IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 5, Issue 2, May 2021
Impact Factor: 4.819

Framework For Image Forgery Detection And Classification Using ML

Harshada Nande[1], Akash Mhaske[2], Sonali Gadakh[3], Jayshri Pawar[4], Prof. Pravin Avhad[5]
Students, Department of Computer Engineering[1,2,3,4]
Professor, Department of Computer Engineering[5]
Shri Chhatrapati Shivaji Maharaj College of Engineering, Nepti, Maharashtra, India

Abstract: *In the recent times, the rates of cybercrimes have been surging prodigiously. It has been proven incredibly easy to create fake documents with powerful photo editing soft-wares being as pervasive as ever. Documents can be scanned and forged within minutes with the help of these softwares that have tools readily available just to do that. While photo manipulation software is handy and ubiquitous, there are also means to deftly investigate these morphed documents. This paper lays a foundation on investigation of digitally manipulated documents and provides a solution to distinguish original document from a digitally morphed document. A Graphical User Interface(GUI) was created for detection of digitally tampered images. This method has accuracy of96.4 % and has proven to be efficient and handy.*

Keywords: Artificial Neural Networks; GLCM features; Graphical User Interface; Machine Learning ; Support Vector Machine.

I. INTRODUCTION

With increasing accessibility to a wide range of printing devices, printed documents have paved their way into our daily transactions. These days most of our legal and official documentation is maintained in printed mode. Printed records are not only employed for maintaining financial contracts and judicial testimonies, but have a widespread use in personal identification documentation as well. Nevertheless, with easily accessible technological support, modification and alteration of printed documents, for malicious purposes, has become a frequent event. As a consequence, organizations maintaining important documents are promoting strict protocols to counter measure against document forgeries. A common procedure is to restrict official document printing to specific formats and printers. This enables document analysts to employ source printer identification techniques, to discriminate between documents printed by different printers. However, with the amount of printed documentation in use, nowadays, manual examination of suspicious documents is becoming a tedious and cost inefficient task.

Now a days, people are using fake documents such as voting cards, driving licences, PAN cards, Aadhar cards, and passports etc. for there benefits. This paper is used to create a false identity in order to commit fraud and scam. As a result, we are developing this device to detect fraud documents and reduce fraud and scam.

Image forgeries may be classified into many types- such as copy-move forgery, splicing and many more. Research has been going on in this field for years now and many effective methods have been proposed to detect such forgeries. Xudong Zhao et al. proposed a method for colour channel design to find the most inequitable channel, which they called the optimal chroma-like channel, for feature extraction.

We are going to use following UML Diagrams:

1. Use Case Diagram
2. Activity Diagram
3. Sequence Diagram
4. Class Diagram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 5, Issue 2, May 2021
Impact Factor: 4.819

II. Related Work

In this paper, we present a new image forgery detection method based on deep learning technique , which utilizes a convolutional neural network (CNN)to automatically learn hierarchical representations from the input RGB color images. The proposed CNN is specifically designed for image splicing and copy move detection applications. Rather than a random strategy, the weights at the first layer of our network are initialized with the basic high-pass filter set used in calculation of residual maps in spatial rich model(SRM), which serves as a regularizer to efficiently suppress the effect of image contents and capture the subtle artifacts introduced by the tampering operations. The experimental results on several public datasets show that the proposed CNN based model out perform some state of the art method.

III. Proposed Work

In this section, we elaborate the proposed technique for feature extraction and classification, for source printer identification. As discussed in previous sections, most of the proposed solutions in literature are text-dependent (i.e. same content for training and testing is required). This approach has limitations in most real-world scenarios. Therefore, we selected a text independent approach towards source printer characterization, by employing random patches from document for feature extraction purposes. By employing transfer learning on pretrained CNN architecture, the extracted patches of a printed document are classified .Both feature extraction and fine tuning are applied, for this purpose. A set of popular textural features is also extracted from patches and used to train a number of classifiers.

