for the transmission of incorrect knowledge. For the purpose of defining The dignity of the photos that we use to describe any transition About the picture. Digital Image Forensic is the science division. It focuses on revealing the malicious exploitation of images.

The technique of digital image forgery detection is commonly divided into two categories: first, active technique and second, passive technique. In active technique, there is some disadvantage in embedding some information in images either during production or before it is broadcast to the public. We could not incorporate any details or information compared to the blind technique . We have addressed these two methods of detecting image forgery below.

ACTIVE AUTHENTICATION:

The digital image can be easily produced or manipulated by several instruments. As a consequence, image validity can not be granted, now a day digital image is used for legal photographic evidence in that situation in which we can not easily trust any digital record. Picture manipulation consists of multiple processing operations, such as scaling, rotation, blurring, modification of brightness, Contrast shifts, etc., or some combination of these operations. Image doctoring means that one part of the image is skilfully pasted onto another part of the image, without any trace. Digital signature and watermarking are significant methods for digital image authenticity.

1.1DIGITAL WATERMARKING:

For image forgery detection, watermarking is often used. Several methods of watermarking have been suggested. One uses a checksum schema to apply data to the last most significant pixel bit[3]. Others add to the pixel data a maximum length linear shift register sequence and then define the watermark by computing the sequence's spatial cross-correlation function and the watermarked image.

These watermarks are intended to be invisible or to blend in with natural noise from the camera or scanner. Often, visible watermarks exist. In addition to this, there is also a visually undetectable watermarking schema that can detect the change in single pixels and locate where the change takes place[3]. Embedding watermarks during digital image production can restrict its application if the mechanism for generating digital images has built-in watermarking capabilities. There are some drawbacks to these active strategies since they require some human interaction or specially designed cameras. Passive authentication has been suggested to resolve this problem.

1.2 DIGITAL SIGNATIURE:

Signatures consist of a combination of individual written features; they are therefore, among the most significant means of obtaining the endorsement of the content of a document by a person. Significant progress has been made in the use of automated signature verification systems to date, particularly in the fields of biometrics and forensics. Automated signature verification techniques in forensics allow forensic document examiners (FDEs) to carry out their work in a faster and more objective way. Data acquisition can be divided into offline or online methods in automated signature verification systems. Offline methods verify the identity of an individual by comparing the signature of a current query with reference signatures that have already been written. Dynamic knowledge related to the script is analyzed by online methods in Question in the form of digital devices (e.g. digitizing laptops, tablet PCs, and smart phones) (e.g., pen position, pen pressure, and pen inclination angle). Online methods tend, therefore, to be more stable than offline methods and to have greater accuracy.

For the purpose of authenticating writing in a permanent form, a handwritten may be described as the scripted name or legal mark of a person or executed by hand. The actions of signing with a writing or marking instrument such as a pen or stylus are sealed on a document. Ordway Hilton (who presented a study of Dynamic Handwritten signature verification) introduced that the signature has at least three shape, motion and variation attributes. Because the signatures are produced by moving a pen on a paper, the most important feature of the signature may be movement. These nerve impulses are managed by the brain without any careful attention to information until an individual is used to sign his signature. Signature verification and forgery detection is the method of automatically and immediately checking signatures to determine whether or not the signature is authentic.

There are two main forms of signature verification:

1.Static2.Dynamic

Static or offline verification is the method of verifying an electronic or paper signature after it has been produced, whereas dynamic or online verification takes place as the subject produces its signature on a digital tablet or similar computer.The signature in question then constitutes the database or knowledge base, compared with previous samples of target signatures. The machine allows the sample to be scanned for analysis in the case of an ink signature on a document, where the digital signature that is already stored in data format can be used for signature authentication.

Signatures are now one of the most commonly recognized personal characteristics for identity controversy, whether from the banking or business field. Thus in some cases, these signatures are easy to counterfeit. Four types of forgeries are possible in this situation.

