Medical Prescription Recognition using Deep Learning

Krishna Sharma(Roll No. 2101104)
Guide: Dr. Ferdous Ahmed Barbhuiya

Indian Institute of Information Technology, Guwahati
CS300

May 2, 2024

Outline

Motivation and Problem Statement

Motivation and Problem Statement (1/2)

Motivation:

- **Patient Safety:** Medication errors due to misinterpreted prescriptions can cause serious harm or even be fatal. An automated system can significantly reduce these errors, ensuring that patients receive the correct medication and dosage.
- **Efficiency:** Pharmacists often spend a considerable amount of time deciphering doctors' handwriting. An automated recognition system can save time, allowing pharmacists to serve more patients and focus on other critical tasks.

Aim: To develop an AI-powered Medical Prescription Recognition System that accurately interprets handwritten and printed prescriptions, thereby reducing medication errors, improving patient safety, and streamlining the pharmacy workflow for efficient healthcare delivery.

Literature Review/Related Work

Literature Review/Related Work(1/2)

- **Recognition of Doctors' Cursive Handwritten Medical Words by using Bidirectional LSTM and SRP Data Augmentation (2021)(1)**
  The study developed a primary "Handwritten Medical Term Corpus" dataset with 17,431 data samples comprising 480 words from 39 Bangladeshi doctors. On the preprocessed pictures, a new data augmentation technique called SRP(Stroke Rotation and Parallel-shift) is used to increase the number of data samples. Following this, a sequence of line data is extracted from both the original and augmented image data. Bidirectional LSTM is applied to the sequential line data derived from the augmented handwritten images to produce complete end-to-end recognition.
- **Handwriting Recognition for Medical Prescriptions using a CNN-Bi-LSTM Model (2021)(2)**
  In this paper, they used neural network techniques such as CNN and BI-LSTM for predicting doctor's handwriting from medical prescriptions . The CTC loss function is used for normalization. This model built on the IAM dataset.

Literature Review/Related Work(2/2)

- **Medical Handwritten Prescription Recognition Using CRNN(2019)(3)**
  The approach established a Convolutional Recurrent Neural Network (CRNN) technology using Python that can interpret handwritten English prescriptions and translate them into digital text . For this, datasets with 66 different classes, including alphanumeric characters, punctuation, and spaces, were used. Since prescriptions generally contain two or three words, the training was carried out using short texts.
- **Intelligent Tool For Malayalam Cursive Handwritten Character Recognition Using Artificial Neural Network And Hidden Markov Model (2017)(4)**
  The approach uses the Hidden Markov Model (HMM) to recognize cursive handwritten Malayalam characters . By employing a median filter, the algorithm used here helps to avoid errors caused by noise in the scanned image. Furthermore, Artificial Neural Network (ANN) aids in the acquisition of better classification and provides the best matching class for input.

Methodology

Methodology(1/5)

- The project at hand encompasses two distinct yet interconnected components: word detection and recognition
  - **Detection:** The initial phase, word detection, revolves around the segmentation of individual words from the scanned prescription image. Each segmented word is fed to the recognition model for prediction.
  - **Recognition:** Here, the system performs the task of deciphering the actual meaning of each isolated word. Each segmented word is treated separately and appropriate predictions are made.
- **Data collection and Preprocessing**
  - Data plays a pivotal role in the development and evaluation of our deep learning model. We have leveraged the IAM dataset, a widely recognized resource known for its comprehensive collection of handwritten text samples, to fuel our project.
  - **Gray-scale conversion:** Gray scaling is the first step in digital image pre-processing that must be done. Here each pixel's value solely encodes the light's intensity information. There are just three colors used in grayscale images: black, white, and gray, which comes in a variety of tints.
  - **Normalization:** Following grayscale conversion, the images undergo normalization—a critical process aimed at refining the data for seamless integration into the model. Normalization involves scaling down the pixel values from their original range of 0 to 255 to a standardized range. This uniformity ensures consistency across the dataset, reducing potential discrepancies and facilitating effective model training.

