

Things in common from all proposals

1. Technical Approach:

- Two methodological:
 - **Detection + Classification pipeline** (Luca): detects characters and then classifies them.
 - **Sequence recognition (CRNN + CTC)** (Kengo, Nitish, Muhid): directly predicts text sequences without segmentation

2. Model Design:

- **CNN backbones** (ResNet-18, VGG)
- **BiLSTM layers** are included in the CRNN-based approaches (Nitish, Muhid), while Luca's model does not use RNNs
- **CTC loss** is adopted by the sequence recognition approaches (Kengo, Nitish, Muhid), while Luca combines **IoU loss + Cross Entropy** for detection/classification

3. Data Preprocessing & Augmentation:

- Everyone has similar ideas:
 - Preprocessing:
 - grayscale conversion, resizing, normalization, mapping labels to IDs for training
 - Augmentation:
 - Random rotations ($\pm 10\text{--}15^\circ$)
 - Contrast and brightness adjustment
 - Gaussian noise addition
 - Blurring
 - Line/shape occlusions

4. Loss Function & Training Strategies:

- **CTC loss** for Kengo, Nitish, Muhid
- **IoU + Cross Entropy** for Luca
- Optimizers:
 - mostly **Adam or SGD** with learning-rate scheduling (cosine annealing or step decay)
- Regularization:
 - **dropout and weight decay** are common for all

5. Approach to Part 3 & Part 4:

- Part 3:
 - Everyone proposes **stronger data augmentation** (more noise, distortions, rotations, background changes)
 - Nitish and Muhid propose **ensembling** or **uncertainty estimation**
- Part 4:
 - Luca, Kengo, and Nitish plan to use **oriented bounding boxes** or angle prediction
 - Muhid considers both option 1 and 2

Things can be discussed together

1. Overall Model Architecture (2 options)

- **Detection + Classification pipeline** (Luca)
- **End-to-end Sequence Recognition (CRNN + CTC)** (Kengo, Nitish, Muhid)

2. Backbone and Detailed Design

- **CNN backbones** (ResNet-18? VGG? or other?)
- **BiLSTM layers** are used in the CRNN-based models (Nitish, Muhid) for sequence modeling, while Luca's approach does not require them
 - Whether **RNN layers** should be included? or whether a purely CNN-based approach is enough?

3. For Bonus Tasks

- Part 4:
 - decide option1 or option 2

4. Task distribution idea