

(a)



(b)



(c)



(d)

# Exploratory Data Analysis on Helmet Detection and Bangla License Plate Datasets

S.M Hasnan Monir 24341278

HASNAT RAFI UDDIN 24341277

MAZBHA UL HAQUE 24341266

ASRAR AZIM 21301634

# Introduction to the

## Datasets

### Helmet Detection

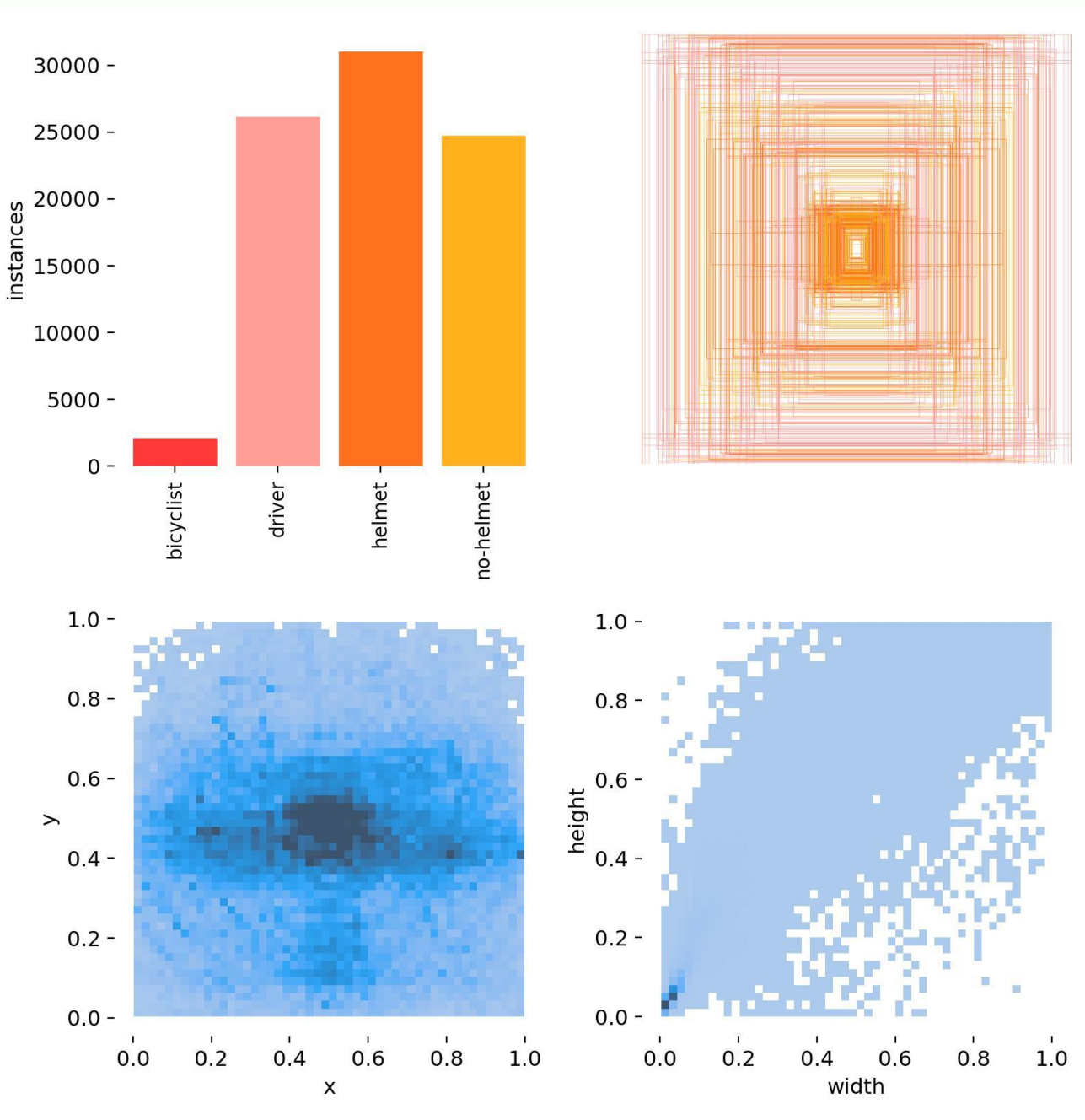
The Helmet Detection dataset contains images of people on motorcycles. The goal is to build a model that can accurately detect whether a person is wearing a helmet.

<https://universe.roboflow.com/imagerecognition-43zpb/helmet-detection-ntbfz/dataset/18>

### Bangladeshi License Plate

The Bangladeshi License Plate dataset contains images of license plates. The goal is to build a model that can accurately extract the license plate number from the image.

<https://universe.roboflow.com/kaiser-durden/bangladeshi-license-plate-fonfq>



# Data Exploration: Helmet Detection

## 1 Dataset

The dataset includes a picture with 640X640 Pixels.

