Pedestrian Detection

AI Project Presentation

Anita Chia | Bernadine Lye | Woojin Park

About Us

Woojin Park

BIDA-12

woojinpa@andrew.cmu.edu

Bernadine Lye

MISM-16

slye@andrew.cmu.edu

Anita Chia

MISM-16

sipeianc@andrew.cmu.edu

PROBLEM STATEMENT

Pedestrian Identification for Autonomous Vehicles

Problem Statement

Importance

- Safety for Driver
- Safety for Pedestrians
- Liability for Autonomous Vehicle Manufacturers

DATA & ENVIRONMENT

Pedestrian Dataset

Pedestrian Dataset

WiderPerson Dataset

- 13,382 Images
- 9,000 Annotations
- 5 main classes:
  Pedestrians, Riders, Partially-Visible Persons (PVP), Ignore, Crowd

Annotations & Images

Annotations

Data format:
[class_label, x1, y1, x2, y2]

class_label: 1 out of 5 classes

x1, y1, x2, y2: Coordinates to plot bounding boxes

```
1 45 235 79 318
1 60 209 120 356
1 119 214 168 336
```

Images

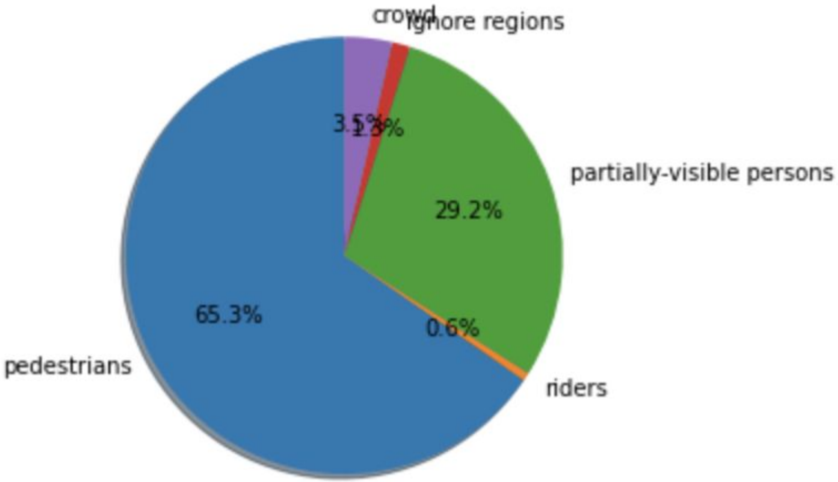
Our Environment

Local	Google Collab	AWS
Environment settings were unfavourable for our project	Could not handle the amount of pictures we had	AWS EC2 p2.XL

