Methods of Detecting Image forgery using convolutional neural network

Alka Leekha¹, Arpan Gupta, Amit Kumar² and Tarun Chaudhary

Department of Information Technology, Bharati Vidyapeeth’s College of Engineering
¹E-Mail: alka.leekha@bharatividyapeeth.edu
²E-Mail: kumaramit9740@gmail.com

Abstract

Image Forgery means manipulation of the digital image so as to hide the necessary information or to convert the image into something which can be used for illegal motives. Differentiating between an original image and a tampered image is a strenuous work for a human being. Machines can do that easily for large amount of data with the help of many artificial intelligence technique and mathematical approaches that are used to develop a system which will help in recognizing whether the image is tampered or not. In this paper Various models like CNN, Hashing, DCT, DWT will be discussed. This paper also discussed about the implementation of CNN done on a data set to figure out a model which is best suited for the development of an image forgery detection system.

1 Introduction

Image Forgery is the manipulation of digital image so as to conceal some important information or to convert the image into some other one so as to use it for illegal motives. Image forgery can’t be detected by the naked eye or in other words the human being can’t distinguish between an original image and a forged one. But through a machine the detection of image forgery is now possible. Development of a model with the help of deep neural network or using machine learning algorithm is a challenging task. But the recent trend shows that the convolutional neural network shows far good accuracy is detecting the forgery.

Nowadays deep neural network due to its advantages is used over machine learning for image forgery, the features like image prep processing, feature extraction and training the machine more accurately. Detection of image forgery finds great use in banking system to detect the photo uploaded by the user is tampered or not also whether the signature uploaded by the user are authentic or not. Various security issues can be solved using the proposed system.

Image forgery detection system development can be classified into 3 categories :

1. Image pre processing
2. Building the model
3. Predicting the forgery

In image pre processing phase the image is converted from grey scale image to binary image to detect the boundaries in the image which will only be present in the fake ones. To convert the image into binary image machine learning library is being used. After that image is rescaled between 0 and 1. keras image pre processing is used which can sharpen the image and also features like zoom, brightness and many more. Through image data generator we can fit the CNN model to our image dataset.

The CNN architecture as shown in figure 1 consists of several layer i.e. convolution layer, max pooling layer, flatten layer, fully connected layer. The convolution layer generally consists of number of feature maps. A convolution layer is applied on the input image in order to extract features from it. In this
transformation, the image is convolved with a kernel (or filter). The feature map is nxn matrix. This matrix is always smaller than the image matrix and this matrix is also known as convolution matrix. The length with which the feature map slides over the image matrix is called stride. Mac Pooling layer is used to reduce the size of the input image. It is made in such a way that the features extracted does not get lost in the process of pooling also somewhere making the features more robust. There are 3 types of pooling:

1. Max pooling
2. Min pooling
3. average pooling

Flatten layer is used to convert the output of the max-pooling layer into 1D array for the input to the fully connected layer to generate the output. Flatten layer basically generates number of rows of the max pool matrix.

Fully connected layer captures the relationship between the high level features and further generates the output. The output of this layer is a one dimensional feature vector. In this the activation function is generally sigmoid function.

To classify a image whether it is forged or not fully connected layer or SVC can be used but recent trend shows that the accuracy is high in case of fully connected layer in comparison with SVC.

This paper is divided into various sections as follows: Section 2 discusses about various models which can be used to implement the image forgery detection. Section 3 discusses about the various dataset used. Section 4 discusses about the process of implementation of cnn and hashing. Section 5 discusses about the various parameter used to compare the accuracy of all the models discussed. Section 6 discusses the comparison between the various models for the image forgery. Section 7 discuss about the conclusion of the comparison of all the models.

2 Models

In this section we will discuss some of the models which are used in developing an image forgery detection system. Some of the models like CNN, Hashing, DCT, DWT are discussed. All these methods are totally different from each other and have different preprocessing approaches. Some models share common dataset like DVMM and CASIA which contains spliced images and copy-move forged images respectively.
1. Image Pre-Processing

(a) According to Al-Qershi [2–4] there are some factors regarding the image which can help in easy detecting of the forgery

i. Image size is a very important factor in determining the dataset for training and testing since processing time is directly proportional to the image size. So image with smaller size will take less processing time. Sometimes authors tends to downscale the image size according to the need.

ii. Size of copy move Region plays an important role in producing the accuracy since the number of features extracted is directly proportional to the area of the copy move region. According to the recent trend there is no boundary to the duplicate region so as to nullify this disadvantage, the dataset should be reduced into three categories according to the size of the tampered area i.e. small, medium and large.

iii. Type of attack-To be able to detect forgery efficiently the dataset should contain images containing all the types of tampering whether it is copy move, adding noise like Gaussian, blur or splicing.

iv. Number of image- The dataset should contain large amount of images to train our machine effectively so as to produce good accuracy.

