

---

## Use of Deep Learning in Digitization of Libraries

---

<sup>1</sup>Gajendra Muley, <sup>2</sup>Mangesh Ambhure, <sup>3</sup>Bhavesh Pawar, <sup>4</sup>Simran Khiani

<sup>1-4</sup> Dept. of Information Technology, G H Raisoni College of Engineering and Management Pune, India.

Email: nmuley0@gmail.com

Received: 22<sup>nd</sup> December 2018, Accepted: 13<sup>th</sup> February 2019, Published: 30<sup>th</sup> June 2019

### Abstract

This is a proposed idea for the formation of digital databases, which contain information about all the books present in a library, without human intervention. We propose that deep learning is used to scan a book, extract its title and author and store the obtained information in a database which can later be accessed by the members of the library. An idea of how Hough Transform can be used to detect the boundaries of a book is elaborated. The main Deep Learning Techniques used are Convolutional Neural Networks, Recurrent Neural Networks. We have briefly explained the basic Artificial Intelligence concepts of Optical Character Recognition and Natural Language Processing. The complete mechanism of the proposed system and the problems or shortcomings of the system are discussed before the conclusion. We would like to conclude that by using the aforementioned practices we can develop a smart system to digitize book inventories and categorize the books on various fronts, without human intervention.

### Keywords

*Convolutional Neural Networks, Recurrent Neural Networks, Natural Language Processing.*

### Introduction

A large number of literature sources are available in various libraries of India. These books are means of crucial information for various sectors. Due to a large number of books in a library the right book is often not found. This problem can be solved if we create a database of all the books present in a library and provide the user with the option to search through the database for a specific title or an Author of their choice. The formation of a digital database of thousands of books is very time and labor consuming.

To reduce and minimize this effort, we can use Deep Learning and Artificial Intelligence techniques to identify the title of a book and store it in a digital database. To achieve this, we will use Optical Character Recognition (OCR). Most OCR algorithms perform poorly in low light or on distorted images. We will discuss techniques such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) that are used to recognize the text on book covers or the spines of books. These deep neural networks also help in improving the results of OCR. The support of techniques such Hough Transformation helps in early stages on text recognition and is also explained. We can use Natural Language Processing (NLP) and further query the obtained title on the internet and get information about the genre and price of the book.

We establish that using deep neural networks and various deep learning algorithms we can detect the title of a book, its genre and locate it inside a library.

### Related Work

Previous attempts have been made to automate the process of localization and indexing of library books along with provisions of retrieval. Experiments have been carried out to detect the spine of a book in bulk of books when they are kept in a shelf and run character recognition on individual books upon successful segmentation of book spines [1]. Few approaches use a hybrid mixture of text and image analysis methods for detection of book titles [2]. Hough Transformation has been used to detect boundaries of books [3]. Studies for the identification of Devanagari titles, by introducing new datasets are also available [4].

All these experiments use a pre-existing database, of images or a textual database, of books to identify the title of the book. We introduce a different approach. Our system shall, upon recognition of text from the cover or spine of the book, run an internet search and use NLP to gain accurate information regarding the title and the author of the book.

### Concepts Involved

The proposed system is supposed to scan the cover of a book to detect the text on it and then make sense of the obtained result and store it into a digital database, all without human intervention. The various concepts involved in the process are explained.

### A. Hough Transform

We use this basic technique to figure out the boundaries of our book or in some cases, to detect the spine of the book. Hough Transform is an elementary method which, upon figuring out the edges of an object in the image, helps in assigning a line or linear shape to the object.

For example, consider an image consisting of a pole. Once you figure out the edges, Hough Transform considers only the pixels which are already in a perfect line and thus by taking in account the number of aligned pixels it defines the boundaries or shape of the object.

### B. Optical Character Recognition

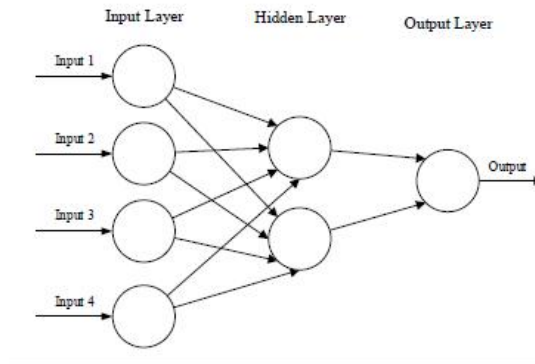
As the name suggests, the process of analysis and identification of the characters in an image and giving the identified text as output is OCR. The image may have printed or hand-written text in it. This is a popular AI technique which can be achieved through various methods.

