Interim Project Presentation

Tool For Converting Hard Document Into Editable Digital Text Documents

Bachelor's of Technology - Computer Science and Engineering

Group Details

Sl.No.	Reg. No.	Name of the student	Department
1	16ETCS002114	Sarah Zeya Anwar	CSE
2	16ETCS002115	Satyam Agarwal	CSE
3	16ETCS002144	Ashish Kumar	CSE
4	16ETCS002161	Saurav Kumar	CSE

Batch : FT-16

- **Title of the Project:**
  TOOL FOR CONVERTING HARD DOCUMENT INTO EDITABLE DIGITAL TEXT DOCUMENTS

- **Supervisors**

  Supervisor : Ms. Pallavi R Kumar

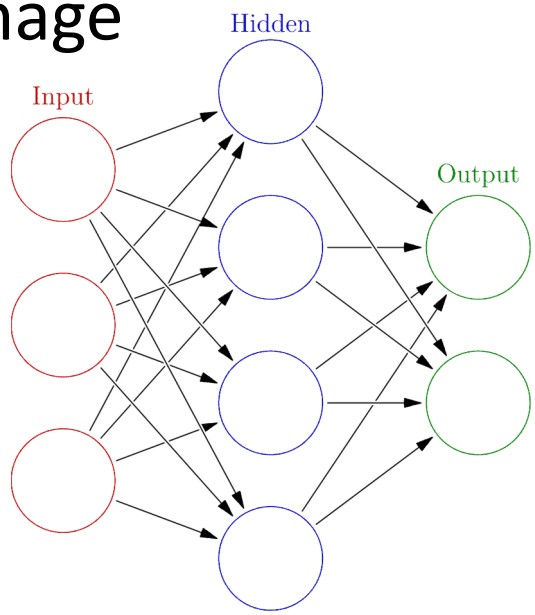
- **Place of Work**
  RAMAIAH UNIVERSITY OF APPLIED SCIENCES

Outline

- Introduction
- Motivation(Project Concept and its relevance)
- Aims and Objectives
  - Title, Aim, Objectives, Methods and Methodology
- Problem Solving
  - Project Concept, Design, Implementation
- Project Costing
- Expected outcomes
- Workload Allocation
- Updated Gantt chart with separate coloring for completed work
- References
- Demonstration (If applicable)
- Team Experience

Introduction

- Supervised machine learning to convert hard copy of document to digital text document using Image processing and Neural Network.
- Artificial neural networks (ANN) are computing systems vaguely inspired by the biological neural networks that constitute animal brains.



- Deep learning is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms.
- Deep learning-based image recognition has become "superhuman", producing more accurate results than human contestants.

Motivation (Project Concept and its relevance)

- Earlier, in field of studies and documenting, if someone were presented with a large amount of handwritten text to edit, they would have to input it manually into the computer which is time consuming.
- The huge amount of data stored in form of papers in the government offices, case details in courts and police stations etc. become hard to search through and interpret.
- These handwritten or printed documents stay in a large pile of pages at workplace and makes the work place messy and put a psychological tension on the person trying to get the data out of those piles.

Motivation (Project Concept and its relevance)

- Handwritten notes and important documents are prone to getting lost or destroyed, saving a digital copy is much more efficient and helpful.
- This project holds great significance since it aims to assist in easing the conversion from manual to digital text type.
- With the help of our tool any amount of handwritten data or printed documented can be turned into digital document.
- Process of editing and searching in the documents get easier.
- Interpreting the data becomes very easy hence a data which need to very specific can also be obtained without any effort.

PROJECT CONCEPT

- Concept of the project is to construct a model which will take handwritten document(image) as input dataset, preprocess the image, train the neural network model from the dataset to recognize the characters and then store the recognized characters into a text document.

Aim

AIM: To develop a tool for conversion of hard copy of document to digital text document.

