Handwriting Authorship Recognition using Convolutional Neural Networks

Shravani Patel, Daanish Baig*, Dasi Sanjay, Ashwin Kothari

Department of Electronics and Communications Engineering, Visvesvaraya National Institute of Technology, Nagpur, India

**shravanipatel9@gmail.com, daanishbaig@hotmail.com, thesanjayoffl@gmail.com, ashwinkothari@ece.vnit.ac.in*

Abstract

With the advent of imaging and communication technology, files consisting of high-quality images of handwritten text grow increasingly common in academia, business, education, medical sciences, forensic sciences and various other fields. It is often necessary to ascertain the authorship of such data, especially in the cases of education and forensics. Human-based authorship recognition is fairly reliable, but time-consuming and tedious. In this paper, we explore the possibility of using pre-existing convolutional neural network architectures like ResNet50 as well as data augmentation and image pre-processing to accurately identify authorship of a given page of handwritten text. As a result of this work, we were able to train the model and achieve an accuracy of over 80% on a subset of the Firemaker dataset. Overall, we conclude that neural networks have strong potential for rapid and accurate author identification, however, long training times and the large networks required for this level of accuracy somewhat hamper the system.

Keywords. Convolutional neural networks, authorship recognition, handwriting, residual neural networks, ResNet50, machine learning.

1. INTRODUCTION

The rapid growth and combination of old methods of data saving (handwriting) with more modern techniques (digital files) has resulted in a situation where handwritten data is being broadcast as an image, with the intent of it being read as handwritten data. This differs from pure textual data as a person's handwriting is dependent on various criteria such as personality, age, upbringing, education, parents, surface and scribing tool, nervous system health etc. [1-2]. Generally, we consider the case of a generic ballpoint pen on paper. As there are a large number of factors that could potentially affect an individual's handwriting, and these factors could not all be the same across every individual, it is possible to identify individuals based on the unique style of their handwriting, similar to how fingerprints are different for every person and uniquely identify them.

Handwriting identification is a lengthy and time-consuming process. First and foremost, one needs examples of handwriting from known individuals — the more available, the better, as it covers more of the fringe and edge cases that could reasonably be expected from natural variations in an individual's handwriting. One also needs a diverse database with as many samples from different sources as possible. Additionally, the data should be uniform so that the analysis takes place without needing to consider variations in scribing tool, surface, noise, damage to text etc. Due to the complexity and intricacy of these techniques, they present a suitable target for automation using convolutional neural networks (CNN), support vector machines (SVM) and other automated machine learning (ML) based approaches [3].

This paper uses an existing architecture (ResNet50) which has previously been used for signature verification by various parties [4]. The accuracy of over 90% in this use case was a strong suggestion that it might be suit-able for usage in a broader authorship-recognition environment. The aim of our work is to develop a system that achieves a reasonably high degree of consistent accuracy (80%) over a suitably challenging training and testing dataset. Additionally, only a single page of handwriting is used for training, and the testing and training datasets consist of different texts. The images are segmented, thresholder and resized before being used to train the CNN, which greatly improves accuracy of the results. Further, data augmentation in terms of angle and noise are performed to simultaneously add resistance to rotation and noise, and also to increase the amount of effective training data available.

The applications of the proposed method range across a variety of fields — for example, in academia, it could be used to verify that students are submitting their own work. In forensics, it could suggest connections between writings and people. In medical science, it might work as a diagnostic tool to detect degradation of fine motor skills or neurological issues.

The rest of the paper is structured as follows - Section 2 consists of an explanation of CNNs and ResNet50. Section 3 pertains to the dataset used, as well as the processing applied to it. Section 4 covers the implementation of the machine learning aspect, and various choices made with regards to parameters, learning rate, etc. Section 5 covers the results of the training, and Section 6 concludes the paper.

2. Convolutional Neural Networks

A convolutional neural network is a specialized form of an artificial neural network (ANN). As stated by O'Shea et al. in [5], The key difference be-tween ANN and CNN is the neuron organization. In a CNN, the neurons are organized into three dimensions, which consist of height and width of input along with depth i.e. activation volume of the ANN. Neurons of any layer may only connect to a reduced subset of the layers neighbouring, unlike the behaviour of ANNs.

These operations can be represented by convolutions, and can quickly and easily be performed using modern computing technology. Apart from that, they are similar in nature to ANNs i.e. the kernels involved are updated based on the cost function (difference between the expected result and the actual result.)