Simulation forgery:

In which the forger has a signature sample that is to be forged. A simulation's quality depends on how much the forger practices before attempting the actual forgery, the forger's skill, and the attention of the forger to detail in simulating the signature. Identified forgeries are graded between unskilled and professional forgeries on the basis of the forger's experience.

Unknown /Random/Blind forgery:

This is because the forger has no idea what looks like the signature that is to be forged. This is the simplest form of forgery to detect since it is typically not similar to a genuine signature appearance. Often this method of forgery would allow an examiner to determine who made the forgery on the basis of the handwriting habits present in the forged signature.

Tracing forgery:

Tracing is the third form of forgery. Tracing can be achieved by keeping the model document and the document in question up to date and tracing the lines of the model signature to the document in question using a pen. To create an impression of the model signature in the text, a tracing can also be performed by using a blunt stylus on the questioned document. To create the appearance of the design signature, this impression is then filled in with an ink. This form of forgery is often difficult to detect from a photocopy if the model signature used by the forger is not identified.

Optical Transfer:

It's one in which, by using a photocopier, printer, facsimile machine, or photography, a real signature is transferred to a paper. An examiner caiuiot unequivocally recognizes a signature as genuine with this form of forgery without having the Original for reference.

Official papers, however are normally passed on successfully until a signature is obtained, which can lead to several problems for the individual concerned. In the field of net banking, the signature forgery detection finds its application, the passport verification system offers a means of identity confirmation for candidates in public examinations, credit card purchases, and bank checks. Therefore, with the increasing demand for individual identity security, it is important to design an automated signature system.

On the other hand, online signatures use the system's static features, which include techniques for image processing to analyze the accuracy of the signatures. They involve a person's original identity via the password. Other multimodal systems use two separate biometric features to strictly authenticate the identity of an individual. Our project falls into the category of online authentication systems for signature verification.

2. RELATED WORKS:

Offline signature authentication is discussed in separate explanations. Model matching and Secret Markov model techniques are commonly employed in offline signature verification. The arrangement of the signature is based on these techniques. Template matching is a digital image processing technique for tiny, fine sections of an image that fit a template image.

When it comes to template matching, it is possible to use metrics such as n square error or structural similarity index, or a warping approach that warps one curve onto another that retains the original form. The use of Deep CNNs to classify to whom the signature belongs and whether it is a forgery is seen by the authors in[7]. This is achieved in a two-phase approach: learning of writer-independent characteristics, and classification based on writer.

3. THEORETICAL ANALYSIS

An individual's signatures are a biometric behaviour that is not dependent on the physiological characteristics of a signature person, but not on behavior that changes over time. Because an individual signature can change at the time of authentication and signature verification, it may take a long period of time to find errors that are large in some cases.

Implementation:

3.1 Data Acquisition:

To build a knowledge base for each and every user, digital signatures are gathered, and some unique features are extracted. For evaluating the performance of signature verification, a standard dataset of signatures for every person is needed. So, the Dataset we collected from Kaggle is A selection of nearly 2000 images of signatures of both genuine and forged signatures of a person is the dataset that has been used in our project. As it is a huge set of images, we have taken a sample of images for 12 persons with both genuine and forged set of images for each individual. The dataset is divided into genuine and forged set of images .All the images are in RGB format which are collected for this work.

Fig . Forged and Genuine Images

3.2 Pre-Processing:

Any RGB image is represented in Layman's terms as a matrix of X, Y and a depth of three planes where the plane consists of Red, Green and Blue values ranging from 0 to 225. Whereas a gray image, only one plane is represented as a Matrix of X, Y and depth. In this project we have used skimage libraries to convert a RGB signature into gray.

Fig . RGB image

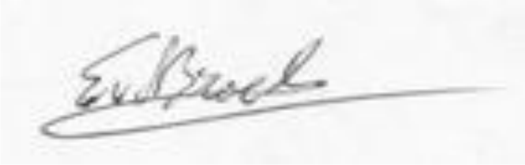


Fig 3.5 Gray Scale Image

3.2.3 Gray to Binary:

To convert our Gray image to binary we need to Import from skimage.filters import threshold_otsu”.Pixel values vary between 0 and 255 as we work over a gray scale picture. Also, because we want to convert the picture to black and white, the value to which we want it to be converted is 255 when the pixel is greater than the threshold.

