Logbook 1

Dataset used:

150 images extracted and manually annotated from the videos. 75 for training and 75 for validation. Just an initial test, more images will be added.

Key points used:

Nose

Left eye

Right eye

Left ear

Right ear

Left front leg

Right front leg

Left back leg

Right back leg

Back

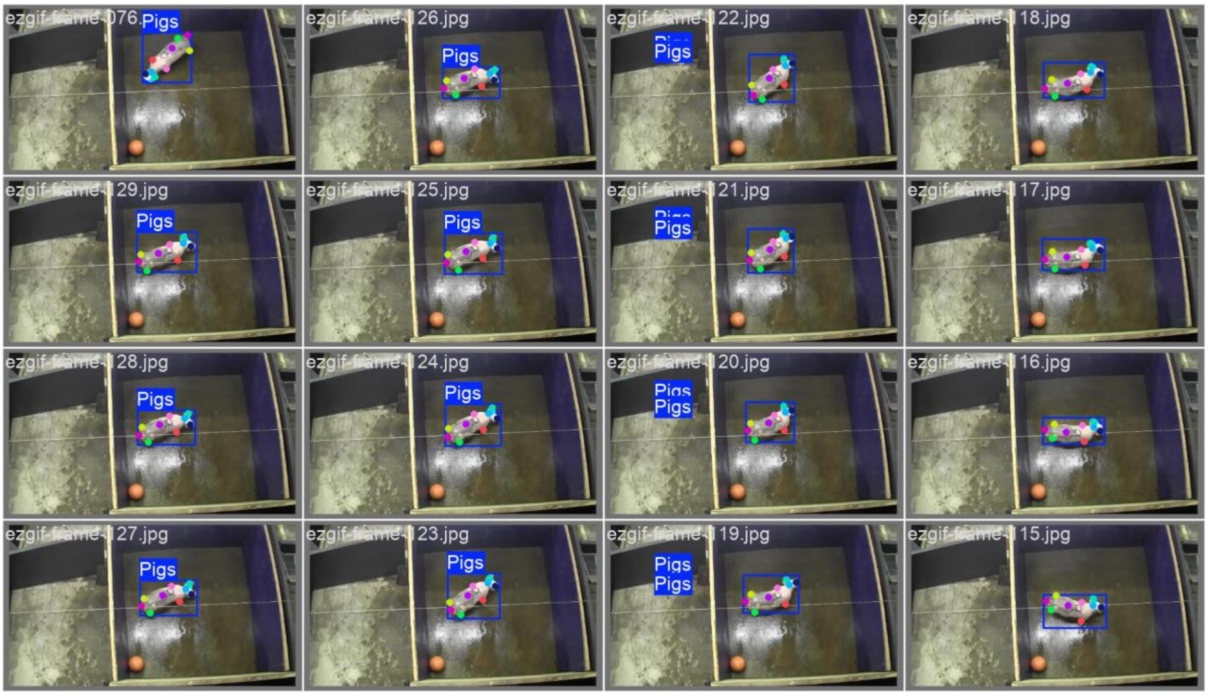
Neck

Tail

