COSE474-2024F: Final Project Proposal
‘Handwritten Character Recognition system’

Hazirah Binti Ahmad Irfan

1. Introduction

This project aims to develop a deep learning model, using Convolutional Neural Networks (CNNs), capable of recognizing not just digits (0-9), but also uppercase and lowercase letters (A-Z, a-z) and potentially special symbols. By leveraging datasets such as the Extended MNIST (EMNIST), which includes both letters and digits, this project will explore how CNNs can be trained to handle more complex character recognition tasks.

In addition to expanding the dataset to include letters and symbols, we will address common challenges such as distinguishing between similar-looking characters (e.g., ’O’ vs. ’0’ and ’l’ vs. ’1’), improving generalization through data augmentation, and enhancing model performance using advanced techniques like dropout and data augmentation.

2. Problem definition & challenges

Recognizing letters and symbols presents unique challenges compared to digits:

Similar-looking characters: Some characters like ”O” (letter) and ”0” (digit), or ”l” (lowercase L) and ”1” (digit) can be difficult to distinguish. Different fonts and handwriting styles: Handwritten letters and symbols can vary more in shape than digits, making it harder for the model to learn consistent features. In addtion, there may be fewer samples for some letters or symbols compared to digits. Use techniques like oversampling or class weighting to handle class imbalance.

3. Related Works

1. LLaMA: Open and Efficient Foundation Language Models
Hugo Touvron, Louis Martin, Kevin Stone, et al. (2023)

2. CLIP: Learning Transferable Visual Models from Natural Language Supervision
Alec Radford, Jong Wook Kim, Chris Hallacy, et al. (2021)

4. Datasets

EMNIST (Extended MNIST)

The EMNIST dataset is an extension of the MNIST dataset and provides several subsets that include both digits and letters. It contains EMNIST Balanced with 131,600 images with 47 balanced classes (digits and uppercase/lowercase letters). Also, EMNIST ByClass with 814,255 images across 62 classes (digits 0-9, lowercase letters, uppercase letters). In addtion, EMNIST Letters with 145,600 images across 26 classes (uppercase letters only). This dataset has the same format as MNIST (28x28 grayscale images), making it easy to transition from digit recognition.

5. Goals to achieve throughout this project

The primary goal of this project is to develop a multi-character recognition system that extends beyond digits to include letters, symbols, and possibly multilingual scripts. By leveraging pre-trained foundation models like CLIP and LLaMA through transfer learning, the system aims to enhance recognition accuracy and achieve robust generalization across diverse handwriting styles. The project also focuses on building an efficient and scalable system that can operate in real-time, making it suitable for practical deployment in resource-constrained environments such as mobile applications or cloud-based OCR tools. In addressing challenges like noisy handwriting, similar character distinctions (e.g., ”O” vs. ”0”), and complex writing patterns, the model will incorporate advanced techniques for data augmentation and error correction.

6. Schedule

Week 1: Project Setup and Dataset Preparation
Week 2: Model Design and Pre-trained Model Integration
Week 3: Model Training (Phase 1)
Week 4: Model Evaluation and Error Analysis
Week 5: Model Optimization
Week 6: Advanced Feature Integration
Week 7: System Testing and Final Adjustments
Week 8: Final Evaluation, Documentation, and Deployment

7. State-of-the-art methods and baselines

Traditional handwritten recognition systems, like those built using simple Convolutional Neural Networks (CNNs), generally start training from scratch or with minimal transfer learning. Some recent models like DenseNet, ResNet, and VGG16 have been pre-trained on large-scale image datasets (e.g., ImageNet) and fine-tuned for character recognition tasks. This proposed system will use CLIP and LLaMA, pre-trained on vast and diverse datasets, enabling the model to better generalize and understand ambiguous characters by leveraging both visual and contextual information. This is a notable improvement over traditional pre-trained models, as CLIP can link images with textual concepts and LLaMA can provide contextual corrections, making the system more robust. This gives better generalization due to leveraging vast, diverse training data and context-aware correction with language models, which many SOTA systems lack.