This process is divided into three steps, acquisition of the image, prior processing to eliminate noise and extraction of text features. Upon identifying the text part of the image, each letter is broken in segments and upon segmentation, it is compared with the dataset to give an output. Traditional OCR systems cannot identify handwritten and slightly distorted and noisy images.

This is a mostly an outdated approach and taking inspiration from the proposal of Oquab *et al* we now use Convolutional Neural Networks to perform OCR [5].

### C. Artificial Neural Networks

The aim of any neural network is to try and mimic the biological functioning pattern of the human brain. Like neurons in our body, the neural networks of deep leaning interconnect neurons with other neurons to divide and process a specific task to yield the desired result. These simple neural networks are known as Artificial Neural Networks (ANN). A weigh is considered on every neuron based the number of layers and the connected neurons in ANN. However, the complexity of computation in a traditional ANN is very high due to the large weighs on each neuron and the number of layers. The first neuron in Fig 1. Has a weigh of 784 ( $28*28*1$ . This is considering a black and white input) If we consider a colored image of  $64*64$  pixels, the first neuron of the first hidden layer would weigh 12,288.



**Figure 1: The Three Layers of a Typical Artificial Neural Network Algorithm.**

The large weighs on neurons in ANN make it impossible to use for large-scale computation due to the availability of limited computational resources and unavailability of time. To overcome these drawbacks, we use Convolutional Neural Networks.

### D. Convolutional Neural Network

A Convolutional Neural Network is an ANN which has been developed specifically to work with the input of images. Since CNN has been specifically designed to work on images, it can function properly with lesser parameters than a typical ANN. Unlike most ANNs, in CNN all neurons from the input layer are not connected with the hidden layer. A section of neurons from the input layer is connected to a single neuron in the hidden layer. All these neurons receive input and process data. The entire CNN will express a single weight. The feature of Transfer Learning is also available in CNN.

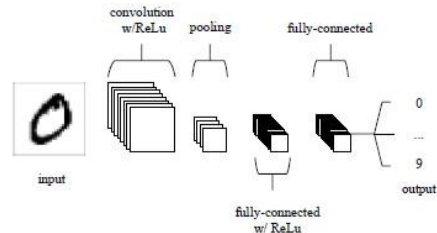
Due to the property of CNN being specially designed for the purpose of working with image inputs we have reduced the number of parameters and in turn, the computation time that will be required to detect or extract the characters from our book cover. The transfer learning also allows us to spend lesser computational time on extraction as we train the model with more and more data.

A typical CNN is in three layers. Fig. 2 shows the architecture of a simplified CNN for classification of the MNIST database.

- The Convolutional Layer: This layer aims to find out the outcome of neurons which are associated to local regions of the input via the calculation of the scalar product of their weights and the area connected to the

input volume. The sigmoid is an ‘elementwise’ activation function. The **rectified linear unit (ReLU)** aims to apply functions like sigmoid to the outcome given by prior layers.

- Downsampling is then performed by the pooling layer to reduce the total parameters in that activation. The downsampling is done across spatial dimensions.
- For the purpose of classification, the fully connected layers will then try to generate the scores of class from the activation. The use of **ReLU** is recommended for improvement of results.

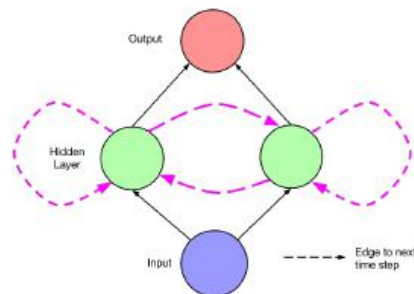


**Figure 2: An Simple CNN Architecture, comprised of just five layers**

### **E. Recurrent Neural Networks**

The word recurrent stands for repeated and as the name suggests, RNNs are deep learning algorithms that are used to process data that has reappearing sequential tasks or functions. Unlike the CNN algorithms, which consider that all neural inputs and outputs are separate and are not dependent upon one another, the RNN algorithms are written considering that the inputs and outputs are related to one another. A typical example where RNNs can be used is in sentence completion. The RNNs, due to their feature of dependency on the output produced by the previous thread can deduce or estimate the context behind the spoken statement and give better accuracy in sentence completion. Though, RNNs only save a few of its previous outputs before moving on to the next neuron.