(b) Vishal [5] proposed a method for image pre processing in which the grayscale images were converted to binary using Otsu’s thresholding (implemented in OpenCV) after denoising using a Gaussian filter.

```python
for grayscale in xtrainmasks:
    blur=cv2.GaussianBlur(grayscale,(5,5),0)
    ret,th =cv2.threshold(blur,0,255,cv2.THRESHBINARY+cv2.THRESHOTSU)
```

(c) Rao [6–9] patches of the image dataset. He prepared the positive samples (tampered). In the tampered image patches were drawn along the edges of the spliced or copy move region. The greater the area of the tampered region in the image, the more the number of positive patches will be present.

For the authentic images the negative patches were drawn randomly. The positive and the negative sample together constitute the training set for the model CNN.

Through this patches he also prevented the over fitting of the data because of the creation of large dataset i.e. the image dataset was increased by the factor of 8.

(d) Christlein [10–12] used grayscale image for the pre processing and to convert it into grayscale image all the channels should be merged into one.

(e) Zhang [13–16] proposed a method for detection of tampering using properties of images containing shadow. Various properties of shadows enlisted by the author are as follows:

i. Shadows are comparatively darker than the surrounding region and the pixel value of the shadows should be in RGB bands.

ii. From the shadows we can locate the position of the light source as well as the strength.

iii. Shadow is always associated with the object that cast and the behaviour of the object.

iv. Shadows does not affect the surface texture of the background covered.

According to Zhang [13–16] difference in the texture of the non shadow region and shadow region indicates forgery in the spliced image. So for an image to be authentic there should be no inconsistency in the shadow and non shadow region. Mask is generated from suspicious region according to figure 2.
Figure 2: Mask Generation

(f) SS Menon [17] proposed a method of hashing to detect the image forgery and for that she pre-processed both the original and forged image. Firstly the images were resized to a specific size. Secondly they were converted into grayscale images. RGB components were used for the conversion where black and white components were extracted from the image. Then only white components were used since they are the visible ones and black components were ignored because they are the invisible components. The process of image processing is depicted in figure 3.

Figure 3: Image Preprocessing

(g) Chun Lu [18] proposed a method to overcome previous disadvantages of hashing techniques which detect geometric transformations only to some small angles. Lu proposed mesh hashing to overcome the disadvantage. Mesh hashing uses geometric properties to generate the hash values. In this proposed method triangles are constructed for hashing. Feature extraction is done using harris detector and so as to increase it’s robustness discrete wavelet transformation is done on the image. DWT transforms the image into lowest frequency sub bands as depicted in figure 4. Using DWT transformation noisy points are also removed from the salient features of the image.
Figure 4: Low Frequency Subbands [18]

(h) Tang [19] proposed a method to overcome false detection of image forgery in case of rotation of image in which we only rotate the image according to the ground level so as to balance the image. He proposed a method of ring partition to make a rotation independent matrix. Then combined with the concept of NMF (Non Matrix Factorization) which is used to reduce the dimensionality of the matrix to produce hash values. NMF is a procedure in which a matrix is reduced into base matrix and a coefficient matrix for further processing. Ring partition is process in which we divide an image into n number of rings of equal area so as to construct a secondary image which will help in making matrix for NMF. Hence the proposed system will comprise of three steps:

i. Preprocessing
ii. Ring Partition
iii. NMF

First step in the proposed system will be to normalize the image in mxm size using bipolar interpolation. This will make hashing buoyant to scaling operation. After normalization we convert RGB value of the image into YCbCr where Y = Luminance Cb = Blue Difference Chroma Cr = Red Difference Chroma For further processing we only take Y component of the image for ring partition. Conversion of RGB into YCbCr will be done using the given formula:

\[
\begin{pmatrix}
Y \\
C_b \\
C_r
\end{pmatrix}
= \begin{pmatrix}
65.481 & 128.553 & 24.966 \\
-37.797 & -74.203 & 112 \\
112 & -93.786 & -18.214
\end{pmatrix}
\begin{pmatrix}
R \\
G \\
B
\end{pmatrix}
+ \begin{pmatrix}
16 \\
128 \\
128
\end{pmatrix}
\] (1)

(i) Bajaj [20–22] acquired the signature images. Binary images at a resolution of dpi=200 were generated using scanner. But due to the presence of noise in the images, they got corrupted. There were intermediate discontinuities in the image due to the formation of isolated white pixels.