IV. System Architecture

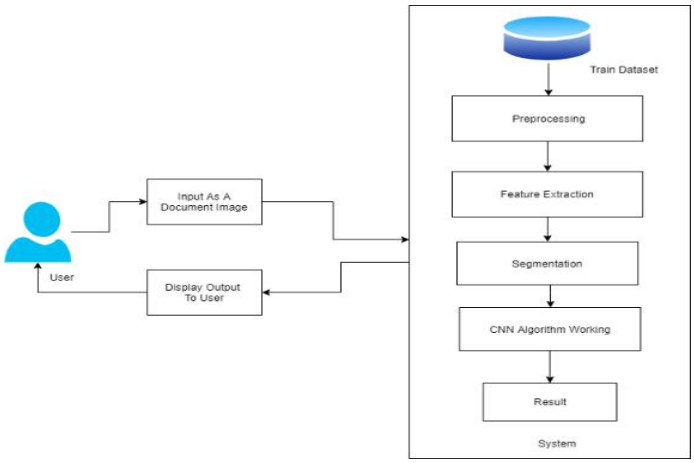

Figure 1: System Architecture

4.1 Creation of Dataset

The images used for this project were collected from various internet sources and morphed using photo editing tools. These images were edited using Adobe Photoshop CC 2017 to create a dataset with 120 pairs of images- one original and its edited version. These images (total 240 in count) were used in further analysis using MATLAB R2015a.

4.2 Pre- Processing of the Images

To make the details of the images stand out more, the query image was enhanced using histogram equalization. It is a necessary step because sometimes minute forgeries go undetected through the entire process. It is important that the machine gets most of the details in one go. Histogram equalization, as the name suggests, is a method, where the intensities are adjusted using the histogram of the image. This technique is used here for contrast enhancement. Another essential stage in the pre- processing of an image is the removal of noise. De-noising is again done so that the details of

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 5, Issue 2, May 2021
Impact Factor: 4.819

the image are sharper and are not missed while extracting the features of the image. In this paper, de-noising is done using the median filter in MATLAB.

4.3 Segmentation

The image is segmented into 3 clustering by using k-means clustering. K-means clustering is a technique for quantizing vectors. This method divides the image into k segments, each containing mutually exclave data. This is a common method when it comes to pattern recognition and machine learning. One of the segmented images is chosen on the basis of the information contained in it. To determine this, the GLCM features of each segment are calculated and the segment with the highest mean is chosen. The GLCM of the segmented image are then compared with the original image using cross-validation, which gives another array, which is studied to determine whether an image is morphed or not, and function for the final result is added on the basis of that.

4.4. Extraction of GLCM Features

Out of all the methods to analyse an image, extraction of GLCM features has proven to be efficient time and time again. The gray level co-variance matrix is a tabulation that provides with statistical measures for texture analysis. This method takes into account the spatial relationship between the intensities of pixels in a gray-level image. In this paper, the GLCM features were calculated to study the differences in the original image and the digitally forged image. This gave 22 texture values (for each image) to work with, most of which were similar when it came to an image and its fraudulent counterpart. In practice, this would lead to redundancy and would also increase the time to run the algorithm.

4.5 Classification

Initially, the classifier used for classification of dataset into two parts as original or morphed was linear kernel SVM. A linear kernel SVM is the most suitable classifier for two-class classification problems. It finds an equivalent hyper plane which separates the whole data by a specific criterion that depends on the algorithm applied. It tries to find out a hyper-plane which is far from the closest samples on the other side of the hyper plane while still classifying samples. It gives the best generalization techniques because of the larger margin. The accuracy obtained through linear SVM was low in this case (87.6%). So, Artificial Neural Network (ANN) classifier was applied on the dataset. ANN networks are basically a system of interconnected neuron like layers. The interconnection of the network can be adjusted based on the number of available inputs and outputs making it ideal for a supervised learning. Hidden layers were chosen as 5 for our dataset.

4.6 Creation of GUI

Designing of a Graphical User Interface (GUI) was deemed necessary because one had to repeatedly check whether an image had been tampered or not. To do this, the drag-and-drop GUIDE Layout Interface in MATLAB was used. Once, the front-end design was complete, a modified the backend code of the interface was coded, which allowed to program the functions into the push buttons to give the required results. For example, the first push button was used to load the image onto the GUI; therefore, few lines were added to the code which allowed us to upload an image at the front-end.

A. Use Case Diagram with Necessary Information

Use case diagram is used to show which operations are performed by the user and which operation are performed by the system.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 5, Issue 2, May 2021
Impact Factor: 4.819