As a result of these changes, the CNN is much more suitable to image-based applications, especially in cases where a large number of layers are required for complex operations as the problem of overfitting is greatly reduced. In this case, we require the high number of layers and complexity in order to perform the task of identifying the author of a particular handwritten text based purely on images of the text, as the only information available is image-based i.e. no online stroke or path information is given.

2.1. Residual Networks (ResNet)

CNNs are powerful and exceedingly useful for many applications. Adding additional layers to a CNN would logically improve performance, as the model gains more flexibility and ability to adapt to a particular dataset. However, in practical applications, after a certain level of depth is reached, performance actually starts to degrade. This is due to the 'vanishing gradient problem', wherein gradients from the loss function shrink to zero and the system is no longer able to update any values, effectively stopping the learning process. Residual networks or ResNets are a further specialized form of CNNs, designed to resolve this problem, developed by He et al. [6]. The primary feature of ResNets is the skip-connection feature spanning across layers, al-lowing gradients to flow through them or for inputs to skip past layers of low importance, as seen in Figure 1. Further, it should be noted that the skips occur over intervals of 2 layers in the case of ResNet34, however, this is instead over 3-layer intervals in ResNet50. As a result of this modification, significantly deeper and more powerful neural networks can be used in order to extract features and generate meaningful information from the provided data. In this paper, we have selected ResNet50 as it provides a good balance between nuance and accuracy, and time taken for training and testing.

3. DATA USED

The Firemaker dataset [7] was used for both training and testing the proposed model. Firemaker consists of 4 sets of handwritten text collected from 250 writers. The 4 sets of text are - P1 (sentence case, fixed text), P2 (uppercase, fixed text), P3 (attempted disguising of handwriting) and P4 (description of a cartoon image). Of these, P1 and P4 were selected for training and testing sets respectively, following [8]. This was done so as to simplify the training process, since the handwriting was standardized to some extent. Further, a testing set completely disjoint from the training set was used to mimic practical applications where exact handwriting samples of identical text for comparison would not be available.

3.1. Image Pre-processing

The pre-processing pipeline consisted of the following steps, taken sequentially.

1. 25 writers out of the whole dataset were arbitrarily chosen. Images from set P1 were manually segmented into five paragraphs of roughly equal length.
2. The adaptive thresholding method proposed by N. Otsu [9] was used to improve contrast between the actual text and the background.
3. Data augmentation techniques were used to increase the amount of data available as well as add rotation and noise resistance. This step consisted of two substeps -
   a. Angle augmentation, wherein the data was rotated in increments of 2 degrees, from -10° to +10° offset from the given sample. This boosted the five samples per writer to fifty-five.

b. Next, noise augmentation was performed. Various forms of noise (speckle, salt-and-pepper, Gaussian) were randomly added to the image, based on the findings of Joshua et al. [10] This resulted in the generation of 330 images per writer.

4. The resultants were resized into a standard, uniform shape that was identical for all samples.

5. Similar steps were followed for the P4 set that was used for testing, with the primary difference being that the text was segmented into only two paragraphs.