Non smooth surface of paper or the uneven spread of the ink were basically the reason for this imperfection in the image. Total number of black pixels were calculated in the each pixel of size 7 by 7 and if the threshold is less than the total number of the pixels then it is classified as 1. A lower threshold was used to fulfill the basic objective of bridging the noisy gaps. For that the value of 15 was chosen for the threshold.

2. Feature extraction

(a) Christlein [10] used block based method for feature extraction in which the image is subdivided into rectangular regions. Feature vector is evaluated for each of the region and is matched
Whereas in key point method there is no image subdivision but the feature vector is computed only on image with high entropy and then subsequently matched. Block diagram for feature extraction is shown in figure 5.

![Block Diagram of Feature Extraction](image)

**Figure 5:** **Block Diagram of Feature Extraction**

(b) Rao [6] used CNN (convolutional neural network) for image feature extraction as shown in figure 6. He used several feature maps of kernel – 4,5,3 and stride 2 to extract the features. Several convolutional layers followed by max pooling layer were present in the CNN architecture.

![CNN for Feature Extraction](image)

**Figure 6:** **CNN for Feature Extraction**

(c) Vishal [5] used custom CNN architecture as shown in figure 7 which comprises of 3 convolutional layers, 2 max pooling layer and then fully connected layer. The proposed CNN takes 128x128 as input image for feature extraction. A dropout rate of 0.2 was applied to the flattened output of 20 units. We used Adam optimizer with a default value of learning rate (0.001) and $\beta_1, \beta_2$. 

![Custom CNN Architecture](image)
Wei Lu proposed a framework as shown in Figure 8 in which he used physical methods for feature extraction. He not only extracted features in DCT (Discrete Cosine Transformation) but also used DWT (Discrete Wavelet Transformation). In this framework, expanded Markov features are extracted so as to include both intra-block correlation and inter-block relations between DCT coefficients. Extra Markov features are developed in DWT domain with the help of DWT transformation whose aim was to characterize the residual correlation by tuning the three varieties of dependency among orientations, scales, and positions.

The process of extracting the expanded features of image in DCT domain is divided in 6 parts:

i. First by applying 8x8 Discrete Cosine Transform on the original array of pixels and DCT coefficient array is generated as the output.

ii. Round off the coefficient array to integers and take only absolute values in consideration.

iii. Vertical and horizontal differences are to be calculated using formulas:

\[ F_h(u,v) = F(u,v) - F(u+1,v) \]  
\[ F_v(u,v) = F(u,v) - F(u,v+1) \]

iv. Now a threshold value T is declared \((T \in \mathbb{N})\), where if the value in \(F_h\) (or \(F_v\)) will be either greater than T or smaller than \(-T\).

v. Now vertical and horizontal transition probabilities were calculated using:

\[ P1_h(i,j) = \frac{\sum_{u=1}^{S_U-1} \sum_{v=1}^{S_V} \delta(F_h(u,v) = i, F_h(u+1,v) = j)}{\sum_{u=1}^{S_U-2} \sum_{v=1}^{S_V} \delta(F_h(u,v) = i)} \]  
\[ P1_v(i,j) = \frac{\sum_{u=1}^{S_U-1} \sum_{v=1}^{S_V-1} \delta(F_v(u,v) = i, F_v(u,v+1) = j)}{\sum_{u=1}^{S_U-1} \sum_{v=1}^{S_V-2} \delta(F_v(u,v) = i)} \]

\[ P2_h(i,j) = \frac{\sum_{u=1}^{S_U-1} \sum_{v=1}^{S_V-1} \delta(F_v(u,v) = i, F_v(u+1,v) = j)}{\sum_{u=1}^{S_U-1} \sum_{v=1}^{S_V-1} \delta(F_v(u,v) = i)} \]

\[ P2_v(i,j) = \frac{\sum_{u=1}^{S_U} \sum_{v=1}^{S_V-2} \delta(F_v(u,v) = i, F_v(u,v+1) = j)}{\sum_{u=1}^{S_U} \sum_{v=1}^{S_V-2} \delta(F_v(u,v) = i)} \]
vi. Every element in transition probability is used for feature extraction.

For detecting Markov features in DWT domain the process was divided into 3 parts:

i. He started with by applying a 3 level Discrete Wavelet Transformation to the pixel based image so as to obtain coefficient array(12 sub bands) and then rounding off the value to the closest integer and took the absolute value.

ii. 48 Markov features can be obtained by using one of the dependency i.e. position which is similar to DCT transformation. 12 sub bands when replaced by f in equation 1 then 12x2x2 features will be formed.

iii. Now another dependency will be considered i.e. of scale and horizontal and vertical difference like array will be formed using sub band $H_i$ and $V_i$

$$H_{i+1}(x,y) = \text{round} \left( \frac{H_i(2x-1,2y-1) + H_i(2x-1,2y) + H_i(2x,2y-1) + H_i(2x,2y)}{4} \right)$$

(8)

$$HV_i(x,y) = H_i(x,y) - V_i(x,y)$$

(9)

$$VD_i(x,y) = V_i(x,y) - D_i(x,y)$$

(10)

$$DH_i(x,y) = D_i(x,y) - H_i(x,y)$$

(11)

(e) Zhang [13] used Local Binary Pattern (LPB) and some aspects considered into choosing this method are texture discriminative property and low computational cost. Other than LPB there are several other methods like Gray Level Concurrence, Difference Matrix and Gabor Filtering. Through LPB we convert each pixel is converted into binary image as shown in figure 9.

