Image Forgery Detection System

Gokul Ravi Iyer, Maitrey Gholap, Prathamesh Ghude, Harsh Gaikar
Guide: Mrs. Aasha Chavan
Department of Computer Engineering,
Ramrao Adik Institute of Technology, DY Patil, Deemed to be University,
Nerul, Navi Mumbai, Maharashtra, India
Email: gok.iye.rt23@dypatil.edu, mai.gho.rt23@dypatil.edu, har.gai.rt23@dypatil.edu,
pra.ghu.rt23@dypatil.edu, aasha.chavan@rait.ac.in

***Abstract* -- Document forgery is a growing issue for private enterprises and public administrations. It can be said that it is a sign of the wastage of time and resources. There are plenty of classical solutions to such issues like the detection of an integrated security pattern. In such a scenario, it is necessary that we use forensic techniques for the detection. The concept of using these forensic techniques can also be applied using artificial intelligence/machine learning which is of lower cost and can give the same or better result. The experimental result shows that different models have good detection ability to detect multiple forgeries. In this paper, we have proposed a different approach of forgery detection in a document. The forgery that we are detecting can be considered as hand-written signature forgery and copy-move forgery of any photo, text, or signature. We have proposed a new approach using capsule layers to detect a forgery in handwritten signatures. We also use ELA (Error Level Analysis) to detect any error in the compression levels of the image.**

Keywords: Document, Forgery Detection, Capsule Neural Networks, CNN, Copy-Move Forgery, Signature Forgery

I. INTRODUCTION

We now live in an era of technology where it has become a focal point of our everyday life, from creating, processing, and storing information to displaying user wants and needs. Knowledge representation is multidimensional and is kept in bits and bytes in various forms such as texts, pictures, and films. Technology is not always a good thing. People are misusing technology nowadays in a way that is harmful to society. Changing data without traces and authentic proof of intervention has never been simpler. With this, there are many opportunities as well as risks, and the seriousness of the threats cannot be overemphasized.

Impersonating or modifying artifacts, like important papers, photographs, news reports, artwork, etc., is forgery. It is always accompanied with other fraudulent behaviour like check fraud, insurance fraud, identity takeover, and smuggling. Fake social media profiles and adapted email correspondence are some examples of technological forgeries which are not always physical. We have observed a huge hike in demand for online document authentication in e-commerce and e-government scenarios because of the pandemic's ongoing problem. Documents are uploaded to internet platforms for a number of reasons. But some editing software or other technology can modify the document's content. It is necessary to detect the modifications made to document photocopies. Because pictures can be small fragments of forensic evidence or can be taken to a court. Authentication of documents is extremely necessary. Forgery process of reproducing or modifying objects which can be important documents, photographs, news, artwork, etc. It is always accompanied by other fraudulent behaviours like check fraud, insurance fraud, identity takeover, smuggling, and many more. Forgeries need not always be physical; they can be electronics too like forged social media pages or adaptation of email content. Since the worry of our subject matter is documented forgery, we will exhibit the limelight on various types of document forgeries. Real passports with no personal information or with no stamp, or any signed memorandum letters with no content can be called real documents because they are blank documents. The forger fills up the data to carry out the fraudulent act. The forgeries in these kind

Documents can be tricky to check for changes. Spotting any tampering in document images matters a lot because these images might be used as legal proof in investigations or other situations. So, it's really important to confirm that a document is real.

II. Types of Image Forgery

- Copy-Move Forgery:

Copy-move forgery is a common way to mess with digital images. It happens when someone takes a part of an image and sticks it somewhere else in the same picture. This trick is usually used to hide things or copy parts of the image to mislead people while keeping the rest of the image looking the same, like the lighting and texture. Because both the original and copied parts look alike, it can be tough to spot the changes. To catch this kind of forgery, experts use various methods like comparing blocks of pixels or analysing key points, along with some advanced learning techniques.

- Splicing:

Then there's splicing, which is when two or more pictures are combined to make a fake image. Unlike copy-move, splicing mixes in different content, so it's often easier to spot. But it can still be tricky since the lighting, noise, and resolution from the original images might not match perfectly.

