Amazon ML Challenge 24'

1. Problem Statement:
Feature Extraction from Images

2. Dataset:
On exploring the data, the following was observed:

- The dataset was clean with no missing values.
- It included eight possible entity values, such as - 'item_weight', 'item_volume', dimensions ('depth', 'width', 'height'), 'voltage', 'wattage'
- The data was highly imbalanced, with an overwhelming number of records representing 'item_weight'.

Our primary focus was to carefully sample the training images to address data imbalance and minimize the strain on our hardware resources.

3. Outline of Approach:
In order to retrieve the target value (entity_value), the text from the image had to be extracted and processed to obtain the numeric value with the correct unit. After Exploratory Data Analysis & data preprocessing, we employed the following methodologies to achieve the expected results:

3.1. Entity Processing:

Step 1: The 'entity_name' is represented as an integer, which is then passed through an embedding layer for dimensionality reduction.
Step 2: The output embedding is flattened to prepare it for concatenation with the other embeddings.

3.2. Text Processing:

Step 1: OCR and noise reduction are applied to the image, extracting and cleaning the text.
Step 2: The extracted text is tokenized and encoded using BERT for semantic understanding.
Step 3: The BERT embeddings are prepared and flattened to make them ready for concatenation with other inputs.

3.3. Image Processing:

Step 1: The image is resized for standardization, and feature extraction is performed using the VGG16 CNN model.
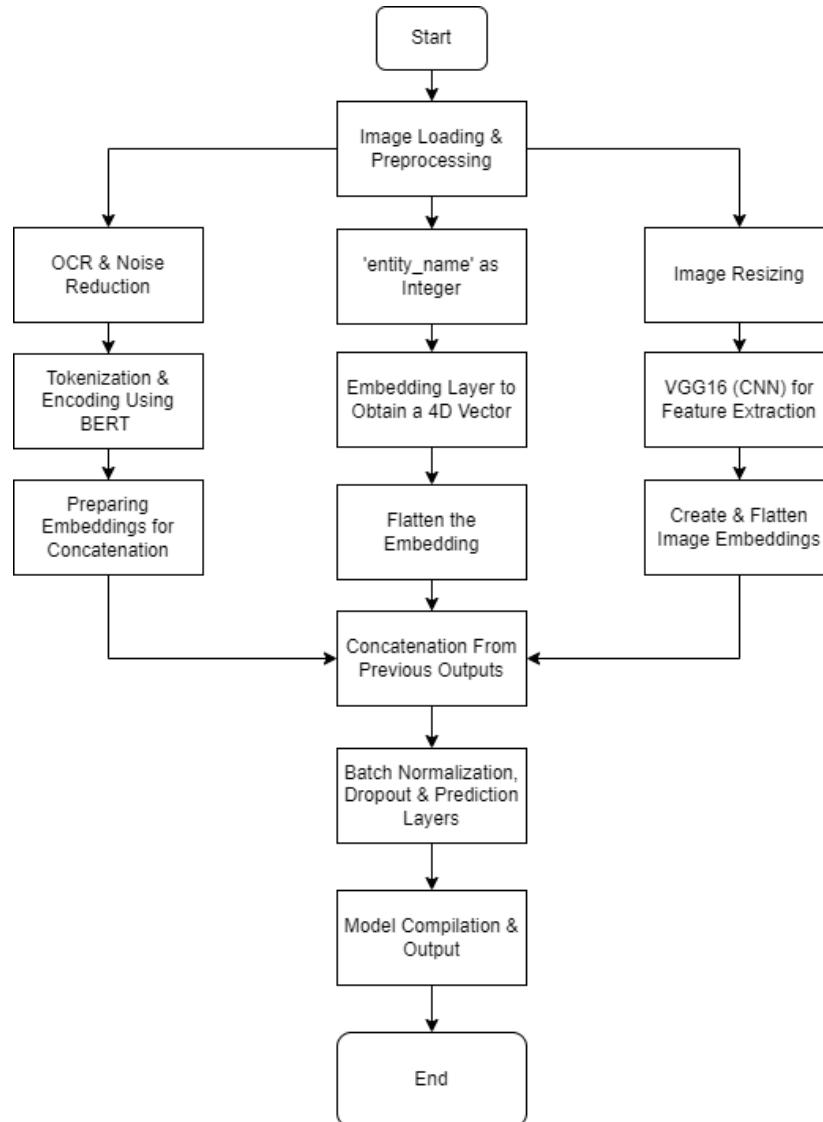**Step 2**: The extracted image embeddings are flattened for further concatenation.

Fig. Flowchart

3.4. Combined Processing:

Step 1: The embeddings from all three pipelines (entity, text, and image) are concatenated together.
Step 2: Batch normalization, dropout, and prediction layers are applied to regularize and improve the model.
Step 3: The model is compiled, and the final output is generated.

4. Conclusion:

In conclusion, by leveraging OCR, BERT, and VGG16 in a combined pipeline, we successfully extracted entity values from images with improved accuracy. Addressing data imbalance and hardware constraints, we ensured an optimized approach for entity-value extraction across multiple categories.