## 2 Train and Validation

We have 12163 images to train and 1540 pictures to validate the dataset after training the model.

## 3 Image Quality

The dataset may contain images of varying quality, potentially affecting model performance.

# Data Exploration: Bangla License Plate Dataset

## Plate Variations

Different plate designs, colors, and fonts exist, impacting character recognition.

## Background Complexity

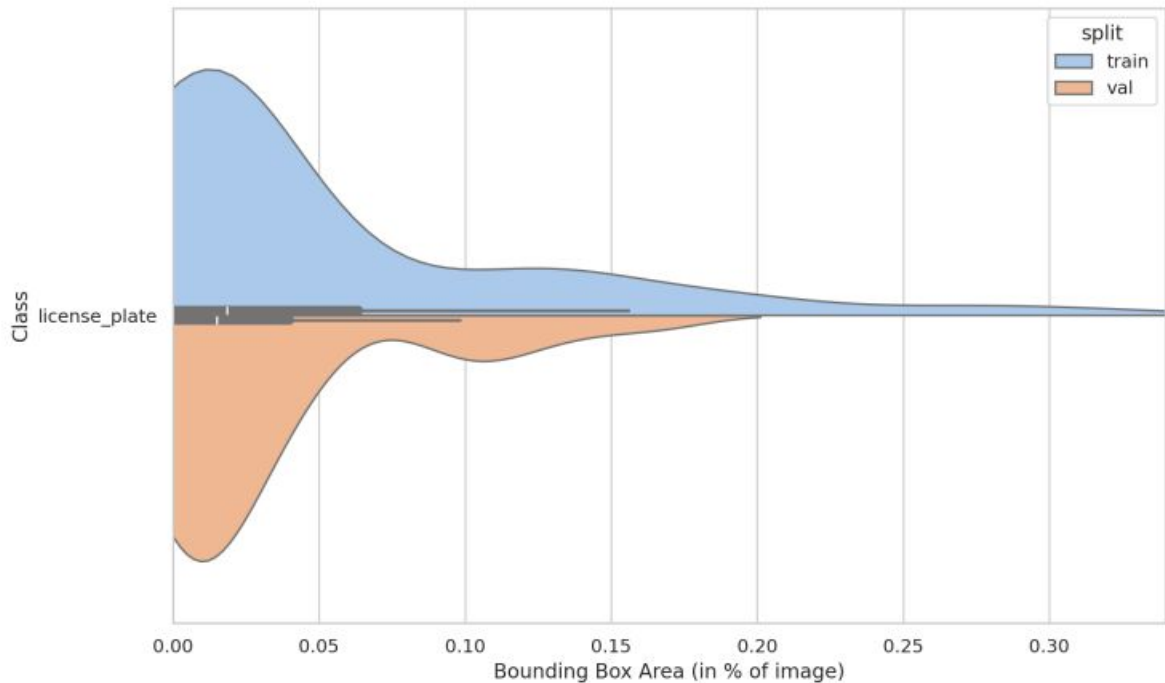
Backgrounds range from simple to complex, potentially hindering plate detection.

## 1. Image Features

### 1.1. General Statistics

	Train	Validation
Images	5501	1330
Classes	1	1
Classes in use	1	1
Annotations	5501	1330
Annotations per images	1.03	1.00
Images with no annotations	0	0
Median image resolution	416x416	416x416
Smallest annotation	0	0
Largest annotation	588	286
Most annotations in an image	4	1
Least annotations in an image	1	1

### 2.3. Distribution of Bounding Box Area





# Data Visualization and Insights



## Confusion Matrix

The matrix shows true versus predicted labels, with diagonal values representing correct predictions