Methodology (2/5)

- **Word Detection**
  - In this part we try to extract the words from an image.
  - The model operates by meticulously classifying each pixel within an image, discerning whether it belongs to the inner part of a word, the surrounding area of a word, or constitutes background.
  - For pixels identified as part of the inner word, the model predicts an axis-aligned bounding box (AABB) encompassing the word. Given the potential for multiple AABBs to be predicted for the same word, a sophisticated clustering algorithm is employed to effectively group them
  - The output maps of the model encode the AABBs and their respective classifications. These output maps include:
    - Three segmentation maps utilizing one-hot encoding to designate:
      – The inner part of a word
      – The surrounding area of a word
      – Background pixels
    - Four geometry maps that encode distances between each pixel and the edges of the AABB, specifying:
      – The distance to the top edge
      – The distance to the bottom edge
      – The distance to the left edge
      – The distance to the right edge

Methodology (3/5)

- ResNet18 serves as the feature extractor, following the widely recognized U-shape architecture commonly used in segmentation tasks.
- During the training process, the input images are resized to dimensions of 448×448 pixels. After traversing through the final layer of ResNet18, the feature maps undergo a downscaling process, resulting in dimensions of 14×14.
- Subsequent layers progressively upscale these maps, amalgamating intermediate results from ResNet18 and extracts the predicted bounding box. Ultimately, the output of the neural network yields all the predicted bounding boxes in the image.
- The loss function comprises two integral components essential for effective training:
  - **Segmentation loss:** To tackle the pixelwise classification problem inherent in segmentation tasks, cross-entropy loss is utilized to ascertain the accuracy of segmenting each pixel.
  - **Geometry loss:** Rather than relying on sum-of-squared errors on the geometry, which may disproportionately weigh larger bounding boxes, the intersection over union (IOU) metric is adopted. This metric ensures a fair assessment of the overlap between predicted bounding boxes and ground truth boxes.
- To address situations where multiple AABBs are predicted for the same word, a robust clustering algorithm named DBSCAN is deployed.
- This algorithm calculates AABB clusters based on the Jaccard distance between AABB pairs. Subsequently, the resulting AABB for each word cluster is meticulously computed.
- This meticulous process ensures accurate localization of words within the image, laying a strong foundation for effective word recognition. The output of this network are fed to the predicting model.

Methodology(4/5)

