CERTIFICATE FORGERY DETECTION FOR CYBERSECURITY

Dept of CSE- BNMIT

Name of Project Guides:

1. Name: Dr. Chayadevi M L
   Email id:chayadevi1999@gmail.com
   Contact No.:9901794101

2. Name: Dr. Jalaja G
   Email id:jalajag.shaker@gmail.com
   Contact No.: 9019431845

Name of Team Members:

1. Name: SUJAY L GOWDA
   USN No.: 1BG19CS108
   Email id:sujaylgowda@gmail.com
   Mobile No.:8970750189

2. Name: SRUTHI S
   USN No.: 1BG19CS106
   Email id:sruthish01@gmail.com
   Mobile No.:7823978829

3. Name: SHREYAS K
   USN No.: 1BG19CS098
   Email id:shreyas14dec@gmail. com
   Mobile No:9066306435

Keywords: Certificate, Forgery, MobileNet v2, ANN, SVM, Real, Fake, Detection.

Introduction

Certificate forgery is a growing problem in today's digital age. With the widespread use of online platforms for education and professional certifications, it has become easier for fraudsters to create and distribute fake certificates. One machine learning method that could be used to develop a certificate forgery detection system is a Convolutional Neural Network (CNN). CNNs have been shown to be effective in image-processing tasks, such as object recognition. They can be trained to identify patterns and features in images that are indicative of forgery or manipulation. The real certificates were collected and fake ones were made. The data was preprocessed and were made sure that they were in the required specific format (jpg, jpeg, png).

The system could also be designed to flag suspicious certificates for manual review by a human expert. Overall, the development of a certificate forgery detection system using machine learning has the potential to greatly improve the security and reliability of certification processes, helping to ensure that certificates are only issued to those who have earned them through legitimate means.

Objectives

The objectives for our project are as follows:

1. Collect and curate a large dataset of both authentic and forged certificates for training and testing the system.
2. Develop a machine learning-based algorithm that can accurately detect forged certificates.
3. Implement the certificate forgery detection system using state-of-the-art machine learning tools and frameworks.
4. Design an intuitive and user-friendly interface for the system to enable easy integration with existing certification processes.
5. Evaluate the performance of the system in terms of accuracy, efficiency, and scalability, and optimize it for real-world use cases.

By achieving these objectives, we can significantly contribute to the reliability and integrity of certification processes, helping to prevent fraud and misrepresentation, and enhancing the reputation of institutions and organizations that issue certificates and degrees.

Methodology

The forgery detection system involves the following steps: data collection and preprocessing of genuine and forged certificates, feature extraction and selection to identify relevant discriminative features, and training a machine learning model using selected features and labeled data as shown in Fig 1.

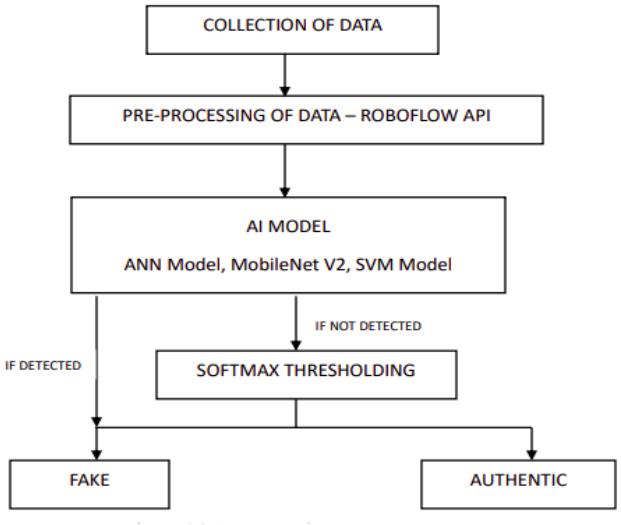

Fig 1 Methodology for the Forgery Detection System

Once the relevant features are selected, various machine learning algorithms such as SVM, Random Forest, or CNN can be used to train the model. The model is trained using the selected features and labeled data.

The dataset is captured and stored in our system by using a Raspberry Pi board and a PiCamera sensor. Raspberry Pi boards are excellent for capturing images using the Raspberry Pi Camera Module or external USB cameras.