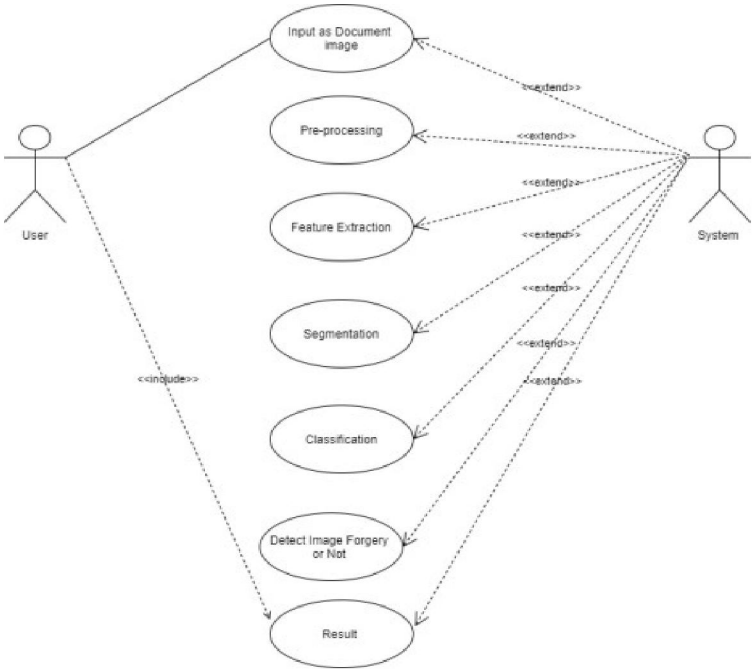

Figure 2: Use Case Diagram

B. Activity Diagram with Necessary Information

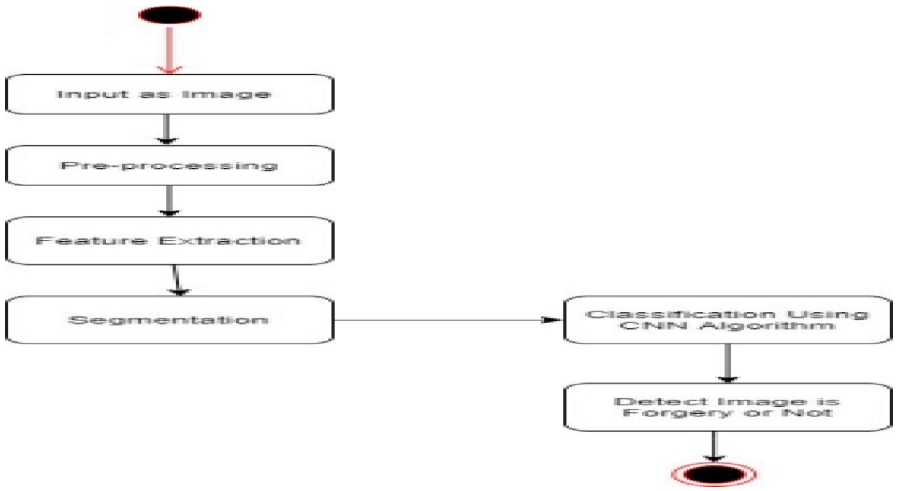

Figure 3: Activity Diagram

Activity Diagram shows the active flow of the system. In above diagram the flow of our project is shown actually how the data flow.

C. Sequence Diagram with Necessary Information

In sequence diagram step by step sequence of steps is shown. In above diagram first preprocess all train data and test data. Then by applying the train data train the machine and build the module and at the last apply machine learning algorithm on it . For testing purpose apply the test data on module and see the classification either fake or real.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 5, Issue 2, May 2021
Impact Factor: 4.819