(f) SS Menon [17] after the pre-processing of the both the original and forged images the hash value of the original image is generated and stored in the database. Then the the hash value of the forged image is computed. The algorithm used for the generation of the hash value is pHash function. pHash algorithm is computed in 7 steps:

i. Size Reduction

ii. Grayscale Conversion

iii. DCT Computation

iv. DCT Reduction

v. Average Computation

vi. DCT Reduction
vii. Hash Construction

(g) Chun Lu [18] after image processing is done on the image, now Delaunay tessellation is done on the image to reduce the image into packets of distorted triangles where each triangle is known as mesh. Now this mesh will be used to generate the hash value for the image to be processed as shown in figure [10]. Now for the comparison of the hash value it is mandatory that the mesh size should be of constant size. So as to convert the mesh into constant size we normalise the mesh and this process is known as mesh normalization. They used concept of affine transformation and interpolation to implement normalization.

![Figure 10: Generation of Hash Sequence](image)

(h) Tang [19] after the preprocessing of the image, Tang divided the Y component into rings of equal size. In this if an image is of nxm size then there will be n number of rings. Each ring will contain a pixel and set of these pixels will be known by \( R_k \) where k is any integer. The rings formed are of equal size because in the secondary image to be constructed, each column will be the ring formed in the Y component. Now they calculated the radius of each and every ring that is formed. Consider the radius of the \( n'th \) ring be \( r_n \), now the area of that ring will be \( A = \pi r^2 \) and the average area of each ring will be \( \mu_A \) where \( \mu_A = [A/n] \) where n is the number of partitions of ring. Now that they have the average area, they can now calculate the radius of each ring using the formula \( r_1 = \sqrt{\mu_A/\pi} \) and for any ring \( k \) the equation will be \( r_k = \sqrt{\mu_A + \pi r_{k-1}^2 k - 1}/\pi \). Now they calculated the pixel value and also the distance between the center coordinates and the pixel coordinates so as to form the set of pixel values. The number of sets formed will be n.

\[
R_k = \{p(x, y)|r_{k-1} < d_{x,y} \leq r_k\} (k = 2, 3, \ldots, n) \tag{12}
\]

Now this pixel value set is sorted in ascending order so as to make it unrelated to rotation. The sorted vector \( u_k \) will be mapped into the secondary image with the help of interpolation and final vector V will be formed.

(i) Bajaj [20] basically used three types of features to eliminate distinct shape characteristics and to draw out a more smooth structural aspects of the signature which is necessary for the detection of the forgery.

i. Horizontal and vertical projection moments
ii. Upper envelope of the signature component
iii. Lower envelope of the signature component

The characteristic of a signature is determined by the writer’s vertical and horizontal flourish at the characters or loops at the end of the signature. The nature of the X and Y point distribution is determined by both the projection image. Numeric measures for the distribution characteristic points can be provided by the skewness and kurtosis features.
3. Prediction of Forgery

(a) Vishal [5] used VGG16 architecture (figure 11) to implement the prediction of the forgery system. The predictions were done on the dataset created by converting the images into binary form and splitting the data into test and training dataset.

![VGG16 Architecture](image)

Figure 11: VGG16 Architecture [4]

(b) Rao [6] proposed a method of feature fusion to be implemented at the output of the feature extraction step. Through fusion we learn a function $f$ that transforms an input $X$ into $Y$ where $X$ is the patch in the feature extracted image and $Y$ is the condensed representation i.e. $Y=f(X)$. We compute $Y$ using a scanning window of size $p \times p$ having stride $s$ which will compute the K-D local descriptor $Y_i$. For every $i$ they calculated the K-D vector and then by applying pooling a K-D feature vector as shown in figure 12 is obtained for SVM classification.

$$Y[k] = \text{Max}\{Y_1[k]...Y_T[k]\}$$  \hspace{1cm} (13)

Where $T$ is the total number of Descriptors.
Figure 12: Generation of Feature Vector

(c) Wei Lu used SVM-RFE which uses the weights magnitude as the ranking criteria for the appropriate features to be used for the classification. The algorithm used for the classification is shown in figure.
Figure 13: **Algorithm 23**

(d) Christlein [10] after the extraction of features, matching process was used to detect the duplicate region as shown in figure 14. Similarity between the two feature descriptor indicates the presence of a duplicate region. Block based technique used the approach of lexicographic sorting. In this type of sorting the feature vector is converted into a matrix i.e. feature vector is made a row of a matrix and then sorting algorithm is applied on the matrix such that the two similar feature vectors are in consecutive rows. For key point based method Best-Bin method is used. In this method nearest neighbour algorithm is used to determine the similar features. The distance used to calculate the centroid is Euclidean Distance.