- Image Retouching:

Image retouching is all about making small tweaks to photos to improve how they look or to change certain details without adding or taking away a lot of content. Spotting these changes can be trickier compared to other types of photo manipulation because the edits are typically, global and fine-grained. These methods check for issues like differences in image statistics, contrast levels, and lighting effects. Lately, deep learning techniques are being used more often to spot editing patterns that we might not notice right away.

- **Deepfake Generation:**

AI can now swap faces and edit videos in a pretty realistic way. To spot deepfakes, experts look at things like how the face moves, blinking, head position, and any weird spots in texture or lighting. The best detection tools use deep learning models that have been trained on tons of real and fake media to catch those small mistakes made when creating deepfakes.

III. DATASET

Finding the right dataset for training and testing a machine learning model can be tough. The performance of the model really relies on the quality of the data used. Unfortunately, it’s not easy to find a dataset with forged document images. So, we decided to create our own by gathering real document images from the internet and then altering them to build our forged image dataset. This dataset will serve two purposes: one part is for training a model to detect copy-move forgeries, and the other is for detecting signature forgeries. Since we made the dataset manually, it needed some preprocessing. We used OpenCV for this, as image noise can really impact the model's results. The preprocessing steps included correcting slants and removing noise. We also created different models to tackle different types of features, so we had to extract those specific features too. The signature forgery detection model takes the extracted signature from the uploaded image for analysis.

IV. PROPOSED WORK

The implemented system consists of preprocessing, capsule network, error level analysis, and CNN.

a) Capsule neural network: To detect the forgery of signature, a capsule neural network is used instead of CNN. Despite the benefits of CNN, there are always challenges in using them. CNNs do not have the ability to model the changes in new fields and dimensions and cannot distinguish between similar components that are placed in different locations of an image. The reason to use CapsNet is that it can reduce the number of layers and parameters in the network architecture and decrease complexity. CapsNet is capable of detecting the details of angular and spatial changes in components. Not only is a capsule network able to represent the existence of particular visual features, but also it can detect the transformations that might have occurred in the features. In capsule networks, spatial details are completely differentiated.

b) Data preprocessing: preprocessing is done to reduce the noise in the uploaded image, the presence of which might result in false positives or lesser efficiency of the detection model. Preprocessing phase plays a vital role in the product implementation since the errors or unwanted data present in this phase will be carried out till the final stages of the product development. Cleaning the data in the earlier stages will improve the system. The pre-processing stage consists of slant correction, cropping of the unwanted region from the input document image, and image denoising. The slant correction and denoising are done using the OpenCV technique

c) Error level analysis and CNN: To convert an image to ELA, the preprocessed images must be reframed at a certain level of quality. The image is whitened or brightened as a result of this technique. In order to reframe the images, forged and real images are considered that have been processed in preprocessed. Finally, the preprocessed image and reframed image are compared to see the difference between them. The tampered parts of the image in the forged ELA framed image are brighter than the corresponding original segments of the image.

Having ELA analysis before the processing of the neural networks has a huge advantage as the ELA reframed image contains only non redundant information and nearby pixels are compared based on similar intensity. The reframed image needs to be resized to make the RGB values lie between 0 and 1 so each cell value can be normalized by dividing the values by 255. The resizing helps the neural network converge faster than usual. CNN model consists of 2 layers of a convolutional network. The first layer of the CNN is a convolutional layer with 32 filters. Dropout is used as a regularization technique for reducing overfitting in neural networks, preventing complex co-adaptations on training data. The CNN consists of 2 convolution layers, max pooling, and 2 dropout layers

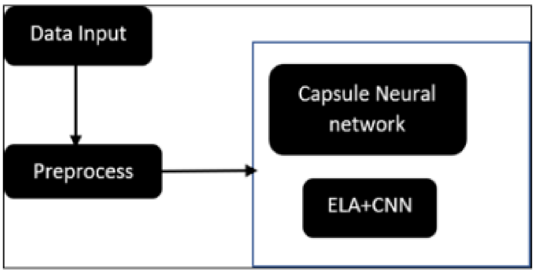


Fig1.System architecture

The forgery detection results come from looking at what each model finds. These models work together in an ensemble. Since each one checks for fakeness in different features, we need to put their results together to get a full picture of how fake the document is. When we combine everything from these models, we get the overall fakeness score as the final output.

