ML Study Design - Google Street View Blurring System

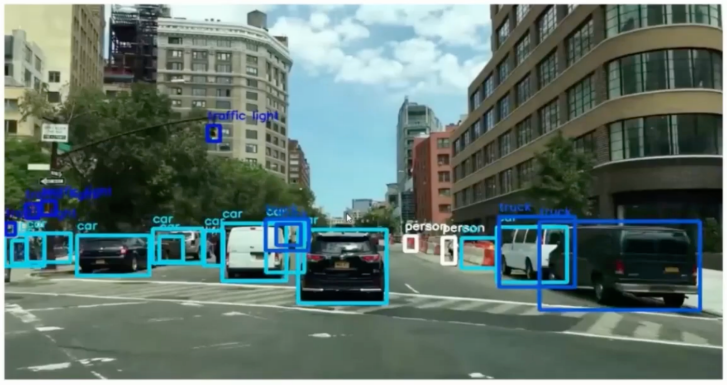
Objective:

ML System Design -02
Google Street View Blurring System

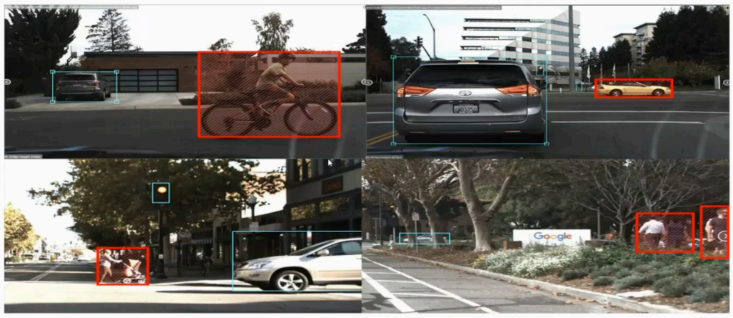
For Privacy protection blurring faces & Car License plates

1. Clarifying the Requirements: Designing a Street View Blurring System which blurs human faces and license plates. B.O. is to protect user privacy. We will also be using training dataset of 1M annotated images of human faces and license plates.

2. Framing the problem as an ML task: To accurately detect object of interest in the image. After detecting the objects we can blur them before displaying it to users.



A. Specifying the i/o of the system: An image with zero to multiple objects at different locations within it. The model would detect and outputs those locations.

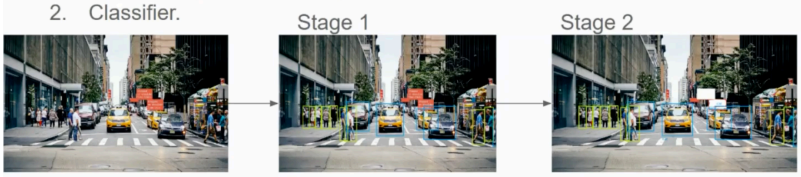


3. Choosing the right ML category.

1. Predicting the location of Object in the image.
2. Predicting the class of each bounding box (person, car, people on bike/cycle etc.)

- The first is regression problem, where the location can be specified by (x,y) coordinates. The second one is multi-class classification.
- Generally, object detection are divided in two parts:
  1. Two Stage Network
  2. One Stage Network

1. Two stage Network: (RCNN, Fast RCNN, Faster RCNN)
   1. Region Proposal Network (RPN).
   2. Classifier.



2. One-Stage Network: (YOLO, SSD)
   - Both the stages are combined. Bounding boxes and object classes are generated simultaneously, without explicit detecting for regional proposals.

One Stage vs Two Stage:

- Two stage networks perform the operation in two sequential steps (slower but accurate).

We will work with Two Stage.
We are dealing with dataset of 1 million images which is not huge by modern standards. And when the training data increases or the need for more quicker predictions arises in future, we can later shift to One Stage.

Data Preparation:Data Engineering:

- Annotated Dataset.
- Street View Images.

1. Annotated Datasets: We have 1 million annotated images with each having bounding boxes and associated Object classes.

Image path	Objects	Bounding boxes
dataset/image1.jpg	human face human face license plate	[10, 10, 25, 50] [120, 180, 40, 70] [80, 95, 35, 10]
dataset/image2.jpg	human face	[170, 190, 30, 80]
dataset/image3.jpg	license plate human face	[25, 30, 210, 220] [30, 40, 30, 60]

2. StreetView Images:

The ML system will process these images to detect human faces and license plates.

Image path	Location (lat, lng)	Pitch, Yaw, Roll	Timestamp
tmp/image1.jpg	(37.432567, -122.143993)	(0, 10, 20)	1646276421
tmp/image2.jpg	(37.387843, -122.091086)	(0, 10, -10)	1646276539
tmp/image3.jpg	(37.542081, -121.997640)	(10, -20, 45)	1646276752

Data Engineering : Feature Engineering

- Applying standard Image Preprocessing operations like Resizing, normalization etc. we will be using data augmentation to increase the size of the dataset.

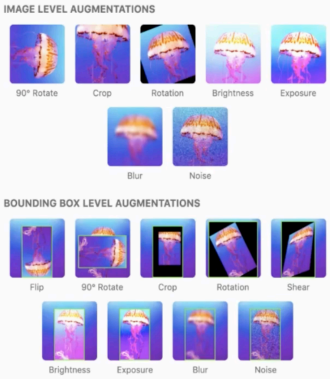
Data Augmentation: Adding a slightly modified copies of original images or creating new images from the original image.

- Helps learn complex patterns
- Helps with imbalanced dataset.

Careful with rotation, flipping and bounding boxes In rotation.

Performing Augmentation: Offline vs Online

Offline - Fast training time but takes more storage
Online - Slow training time but doesn't take additional storage.

