PART A:

1.
The initial dataset I brought to class on October 31st was composed of 152 images separated into three categories. The first 50 pictures were of various angle of a fork, another 50 pictures were of a kitchen cloth, and the rest was composed of random objects from around my apartment. When selecting the objects, I kept in mind that we had to bring them to school so it could not be too big or too heavy, that is why I chose the fork. I also thought it would be interesting given the curves a fork has. I then chose the kitchen cloth because, first it is a very cute cloth, and second, I thought it would be interesting to see if the fabric texture would mess with the object recognition that we would create given that it can be folded and placed in many different ways. It also has patterns on it which I thought might influence the program. The random objects of the dataset were thing I noticed in my apartment like a flower bouquet, a speaker, a lamp, books, etc.

2.
What we were asked to do in class was to train an AI with our dataset to recognise the objects we chose. It was explained that by uploading our images to edge impulse that we used in class, 80 percent of our dataset would be used as training data and the other 20 percent for testing.

3.

a. We created an account with edge impulse.
b. We uploaded the images of our first object with a label (“fork” for my images of fork, “cloth” for my images of kitchen cloth, and “other” for my images of random objects).
c. We then chose the resolution and size we wanted for our images. The better the resolution the easier it is for the AI to recognise an object, same for the colour: if we decide to change our images to a grey scale instead of RGB, the chances of the AI being accurate will drop.
d. The second step was the colour that I just mentioned, and we chose RGB as the colour depth.
e. We then chose a model for the neural network. We had to look at a model that worked with our images. Since we set them to be 160x160, we chose MobileNetV2 160x160 0.5.
f. We started the training and eventually got back results with the training performance, a confusion matrix (that we looked at in the presentation before starting), and a data explorer that displayed the test images in clusters.

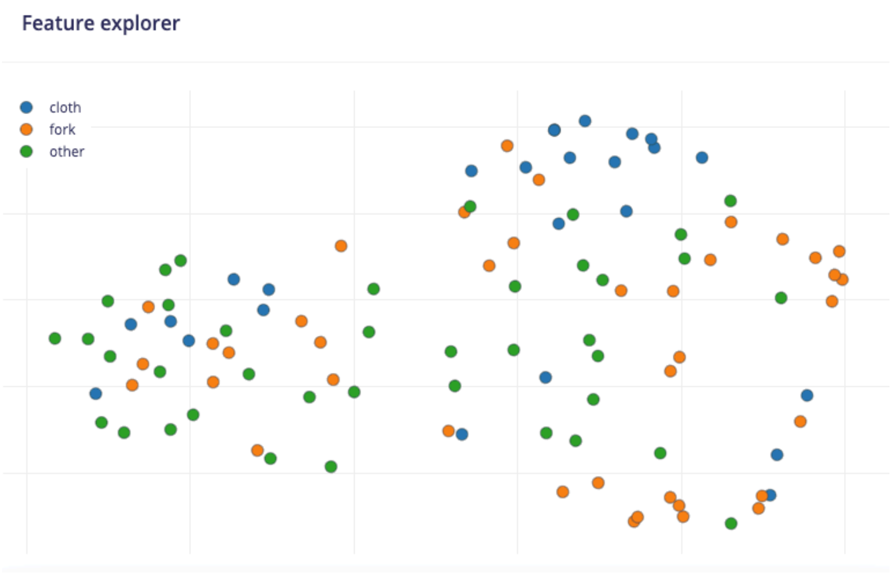
4.

The accuracy was 85.7% and the precision and recall for each category was as follows:

- Cloth
    - Precision: 1.00
    - Recall: 0.60
- Fork
    - Precision: 1.00
    - Recall: 0.91
- Other
    - Precision: 0.63
    - Recall: 1.00

5.

