

Author Identification using Deep Learning

Prof. Deepa Bendigeri¹, Krishnaprasad Mahale², Praveen Kalakoti³, Abhilash Hiremath⁴,
Kushal Harti⁵

¹Prof. Deepa Bendigeri, Professor, Department of Information Science and Engineering, SDM CET Dharwad, Karnataka, India

²Krishnaprasad Mahale, Student, Department of Information Science and Engineering, SDM CET Dharwad, Karnataka, India

³Praveen Kalakoti, Student, Department of Information Science and Engineering, SDM CET Dharwad, Karnataka, India

⁴Abhilash Hiremath, Student, Department of Information Science and Engineering, SDM CET Dharwad, Karnataka, India

⁵Kushal Harti, Student, Department of Information Science and Engineering, SDM CET Dharwad, Karnataka, India

Abstract - The research paper offers a new solution and an improved and more efficient technique for Writer Identification. Writer Identification is the process that helps to find the author of a specified document by comparing it to other documents of the specific writer, which have been previously stored in a database. Convolutional Neural Network (CNN) is used for text classification and detection. Various experimental methods show that this approach produces efficient and nearest to accurate results than other Writer Identification methods.

Key Words: Writer Identification, writer retrieval, Convolution Neural Network, K Nearest Neighbor, feature networks.

1. INTRODUCTION

The process of identifying the author of the handwritten document and comparing handwriting with others that are stored in a database is known as **Writer Identification**. It is a must that the authors of the documents with the most resemblance handwriting are fetched usually these are the documents are written by the same authors.

The documents are sorted according to the distance and for identification, the authors of the documents with the highest resemblance (nearest distance) are then assigned as the writer of the documents this process is followed for retrieval of the writer. Handwritten author identification is a very well-researched subarea and vital benchmark task within the research field due to its various practical applications and financial implications. Because of a variety of potential applications such as the reading of postal codes, medical prescription reading, interpreting handwritten addresses, processing bank cheques, credit authentication, social welfare application forms.

For instance, when some anonymous pieces of handwritten characters are found at a crime site, and it is possible to automatically identify that the writer might be

left-handed which reduces the set of suspects to be investigated. One clear example of this happens in the classification of gender. Even though feminine writing is more circular and uniform than masculine one, there are some examples masculine writing may exist with a feminine appearance. This could be another exact topic in the field of Handwritten Author Identification for future work.

The exceptions for writer identification and writer retrieval consist of a change in the ink of the pen or change of the pen, change in the writer's style of writing, change in the environment of writing, many distractions like noise, unusual position of the writer. These are some of the challenges faced which makes the retrieval of data and hence alter the fetching of the documents from the database in which the previous handwritten documents of the specific writer are stored. The following figure shows the change in the writing style of the writer due to disturbance or distraction.

2. Tools and Technologies used

We are making use of Python for System design.

Front End Tools:

1. Tkinter: This package is used for building GUI

Back End Tools:

TensorFlow

Keras

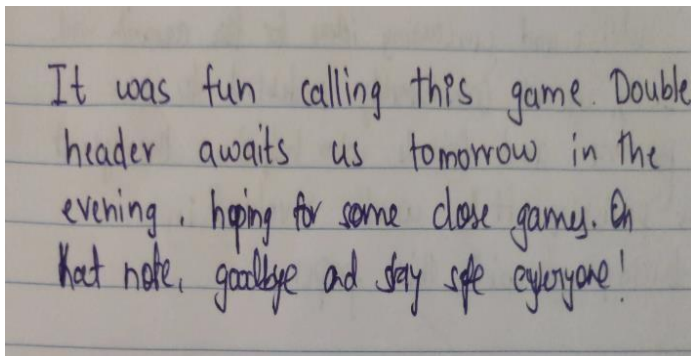


Fig 1. Image shows the change in the handwriting when the writer got distracted

3. Methodology

Our project is designed by making use of Convolution Neural Network which is one of the major research areas in machine learning. We need to generate the feature vector of images in every document to compare the handwriting of different authors with the one that is already there in the database which we have stored as the datasets. We have to undergo different steps or processes before comparing the image with the image that is stored in a database. We have to undergo different filtering techniques to preprocess the image. We need to give an input image to CNN which is of standard size, if the image is large or small, we need to convert it to standard size using different processes. Then we are training the CNN with IAM datasets. We have taken 5000 images of 50 different authors for the training the Convolution Neural Networks. For classification of the images we are using K Nearest Neighbor algorithm which is also called as the KNN algorithm.

