Indian Currency Detection Project Data Report

1. **Data Collection**
   Raw Data:
   o Collected images of Indian currency notes from publicly available datasets or proprietary sources.

   o **Creation Date:** December 2024.

   o **Data Source:** GitHub repository (open source).

   o **Owner:** Vijay S P

   o **Structure:** Dataset consists of images in JPG and PNG formats.

2. **Dataset Features**
   Attributes/Features:
   o **Image Dimensions:** Resolutions vary depending on source.
   o **Image Format:** PNG and JPEG.
   o **Currency Features:** Includes denomination numbers, watermark areas, and security features such as color-shifting ink and micro-text.

   o **Labels/Target Variables:**Bounding boxes representing regions of interest on the currency notes.

3. **Preprocessing**
   Data Cleaning:
   o Verified image integrity and removed corrupted or unrelated files.
   o Standardized file naming conventions for processing ease.

Normalization/Scaling:

- Resized images to 640x640 pixels to meet YOLOv5 requirements.
- Scaled pixel intensity values between 0 and 1.

Encoding:

- Converted annotations to YOLO format (x_center, y_center, width, height).

4. Data Splits

- **Training Set:** 70% (approx. 3500 images).
- **Validation Set:** 20% (approx. 1000 images).
- **Test Set:** 10% (approx. 500 images).

5. Documentation

Data Dictionary:
- **Image File:** String, unique identifier for each image.
- **Annotations:** Bounding box coordinates in YOLO format.
- **Data.yaml:** Contains IDs and names of all labeling classes used.

Codebook:

- Labels generated using the LabelImg tool.Data augmented using Albumentations for adjustments like lighting, flips, and rotations.

Usage Guidelines: The dataset is optimized for integration with YOLOv5 for object detection and inference.

6. Annotations
Labels:
- Manually drawn bounding boxes for currency note features using LabelImg.

Tags/Metadata:

- Class tags for features like "Denomination," "Watermark," "Security Thread."

7. **Storage**

   File Formats:

   - Images in PNG and JPEG.
   - Annotations in plain text format compatible with YOLOv5.

   Location:

   - Stored locally in the system's disk or cloud-based storage.

8. **Versioning**

   - Dataset version-controlled via Git, with changes logged for reproducibility.