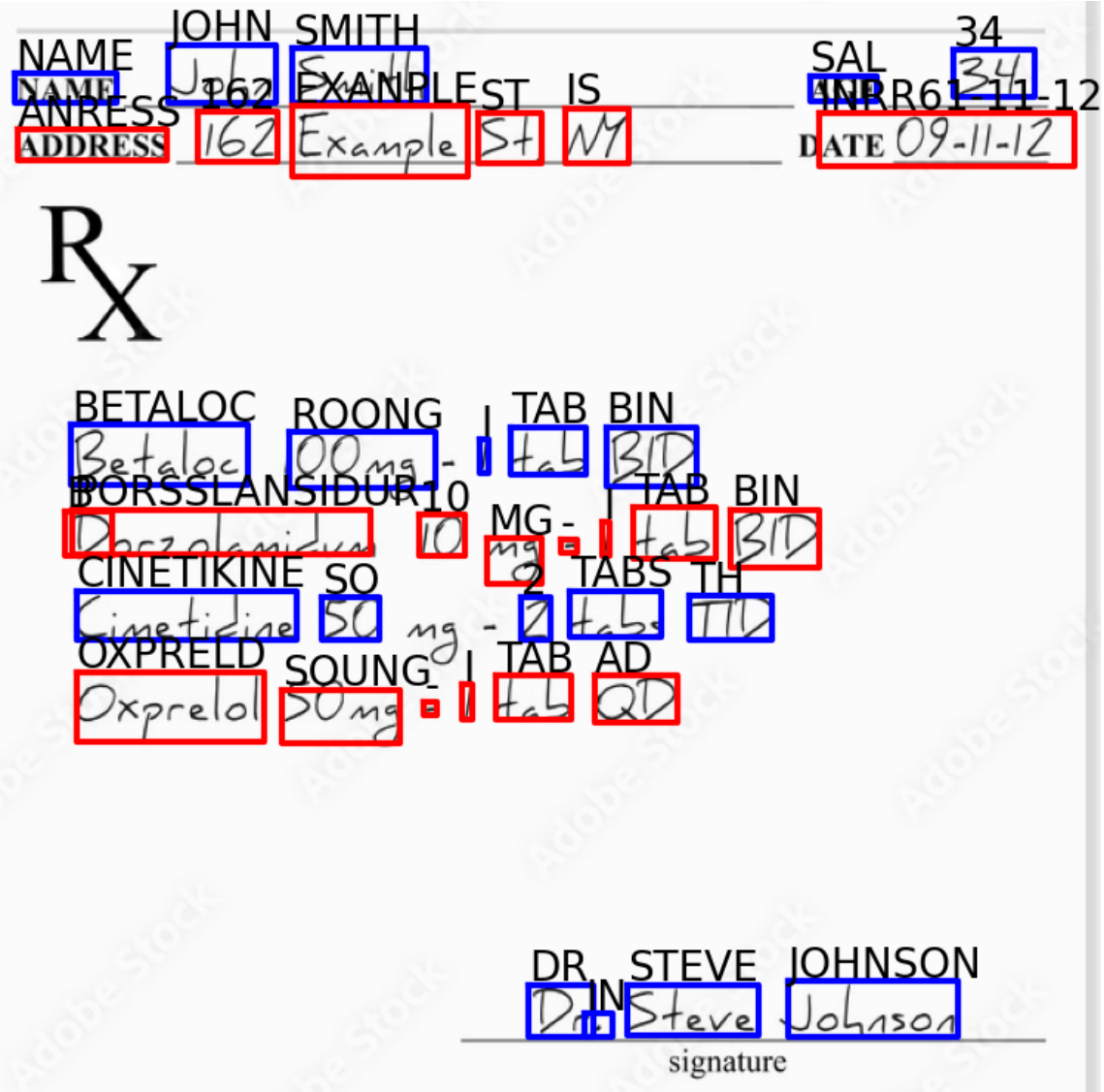
- **Recognition**
- In this part we try predict the accurate text from the image. Our neural network (NN) ar- chitecture comprises convolutional NN (CNN) layers, recurrent NN (RNN) layers, and a final Connectionist Temporal Classification (CTC) layer, each playing a crucial role in predicting the word from the input image.
- The model tries to extract the word from the image. It consists of 5 CNN layers, 2 RNN layers and a final CTC layer.
- **CNN Layers**
  – The input image undergoes processing through the CNN layers, which are adept at extracting pertinent features from the image
  – Each CNN layer consists of three operations: convolution, rectified linear unit (RELU) activation, and pooling
  – Through convolution, filters of varying sizes are applied to the input, capturing spatial patterns
  – RELU activation introduces non-linearity, enhancing the network's ability to capture complex relationships within the image.
  – Pooling layers summarize image regions, downsizing the input while preserving essential features.
  – The output of the CNN layers is a feature map or sequence, where each element represents a feature extracted from the image

Methodology(5/5)

