Image-based Entity Value Extraction

# Project Overview

This project focuses on extracting key entity values (such as weight, volume, and dimensions) from product images through a hybrid approach combining Optical Character Recognition (OCR) and Convolutional Neural Networks (CNN). By leveraging machine learning techniques, we analyze both the textual and visual elements from the images to predict the entity values accurately.

# Setup and Installation

1. Clone the repository:

git clone https://github.com/username/repository.git

2. Install dependencies:

pip install -r requirements.txt

3. Prepare the dataset:

Download and preprocess the dataset using the provided script:

python data\_preparation.py

# Usage

1. Data Preparation:

python data\_preparation.py

2. Feature Extraction:

python feature\_extraction.py

3. Model Training:

python model\_training.py

4. Prediction:

python predict.py

# Model Architecture

The hybrid model architecture consists of two branches:  
- OCR Branch: Input textual features are processed by an Embedding layer, followed by an LSTM to capture sequence information.  
- CNN Branch: Visual features are pre-extracted and passed through fully connected layers for further processing.  
- Combined Output: The concatenated features from both branches are passed through fully connected layers to produce the final entity value predictions.

# Performance

Validation Accuracy: 87%

F1 Score: 0.85

# Future Improvements

1. Data Augmentation: Implement advanced augmentation techniques to improve generalization.  
2. Advanced OCR Methods: Explore state-of-the-art OCR techniques for improved text recognition from images.  
3. Hyperparameter Tuning: Apply Bayesian optimization or other advanced methods to fine-tune the model’s performance.

# Contact

If you have any questions, feel free to open an issue in the GitHub repository or reach out directly via email.