3.1 Preprocessing

In the beginning, have to check if the image that is given as an input is grayscale-like, then binarization. For binarization Otsu method is used because the input images have documents that don't contain noises and the segmentation process is done after that. All the images are reduced to some standard size in preprocessing using which feature vector is generated. Making use of lines for calculating handwriting's height and have to mark the upper point of the handwriting and the lower point of handwriting based on that images are reduced or enlarged to standard size. In preprocessing slanted or slightly angled handwritten images are straightened first and then feature vectors are generated. Here the main process is concerting the colored images to black and white images also called grey images for making it easy for predicting the results for the next steps.

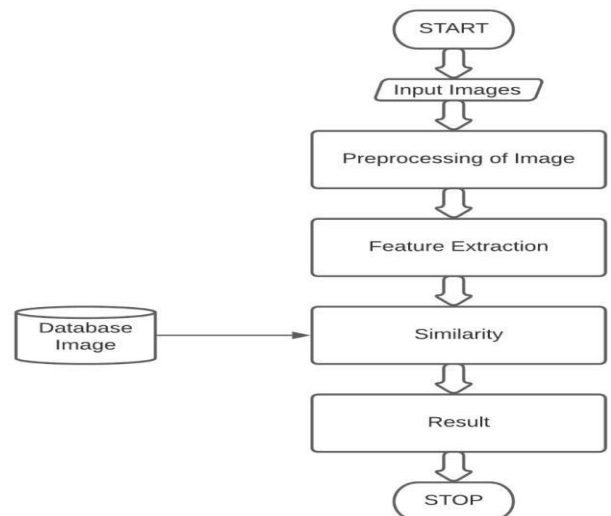


Fig 2: Flow chart of the design

3.2 Feature Extraction

We are making use of the Convolution Neural Networks for generating feature networks. Images are divided into small pixels with uniform sizes. We are training CNN model with the datasets. Data sets with different handwriting styles and different angled handwriting images are trained to CNN. Then the new images are given as the input to CNN for testing.

4. CONCLUSIONS

This paper helps to identify the writer by his handwriting using neural networks. The handwriting of the writer must be present in the database before identification. The input data which has the most similarity with that of in the database is displayed

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to several individuals and organizations for supporting us throughout our project phase. First and foremost, we would like to express our sincere thanks to our supervisors Dr. Jagadeesh Pujari and Dr. Vandana S Bhat, Department of Information Science and Engineering, SDMCET Dharwad. We also wish to express our sincere thanks to our project coordinator Prof. Deepa Bendigeri for her invaluable suggestions, practical advice, and ideas for research work. Lastly, we are considerably indebted to our parents and friends who helped us throughout in publishing the work successfully.

REFERENCES:

- [1] Salma Shofia Rosyda and Tito Waluyo Purboyo "A Review of Various Handwriting Recognition Methods," International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 2 (2018) pp. 1155-1164
- [2] Vincent Christlein, Martin Gropp, Stefan Fiel†, and Andreas Maier "Unsupervised Feature Learning for Writer Identification and Writer Retrieval" Pattern Recognition Lab, Friedrich-Alexander-Universitat Erlangen-N umbernberg, 91058 Erlangen, Germany " Austria
- [3] Vincent Christlein, David Bernecker, Andreas Maier, Elli Angelopoulou "Vincent Christlein, David Bernecker, Andreas Maier, Elli Angelopoulou" Pattern Recognition Lab University of Erlangen-Nuremberg, Germany.
- [4] Vincent Christlein, Andreas Maier "Encoding CNN Activations for Writer Recognition" Pattern Recognition Lab, Friedrich-Alexander-Universitat Erlangen-N umbernberg, 91058 Erlangen, Germany

BIOGRAPHIES



Krishnaprasad Mahale
Student, Department of Information Science and Engineering, SDM CET Dharwad, Karnataka, India



Praveen Kalakoti
Student, Department of Information Science and Engineering, SDM CET Dharwad, Karnataka, India



Abhilash Hiremath
Student, Department of Information Science and Engineering, SDM CET Dharwad, Karnataka, India



Kushal Harti
Student, Department of Information Science and Engineering, SDM CET Dharwad, Karnataka, India



Prof. Deepa Bendigeri
Professor, Department of Information Science and Engineering, SDM CET Dharwad, Karnataka, India