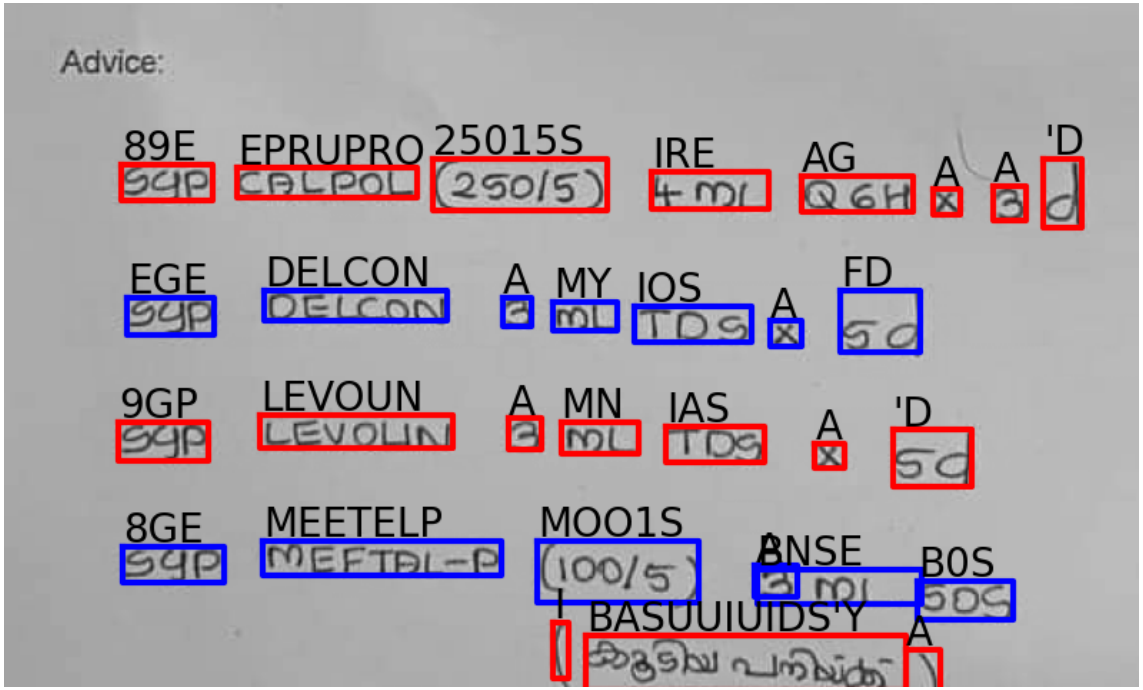
- **RNN Layers**
  – Operating on the feature sequence derived from the CNN layers, the RNN propagates relevant information through the sequence
  – Leveraging Long Short-Term Memory (LSTM) cells, the RNN can effectively capture dependencies over longer distances and exhibits robust training characteristics
  – LSTM cells are designed to overcome the vanishing gradient problem commonly encountered in traditional RNNs. This problem arises when gradients become extremely small during backpropagation, hindering the learning process, especially for long sequences
  – It employs gating mechanisms, including input, forget, and output gates, to regulate the flow of information within the cell.
- **CTC Layer**
  – During training, the CTC layer receives the output from the RNN layers and the ground truth text, computing the loss value.
  – In inference mode, the CTC layer decodes the output into the final predicted text.
  – CTC layer processes input sequences and calculates probabilities of target output sequences given input, allowing for variable-length inputs and outputs.
  – Notably, both ground truth and predicted texts are constrained to a maximum length of 32 characters.
- Each component of the NN architecture collaborates synergistically to process the input image, extract relevant features, capture temporal dependencies, and decode the output into the predicted text, ultimately enabling accurate word prediction.