Exploratory Data Analysis

Pedestrian Dataset

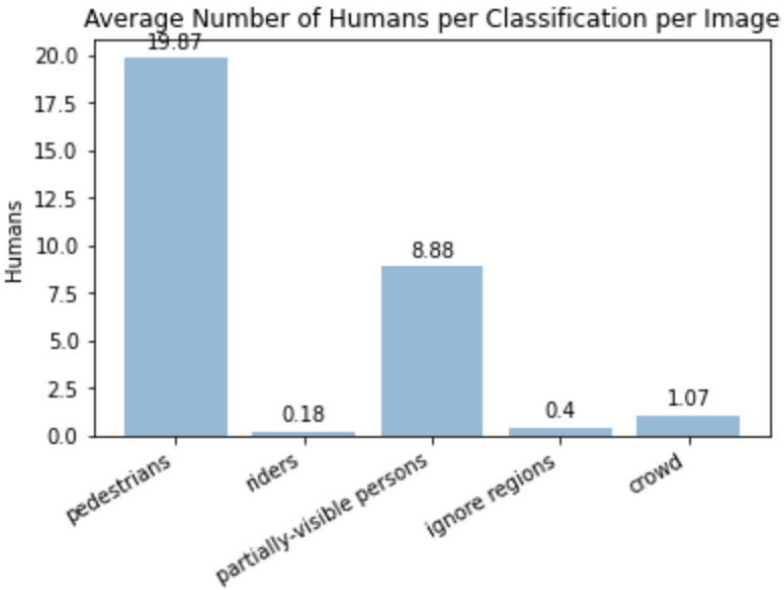
Exploratory Data Analysis

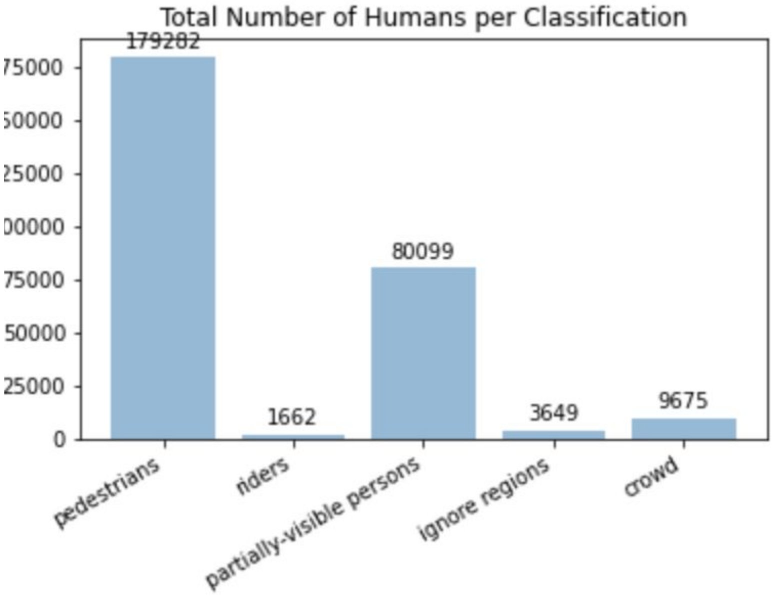


Class Breakdown

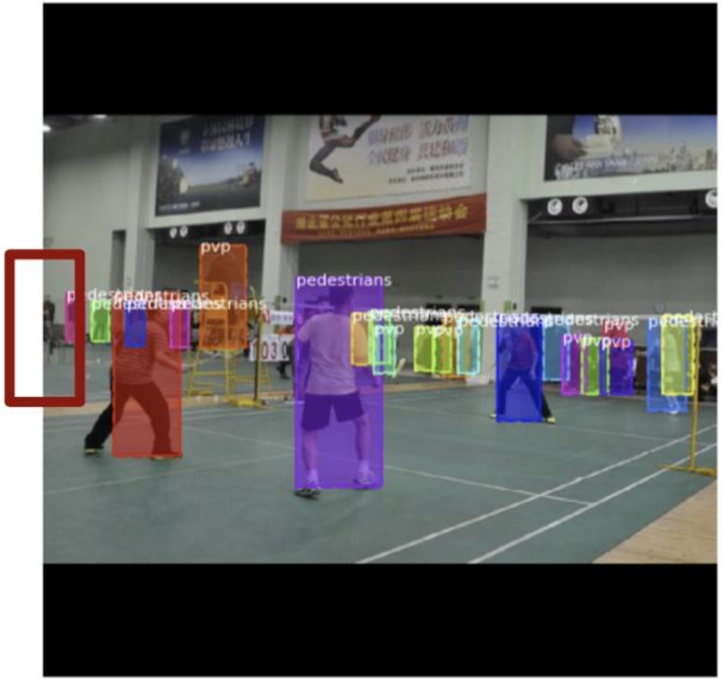
- 0.6% Riders
- 1.4% Ignore regions
- 3.5% Crowd
- 29.2% Partially-visible
- 65.3% Pedestrians

Exploratory Data Analysis





Exploratory Data Analysis



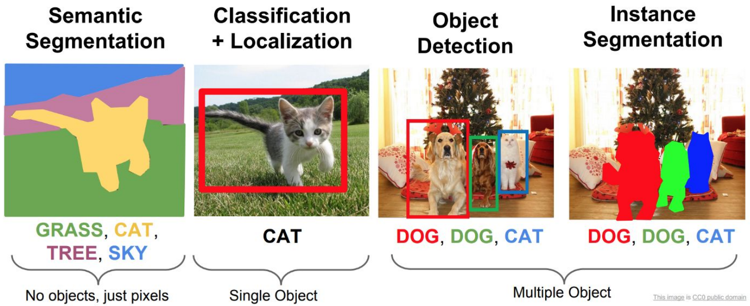
Our Approach

Mask R-CNN

Mask R-CNN

Instance Segmentation

There are four main classes of problems in 'detection' and 'segmentation' in Computer Vision as described in the image.