9. **Accessibility**
   APIs:
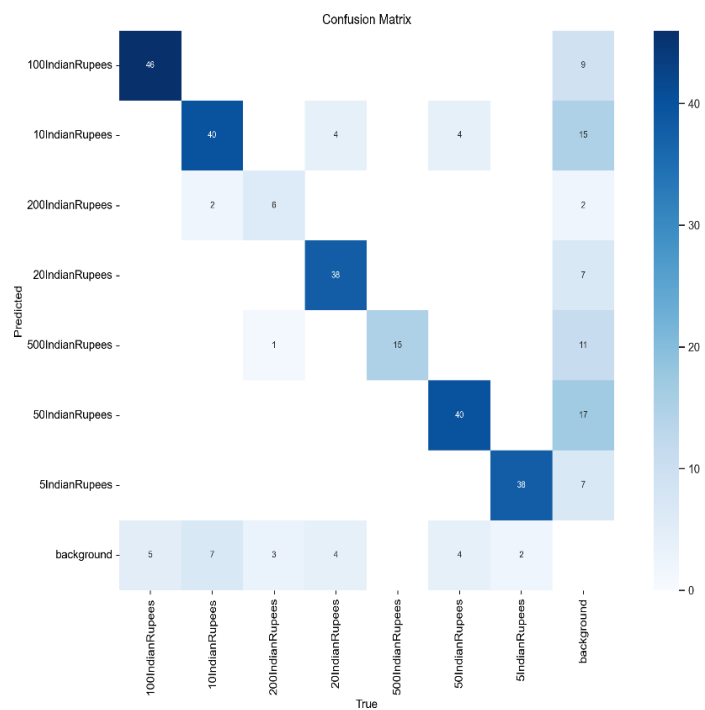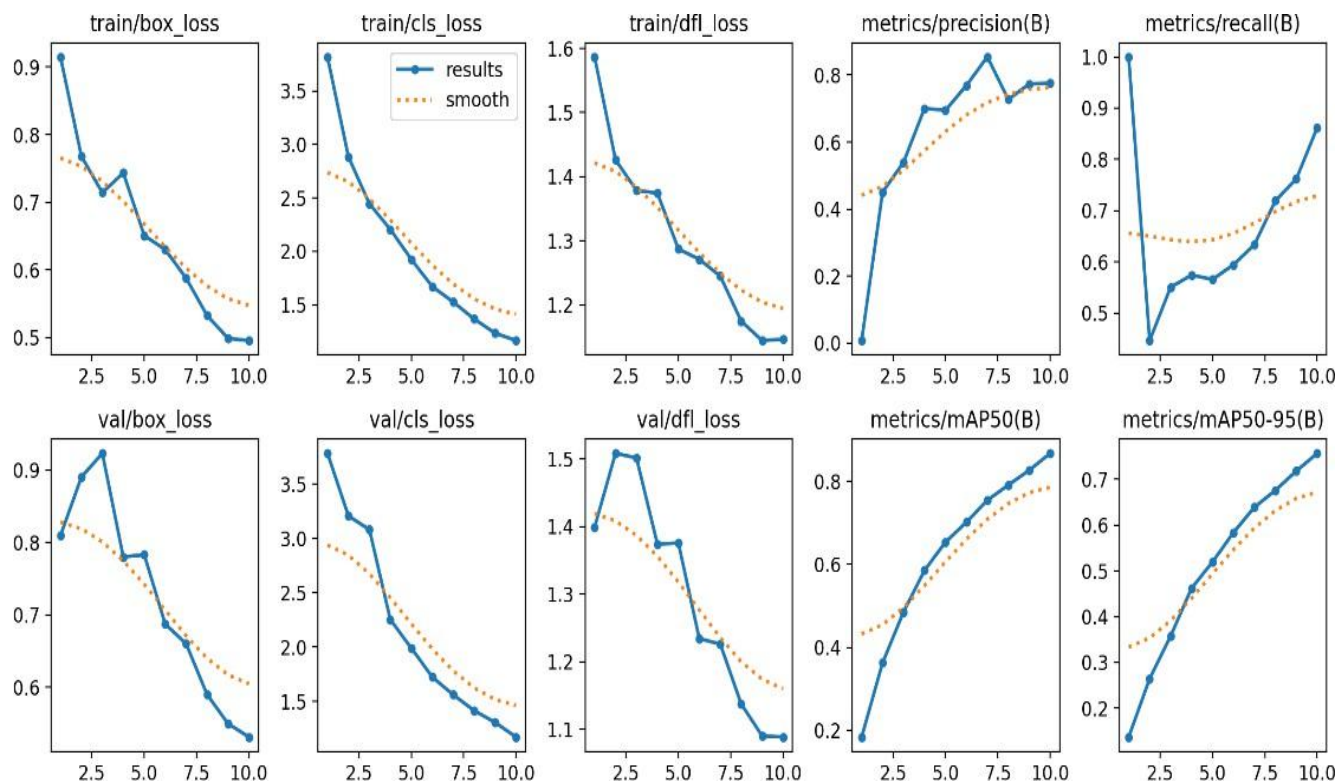   - Data accessible via Python scripts interfacing with YOLOv5.
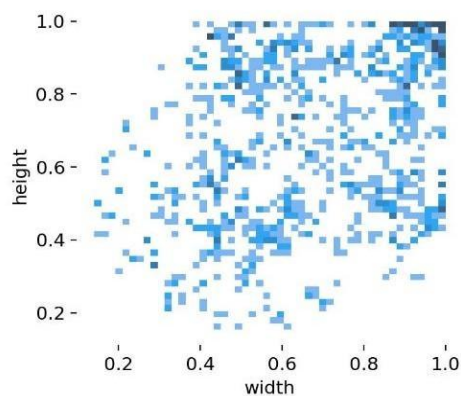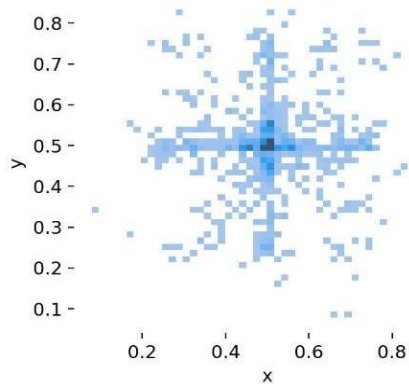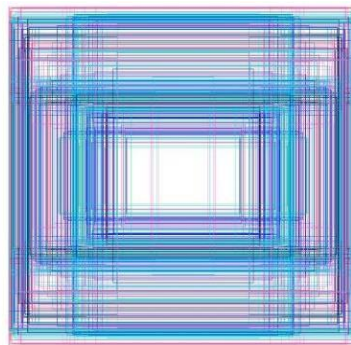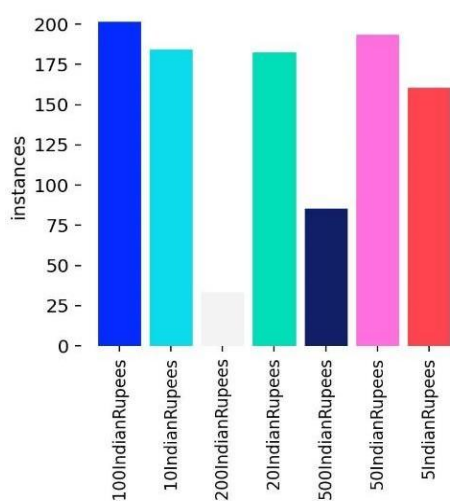
   Licensing:

   - Dataset subject to specific licensing terms (e.g., open source or proprietary use).

10. **Visualization**
    Summary Charts:
    - Histograms of class distribution.
    - Scatter plots for bounding box dimensions to ensure annotation quality.

Confusion Matrix

F1-Confidence Curve

train/box_loss   train/cls_loss   train/dfl_loss   metrics/precision(B)   metrics/recall(B)

results
smooth

val/box_loss   val/cls_loss   val/dfl_loss   metrics/mAP50(B)   metrics/mAP50-95(B)
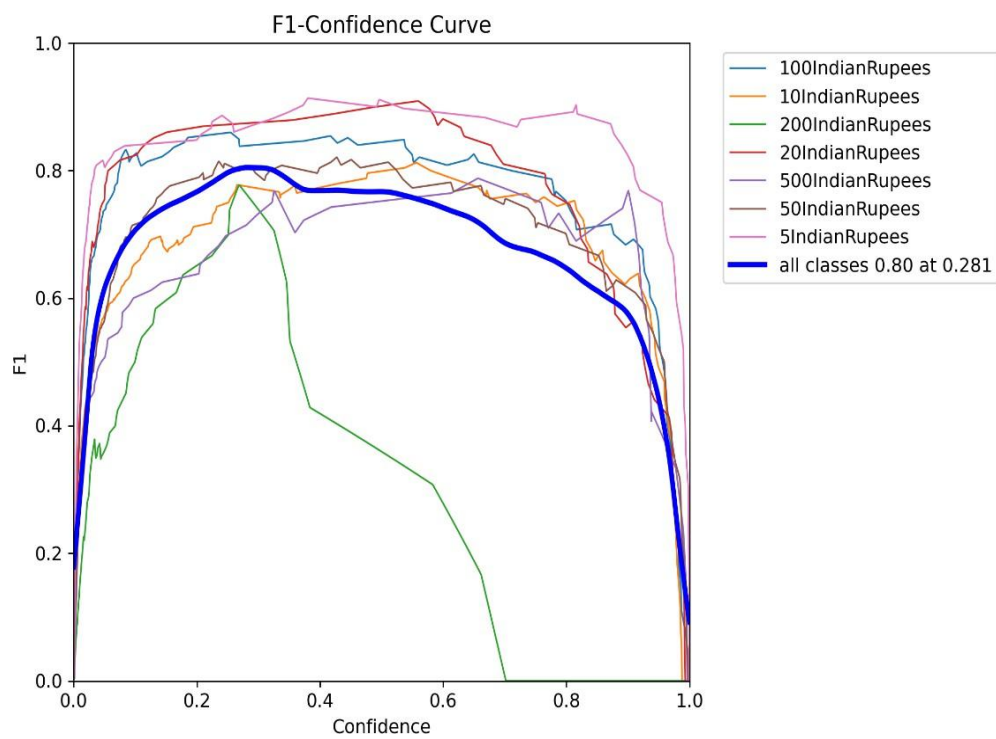
Exploratory Analysis:

o Visual inspection of sample images with bounding boxes.

11. **Security & Privacy**
    Anonymization:
o Dataset excludes any personally identifiable information (PII).
    Data Encryption:
o Not applicable for publicly available, non-sensitive data.

12. **Validation and Testing**
o **Dataset Tested For:**
  ▪ File integrity and accessibility.
  ▪ Annotation correctness through visual inspection.
  ▪ Consistency across training, validation, and test splits.