Document Classification and Information Extraction with Deep Learning techniques

Gunti Lahari
IIT Hyderabad
ai22btech11008@iith.ac.in

J Hima Chandh
IIT Hyderabad
ai22btech11009@iith.ac.in

S Divija
IIT Hyderabad
ai22btech11026@iith.ac.in

K Anuraga Chandan
IIT Hyderabad
ai22btech11011@iith.ac.in

***Abstract*—In today's information-driven era, automating the process of extracting information from identity cards has become essential across various sectors, including banking, telecommunications, healthcare, and hospitality. This report focuses on leveraging computer vision techniques to develop a system for identity card classification and information extraction. By employing basic image processing methods, the system achieves efficient identification of identity card types and accurate localization of essential information, enhancing the overall speed and reliability of the process. Key challenges include mitigating noise interference from camera hardware, lighting conditions, and shadow grid lines, which obscure critical details. Additionally, extracting foreground content from complex backgrounds remains a significant concern. This study addresses these challenges, proposing a robust pipeline that streamlines identity card recognition and facilitates accurate information management, particularly for widely-used identity certificates such as ID cards. The proposed system demonstrates potential for real-world applications, ensuring precise and rapid document handling.**

***Keywords*—Classification, Information Extraction, CNN (Convolutional Neural Network), Hough transform , Canny edge detection,Rotation Correction and Face Detection Gaussain Blur, Haar Cascade Classifier Haar Features, Adaboost Algorithm , Attention, Data Augmentation, Padding, ReLU, Dropout, Cross Entropy Loss, momentum, SGD**

I. Introduction

In the digital age, identity verification and information extraction from ID cards have become critical tasks across various domains such as telecommunications, healthcare, banking, hospitality, and government services. Automating this process enhances operational efficiency, reduces human effort, and minimizes errors. With advancements in computer vision and machine learning, systems for identifying and extracting relevant information from identity documents have gained widespread attention. These systems help streamline workflows, improve user experience, and ensure accurate handling of sensitive information.

Identity card classification and information extraction involve several challenges, such as dealing with noise introduced by camera hardware, uneven lighting conditions, shadow grid lines, and complex backgrounds that obscure important details. Additionally, the presence of varying card designs, fonts, and languages necessitates robust methods to ensure adaptability and accuracy across different identity document types.

This work aims to address these challenges by developing a system that employs computer vision techniques to classify identity cards and extract relevant information efficiently. The system leverages basic image processing methods for faster processing while maintaining accuracy, making it suitable for real-time applications. The scope of this study includes tackling problems such as noise interference, shadow removal, and foreground extraction in complex backgrounds, particularly focusing on widely-used identity certificates like Chinese ID cards.

The research presented in this report lays the foundation for future advancements in identity document automation. By addressing existing challenges, the proposed system has the potential to enhance data management, improve user interactions, and contribute to the development of secure, reliable identity verification solutions

II. Our Work

Firstly we tried to implement paper[1]. In this, firstly ID card is detected and cutout of the background then standardized in direction and size . This is based on image processing techniques to speed up the model training. The steps involved are :

1) Performing the detection of straight lines (edges of rectangle)
2) Expanding the straight lines in the step 1
3) Detecting the corners of the rectangle based on step 2

This involves a two-step approach to process identity cards: **locating the ID** and **standardizing the ID**. The ID is located by applying HSV-based color filtering ($84^\circ \leq H \leq 141^\circ$, $8\% \leq S, V \leq 100\%$) to generate a binary image, followed by morphological transformations to link clusters and identify the largest connected region as the ID. The smallest bounding rectangle enclosing this region is calculated using the Ramer-Douglas-Peucker algorithm, minimizing the area by iteratively constructing lines around the polygon. To standardize orientation, red color regions ($-15^\circ \leq H \leq 15^\circ$, $50\% \leq S, V \leq 100\%$) are filtered, and morphological erosion reduces noise. The orientation is determined by analyzing the average distances of key features (e.g., text and emblem) from the long edges of the ID, with a perspective transformation aligning it upright. Finally, the image is cropped and resized to standard dimensions, ensuring uniformity for further processing.