PiCamera is a Python library that provides a simple and convenient interface for accessing and controlling the Raspberry Pi Camera Module. It allows you to capture images and record videos using the camera module connected to your Raspberry Pi.

PiCamera simplifies the process of interacting with the Raspberry Pi Camera Module, allowing you to easily capture images and record videos using Python code.

The dataset for feeding into the model consists of genuine and forged certificate images. Genuine certificates were collected from our friends, while the forged certificates were created by editing the genuine certificates in image editing software such as Adobe Photoshop. The dataset includes a range of forgery techniques such as VTU Emblem change, signature alteration, image manipulation, and many other features. The images were preprocessed to ensure that they were of a consistent size and format. The dataset was then split into training and testing sets for model development and evaluation. Fig 2. and Fig 3. is the dataset gathered by our team.

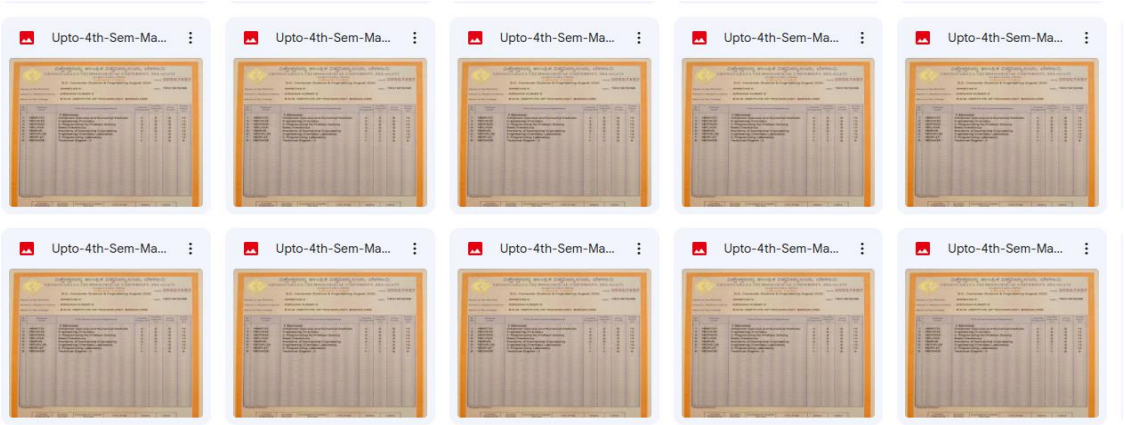

Fig 2. Collection of Real VTU Result Cards

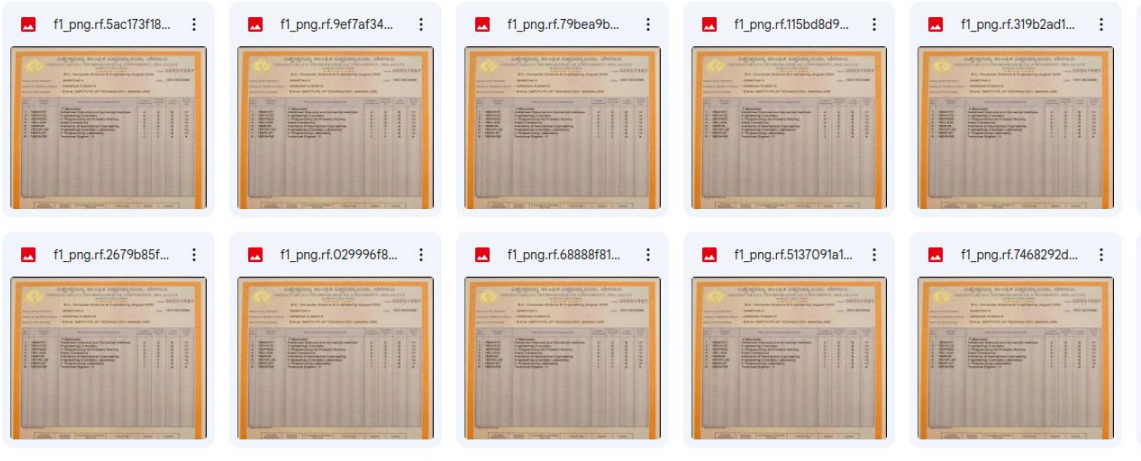

Fig 3. Collection of Fake VTU Result Cards

Roboflow is a data preprocessing tool that allows us to prepare our image data for machine learning tasks such as object detection, classification, and segmentation. The tool offers a range of features for data preprocessing, including image resizing, data augmentation, format conversion, and annotation as shown in Fig 4.