This functionality of RNNs makes it very hard to train them. A huge number of arguments have to be taken into consideration and the computational time consumed is also much larger than that of a CNN. New advancements in RNN have provided us with better options such as Long Short-Term Memory (LSTM) and Bidirectional Recurrent Neural Networks (BRNN) [6]. These new algorithms have exponentially improved the results of Natural Language Processing experiments. We use RNNs with the help of CNNs and Hough Transform to extract the text from the label of book covers. It is essential to combine more than one technique to maximize the possibility of accurate outcome since many book titles are printed in distorted patterns. A simple RNN is displayed in figure 3.



**Figure 3: A Simple Recurrent Neural Network where Data is Passed from Edge to the Next Time Step in the Hidden Layer.**

### **F. Natural Language Processing**

For a computer to understand a language like a human by using a linguistic input in form of text, voice or image and make sense of the input in order to perform the informed task certain algorithms were developed. This feature of understanding user’s language like any human would in a day to day life is known as Natural Language Processing. RNNs have been a popular choice for NLP.

While Tech Giants are working to build a sustainable NLP to perform an interactive dialogue, we will use NLP to scan through the search results on the internet upon querying the extracted text from our book cover. The

extraction process would have given us a set of words which could be the title or the name of an author. Using NLP on the search results, we will determine the crucial information such as author and genre about the book. NLP can also be used in retrieval of a specific book title upon voice command by the user.

### **Process of Digitization**

By using Hough Transform, we first isolate each book spine or set boundaries for the cover of a book. Then we detect the text on the spine or the cover of each book with the help OCR via CNN. We query the obtained results on the internet and use NLP that functions on the principles of RNN to acquire information about the author and genre of the book. The location of each book is also allotted sequentially and is stored in the database. Upon requirement, the user can search the database for an author or a title and can obtain the location of the desired book.

### **Problems Faced**

- The computational time required for obtaining information from the internet be high and, in some cases, the search might be unfruitful
- The system, upon extraction of text from the book, needs an algorithm which will help it differentiate between the name of the author and title of the book.
- A way to teach the system to differentiate between the title on the book and the name of an author in a case where the title is also a name of an author is to be figured out.

### **Conclusion**

By using various deep learning techniques such as Hough Transform, Convolutional Neural Networks for OCR, Recurrent Neural Networks for NLP and connecting the obtained output to the internet we can accomplish the process of creating a digital database of the books present in a library and localize this information. This study shall be complete upon the solving the problem of inability of the system to differentiate between the title and author that might appear on the cover page of the book.

### **Acknowledgment**

We would like to thank the Head of The Information Technology Department and the faculty members for their constant support throughout this research.

### **References**

- [1] Xiao Yang, Dafang He, Wenyi Huang, Alexander Ororbia Zihan Zhou, Daniel Kifer, C. Lee Giles on “Smart Library: Identifying Books on Library Shelves using Supervised Deep Learning for Scene Text Reading” The Pennsylvania University, University Park, PA 16802, USA.
- [2] Sam S Tsai, David Chen, Huizhong Chen, Cheng-Hsin Hsu, Kyu-Han Kim, Jatinder P Singh, and Bernd Girod. 2011. Combining image and text features a hybrid approach to mobile book spine recognition. In *Proceedings of the 19th ACM international conference on Multimedia*. ACM, 1029–1032.
- [3] David M Chen, Sam S Tsai, Bernd Girod, Cheng-Hsin Hsu, Kyu-Han Kim, and Jatinder Pal Singh. 2010. Building book inventories using smartphones. In *Proceedings of the 18th ACM international conference on Multimedia*. ACM.
- [4] Shailesh Acharya. Ashok Kumar Pant and Prashna Kumar Gyawali on, “Deep Learning Based Large-Scale Handwritten Devanagari Character Recognition” in 2015 9th International Conference on Software, Knowledge, Information Mangement and Applications (SKIMA).
- [5] Tan Chiang Wei, U. U. Sheikh, Ab Al-Hadi Ab Rahman on, “Improved Optical Character Recognition with Deep Neural Network” in 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA 2018), 9-10 March 2018, Penang, Malaysia.
- [6] Achary C. Lipton, John Berkowitz and Charles Eklan on “A Critical Review of Recurrent Neural Networks for Sequence Learning”.