Objectives

1. To conduct literature survey on Character Recognition using neural network and textual image processing.
2. To design the methodology to be followed for developing the text documenting tool.
3. To develop and implement the algorithm that recognizes a hard text from the image and convert into a digital text document.
4. To test and validate the developed tool for various textual images.
5. To document the report by unifying all the results and outcomes.

Methods and Methodologies

Objective 1: Literature Review

1. *Machine Learning in Document Analysis and Recognition*

→Summary: The objective of this paper is to understand principal components required for the implementation of a pattern recognition that are pre-processing, object-segmentation, object recognition and post processing in significant details. The paper explores the application of different pattern recognition outputs including handwriting and text recognition.

Literature Review

2. Neural Network based Handwritten Character Recognition system

→Summary- The effectiveness of their work emphasizes using feature extraction using character geometry and gradient technique from scanned images containing handwritten characters. Their proposed methodology has produced good results for images containing handwritten text written in different styles, size and alignment.

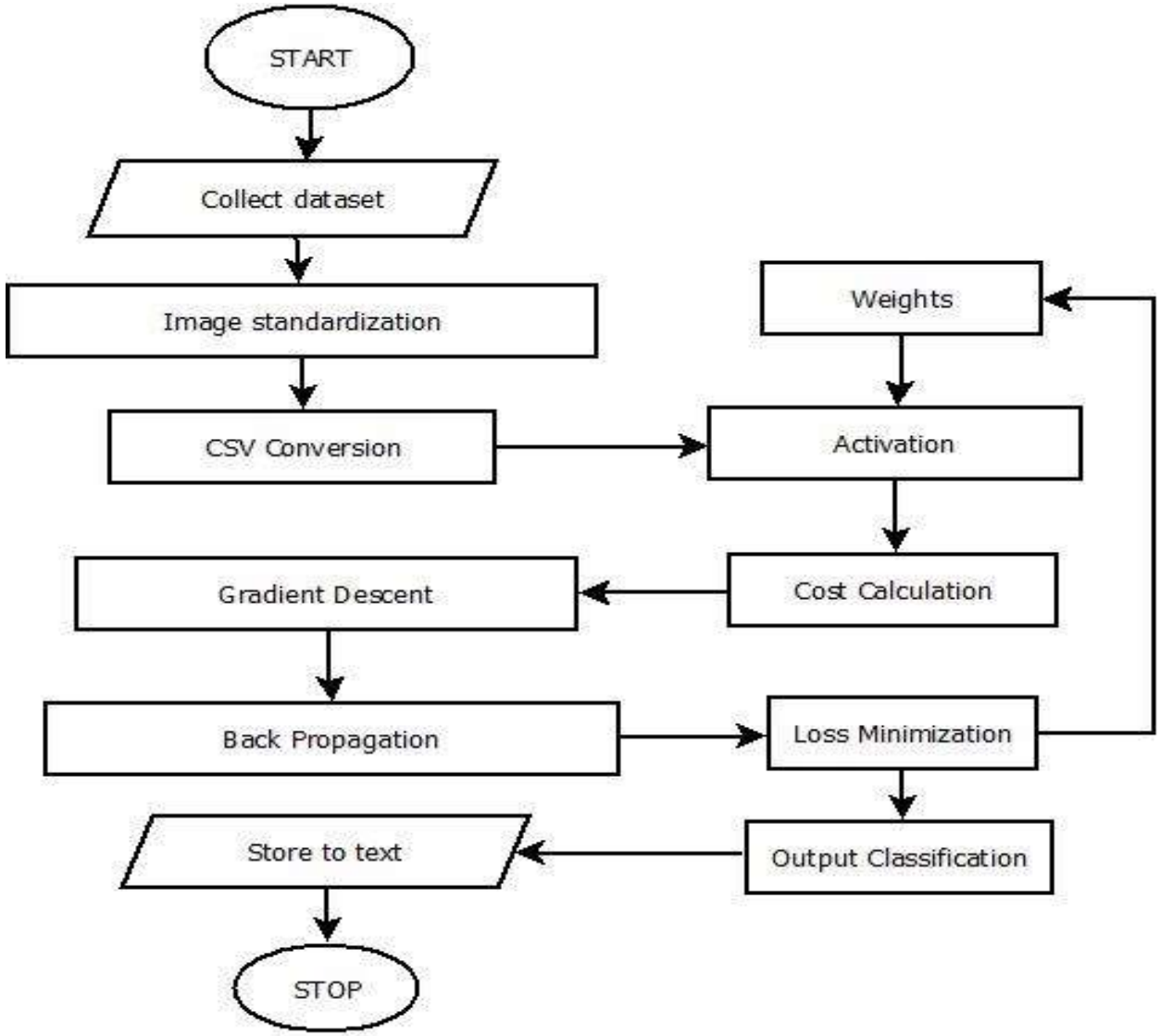
Method and Methodology Contd..