V. RESULTS AND DISCUSSIONS

The training images for the copy-move model, CASIA-2 and MICC dataset have been considered. The dataset consists of images from these two datasets which are split into training and testing datasets and then passed onto our model to classify them into two classes i.e. authenticate and forged.

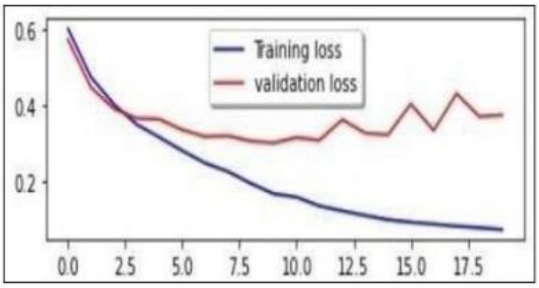


Fig 2: Training loss and Validation Loss of Copy Move Forgery Detection Model

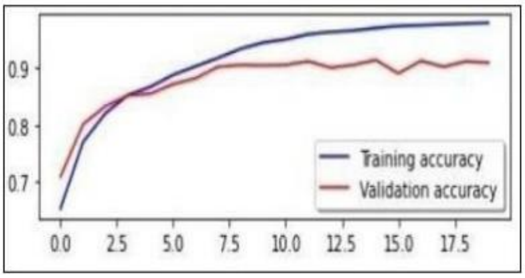


Fig 3: Training Accuracy and Validation Accuracy of Copy Move Forgery Detection Model

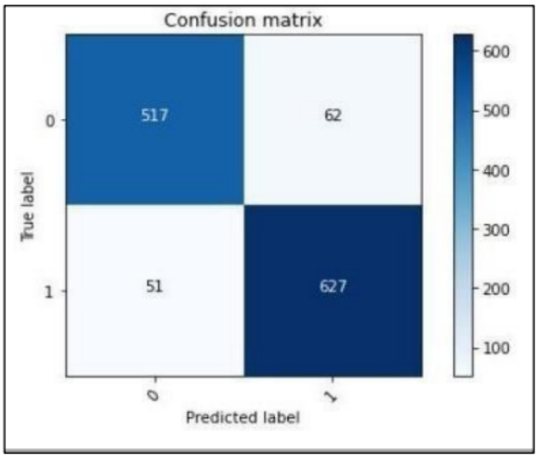
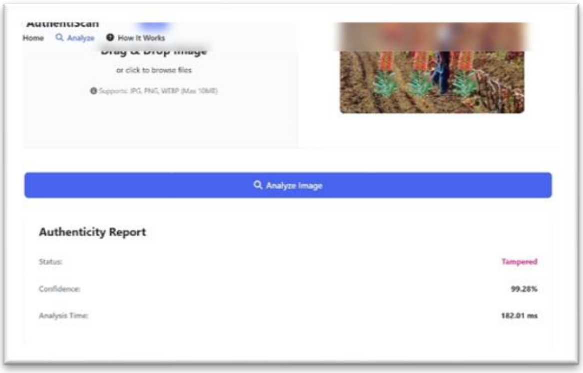