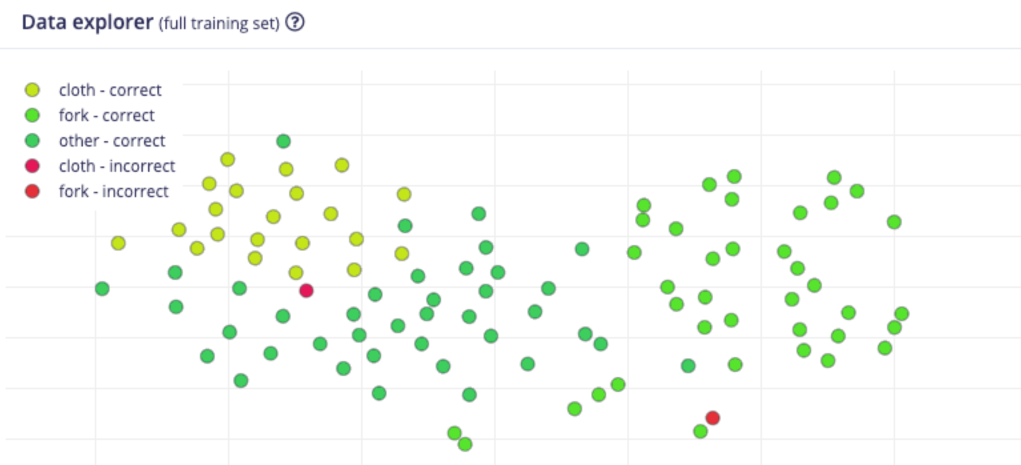
This is the graph I get from the feature explorer. Each point represents an image and how similar they are to the points beside them. What I get is not very organized when it probably should be. I expected to get forks closer together and a little more separated from the images of cloth. I guess it is to be expected for the “other” group to be a little everywhere as they are simply random objects that don’t necessarily have any similarities. It is also quite strange to see that there seems to be a separation in the middle, but each side is a mix of all three labels.



The second thing we can look at is a confusion matrix that gives us information on the predicted result compared to the actual result. The "predicted" side is the top horizontal while the "actual" side is the left vertical, which means that the red cell with 40% in it means that the algorithm predicted that the image was part of the "other" category when it is actually part of the cloth category. The F1 Score row contains numbers that are calculated based on the precision and recall of each category as follows: $F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$.

Confusion matrix (validation set)

	CLOTH	FORK	OTHER
CLOTH	60%	0%	40%
FORK	0%	90.9%	9.1%
OTHER	0%	0%	100%
F1 SCORE	0.75	0.95	0.77



This graph shows an arrangement of the images into clusters or where the neural network would place them. The three greens that are used to show the different categories are quite similar so it is a little difficult to separate them, but we can see that all the images of cloth are top left part close to the "other" category in a darker green. The "fork" category is placed more to the right of the graph. The red and orange circles are the mistakes made respectively for the cloth and the fork these were false negatives as the training set predicted them to not be a cloth/fork when they were.

6.

When thinking of a better performance, I first think of getting a better accuracy, precision and recall. I also think of getting the model to work with images that are not similar: what if the object is outside with a different lighting? A better performance would be that it recognises the object in the image when it is shown in a different environment for instance.

Basing my definition on that, I think that to improve my model's performance, I need more images and images that are more varied in their context. So far, all the images of my fork and cloth were taken on the same table with the same light which means that even without the object, the background is already similar. I also noticed that the images that were wrongly interpreted where images that I took using a very different angle like the following:

Therefore, the easy fix would be to have very similar images so the algorithm can easily recognise them, but that would not actually improve the model's performance, it would only improve it in that closed environment.

PART B:

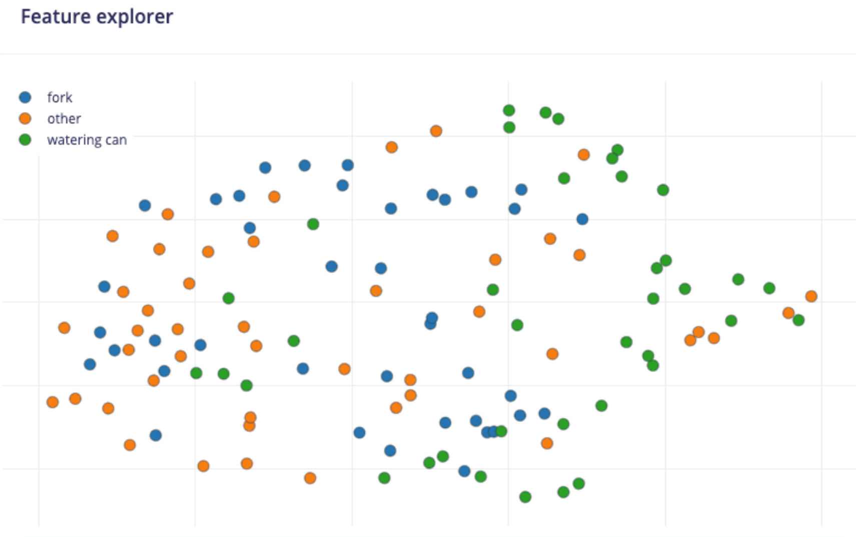
1.
For this second model, I used Shayne's 51 images of a red watering can and my own images of a fork and of other objects that I talked about in PART A. For the images of the watering can, I decided to go with this one as it has an interesting shape and also because I thought the bright colour red might increase the chances that the system will recognize it.