The threshold function naturally helps one to define these parameters. The first input of the function, therefore is the gray scale picture to which we want the procedure to be applied, which receives the threshold value.

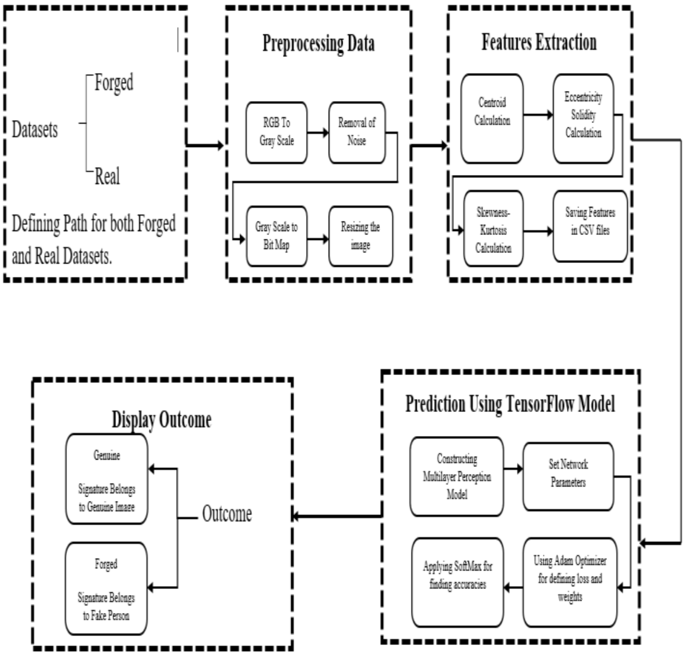
The value of 127, which is in the center of the scale of the values that can be taken by a pixel in the gray scale (from 0 to 255), is considered. The feature receives the user-defined value to which a pixel should be converted if its value is greater than the threshold.

We're going to use the value 255, corresponding to white. Recall that we want to convert the image to black and white, which means that at the end we want an image with pixels with either the value 0 or 255.

3.2.4 Noise Removal and Resizing:

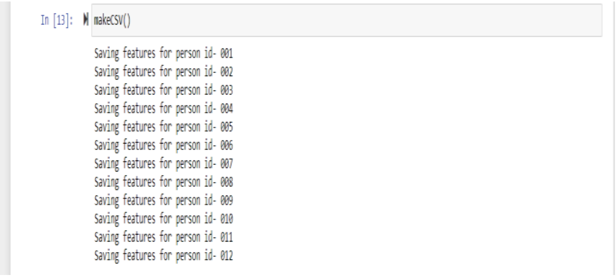
To remove small components or noise from the image we have used “img = ndimage.gaussian_filter(img, blur_radius)”,with a blur radius of 0.8. And next to that ,we will make a bounding box with the boundary as the position of pixels on extreme.Thus we will get a cropped image with only the signature part.

Fig Architecture Diagram


Datasets
Forged
Real
Defining Path for both Forged and Real Datasets.
Preprocessing Data
RGB To Gray Scale
Removal of Noise
Gray Scale to Bit Map
Resizing the image
Features Extraction
Centroid Calculation
Eccentricity Solidity Calculation
Skewness-Kurtosis Calculation
Saving Features in CSV files
Prediction Using TensorFlow Model
Constructing Multilayer Perception Model
Set Network Parameters
Applying SoftMax for finding accuracies
Using Adam Optimizer for defining loss and weights
Display Outcome
Genuine
Signature Belongs to Genuine Image
Forged
Signature Belongs to Fake Person
Outcome