Why Mask R-CNN Instance Segmentation?

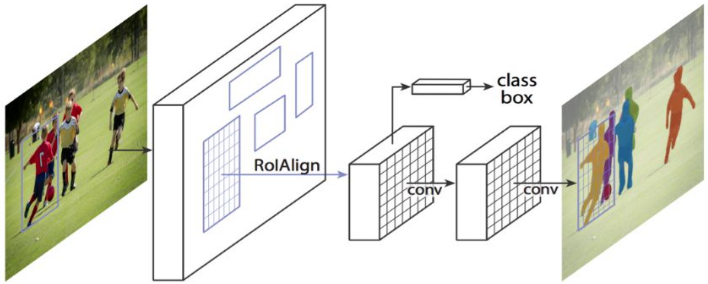
Our dataset contains dense human features with various kinds of occlusions

1. Object detection of all objects in an image
2. Segmenting each instance with high resolution

Mask R-CNN

3 main stages

1. **Regional Proposal Network (RPN)** - propose candidate object bounding boxes
2. Project this values into original image to draw anchor boxes: **Apply ‘Non-max-suppression’** to choose the highest score anchor box
3. **‘ROI Align’** to re-scaling the Anchor Boxes and pass anchor box through ‘Classification’, ‘Bbox regression’ and ‘Mask’ branch respectively



Mask R-CNN framework for instance segmentation. Source: https://arxiv.org/abs/1703.06870

Mask R-CNN

Mask R-CNN, extends Faster R-CNN Predicting

1. ‘Object mask’ in parallel with
2. ‘Bounding box’ recognition and ‘Confidence score’ which is the probability that an anchor box contains an object.
3. Class label

Mask R-CNN enables pixel-to-pixel alignment to have high-quality segmentation masks for each instance.

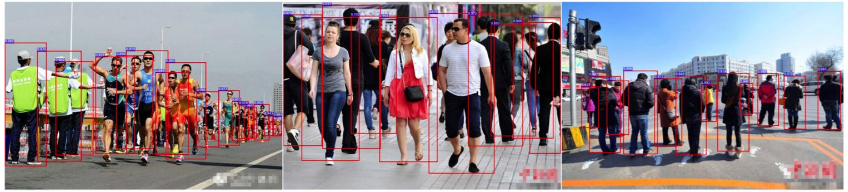


Fig1. Faster R-CNN Model result of Wider Person Journal

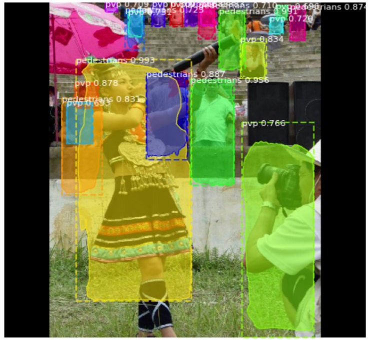
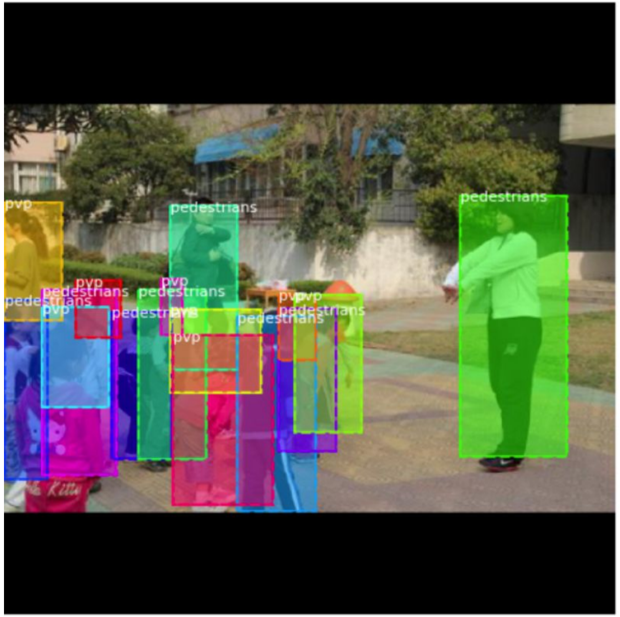