**Figure 14: Feature Matching and Filtering**

To minimize the probability of false matches, christlein used the process of filtering. Pixels
which are closer to each other sometimes have identical intensities which could lead to prediction of false forgery. To avoid such imperfections various other distance measuring techniques were tried.

Now to make cluster of similar matches they defined separate methods for separate feature extraction blocks they defined:

i. For block based approaches they used threshold value according to SATS connected area to remove the spurious matches in the matching process.

ii. For key point approach clusters are made according to the distance between the neighbours.

Now if the image has connected regions then the image is said to be tampered.

(e) Zhang [13] used a simple method to compare the two texture features extracted i.e. shadow feature and non shadow feature for the forgery detection. To compare the two texture he defined a correlation coefficient $r$ between $X$ and $Y$ which is two dimensional

$$
 r = \frac{\sum_n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{(\sum_n (X_i - X)^2)(\sum_n (Y_i - Y)^2)}}
$$

(14)

Where $X$ is the texture feature of the shadow region and $Y$ is the texture feature of the non shadow region.

Now if the value of $r$ is not close to 1 then it indicates that there is inconsistency in the shadow region and it is most likely to be a tampered region.

(f) SS Menon [17] The computed value of the forged image is now compared with the original image hash value in the database. The hash value generated is a very big number hence it is converted into binary number from hexadecimal number. If the hash value of the forged image is greater than 0.65 then it is said to be forged otherwise it is original value. A block diagram for forgery detection is shown in figure 15.

![Block Diagram for Forgery Detection](image)

(g) Chun Lu [18] converted the normalised mesh in hash value in 8x8 block DCT domain. Now to obtain the hash value the mesh is first flipped and padded to form 32x32 block. Now mapping is done from this 32x32 block to 8x8 one to generate the hash bits as shown in figure 16.
Now to classify the image into original or tampered we use bit error rate function. If the bit error function between the hash value of tampered image and hash values of the original is less than the threshold value(T) whose value lies between 0<T<1. The bit error rate function is:

$$BER(H_1, H_2) = \frac{\{t | H_i^1(t) \neq H_j^2(t)\}}{|H_i^1|} < T$$ (15)

Where $H_i$ and $H_j$ are the hash value of original and tampered image.

(h) Tang [19] used NMF on the secondary image constructed with the help of ring partition. The output of the NMF are the 2 matrix which are base matrix and coefficient matrix and they used only the latter matrix to derive the hash values. The length of the hash value is $L=nK$ where n is the number of rings and k is the rank of the NMF. Now the two images i.e the original and the suspected image will be compared to determine whether the image is tampered or not. The metric used to determine is the correlation coefficient.

$$S = \frac{\sum_{i=1}^{L} [h_1^{(1)} - \mu_1][h_2^{(2)} - \mu_2]}{\sqrt{\sum_{i=1}^{L} [h_1^{(1)} - \mu_1]^2} \sqrt{\sum_{i=1}^{L} [h_2^{(2)} - \mu_2]^2} + \epsilon}$$ (16)

Here $h_1$ and $h_2$ are the hash values of the two images and $\mu_1$ and $\mu_2$ are the means of the hash values. $\epsilon$ is used to neutralize the singularity effect in the equation. The value of s is bounded in the range of [-1,1]. The greater the value of s means that the image is not tampered and lesser the value means the image is tampered.

(i) Bajaj [20] used different classifiers for the recognition based on the feature vector. Redundancy is introduced in the scheme due to the presence of multiple classifier as shown in figure 17. With the help of multiple classifier erroneous classification by the single classifier can be compensated when the rest of the classifier is giving correct output.
Single layer feed forward net is used to combine the output of individual layer. Output of the classifier net is used as an input for the ADALINE (Adaptive Linear Element) which is present in the classifier net.

Table 1: Percentage of Error

| Error Type | Moment | Upper Envelope | Lower Envelope | Equal Weight |
|------------|--------|----------------|----------------|--------------|
| Type II    | 10.73  | 7.85           | 12.34          | 6.5          |
| Type I     | 3.33   |                |                | 3.33         |

The metrics chosen for analysing the accuracy of the system is False Acceptance and False Reject. It is applied on the different classification schemes. If a signature belongs to a person y but classified as x then the classification is considered as misclassification as shown in table 1.

