# Form 2:Literature Documents

1.Team No: 2

2. Project Title: LSTM based OCR

Comparison of Existing Methods

| SI no | Author[s] | Method | Advantages | Disadvantages |
|---|---|---|---|---|
| 1 | Yiyi Liu<br>Yuxin Wang<br>Hongjian Shi | 1. DBNet:<br>DBNet is a novel network architecture for real-time scene text detection with differen-tiable binarization.<br>2. Text Direction Classification: Paddle text direction classifier is a module that is added between the text detection and recognition modules to deal with text in different directions.<br>8. Prediction: Classify new websites based on extracted features using the trained algorithms. | 1. Model can effectively segment and recognize text in various backgrounds and orientations by applying the affine transformation, text direction classifi-cation, and clarity evaluation. | 1. Low accuracy for short texts with significant deformations, such as art terms or texts describing scenes in the natural world.<br>2. Limitations of CRNN in complex and multi-oriented text scenes. |
| 2 | Rakesh Chandra<br>Surajit Mohanty<br>Marimuthu Karuppiah<br>Debabrata Samanta | 1. Skew correction, Word segmentation, Character segmentation, Recognition, Inspiredby the success of deep neural networks for feature learning, we have explored CNNs to classify the characters and proposed a new architecture for the same. | 1. 94.32% Model Accuracy | 1. segmen-tation algorithm can be improved so that every character is segmented together with its vattu and gunintham. |

| | | | | |
|---|---|---|---|---|
| 3 | Ekraam Sabir<br><br>Stephen Rawls<br><br>Prem Natarajan | 1. Preprocessing<br><br>To ensure a constant input size to the model, images are<br><br>scaled to a constant height of 30 pixels while conserving the<br><br>aspect ratio.<br><br>2. Feature extraction: sliding window image frames of 2-pixel width transformed into a 60x1 vector as raw input features for 2x fully connected layers with 60 units each. | 1.Temporal aspect of an LSTM allows it the model to take slices of image across variable width characters and recognize it.<br>2. 2-bidirectional LSTM layers with 256 units in each layer and 700 time-steps, where each time-step is a potential character prediction. | 1.It makes use of up to 5 characters in making predictions. It does not necessarily help in making predictions on current character always.<br><br>2.indifferene in performance on character e in Times font. |

References:
[1]. Yiyi Liu, Yuxin Wang, Hongjian Shi (2023) A Convolutional Recurrent Neural-Network-Based Machine Learning for Scene Text Recognition Application.
[2]. Rakesh Chandra Balabantaray, Surajit Mohanty, Marimuthu Karuppiah, Debabrata Samanta (2022). Approach for Preprocessing in Offline Optical Character Recognition (OCR).
[3]. Ekraam Sabir, Stephen Rawls, Prem Natarajan (2020). Implicit Language Model in LSTM for OCR.

Signature of the Supervisor