2.
Here, we compare how the model works with our own images and someone else's images and see how well the model works for each. I think it is also interesting to see how someone else collected the data, how they took their pictures and how it might influence the results.

3.
The accuracy was 92.0% and the precision and recall for each category was as follows:

- Watering can
  - Precision: 0.89
  - Recall: 1.00
- Fork
  - Precision: 0.88
  - Recall: 1.00
- Other
  - Precision: 1.00
  - Recall: 0.80

4.

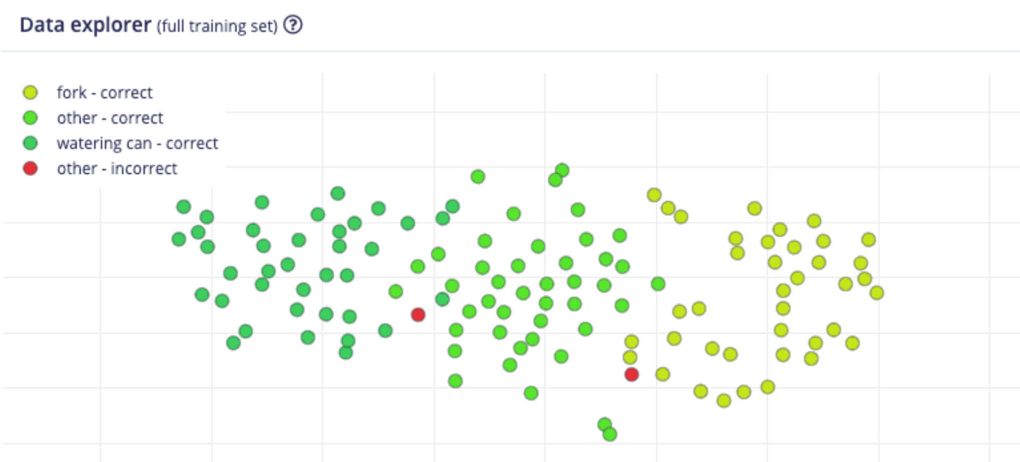


This visualisation of the images show how similar some elements are and although still very messy like the one for PART A, it is interesting to notice that there are more green points/watering can images towards the right of the graph, and more orange/other and blue/fork images on the left of the graph.

This confusion matrix gives a good visual of where the AI failed and where it succeeded. It is interesting to see that the AI was quite good true positives on the fork and watering can but failed with false positives for those two categories. Overall, the accuracy of the AI seems pretty good considering the number of images used.

Confusion matrix (validation set)

	FORK	OTHER	WATERING CAN
FORK	100%	0%	0%
OTHER	10%	80%	10%
WATERING CAN	0%	0%	100%
F1 SCORE	0.93	0.89	0.94



In this graph, we can notice that the positioning of the points according to their colour/label, is more organised than the graph from the Feature Explorer. The forks are on the right, the watering cans on the left, and all the rest is mostly in the middle but still overlapping a little bit with the other points. It is also interesting to see where the errors were made, the red point on the right is right at the limit between the forks and the other and the AI thought it was a fork when it was in fact part of the "other" category. The red point on the left, the AI predicted that it would be a watering can, but it is in the "other" category.