3 Dataset

This section includes some of the dataset which are used in the image forgery detection system. These dataset includes several attacks like splicing, copy move, rotation, JPEG compression.

1. MICC

MICC is the oldest and most used dataset previously. It consists of 4 subsets as shown in table 2. First two subsets consists of images in which a square or a rectangle area is selected and copy pasted after several changes done to the image like scaling, rotation. Some disadvantages of this dataset are:

(a) There is no post processing like adding noise or JPEG compression.

(b) The dataset only contains the tampered image so we can’t actually generate the actual accuracy.
Table 2: MICC Dataset

| Subset         | Contents                                                                 | Description                                                                 | Size of images                  |
|---------------|---------------------------------------------------------------------------|------------------------------------------------------------------------------|---------------------------------|
| MICC-F220     | Composed by 220 images; 110 images are tampered and 110 originals         | used by the authors who build this dataset to properly set the threshold in detection algorithm | varies from 722x480 to 800x600 pixels |
| MICC-F2000    | Composed by 2000 images; 700 are tampered and 1300 originals              | Prepared for complete evaluation against different types of modifications    | 2048x1536 pixels                |
| MICC-F8multi  | 8 tampered images with realistic multiple cloning                          | Prepared to imitate real life copy-move image forgery                        | varies from 800x532 to 2048x1536 pixels |
| MICC-F600     | composed of 440 original images and 160 tampered images and 160 ground truth images | only subset with ground truth for block based detection algorithm             | varies from 800x532 to 3888x2592 pixels |

2. Image Manipulation Dataset

This dataset consists of several dataset particularly containing original images, copy move images, noise images, scaled images, rotated images, combined attack images. One of the advantage of this dataset is robustness and can be effectively used for copy move forgery. Some disadvantages of using this dataset are:

(a) Large size of the dataset which will increase processing time.
(b) Lack of combination of attacks and lack of blurred images.

3. CoMoFoD

This dataset is comparatively better than the other above mentioned dataset because of the inclusion of more intermediate attacks like distortion or post processing attacks like colour reductions as shown in figure [18]. This dataset consists of two sets i.e. small and large dataset containing 3000 and 10000 images. Main advantage of this dataset is the its small size category of image which will less time to process. Another factor which makes it a good dataset is the inclusion of the variety of the attacks used to create the dataset. Some disadvantages of the dataset are:

(a) Parameters of the intermediate attacks which are used to make the tampered image dataset is not specified. Which makes it difficult to analyse the interaction between the intermediate attacks.
(b) Size of the copy-move region varies from small to very large which will lead to high threshold value.
4. Coverage

In this dataset the forgery is evidently visible to the human being as shown in figure 19. This dataset contains images which are tampered using scaling, rotation, illumination and combined attacks. Some disadvantages of this dataset are:

(a) Large size of the dataset
(b) Area of the copy-move region is sizable which will lead to less challenging detection of forgery.

5. DVMM

This dataset contains images on which splicing is done as shown in figure 20. The number of
authentic and tampered images is same. Various types of tampering is done to the image. Several subset of this dataset are smooth vs textured , arbitrary object boundary vs straight boundary.

Figure 20: DVMM Dataset

4 Implementation

In this section we will discuss about the steps included to implement the image forgery using convolution neural network. There are 3 steps to achieve the implementation of CNN :-
1. Data Preparation
   Split the dataset into training and testing data for training of the CNN model.

2. Building the architecture of the CNN
   The architecture of the CNN as shown in Figure 21 consists of several layers which are 3 convolutional layers, 2 max pool layer, 1 flatten and 2 dense layers.

3. Prediction of forgery
   Give an image as an input to the trained model and the output will be classified into two categories which are tampered or original.

**System Requirements**

1. Software Requirement
   - Anaconda 2019 or more
   - Python 2.7 or more
   - Tensorflow 1.3 or more

2. Hardware Requirement
5 Metrics

In image forgery detection different parameter are taken into consideration which will help in comparing the performance of various models and will help in deciding the best suitable approach. Different metrics considered are :-

1. Image Based Evaluation

This type of metrics is used when we don’t have original or authentic image to compare our result with. When we have only tampered image dataset we use this metric. Measures used at the image based evaluation are :-

- True Positive (TP)
  Images which are detected correctly as forged image
- False Positive (FP)
  Images which are detected wrongly as forged image.
- False Negative (FN)
  Images which are falsely missed as forged image
- True Negative (TN)
  Images which are detected correctly as original image.