In the case of our project, the data preprocessing step using Roboflow could involve resizing the images to a consistent size and format, converting them to a standard format such as JPEG, JPG, or PNG, and annotating the images to identify the relevant regions of interest such as the signature, VTU Emblem and Karnataka Flag Symbol.

Fig 4. Roboflow UI for labeling

Results

The results obtained will be both in pictographic as well as in word form. The output image obtained will show the detected labeled features in the same and based on that the code will return whether the certificate is real or fake. The input image for example Fig 5 is fed to the model. The required accuracy and other metrics for the model is calculated and the output is given as shown in Fig 7. The output image is shown in Fig 6 with labeled features denoting whether the features are real or fake. The fed image was real and similarly shown in the output.

Fig 5. Input Image

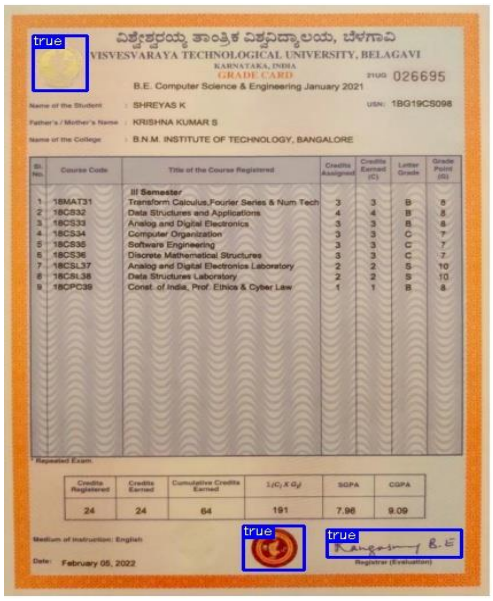

Fig 6. Output Image

```
from PIL import Image
img=Image.open('/content/drive/MyDrive/imgs/1/Upto 4th Sem Markscard_page-0003.jpg')
img=data_transforms(img)
img = img.unsqueeze(0)
output=model(img)

[7] if output[0] > output[1]:
        print("The certificate is fake")
    else:
        print("The certificate is real")

    The certificate is real
```

Fig 7. Code

Conclusion

The following are the conclusions drawn from the project:

- The use of machine learning techniques can effectively detect certificate forgeries with high accuracy.
- Preprocessing techniques such as image resizing, normalization, and enhancement can improve the performance of the models.
- The choice of features and model architecture can significantly impact the performance of the system.
- A combination of multiple models, such as ANN, MobileNet V2, and SVM, can improve the overall accuracy of the system.
- The SoftMax thresholding technique can be used to further improve the accuracy of the system by adjusting the decision threshold.

Innovation

The certificates are generally detected using the Register number or by cross-checking with the database. The method we have used is for alteration or tampering of visible features such as stamp, emblem, or signature. These can be detected using our model which is fast and easy. The best model used is MobileNet v2 which has an accuracy of the 97%. This model is used for object detection and feature detection.

Scope for Future Work

The following are the possible future enhancements that can be made:

- The dataset used in this project was relatively small and limited in scope. A larger and more diverse dataset could be used to further improve the performance of the models.
- The system could be extended to detect other types of document forgeries, such as ID cards or passports.
- Other preprocessing techniques, such as data augmentation or noise reduction, could be explored to further improve the performance of the models.
- The use of deep learning techniques, such as convolutional neural networks, could be explored to further improve the accuracy of the system.
- The system could be integrated with a larger document management system to provide real-time forgery detection and prevention.
- The dataset used in this project was relatively small and limited in scope. A larger and more diverse dataset could be used to further improve the performance of the models.
- The designed system can help in separating the fakes from the authentic certificates. This can be used in many sectors which can be during recruiting, intake of students, teachers and many more.
- The system can be patented and paper can be published on the designed system.
- The hardware used can be enhanced and scaled out to fit many certificates and easy analysis can be done of the same to distinguish between the real and fakes.
- Detection of forgery is of high importance in society. It is for the common good of the public and can be used to detect real certificates so that jobs or seats can be given based on merit.