Objective 2: Design

Assumptions

- The input image is assumed to contain only English characters.
- The input image is assumed to be a text written on un-ruled white paper.
- The input image is assumed to contain text that are not cursive.
- The input image is assumed to contain text that does not contain noise e.g. ~~Ashish .~~

Design - Flowchart



Method and methodology contd..

Objective 3: *Implementation*

- We are using NIST dataset for the implementation which includes about 800,000 image data of characters.
- The assumptions are made to develop a better classifier as the we have dataset limited to English characters and digits only.
- To process the computation we are taking the help of google colaboratory as they provide free GPU and TPU processing.

Method and methodology contd..

Objective 4: *Testing and validation*

- Collecting various test samples for the input.
- Conducting unit testing and system testing.
- Conducting validation check of the tool.
- Checking the performance of the tool for test inputs.

Method and methodology contd..

Objective 5: *Documentation*

- Based on the literature survey done, the requirements are reported.
- Demonstration of the developed algorithm with the test data will be reported.
- After the design, implementation, verification, testing and validation, results of the system are reported.
- Analyze performance of the tool and report the performance test' results.

Expected Outcomes

- Demonstration of Working Model
- Tool

Project Costing

Component	Units	Cost
Man power	16weeks*4people	₹2,00,0000
Laptop (Hardware and software)	4 Units	₹10,000
Internet and Electricity	4	₹5,000
Total		₹2,15,000

Work Load Allocation

	Saurav K	Satyam A	Ashish K	Sarah Anwar
Literature Survey	■			■
Documentation		■		■
Requirements analysis			■	
Designing	■			
Implementation			■	
Testing		■		

Updated Gantt Chart

Project Work (UG) 16weeks																
Week	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Major Activities																
Project Selection																
Literature survey																
Design																
Implementati-on																
Testing																
Documentati-on																

References

- Attigeri, S. (2018). Neural Network based Handwritten Character Recognition system. *International Journal of Engineering and Computer Science*, *7*(03), 23761-23768.
- LeCun, Y., Jackel, L.D., Bottou, L., Cortes, C., Denker, J.S., Drucker, H., Guyon, I., Muller, U.A., Sackinger, E., Simard, P. and Vapnik, V., 1995. Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural networks: the statistical mechanics perspective*, *261*, p.276.

References

- Simone Marinai, Hiromichi *Fujisawa.Machine Learning in Document Analysis and Recognition*, 2008 Springer-Verlag Berlin Heidelberg.

Team Experience

- We have learnt and explored our experience on this group project which includes analyzing, researching, documenting and implementing.
- Sharing and discussing different ideas and amount of knowledge amongst the group gave us a brief insight on how to deal with problems and come up with an applicable solution.
- We experienced how to confront difficulties in a team work and how to resolve any sort of argument or problem by coming up with something which is agreeable by everyone.

Team Experience

- While working on project such as this, we exercised a great opportunity of developing and enhancing our technical skills effectively and efficiently.
- We have also acquired different set of skills from each other that is beneficial to each one of us in the long run.
- In conclusion, We've learnt a lot of things about ourselves while working with our group members in undertaking this project.

Thank You