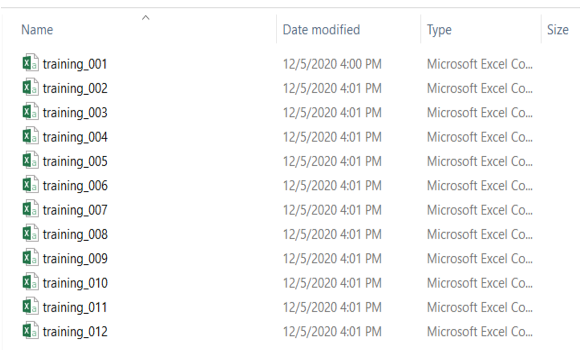
3.3 Features Extraction:

In this process in the project, after the preprocessing of images from RGB to Gray and then from Gray to Binary , resizing the image. We are going extract some features from each image. Features like ratio, centroid, eccentricity, Solidity, Skewness and kurtosis are extracted from each signature image and is saved into Comma Separated Format files for each individual.



```
In [13]: makeCSV()

Saving features for person id- 001
Saving features for person id- 002
Saving features for person id- 003
Saving features for person id- 004
Saving features for person id- 005
Saving features for person id- 006
Saving features for person id- 007
Saving features for person id- 008
Saving features for person id- 009
Saving features for person id- 010
Saving features for person id- 011
Saving features for person id- 012
```



Name	Date modified	Type	Size
training_001	12/5/2020 4:00 PM	Microsoft Excel Co...	
training_002	12/5/2020 4:01 PM	Microsoft Excel Co...	
training_003	12/5/2020 4:01 PM	Microsoft Excel Co...	
training_004	12/5/2020 4:01 PM	Microsoft Excel Co...	
training_005	12/5/2020 4:01 PM	Microsoft Excel Co...	
training_006	12/5/2020 4:01 PM	Microsoft Excel Co...	
training_007	12/5/2020 4:01 PM	Microsoft Excel Co...	
training_008	12/5/2020 4:01 PM	Microsoft Excel Co...	
training_009	12/5/2020 4:01 PM	Microsoft Excel Co...	
training_010	12/5/2020 4:01 PM	Microsoft Excel Co...	
training_011	12/5/2020 4:01 PM	Microsoft Excel Co...	
training_012	12/5/2020 4:01 PM	Microsoft Excel Co...	

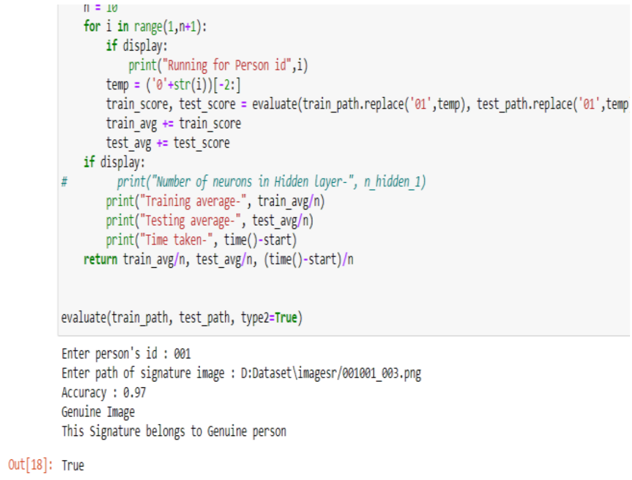
Fig. Saving Features

3.4 TF Model:

We make use of CNNs in this program.A CNN (or ConvNet) is a deep, feed forward artificial neural network class that has been successfully applied to visual imaginary analysis. Biological processes motivated CNNs in that the communication pattern between neurons parallels the organisation of the visual cortex of animals. In our study, we used the Keras library to implement CNN with the "Tensor Flow" backend. The preprocessed image directory is loaded and we then train the model to predict the performance.

In this project we have created a model, called logits three input layers with different weights and biases and 3 hidden layers with set of neurons and a output layer to show the final output (Genuine or Forged).We have also Defined loss and optimizer to minimize the loss.and applied softax to caluclate the accuracy.And finally after evaluating the model we will receive an output, as whether the image is Genuine or Forged.