Fig. 1. Contours detection

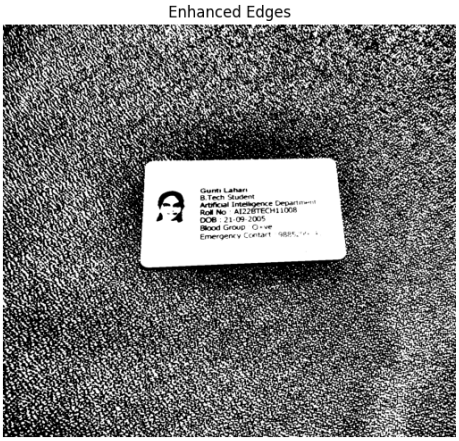

Fig. 2. Binary image

The second stage of the system identifies rectangles containing information fields on a standardized ID card image. By leveraging the consistent font and green background properties, color filtering in the RGB model isolates information pixels, as the background has high G and B values while the content has low R, G, and B values. A grayscale image is constructed, histogram equalization is applied, and thresholding generates a binary image. Morphological operations refine the result, contours are detected, and bounding boxes are drawn around the contours. Large boxes are further normalized by dividing them into reasonable proportions for accurate field detection.

But this we couldn't implement because of our training set images doesn't of different background for detecting contours. We tried to paper[2]. This paper workflow involves

A. Image Rotation Correction

This uses the Hough transform to correct the rotation of tilted ID card images. The method detects straight lines in the grayscale image, calculates their angles, and rotates the image to achieve horizontal alignment. By leveraging the rectangular structure of ID cards, the side with an inclination angle less than 45° is used for correction, ensuring effective alignment for images with tilt angles below 45°.

Fig. 3. Rotation

ID Card Region Positioning Using Face Detection

The process of ID card region positioning involves detecting key elements, such as the national emblem and face, followed by manually identifying the approximate coordinates of text. This technique is crucial for locating the ID card area within a complex background, forming the basis for subsequent text detection.

Haar Features for Feature Extraction: Haar features are fundamental in image processing, primarily for detecting edges, lines, and textures. These features are calculated by comparing pixel intensities in rectangular regions, making them efficient for object detection tasks, such as face detection. The feature value is computed as:

$$\text{Feature Value} = \sum whitepixels - \sum blackpixels$$

The feature value is determined by convolving Haar kernels with a sliding window, where white pixels are assigned a value of $+1$ and black pixels -1. Haar eigenvalues reflect grayscale variations, indicating the gradation changes in the target image. For a given feature x and pixel values in an area p, the eigenvalue is computed as:

$$\text{Eigenvalue} = W_{\text{white}} \sum whitepixels - W_{\text{black}} \sum blackpixels$$

Adaboost Algorithm for Training Object Detectors: The Adaboost algorithm is employed to train the object detector by combining weak classifiers into a strong classifier. Weak classifiers are those with a recognition rate slightly better than random guessing, whereas strong classifiers achieve the desired recognition rate. The algorithm cascades multiple weak classifiers, selecting effective features (e.g., grayscale distributions of the target image) from a large pool of Haar features.

By leveraging Haar features and the Adaboost algorithm, this method efficiently detects objects and identifies regions of interest, such as ID cards, in diverse environments.

Fig. 4. Face Detection

ADABOOST ALGORITHM

- Given: $T = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $y_i \in \{0, 1\}$
- Number of iterations: M
- Initialize weights: $D = \{w_{1,1}, w_{1,2}, \ldots, w_{1,n}\}$ where

$$w_{1,i} = \frac{1}{n}$$

For $m = 1, 2, \ldots, M$:

1) Train the weak classifier $G_m(x)$ using the dataset with D_m.
2) Calculate the error rate of $G_m(x)$:

$$e_m = \sum_{i=1}^{n} w_{m,i} \mathbb{I}(G_m(x_i) \neq y_i)$$

3) Compute the weight of G_m in the strong classifier:

$$\alpha_m = \frac{1}{2} \ln \left(\frac{1 - e_m}{e_m} \right)$$

4) Update the weights:

$$w_{m+1,i} = w_{m,i} \exp(-\alpha_m y_i G_m(x_i))$$

where

$$Z_m = \sum_{i=1}^{n} w_{m,i} \exp(-\alpha_m y_i G_m(x_i))$$

and normalize:

$$w_{m+1,i} = \frac{w_{m+1,i}}{Z_m}$$

Final classifier:

$$F(x) = \text{sign} \left(\sum_{m=1}^{M} \alpha_m G_m(x) \right)$$

This gives the final classifier and face is detected

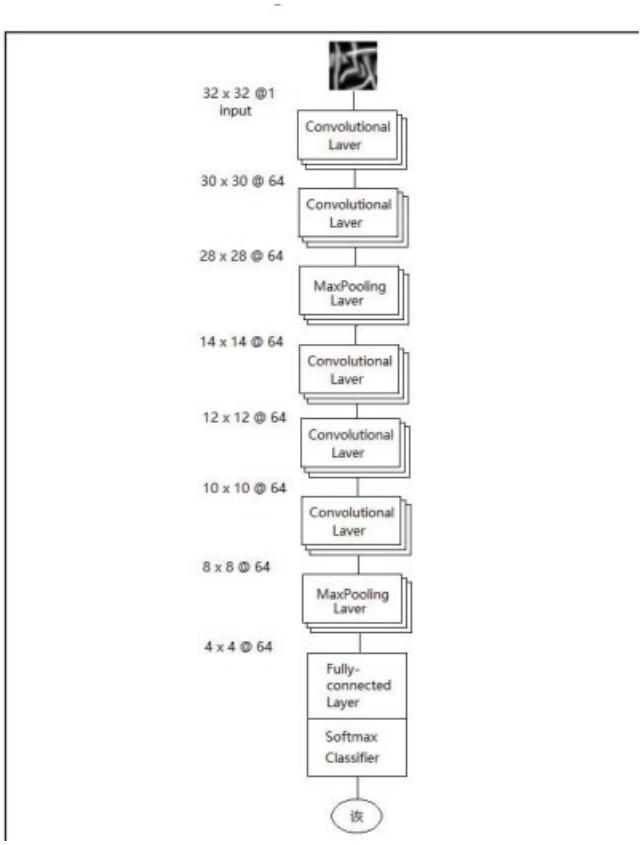

Fig. 5. Enter Caption

B. Text Detection using CNN

After we have the coordinates of the text location we should recognise the text. For this pupose we should train a model that can classify and recognise the Numbers and Alphabets. So we had considered EMNIST data which has english alphabet and numbers to train the model.

We have pre-processed the input image to enchance quality and prepare them for analysis. It invlolves differentiating the background from the foreground content, Removing noise introduced by camera, lighting conditions and shadow grid line, resizing the processed images to the required input size for CNN, ensuring compatibility for classification tasks.

The model architecture is as follows:

Stride 1, Padding 1,Max pooling(kernel size 2*2, stride 2), kernel with size 3x3 for all layers

Network Architecture:

- Convolution layers:
  1) *First Layer*:
     * 64 kernels
     * ReLU activation
  2) *Second Layer*:
     * 64 kernels
     * ReLU activation
- For MLP input, we need to flatten the feature maps of the last convolutional layer.

- MLP (Multi-Layer Perceptron) layers:
  1) *First Layer*:
     * 100 neurons for output
     * ReLU activation
  2) *Second Layer*:
     * 62 neurons for output
     * ReLU activation
- We apply the softmax function to get the label of the input image.

Training:

- We use Cross Entropy loss.
- For optimization, we use Adam (learning rate = 0.001).

III. Results

We have trained our model using 10 epochs and the obtained loss 0.8913 for character recognisation.

References

[1] Jianxing Xu,Wing Wu, School of Computer Engineering and science, Shanghai University "A System to Localize and Recodnize Texts in Oriented ID Card Images ".

[2] Nguyen Thanh Cong, Nguyen Dinh Tuan, Tran Quac Long, Faculty of Information Technology, VNU Univerisity of engineering and Technology "Information Extraction from ID Card via Computer Vision Techniques".