2. Pixel Based Evaluation

This type of evaluation metrics requires authentic image dataset to properly distinguish between the tampered image and the original one. The two categories made in the detection according to the pixel level are copy move and authentic image. The output generated can be assigned to one of the four category depicted through a confusion matrix as shown in figure 22.

![Figure 22: Pixel Based Evaluation](image)

3. Correlation Coefficient

Correlation coefficient is a numerical measure which will determine the correlation between two variables. The two variable can be a piece of data or some multivariate variables. Correlation
coefficient value falls in the range [-1,1]. Where -1,1 are the strong relation agreement and 0 is the strong relation disagreement. Correlation can be visualized as shown in figure 23.

![Correlation Coefficient](image)

Figure 23: Correlation Coefficient

6 Comparison

1. Quantitative Comparison

Below given tables shows the comparison between the various image forgery detection algorithm like CNN, some mathematical transformations. It was noticed that algorithm like CNN was giving more accuracy as compared to other models. Table 3 shows various features used for image processing.

| Reference | Features |
|-----------|----------|
| 23        | Markov features in DCT and DWT domain |
| 6         | Patch based features |
| 10        | Intensity, Moments , keypoint |
| 13        | Shadow Texture |
| 17        | Black and White Components |
| 18        | Salient Points |
| 19        | Luminous Components of RGB value |
| 20        | Moment and Envelope |

Table 4 shows various attacks handled by the authors.
Table 4: Attacks

| Reference | Attack Handled |
|-----------|----------------|
| [23]      | Copy-Move , Splicing |
| [6]       | Copy-Move , Splicing |
| [5]       | Copy-Move |
| [10]      | JPEG Compression , Gaussian White Noise , Rotation |
| [13]      | Copy Move-Splicing |
| [17]      | Morphed |
| [18]      | JPEG Compression , Gaussian Noise , Scaling , Shearing |
| [19]      | JPEG Compression , watermark , Rotation |
| Proposed System | Copy-Move |

Table 5 shows various accuracy of the methods on the DVMM dataset.

Table 5: Accuracy on DVMM dataset

| DVMM Dataset | Approach | Accuracy |
|--------------|----------|----------|
|              | DCT      | 90       |
|              | DWT      | 86       |
|              | DCT + DWT | 93.15   |
|              | CNN      | 96.38    |
|              | CNN + VGG | 94.65   |

Table 6 shows various accuracy of methods on CASIA dataset.

Table 6: Accuracy on CASIA dataset

| CASIA Dataset | Approach       | Accuracy |
|---------------|----------------|----------|
|               | DCT + DWT      | 89.76    |
|               | CNN (Proposed) | 98.2     |
|               | CNN            | 97.19    |

Table 7 shows accuracy based on pixel evaluation.
Table 7: Accuracy based on Pixel evaluation

| Approach                                         | Accuracy |
|-------------------------------------------------|----------|
| DCT                                             | 90       |
| DWT                                             | 86       |
| DCT + DWT                                       | 93.15    |
| CNN                                             | 96.38    |
| CNN + VGG                                       | 94.65    |
| CNN + ResNet50                                  | 95.09    |
| Block Tiling and Keypoint Detection             | 81.62    |
| DWT + Mesh                                      | 90.5     |
| Neural Classifiers                              | 94.65    |

Figure 24 shows a quantitative comparison of all the methods.

2. Qualitative Comparison

The study of the above mentioned models revealed some fascinating advantages of one model over other models.

(a) CNN requires less image preprocessing as compared to other approaches.
(b) CNN gave better accuracy on copy move forgery.
(c) CNN is more easy and flexible to implement.
(d) CNN is domain specific
(e) No post processing is required.

7 Conclusion

From the above discussed models like DCT with accuracy 90 \%, DWT with accuracy 86 \%, DCT+DWT with accuracy 93.15 \%, CNN with accuracy 96.3 \%, CNN+VGG with accuracy 94.65 \% and our proposed implementation done on dataset, the accuracy acquired was 98.2 \%. It was concluded that the CNN is the best among all the methods for the implementation of the forgery detection. The accuracy is thus far better for CNN and has a lot of advantages as compared to other models. The image processing required in CNN is less which makes the system easy to handle.

8 Future Scope

The proposed system currently helps in classifying the image into original and tampered image. In the near future we will be focusing on the detecting the area where the image is tampered and comparing both the original and tampered image on the basis of type of tampering.