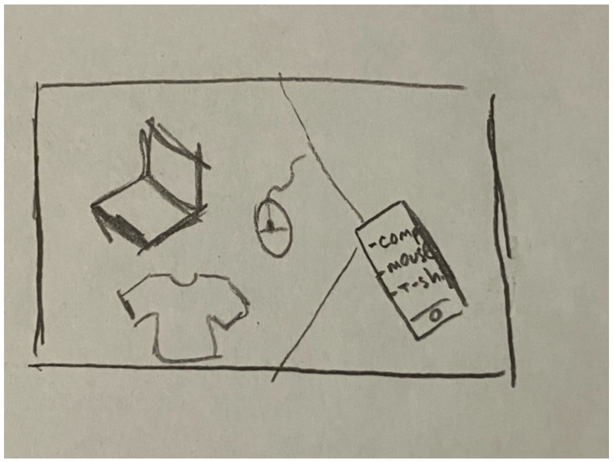
5.
I believe this model performed better than the one for PART A because a watering can is not as malleable as a cloth, so the algorithm was able to more easily identify the object for a similar number of images. When I took images of my kitchen cloth, I moved it a lot, every picture, I folded or mixed up the cloths position which made for a lot of arrangements that a human would recognise, but an algorithm wouldn't. The watering can, although the images were taken with various angles, the object always has the same shape which reduces the options and would be easier for the algorithm to recognise.

PART C:

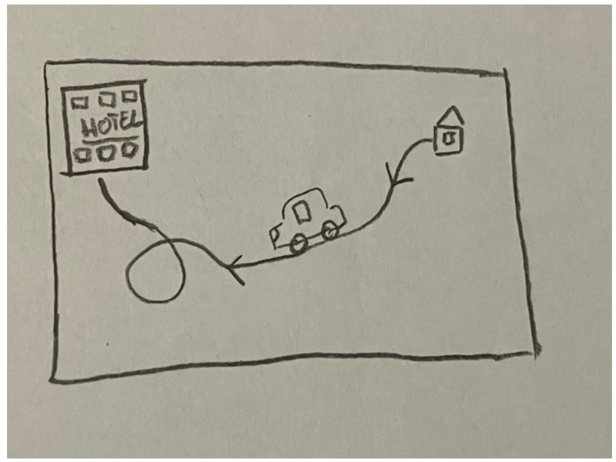
A possible use of Object Detection could be to check if someone has everything with them when they leave for somewhere. For example, whenever I go on a trip, small or big, I always have the feeling that I forget something when I am packing to come back. Object detection could be used to create an app where the user could scan through the things they bring somewhere, and when they are packing up to go back home, they can scan what they are packing and the system would see if they have scanned everything that was scanned before leaving. If anything was not seen, the system could let the user know so they can find it and pack it before heading home.

Storyboard:

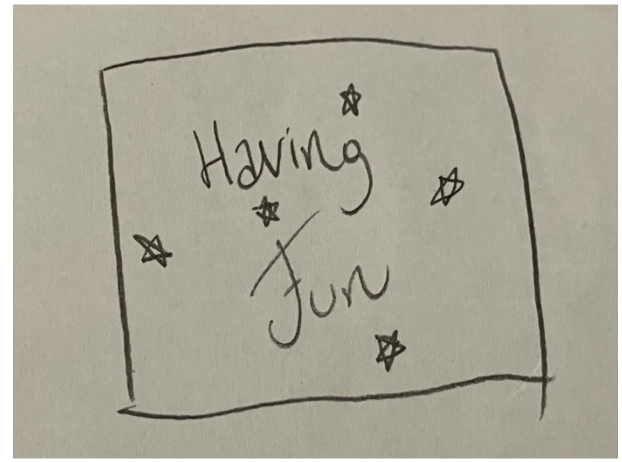
1.



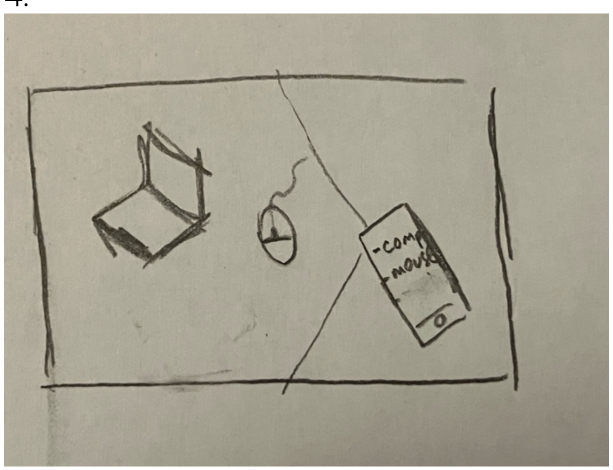
2.



3.



4.



5.