Fig 2. Image Segmentation result of our Mask R-CNN Model

Input Data

Images & Annotations from WiderPerson dataset



Our Results

Mask R-CNN

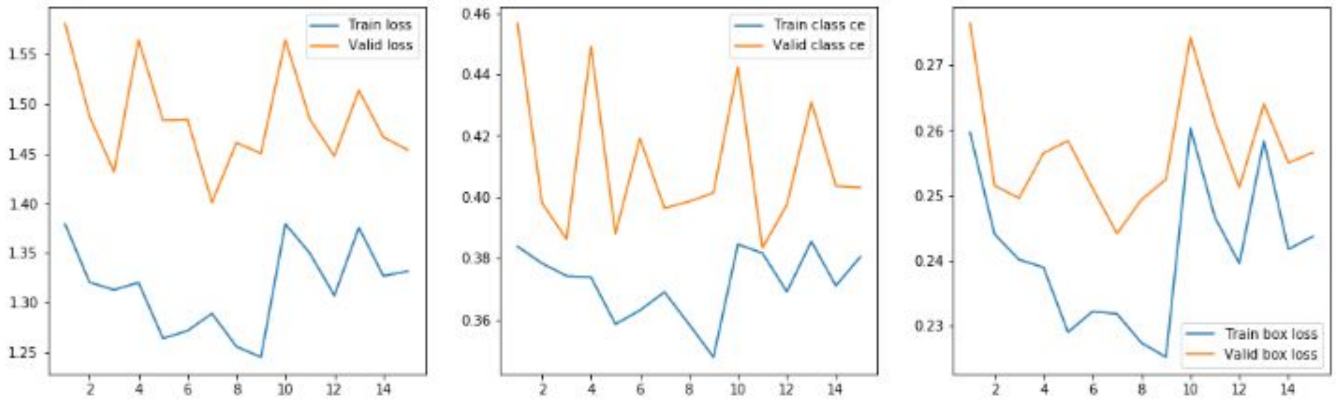
Evaluation Metrics

Optimising Loss

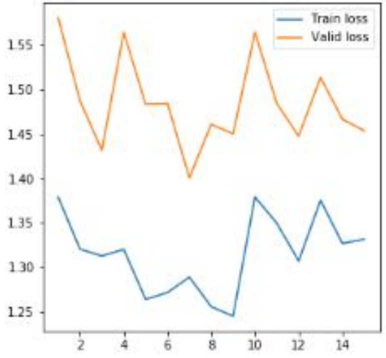
- The lower the loss, the better the model

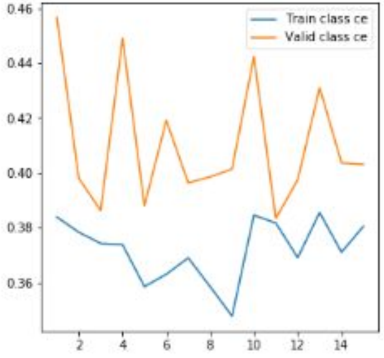
- Loss = Classification Loss + Bounding Box Loss + Mask Loss

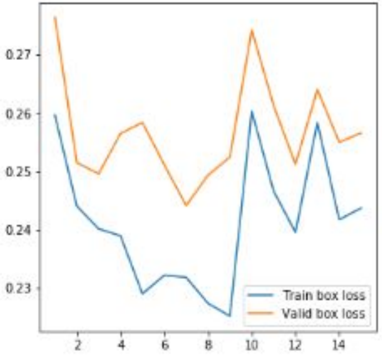
- Comparing Loss across epochs & trainings to find the best model