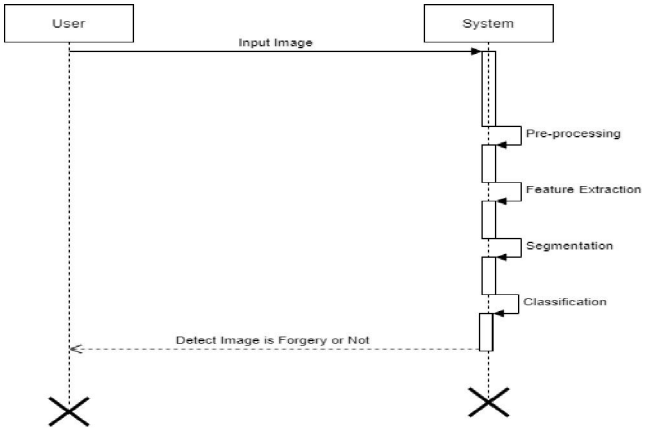

Figure 4: Sequence Diagram

D. Class Diagram with Necessary Information

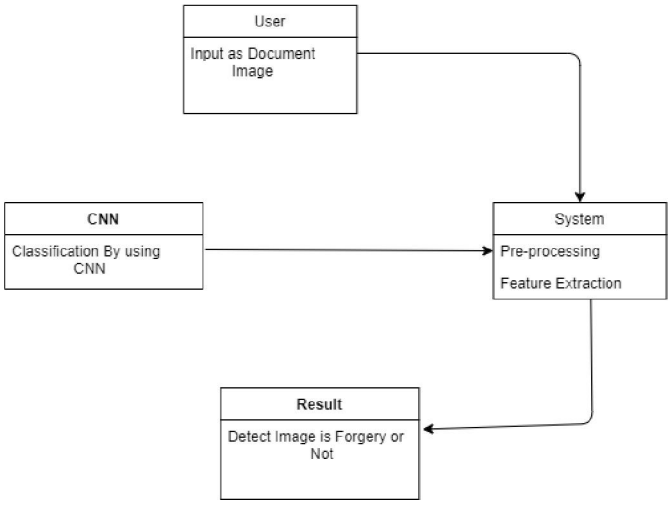

Figure 5: Class Diagram

V. ALGORITHM

5.1 Steps of CNN Algorithm

1. **Convolutional Layer:** Convolutional Layer is the first ConvLayer is responsible for capturing the Low-Level features such as edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features
2. **ReLu Layer:** It is activation function is responsible for transforming the summed weighted input from the node into the activation of the node or output for that input.
3. **Pooling Layer:** Pooling layer is responsible for reducing the spatial size of the Convolved Feature.
4. **Fully Connected Layer:** Adding a Fully-Connected layer is a (usually) cheap way of learning non-linear combinations of the high-level features as represented by the output of the convolutional layer. The Fully-Connected layer is learning a possibly non-linear function. A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered with enough training, convNets have the ability to learn these filters/characteristics.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 5, Issue 2, May 2021
Impact Factor: 4.819

VI. ADVANTAGES

1. One of the main advantages of image is wide availability of powerful digital image processing tools.
2. The advantage of using these features is rotational invariance ansd simplicity.
3. Good efficiency and accuracy.

VII. CONCLUSION

In this paper, We presented a text-independent approach to detect document forgery using source printer identification. The proposed technique relies on extraction of patches from document images which are then employed for feature extraction. We investigated the performance of our patch level technique using both textural and deep learned features. Although classification in both scenarios is performed on patch level, how- ever decision is taken on document level by applying majority voting. Highest classification accuracy on our proposed source printer identification technique, is achieved by applying fine-tuning on a pre-trained ConvNet. We can use these advanced image editing tools in our further extension of the project to implement the required results more easily and instantly. While these tools are mostly used in the creative design related areas, criminals also can easily get access to them and as a result, can exploit them to create fake identities to hide themselves in public, or to commit a crime.

ACKNOWLEDGEMENT

We take this opportunity to express my hearty thanks to all those who helped me in the completion of the project stage-1 on this topic. We would especially like to ex- press my sincere gratitude to Prof. P. S. Avhad, my Guide and Prof. J. U. Lagad HOD Department of Computer Engineering who extended their moral support, inspiring guidance and encouraging independence throughout this task. We would also thank our Principal Dr. M. P. Nagarkar for his great insight and motivation. Last but not least, we would like to thank my colleagues for their valuable suggestions.

Harshada Nande, Akash Mhaske, Sonali Gadakh and Jayshri Pawar

REFERENCES

[1] Francisco Cruz, Nicolas Sidere, Mickael Coustaty, Vincent Poulain ¨ D'Andecy, and Jean-Marc Ogier. Local binary patterns for document forgery detection. In Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on, volume 1, pages 1223– 1228. IEEE, 2017.

[2] Anselmo Ferreira, Luca Bondi, Luca Baroffio, Paolo Bestagini, Jiwu Huang, Jefersson A dos Santos, Stefano Tubaro, and Anderson Rocha. Data-driven feature characterization techniques for laser printer attribution. IEEE Transac- tions on Information Forensics and Security, 12(8):1860–1873, 2017.

[3] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabi- novich. Going deeper with convolutions. In Proceedings of the IEEE confer- ence on computer vision and pattern recognition, pages 1– 9, 2015.

[4] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbig- niew Wojna. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recogni- tion, pages 2818–2826, 2016.

[5] Tao Tsai, Yuadi and Yin. Source identification for printed documents. In 3rd IEEE International Conference on Collaboration and Internet Computing (CIC), pages 54–58, 2017

[6] A. Thakur, N. Jindal, Multimedia Tools and Application, Image Forensics Using Color Illumination, Block and Key Point Based Approach, (2018); 77: 26033.

[7] Anil Dada Warbhe, Rajiv V. Dharaskar, Vilas M. Thakare, "Digital image forensics: An affine transform robust copy-paste tampering detection", Intelli- gent Systems and Control (ISCO) 2016 10th International Conference on, pp. 1-5, 2016.

[8] Badal Soni, Pradip K. Das, Dalton Meitei Thounaojam. (2018) CMFD: a detailed review of block based and key feature based techniques in image copy-move forgery detection. IET Image Processing 12:2, pages 167-178.