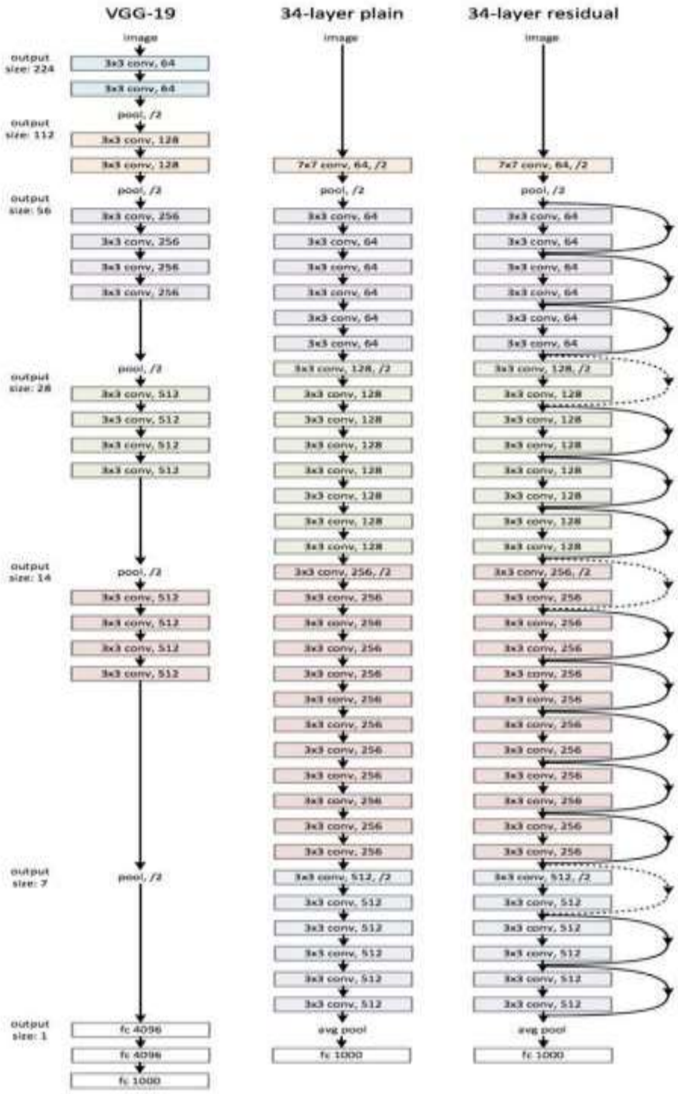

Figure 1. A comparison between a normal 19-layer neural network (VGG-19), a plain 34-layer neural network, and a 34-layer ResNet [3].

4. IMPLEMENTATION

The ResetNet50 model was used to create the proposed model. Of the 23,761,953 trainable parameters, approximately 23,708,833 parameters were used for training, while the remainder were left untrained. The starting weights used in this model were based on those acquired from the ImageNet dataset [11].

This pre-trained model is flattened and sent to a dense layer. Tanh is the activation unit in the fully linked layer, which includes 84 filters. Be-cause our dataset comprises less samples per writer, the dropout layer is employed to avoid overfitting. Softmax is employed as the activation unit for the final linked output. Kernel regularization hyperparameters are also incorporated to make the model less prone to overfitting. The following are the hyperparameters used during the training phase.

Table 1. Parameters used for training the model.

Subset	Learning Rate	L1	Epochs	Validation Split
5 Writer Classification	0.00005	0.001	10	0.10
25 Writer Classification	0.000061	0.00001	15	0.10

The pre-processed images were associated to the corresponding writers and converted into categorical data for use with a CNN. The Adam optimizer [12] was used to increase the rate of convergence towards cost minima as compared to other optimizers. Training was then performed for fifteen epochs, which was the point at which severely diminishing returns per epoch was noted.

5. RESULTS

The model was able to achieve an accuracy of 92% for a set of 5 writers and 82% for a set of 25 writers. This was compared to some other works in terms of Top-1 accuracy on the Firemaker dataset as shown in Table 2.

Table 2. A comparison between various other works and the results of this paper.

Author	Details	Accuracy (%)
Lai et al. [13]	Path signatures	91
Bulacu et al. [14]	Hybrids	83
He et al. [15]	Junclets	80.6
Wu et al. [16]	SDS+SOH	92.4

Khan et al. [17]	DCT features	89.47
He et al. [18]	Quad-hinge	92.2
Proposed method	ResNet50V2 (5 writers)	92
Proposed method	ResNet50V2 (25 writers)	82

The results are on par with the mentioned models, indicating that the proposed model does have the ability to accurately identify writers based only on a training set of one page per writer.

6. CONCLUSION

In this paper, we have introduced a novel data-augmentation and transfer-learning based approach to author identification using existing residual neural network frameworks. The advantages of this proposal are that it uses a pre-trained ResNet as a starting point, greatly cutting down on the length of time required for training to reach a reasonably high level of accuracy. Further, the approach is greatly simplified and easily modifiable to include various other forms of data augmentation, pre-processing or redesigning of the neural network itself. Thus, it is able to be used by individuals with somewhat less experience with machine learning, neural networks or image processing, providing an accessible point of entry for laypersons, academics, students, etc. Limitations of the method include the need for manual data segmentation, which increases the human-time requirement. In future work, the proposed method could be coupled with more extensive data augmentation techniques, enhancing robustness of the input data, as well as automated segmentation of the testing and training data and also training with increased numbers of potential writers.