Evaluation Metrics







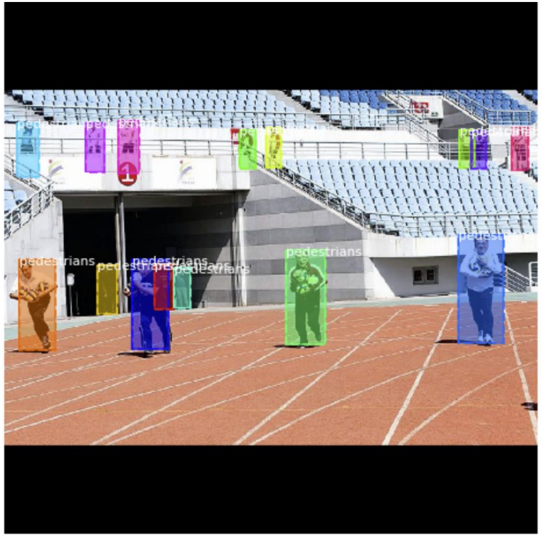
	Val_ loss	Val_ Rpn_ class_loss	Val_ Rpn_ bbox_loss	Val_ mrcnn_class_loss	Val_ mrcnn_bbox_loss	Val_ mrcnn_mask_loss	**loss**	Rpn_ class_loss	Rpn_ bbox_loss	Mrcnn_class_ loss	mrcnn_bbox_loss	Mrcnn_mask_ loss
Epoch 1	1.581092	0.113114	0.407088	0.456547	0.276372	0.327968	1.379078	0.077179	0.336080	0.383837	0.259619	0.322361
Epoch 9	1.540154	0.113099	0.374545	0.401329	0.252447	0.308715	1.244931	0.067287	0.307244	0.347745	0.225188	0.297448

Output

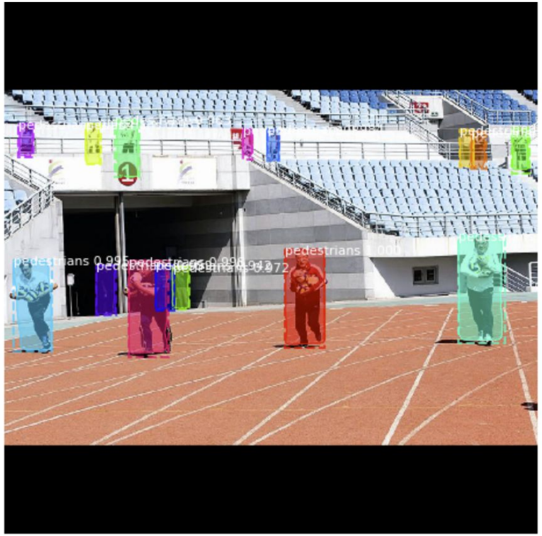
For Each Image:

- Bounding box
- Class label
- Segmentation mask
- Detection Confidence Level

Output Results

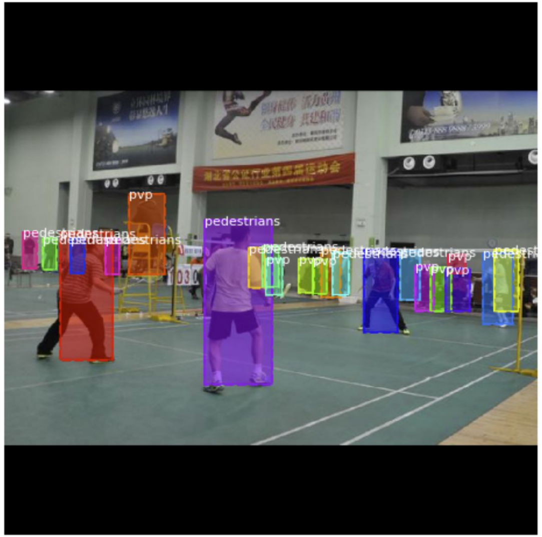


Actual

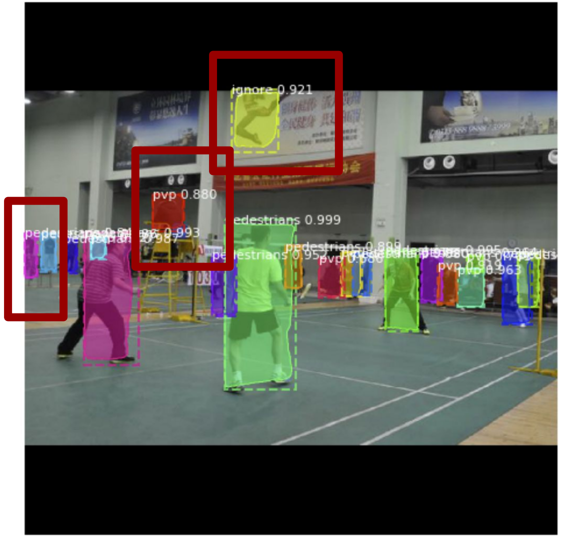


Predicted

Output Results

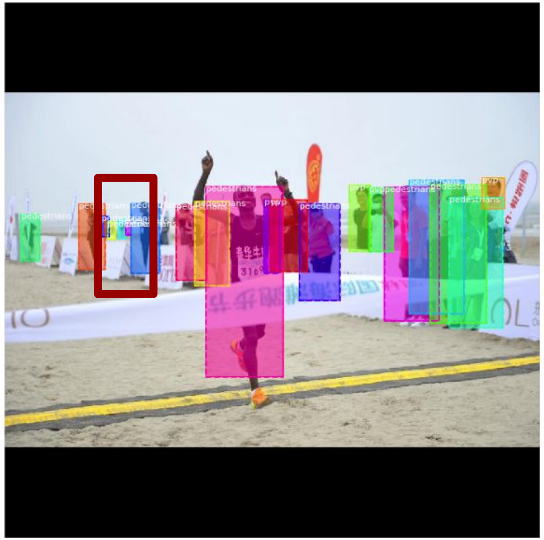


Actual

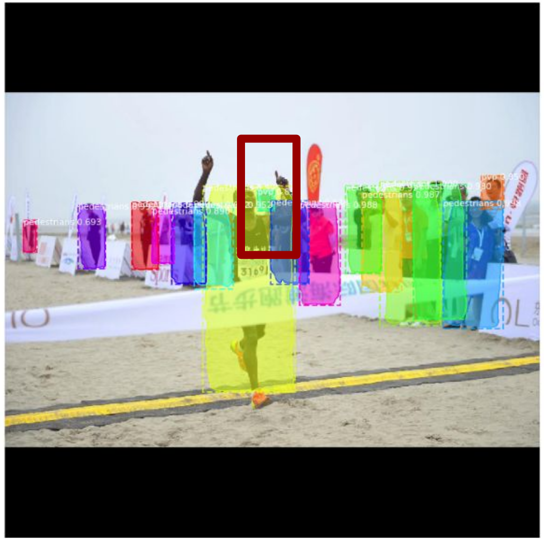


Predicted

Output Results



Actual

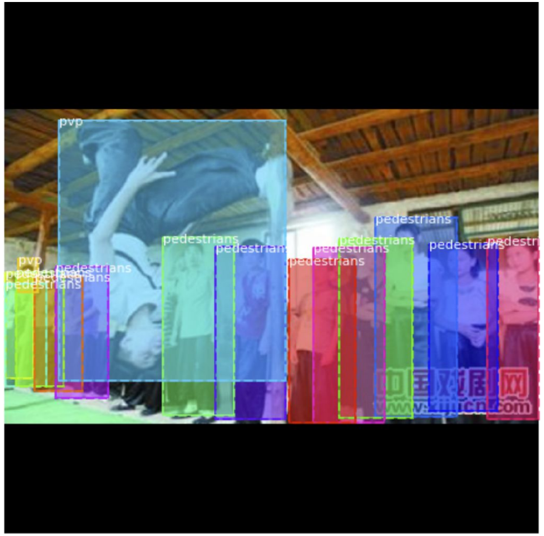


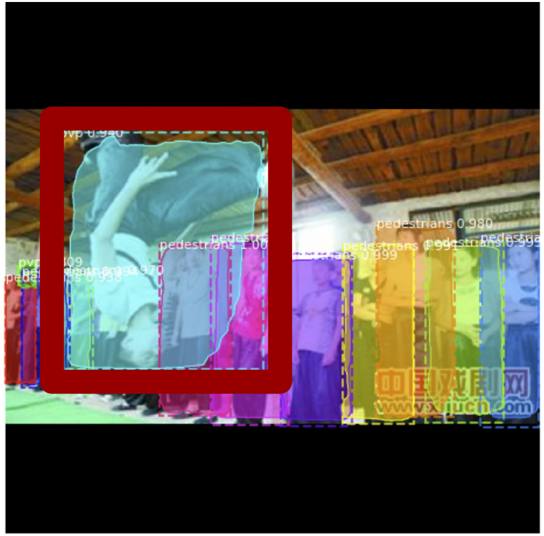
Predicted

Reflection

Our journey & future opportunities

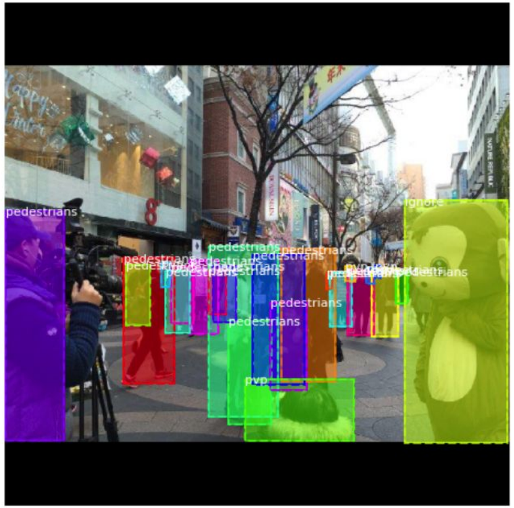
We were surprised!

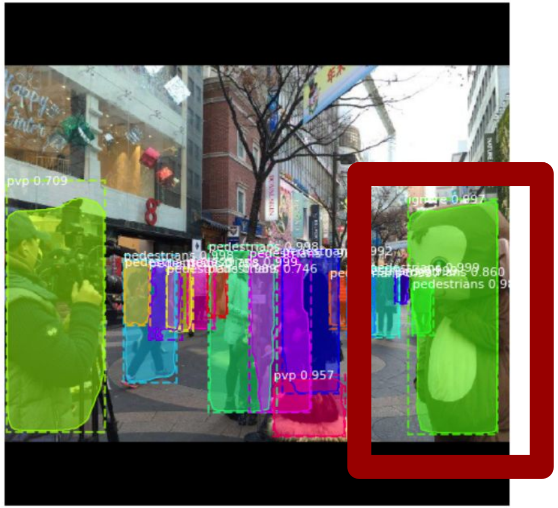




Our model was able to pick up partially visible, overlapped and even humans who were not standing upright.

We were surprised!





IGNORE

Our model was able to pick up non-humans

We were surprised!

Amazon Elastic Compute Cloud running Linux/UNIX		$132.98
$0.9 per On Demand Linux p2.xlarge Instance Hour	147.750 Hrs	$132.98

The computing power needed - as we were working with thousands of images

Next Steps

- Obtain more data for other classes
- Video detection (Live images)
- Compare other Object Detection models (i.e. YOLO)

Lessons Learnt

Computer Vision Techniques:
Object Detection vs Instance Segmentation

Importance of a **balanced dataset**

Was anything especially hard/easy?

Data was well-labelled and relatively clean, with standardised picture sizes

Multi-Class classification was challenging given the imbalanced dataset

Parameter tuning

Computational power was lacking - limiting our abilities to discover the best model

What we would have done differently

Find a more balanced dataset

or

Try to balance the dataset by looking for more pictures that represent these under-represented classes and manually label them

THANK YOU!