Report

Supervisor: Prof. Anand Mishra
Mentor: Anik De
Team Members: Chaitanya Singh (B24CS1088), Pawan Kumar (B24BB1024) & Varnit Sharma (B24EE1085)

1. Introduction

The focus of this project was on collecting natural scene text images, annotating them, evaluating their recognition performance using the IndicPhotoOCR toolkit, and documenting the complete workflow through GitHub and structured reporting. Through this project, we contributed by collecting real-world images and evaluating the system's performance, helping expand resources for low-resource Indian languages.

2. Objectives

- Collect minimum 50 natural scene images containing text in Indian scripts.
- Manually annotate the visible text in each image.
- Run and evaluate the IndicPhotoOCR pipeline on the curated dataset.
- Analyze Word Recognition Rate (WRR) and script-wise performance.
- Organize all data, annotations, and results in a GitHub repository.
- Prepare a structured report and presentation summarizing learnings.

3. Tasks Completed

Data Collection

We collected 61 scene images from Jodhpur city, IITJ campus, and additional contributions from friends/family. We aim to capture images containing Hindi & English scripts.

Annotation

Every image was manually annotated using VGG Image Annotator (VIA). Polygon-level bounding regions and transcription text were added, maintaining accuracy.

Model Evaluation

We ran the IndicPhotoOCR inference pipeline on our dataset. After exporting predictions, we compared them to ground truth using word-level matching, script grouping, and case-normalized evaluation.

4. Results & Model Performance

After evaluating the model on our dataset, script-wise Word Recognition Rate (WRR) was computed. The results below represent the accuracy of IndicPhotoOCR in recognizing ground-truth words:
• Hindi WRR: 28.91%
• English WRR: 4.01% (for case sensitive) & 43.27% (for case insensitive)

GitHub repo here