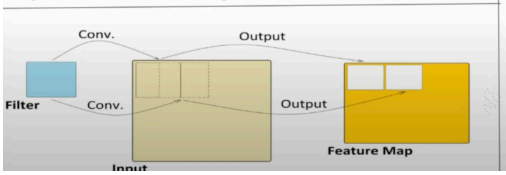


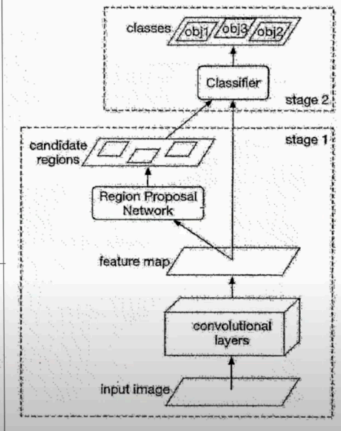
Model Development:

Convolutional Layer - prepares feature map of image by taking in input as an image.

RPN - proposes candidate regions, takes feature map as input and gives candidate regions in the image.

Classifier - Takes the feature map and the proposed candidate regions and assigns an object class to each region.





Model Training:

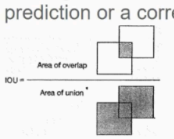
- The model is expected to do two tasks well:
  - The bounding boxes are supposed to highly overlap the ground truth bounding boxes.
  - Predicted probabilities for each object should be accurate.
- We will use regression loss and classification loss - loss functions.

Usually contains: Forward propagation, Backward propagation and loss calculation.

Evaluation:
Generally, the machine learning model has to detect N different objects in the image. So to evaluate the model's performance, we will evaluate each object separately and average the results.

Intersection Over Union (IOU):

- Measures the overlap of the bounding boxes.
- Shows the detected bounding box are aligned with ground truth bounding box.
- IOU=1 indicates they are fully aligned, though rare.
- Higher IOU means more accuracy.
- IOU higher than 0.7 is considered a good prediction or a correct detection.



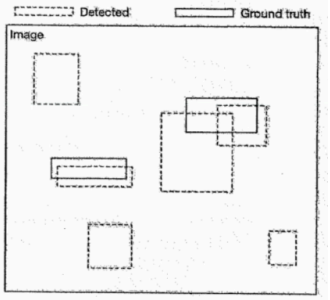


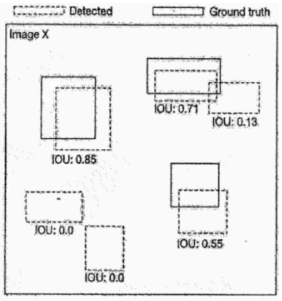
Figure 3.8: Ground truth and detected bounding boxes

Offline metrics:

Precision: P = correct precision/ number of precision
But for different IOU threshold eg. =0.7, 0.5 and 0.1 the value changes.

Average Precision: for measuring the precision of single object being detected by the model.

mAP : for measuring the model's overall performance.



Online metrics:
We will use "User Reports".

Serving:

- When we run an object detection algorithm, it is common to see overlapping of bounding boxes because RPN proposes various regions in the image.
- So, it is important to bring that down to one bounding box. For that we will use an algorithm, NMS.
- NMS will keep highly confident boxes and will remove the overlapping boxes.

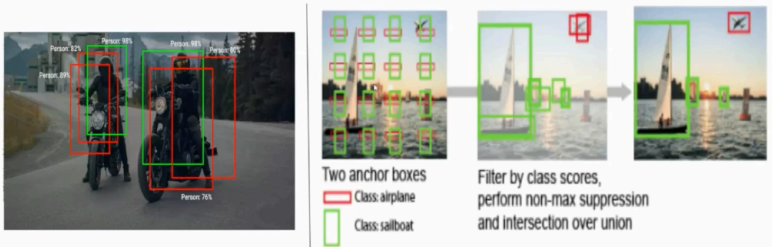
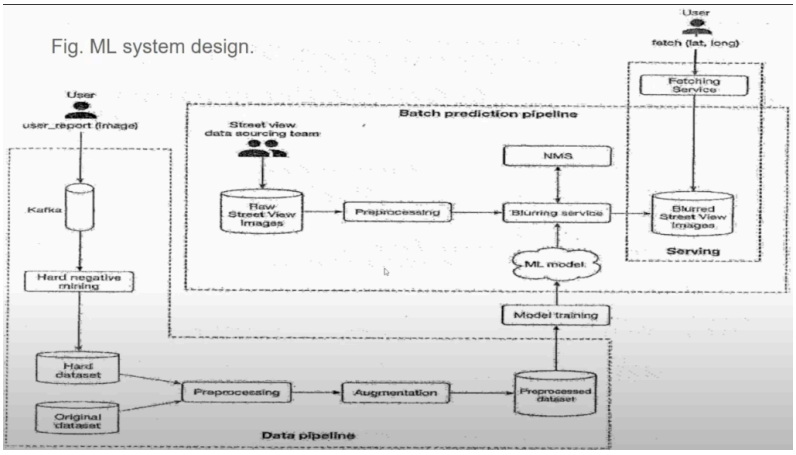


Fig. ML system design.



Batch Prediction pipeline:

1. Raw Street View images.
2. Preprocessing - preparing the images for the model.
3. Blurring system:
   a. Provides a list of detected objects.
   b. NMS gives the final detections.
   c. Blurs detected object.
   d. Stores the image in object storage.
   - The first two processes a & b are CPU bound processes, the next two processes c & d are GPU bound processes.
   - Advantages :
     - Scaling the services independently.
     - Better utilization of CPU and GPU.

Data Pipeline:

User-reports - we collect reports from user, and from hard negative mining we produce a hard dataset.

Combining those with the original dataset we train the model to improve the performance.