Results and Discussions

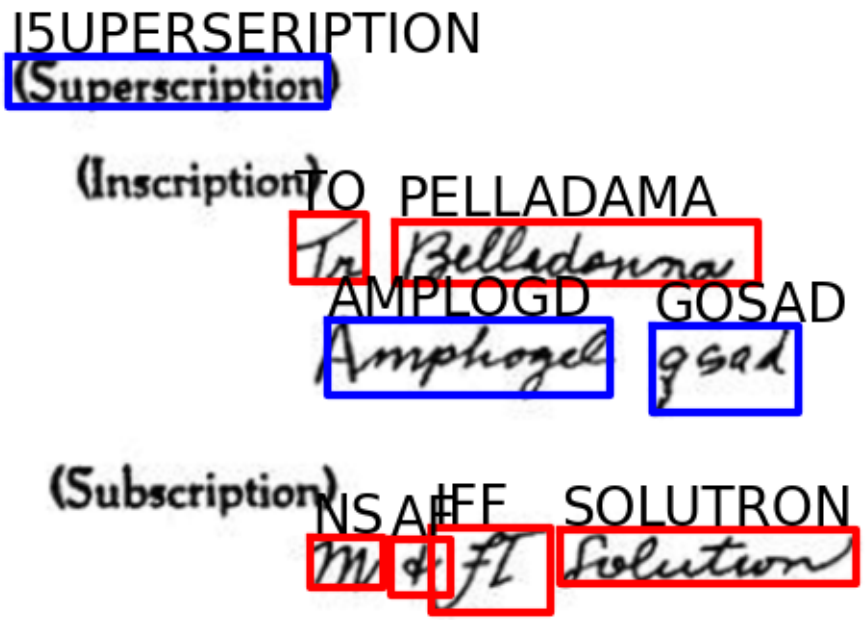
Results and Discussions (1/3)



Results and Discussions (2/3)



Results and Discussions (3/3)



- By using the above mentioned technique we can predict the text with an average accuracy of nearly 57%

Conclusion

Conclusion and Future Work

- In conclusion, the presented approach utilizes a combination of convolutional neural networks (CNN), recurrent neural networks (RNN), and connectionist temporal classification (CTC) to achieve word prediction in handwritten images. By leveraging the capabilities of ResNet18 for feature extraction, CNNs and RNNs for sequence modeling, and CTC for loss computation and decoding, the model demonstrates promising results in predicting words from scanned images of handwritten text.
- Through the implementation of this approach, several key insights have been gained into the challenges and opportunities of handwritten word prediction. By successfully classifying pixels, predicting bounding boxes, and decoding sequences, the model showcases the potential of deep learning techniques in tackling complex image processing tasks.
- Moving forward, there are several avenues for future exploration and improvement.

Future Work

Future Work

- **Model Optimization:** Further optimization of the neural network architecture and hyperparameters could enhance the model's performance and efficiency.
- **Dataset Expansion:** Increasing the diversity and size of the training dataset, possibly by incorporating additional handwriting styles and languages, can improve the model's generalization capabilities.
- **Integration with Applications:** Integrating the trained model into practical applications could facilitate real-world usability and impact.
- **User Interface Enhancement:** Improving the user interface and user experience of the system, including error handling and feedback mechanisms, can enhance usability and accessibility.

Reference

[1] S. Tabassum et al., "Recognition of Doctors' Cursive Handwritten Medical Words by using Bidirectional LSTM and SRP Data Augmentation," 2021 IEEE Technology Engineering Management Conference - Europe (TEMSCON-EUR), Dubrovnik, Croatia, 2021, pp. 1-6, doi: 10.1109/TEMSCON-EUR52034.2021.9488622.

[2] T. Jain, R. Sharma and R. Malhotra, "Handwriting Recognition for Medical Prescriptions using a CNN-Bi-LSTM Model," 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-4, doi: 10.1109/I2CT51068.2021.9418153.

[3] R. Achkar, K. Ghayad, R. Haidar, S. Saleh and R. Al Hajj, "Medical Handwritten Prescription Recognition Using CRNN," 2019 International Conference on Computer, Information and Telecommunication Systems (CITS), Beijing, China, 2019, pp. 1-5, doi: 10.1109/CITS.2019.8862004.

[4] N. P. T. Kishna and S. Francis, "Intelligent tool for Malayalam cursive handwritten character recognition using artificial neural network and Hidden Markov Model," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 2017, pp. 595-598, doi: 10.1109/ICICI.2017.8365201.