Fig 4: Confusion Matrix of Copy Move Forgery Detection Model



Authenticity Report

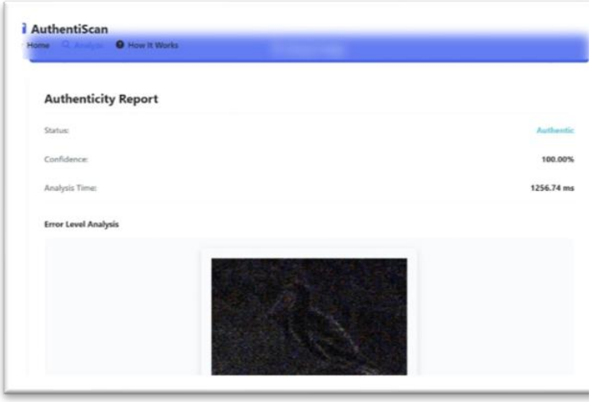
Status: Tampered

Confidence: 99.28%

Analysis Time: 182.01 ms

Fig 5: Result for Tampered image

The system uses methods like Error Level Analysis (ELA) to search for variations in image compression and structure. The tool, upon processing the image, generated an authenticity report with the status labelled as "Tampered" and a high confidence level of 99.28%. The analysis was done within 182.01 milliseconds, convincingly indicating that the image has been digitally manipulated. It is a reflection of how automated forensic tools can effectively detect tampered digital content with accuracy and speed.



Authenticity Report

Status: Authentic

Confidence: 100.00%

Analysis Time: 1256.74 ms

Error Level Analysis

Fig 6: Result for Authentic image

AuthentiScan uses advanced forensics like ELA to detect image manipulation. This image was verified as "Authentic" with 100% confidence in 1256.74 ms, confirming its reliability for digital forensics.

VI. CONCLUSION

In this paper, a method to detect a digital document forgery using a neural network has been presented. The proposed solution uses a capsule neural network for signature forgery detection and a combined model of ELA and CNN is used for copy-paste forgery detection. The proposed method uses features from two datasets of varying difficulty. The experiment results validate that the classification performance decreases when the samples are more challenging. However, the implemented architecture does not easily generalize to datasets with different underlying distributions. The capsule neural network was found to be efficient in detecting the forge-ness of the signatures in the document images. ELP was found to be efficient with detecting copy-move forgery detection. The overall positive and negative forgery detection rate of the designed system was very promising. It is encouraged that future work is required to investigate completely on a wider range of algorithms with even higher efficiency and also to detect the fraudulent use of video [MP4] format files. Overall, the project work has given us an opportunity to dig deeper into forgery detection methods. Image forgery detection is not only an emerging topic in research but an important topic in which faster and more accurate work is needed. The accurate and efficient methods to detect forgery are increasing day by day. Nevertheless, there is surely a lot of work still to be done in the image forgery detection domain and neural networks will be able to detect tampered images regardless of their difficulty. In the future, there is scope to improvise the training model by increasing the variance in the dataset and also it has scope to detect the forgery in video files.

VII. REFERENCES

[1] Rahiche, Abderrahmane; Cheriet, Mohamed (2020). [IEEE 2020 IEEE/Cvf Conference on Computer Vision and Pattern Recognition Workshops (Cvprw) - Seattle, Wa, USA (2020.6.14-2020.6.19)] 2020 IEEE/Cvf Conference on Computer Vision and Pattern Recognition Workshops (Cvprw) - Forgery Detection in Hyperspectral Document Images Using Graph Orthogonal Nonnegativematrixfactorization., (),2823-2831.Doi:10.1109/Cvprw50498.2020.00339 [CrossRef]

[2] S. Jain, M. Khanna, and A. Singh, “Comparison among different CNN architectures for signature forgery detection using Siamese neural network”; 2021 international conference on computing, communication, and intelligent systems (ICCCIS), 2021, pp. 481- 486, DOI:10.1109/ICCCIS51004.2021.9397114. [CrossRef]

[3] Robust forgery detection for compressed images using CNN supervision Boubacar Diallo *, Thierry Urruty, Pascal Bourdon, Christine Fernandez- Maloigne Universite de Poitiers, cnrs, xlim, umr 7252, f-86000 Poitiers, France, volume 2, December 2020, 100112. [CrossRef]

[4] IJCSNS International Journal of Computer Science and Network Security, vol.20no.12 December 2020: https://doi.org/10.22937/ijcsns.2020.20.12.12 [CrossRef]

[5] Zhang, Zhongping & Zhang, Yixuan & Zhou, Zheng & Luo, Boundary-based image forgery detection by fast shallow CNN, conference: computer vision and pattern recognition Doi: 10.11.09/ICPR (2009).