REFERENCES

[1] O'Shea, Keiron & Nash, Ryan. An Introduction to Convolutional Neural Networks. ArXiv e-prints (2015).

[2] K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016 pp. 770-778. doi: 10.1109/CVPR.2016.90

[3] S. Lai and L. Jin, Offline Writer Identification Based on the Path Signature Feature 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019, pp. 1137-1142, doi: 10.1109/ICDAR.2019.00184.

[4] M. Bulacu and L. Schomaker, Text-independent writer identification and verification us-ing textural and allographic features IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 4, pp. 701-717, 2007.

[5] M. Bulacu, L. Schomaker & L. Vuurpijl Writer identification using edge-based di-rectional features. ICDAR '03: Proceedings of the 7th International Conference on Document Analysis and Recognition, pp. 937-941 Piscataway: IEEE Computer, ISBN 0-7695-1960-1

[6] Hemlata, & Sachan, Manoj & Singh, Shailendra. Personality Detection using Handwrit-ing Analysis:Review. The Seventh International Conference on Advances in Computing, Electronics and Communication - ACEC 2018 10.15224/978-1-63248-157-3-33.

[7] S. He and L. Schomaker, ”Beyond OCR: Multi-faceted understanding of handwritten document characteristics”, Pattern Recognition, vol. 63, pp. 321-333, 2017.

[8] X. Wu, Y. Tang and W. Bu, ”Offline text-independent writer identification based on scale invariant feature transform”, IEEE Transactions on Information Forensics and Security, vol. 9, no. 3, pp. 526-536, 2014.

[9] N. Otsu, A Threshold Selection Method from Gray-Level Histograms IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.

[10] Joshua, Owotogbe & Ibiyemi, T & Adu, B. A Comprehensive Review On Various Types of Noise in Image Processing. International Journal of Scientific and Engineering Research (2019). 10. 388-393.

[11] S. He, M. Wiering and L. Schomaker, ”Junction detection in handwritten documents and its application to writer identification”, Pattern Recognition, vol. 48, no. 12, pp. 4036-4048, 2015.

[12] Irjes, Saini Aishwarya Srinivasan Monika and Anup Kumar Kapoor. Inheritance of Handwriting Features.” International Refereed Journal of Engineering and Science (IRJES). ISSN (Online) 2319-183X, (Print) 2319-1821. Volume 4, Issue 4 (April 2015).

[13] Mr. Nikhil R Shrivastva, Dr. Seema. V. Kedar Author Identification of Handwritten Text: A Review International Journal of Management Technology and Engineering. Page No: 493-499. DOI:16.10089.IJMTE.2019.V9I3.19.27486.

[14] F. A. Khan, M. A. Tahir, F. Khelifi, A. Bouridane and R. Almotaeryi, ”Robust off-line text independent writer identification using bagged discrete cosine transform features”, Expert Systems with Applications, vol. 71, pp. 404-415, 2017.

[15] S. Lai and L. Jin, ”Offline Writer Identification Based on the Path Signature Feature,” 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019, pp. 1137-1142, doi: 10.1109/ICDAR.2019.00184.

[16] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, ”ImageNet: A large-scale hierarchical image database,” 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[17] Diederik P. Kingma and Jimmy Lei Ba. Adam : A method for stochastic optimization. 2014. arXiv:1412.6980v9

[18] Deshpande, Santosh & Kulkarni, Mukul. . Writer Identification using Neural Network. Test Engineering and Management. 83. 10529 - 10533. 2020.

Biographies

Shravani Patel. Completed a B.Tech (Electronics and Communications Engineering) at Visvesvaraya National Institute of Technology. General interests include digital image processing, embedded systems, and digital logic.

Daanish Baig. Completed a B.Tech (Electronics and Communications Engineering) at Visvesvaraya National Institute of Technology. General interests include data analysis and visualization, digital image processing and digital logic.

Dasi Sanjay. Completed a B.Tech (Electronics and Communications Engineering) at Visvesvaraya National Institute of Technology. Academic interests consist of broadband communication systems and radar engineering.

Dr. Ashwin Kothari, received a PhD from Visvesvaraya National Institute of Technology in 2010, where he now works as an Associate Professor in the Electronics and Communications Engineering Department. Fields of interest include commbedded systems, signal processing and antennas, with a special interest in applications of rough sets.