4. EXPERIMENTAL RESULTS

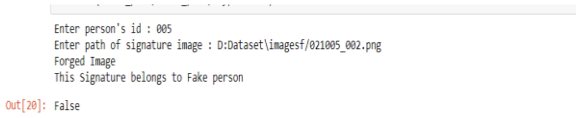


```
    n = 10
    for i in range(1,n+1):
        if display:
            print("Running for Person id",i)
        temp = ('0'+str(i))[-2:]
        train_score, test_score = evaluate(train_path.replace('01',temp), test_path.replace('01',temp
        train_avg += train_score
        test_avg += test_score
    if display:
#           print("Number of neurons in Hidden layer-", n_hidden_1)
        print("Training average-", train_avg/n)
        print("Testing average-", test_avg/n)
        print("Time taken-", time()-start)
    return train_avg/n, test_avg/n, (time()-start)/n


evaluate(train_path, test_path, type2=True)

Enter person's id : 001
Enter path of signature image : D:Dataset\imagesr/001001_003.png
Accuracy : 0.97
Genuine Image
This Signature belongs to Genuine person

Out[18]: True
```



```
Enter person's id : 005
Enter path of signature image : D:Dataset\imagesf/021005_002.png
Forged Image
This Signature belongs to Fake person

Out[20]: False
```

5. FUTURE ENHANCEMENT

As we are dealing with the signature forgery identification, we need to consider every possible parameter that changes the detection of image. Since this method uses CNNs, it can be called extreme. For each user, two classes are generated in the model created in this work (genuine and forgery). If the real and forged signatures of 300 people are given, the model would have 300 classes to predict, which will make the learning process longer. In order to achieve good accuracy and robustness of our application we need to use the best optimizer and best model to minimise the loss and to get maximum accuracy. Therefore, there is a need to consider the group of images to be processed at a time so that our expected outcome will be same as expected like if the signature matches it will give us the Genuine Signature as an output and if the signature does not match it will give us the Forged Signature. Potential creation will be to carry out thorough research on loss functions and to derive custom loss functions (preferably two) that would predict the user to whom the signature belongs and whether or not the signature is forgery

6. CONCLUSIONS

A framework that can learn from signatures and make assumptions as to whether or not the signature in question is forgery has been successfully implemented. This device can be deployed in different government offices where digital signatures are used as a form of approval or authentication. Although this approach uses CNNs, it may be considered extreme. Two classes are created for each user in the model created in this work (genuine and forgery).

If the real and forged signatures of 350 people are given, the model will have 350 classes to forecast, which would make the learning process longer. A potential development would be to carry out extensive research on loss functions and to derive custom loss functions (preferably two) that would predict the user to whom the signature belongs and whether or not it is a forgery.

Such implementation can be deemed serious. Two classes are created for each user in the model created in this work (Real and forgery). We have 12 users, so we have a model that can estimate 60 groups. The highest accuracy we got was 99.7% . The average accuracy is about 97.8%.

7. REFERENCES:

[1] Bailing Zhang. Off-line signature verification and identification by pyramid histogram of oriented gradients. International Journal of Intelligent Computing and Cybernetics, 3(4):611–630, 2010.P.

[2] K. Simonyan, A. Zisserman "Very Deep Convolutional Networks for Large-Scale Image Recognition" in ICLR 2015

[3] E. J. Justino, A. El Yacoubi, F. Bortolozzi, and R. Sabourin, An off-line signature verification system using HMM and graphometric features, in Fourth IAPR International Workshop on Document Analysis Systems (DAS), Rio de. Citeseer, 2000, pp. 211222.

[4] Yanjun Cao a, *, Tiegang Gao b , Li Fan a , Qunting Yang a, "A robust detection algorithm for copy-move forgery in digital images",Forensic Science International journal homepage: www.elsevier.com/locate/for s c iint.