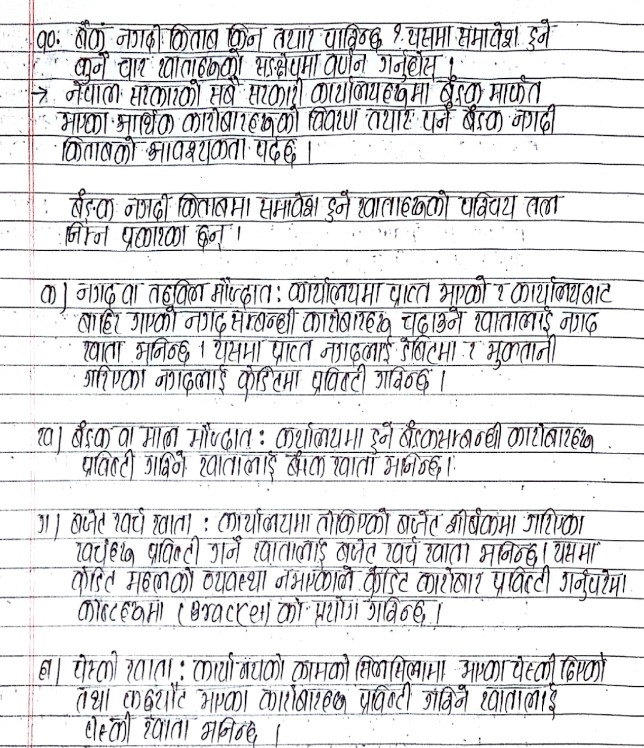
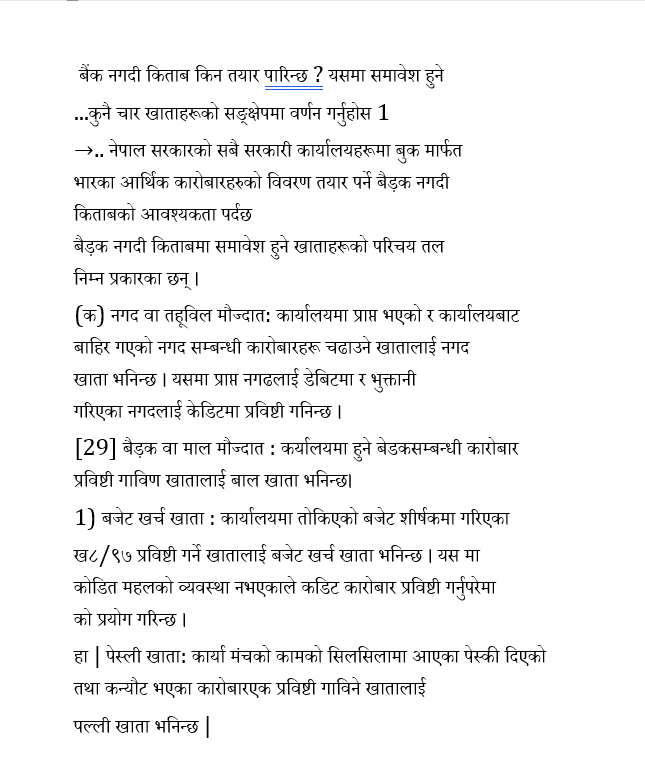
The Nepali OCR model was built and forms the core for detecting and recognizing the Nepali Handwritten Text, which converts them into the digital format. We have use LSTM and CNN. The system recognizes the Nepali Handwritten Text . This approach is driven by limitations in GPU training capabilities, aimed at efficiently utilizing resources while ensuring the quality of the trained model. The models are limited to distance and illuminance from which the images are captured and work with efficiency at distance 10-20 cm and with proper lighting.

