OCR on CAPTCHA Dataset: Report

Exploratory Data Analysis (EDA):

1. The dataset consists of 6,000 training samples and 2,000 test samples.
2. CAPTCHA labels are six-digit sequences.

Modeling:

Model :1

1. Model PyTesserect: url
2. Approach: Utilized a pre-trained model directly for detection. Attempted fine-tuning but faced challenges, so fine-tuning was not completed.
3. Pre-processing: Applied two rounds of pre-processing:
    1. Gaussian blur
    2. Otsu thresholding

Model:2

1. Model PyTesserect: url
2. Approach: Used the pre-trained model for detection. Fine-tuning was attempted but faced difficulties.
3. Pre-processing: Applied two rounds of pre-processing:
    1. Gaussian blur
    2. Otsu thresholding
    3. Gaussian blur
    4. Otsu thresholding

Model: 3

1. Model TR-OCR : URL
2. Approach**: Utilized the pre-trained model "microsoft/trocr-base-printed". Attempted using "microsoft/trocr-small-printed" but encountered issues during inference.
3. Pre-processing**: None. The model takes RGB input, so CAPTCHA images were directly used for inference.

Model:4

1. Model Finetuned TR-OCR model on captcha dataset: URL
2. Approach**: Fine-tuned the pre-trained model "microsoft/trocr-base-printed" on the CAPTCHA dataset. Attempted using "microsoft/trocr-small-printed" but faced inference issues.

By: Manish Kumar

3. Pre-processing**: None. The model takes RGB input, so CAPTCHA images were directly used for inference.

Model:5

1. Model: Fine-tuned TR-OCR with Pre-processing: URL
2. Approach**: Fine-tuned the pre-trained model "microsoft/trocr-base-printed" on the CAPTCHA dataset.
3. Pre-processing:
    1. Converted RGB images to grayscale.
    2. Applied Gaussian blur for noise reduction.
    3. Used Otsu thresholding.
    4. Converted grayscale images back to RGB for model input.

Preprocessing:

1. Converted RGB images to single-channel grayscale.
2. Applied Gaussian blur for noise reduction and Otsu thresholding.
3. Passed the pre-processed images to the PyTesserect model.

Post-processing:

1. Removed non-digit characters and unnecessary spaces from the detected text.

Observation:

1. CAPTCHAs with digits that are not tilted are accurately detected. However, the model struggles with CAPTCHAs containing tilted digits.
2. The model has difficulty detecting CAPTCHAs where two characters are touching.

By: Manish Kumar

Evaluation:

Models:	**Precision (Validation data)**	**Precision with Levenshtein distance(for incorrect detected samples only)**	
Pytessrect	28.2	0.33(1440 samples)	
Pytessrect(Preprocessing twice)	33.1	0.39(1340 samples)	
Pretrained TR-OCR	81.3	0.79 (373 samples)	
Finetune TR-OCR	91.9	0.80(160 samples)	
Finetuned TR-OCR(with preprocessed input)	89.1		

By: Manish Kumar