References

[1] Sumit Saha. A comprehensive guide to convolutional neural networks-the eli5 way. \[towardsdatascience.com/\], Dec 2018.
[2] Osamah M. Al-Qershi and Bee Ee Khoo. Evaluation of copy-move forgery detection: datasets and evaluation metrics. \textit{Multimed. Tools Appl.}, 77(24):31807–31833, 2018.
[3] Antonio Torralba and Alexei A. Efros. Unbiased look at dataset bias. In \textit{Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.}, pages 1521–1528. IEEE Computer Society, 2011.
[4] Charles E. Metz. Basic principles of ROC analysis. \textit{Semin. Nucl. Med.}, 8(4):283–298, 1978.
[5] Vishal Singh. Image forgery detection. \[https://towardsdatascience.com/image-forgery-detection-2ee6f1a65442/\], Feb 2019.
[6] Yuan Rao and Jiangqun Ni. A deep learning approach to detection of splicing and copy-move forgeries in images. \textit{8th IEEE Int. Work. Inf. Forensics Secur. WIFS 2016}, pages 1–6, 2017.
[7] Chih Chung Chang and Chih Jen Lin. LIBSVM: A Library for support vector machines. \textit{ACM Trans. Intell. Syst. Technol.}, 2(3), apr 2011.
[8] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In \textit{MM 2014 - Proc. 2014 ACM Conf. Multimed.}, pages 675–678. Association for Computing Machinery, Inc, nov 2014.
[9] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. \textit{Proc. IEEE}, 86(11):2278–2323, 1998.
[10] Vincent Christlein, Christian Riess, Johannes Jordan, Corinna Riess, and Elli Angelopoulos. An evaluation of popular copy-move forgery detection approaches. \textit{IEEE Trans. Inf. Forensics Secur.}, 7(6):1841–1854, 2012.
[11] Hany Farid. Image forgery detection, 2009.

[12] Ching Yung Lin, Min Wu, Jeffrey A. Bloom, Ingemar J. Cox, Matt L. Miller, and Yui Man Lui. Rotation, scale, and translation resilient watermaking for images. *IEEE Trans. Image Process.*, 10(5):767–782, may 2001.

[13] Yongzhen Ke, Fan Qin, Weidong Min, and Guiling Zhang. Exposing image forgery by detecting consistency of shadow. *Sci. World J.*, 2014(d):1–10, 2014.

[14] Alin C. Popescu and Hany Farid. Exposing digital forgeries in color filter array interpolated images. *IEEE Trans. Signal Process.*, 53(10 II):3948–3959, oct 2005.

[15] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognit.*, 29(1):51–59, 1996.

[16] Marko Heikkilä, Matti Pietikäinen, and Cordelia Schmid. Description of interest regions with local binary patterns. *Pattern Recognit.*, 42(3):425–436, mar 2009.

[17] Sruthi S Menon and Mary N Saana J. Image Forgery Detection using Hash Functions. *International Journal of Recent Technology and Engineering (IJRTE)*, pages 3498–3501, 2019.

[18] Chun-Shien Lu, Chao-Yong Hsu, Shih-Wei Sun, and Pao-Chi Chang. Robust mesh-based hashing for copy detection and tracing of images. In *2004 IEEE International Conference on Multimedia and Expo (ICME)* (IEEE Cat. No. 04TH8763), volume 1, pages 731–734. IEEE, 2004.

[19] Zhenjun Tang, Xianquan Zhang, Shichao Zhang, and Senior Member. Based on Ring Partition and NMF. *IEEE Transactions on Knowledge and Data Engineering*, 26(3):711–724, 2014.

[20] Reena Bajaj and Santanu Chaudhury. Signature verification using multiple neural classifiers. *Pattern Recognit.*, 30(1):1–7, 1997.

[21] Réjean Plamondon and Guy Lorette. Automatic signature verification and writer identification - the state of the art. *Pattern Recognit.*, 22(2):107–131, 1989.

[22] Jon Atlí Benediktsson and Philip H. Swain. Consensus Theoretic Classification Methods. *IEEE Trans. Syst. Man Cybern.*, 22(4):688–704, 1992.

[23] Zhongwei He, Wei Lu, Wei Sun, and Jiwu Huang. Digital image splicing detection based on Markov features in DCT and DWT domain. *Pattern Recognit.*, 45(12):4292–4299, 2012.

[24] Irene Amerini, Lamberto Ballan, Roberto Caldelli, Alberto Del Bimbo, and Giuseppe Serra. A SIFT-based forensic method for copy-move attack detection and transformation recovery. *IEEE Trans. Inf. Forensics Secur.*, 6(3 PART 2):1099–1110, sep 2011.

[25] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Mach. Learn.*, 46(1-3):389–422, 2002.

[26] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S. C. Zhu. On advances in statistical modeling of natural images. In *J. Math. Imaging Vis.*, volume 18, pages 17–33, jan 2003.

[27] Courtney Taylor. Calculating the correlation coefficient. thoughtco.com/how-to-calculate-the-correlation-coefficient-3126228/