Image classifier



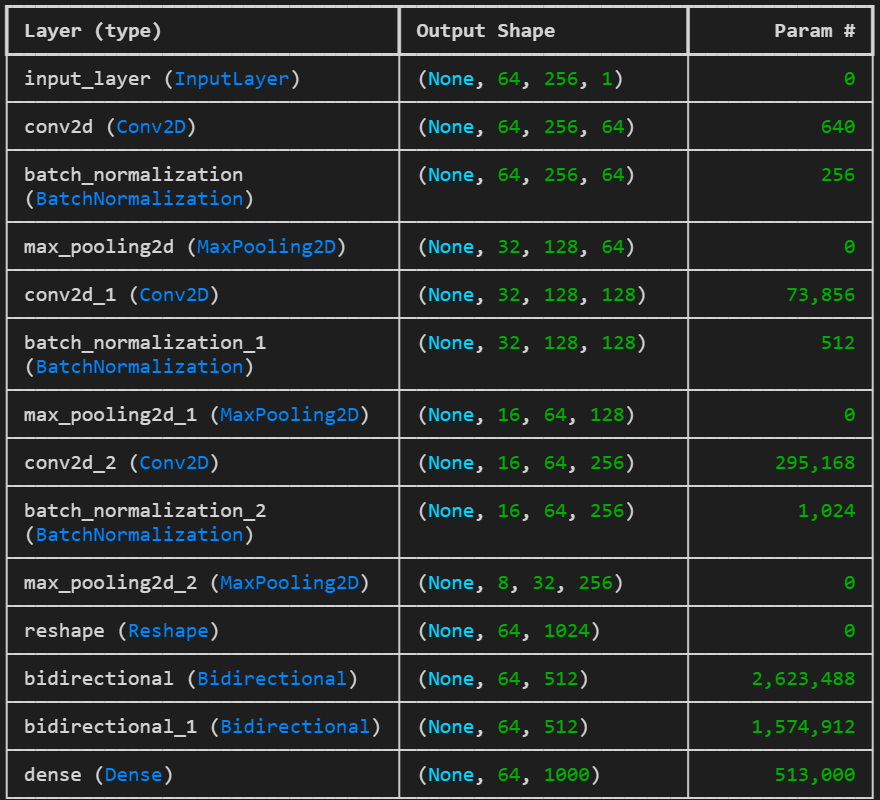
This is the first image of the test that we performed for our OCR system. As it can be seen, that the image is of Nepali Handwritten Text . The system detects the Handwritten Text line by line and recognizes them.



This is the result of the image which was taken for the test purpose and the image above shows the Digital format of the Handwritten Text which was taken for the sample purpose.

Action Recognition with LSTM

The output generated by our ConvLSTM model for Text Detection is presented below, showcasing the intermediate results at each layer of the model. These results were obtained following an extensive and rigorous training and testing process. The training involved leveraging a custom dataset curated specifically to reflect Text detection , complemented by publicly available benchmark datasets to enhance the model's robustness and generalizability. The ConvLSTM architecture, designed to capture both spatial and temporal dependencies, was fine-tuned to handle complex sequences and dynamic behavioral patterns. The layer-wise outputs illustrate the progression of feature extraction, temporal encoding, and decision-making processes within the network, offering a detailed understanding of its predictive capabilities and performance across diverse scenarios.



Total Parameters: 5,082,856 (19.39 MB)

Trainable Parameters: 5,081,960 (19.39MB)

Non-trainable Parameters:896(3.50KB)

The ConvLSTM model utilized for Text detection has a total of 5,082,856 parameters, corresponding to an approximate memory footprint of 19.39 MB. Of these, all 5,081,960 parameters are trainable, indicating that the model's entire parameter set is optimized during the backpropagation process to enhance learning and predictive performance. There are non-trainable parameters that is of 896 (3.50KB), which implies that the model incorporates frozen layers or pre-trained components in its current configuration. This complete trainability ensures that the model can fully adapt to the specific characteristics of the custom and publicly available datasets used during the training phase, maximizing its ability to capture complex temporal and spatial patterns in the Text detection task.