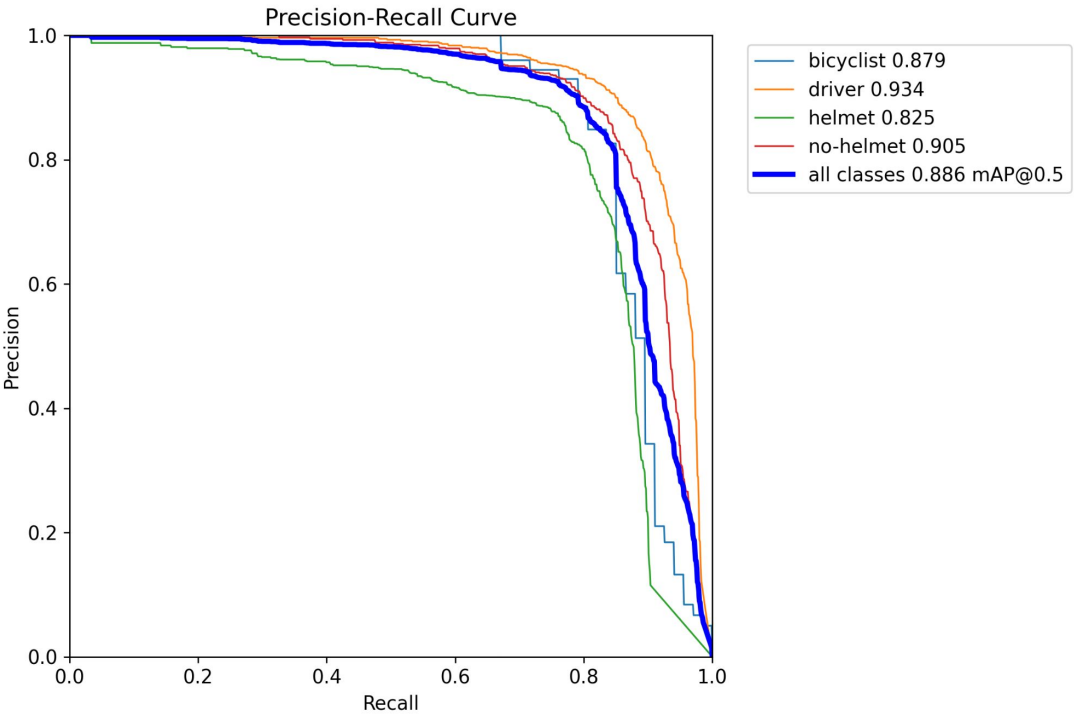
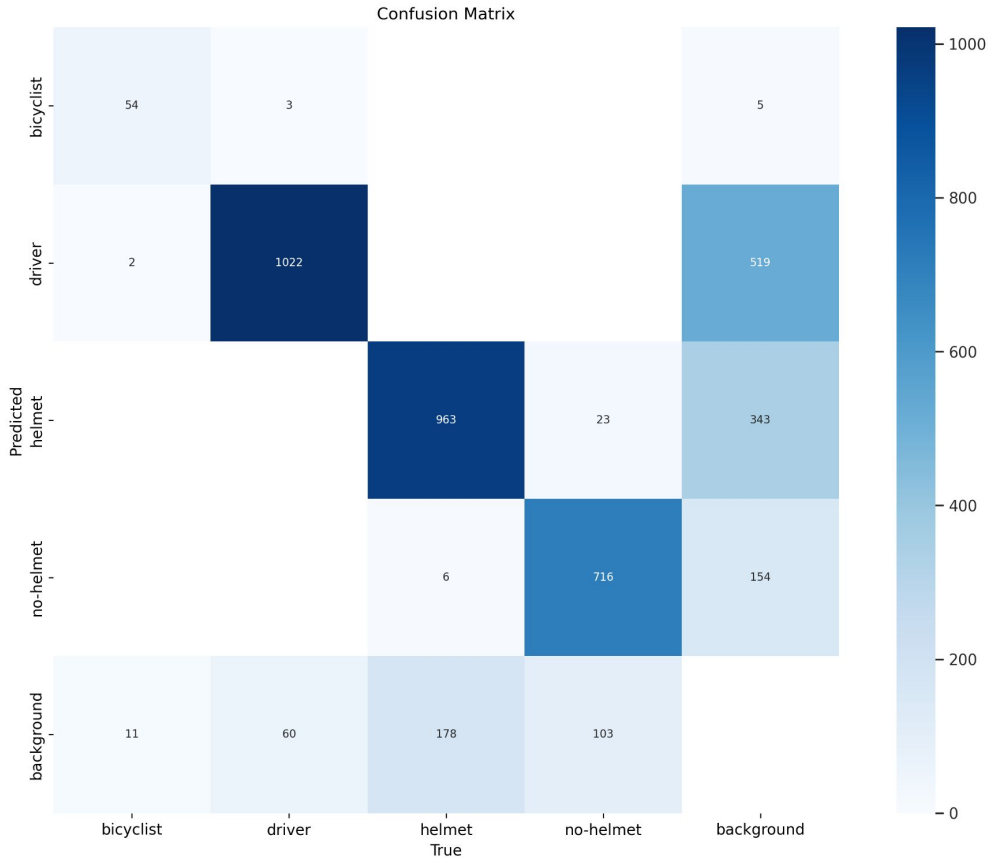
## Recall-Confidence Curve

This plot shows recall at different confidence thresholds, with 0.95 recall achieved at a confidence of 0.000. It assesses model confidence versus recall performance.



## Precision-Recall Curve

The curve balances precision and recall, achieving an overall mAP@0.5 of 0.886. Drivers perform best (0.934), while helmets lag at 0.825.



# Helmet Detection Test

The images show YOLOv8 model detections for people wearing helmets (2), bicyclists (3), and drivers (1), with bounding boxes correctly identifying subjects across various urban settings and lighting conditions.

The model appears effective at distinguishing between helmet/no-helmet cases and detecting multiple subjects in crowded scenes, though some instances show overlapping detection boxes.



**Thank You**