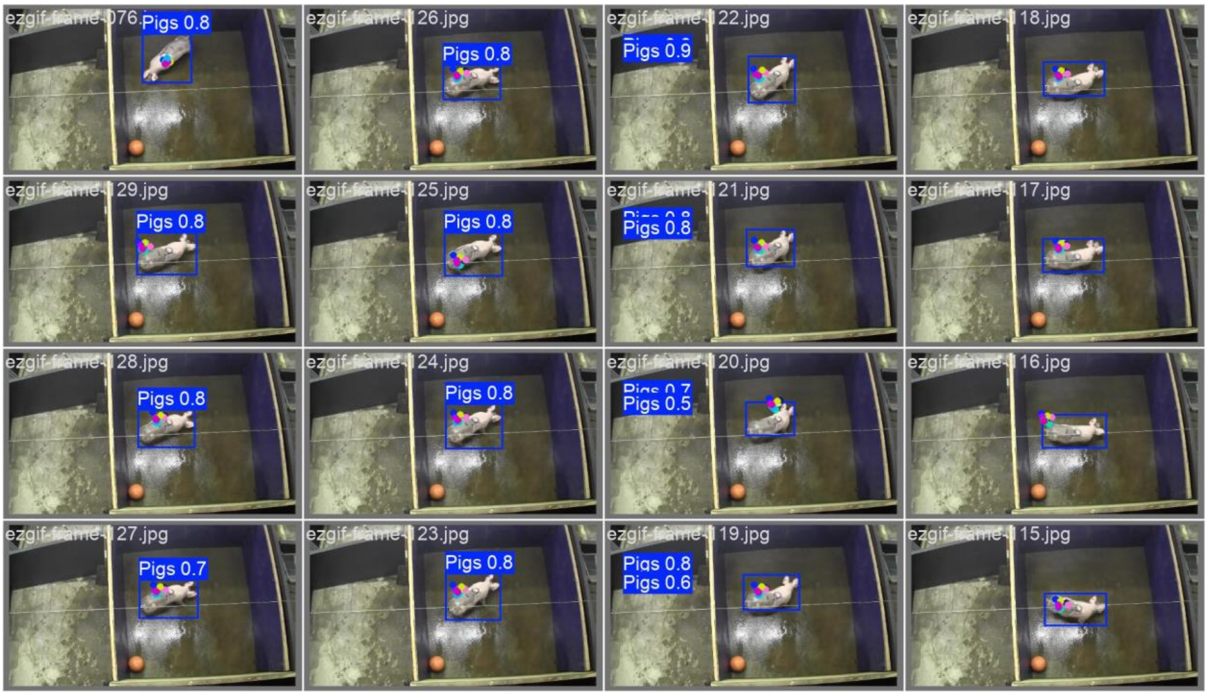
First test:

Using default parameters. 1 epoch, 16 batch, etc.

Examples of the manually annotated val images:

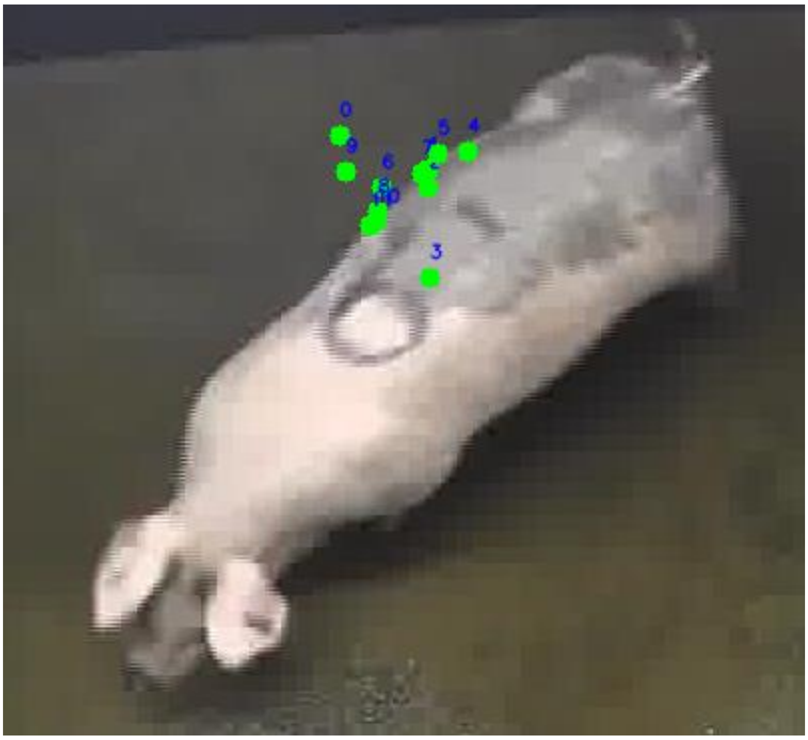


The same images, predicted by the model:



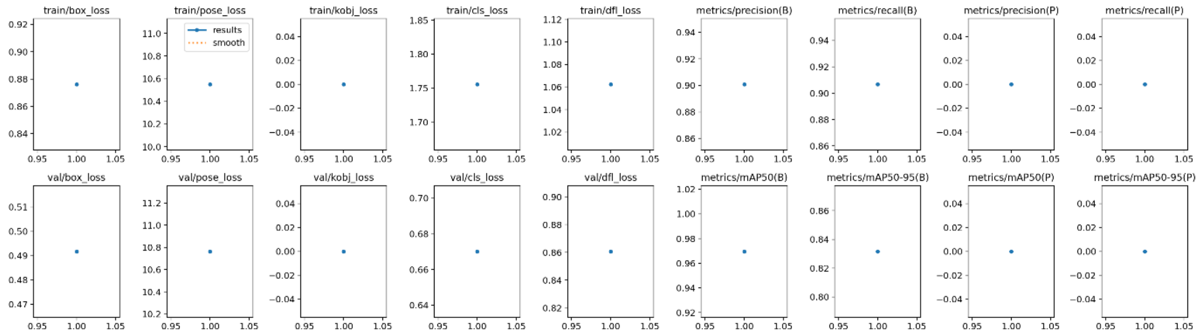
The key points predicted for one of these images with key points numbered for clarity:

Image 76:

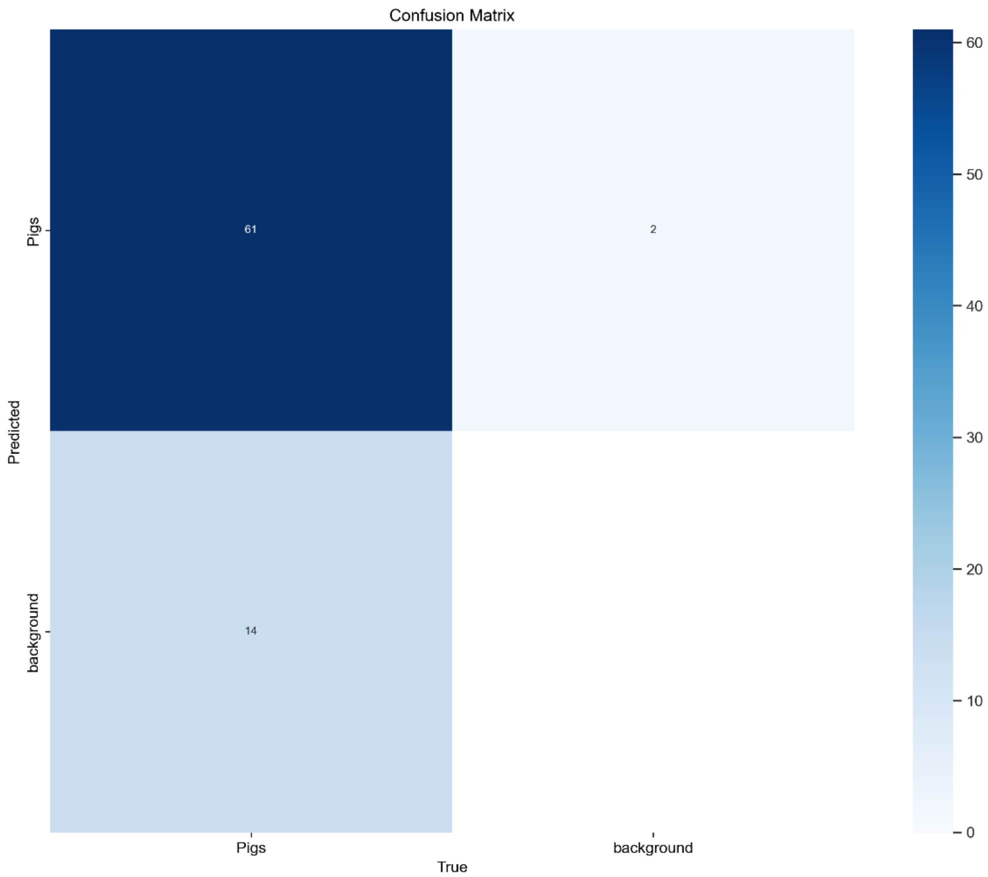


```
"C:\Users\Faisal\Downloads\Internship summer\.venv\Scripts\python.exe" "C:\Users\Faisal\Downloads\Internship summer\VID.3, predict.py"

0: 384x640 1 Pigs, 266.1ms
Speed: 7.0ms preprocess, 266.1ms inference, 2.1ms postprocess per image at shape (1, 3, 384, 640)
Keypoints shape: (1, 12, 2)
Keypoints content: [[[     1303.2      785.38]
  [       1358      822.59]
  [     1356.4      819.78]
  [     1347.2      866.07]
  [     1366.3      808.02]
  [     1362.1      815.55]
  [       1323       824.6]
  [     1343.3      807.39]
  [     1322.4      829.56]
  [       1313       820.5]
  [     1330.8      844.45]
  [     1322.5      842.52]]]
Keypoint 0: x=1303.232421875, y=785.3822021484375
Keypoint 1: x=1357.955322265625, y=822.5852661132812
Keypoint 2: x=1356.4498291015625, y=819.7828369140625
Keypoint 3: x=1347.22119140625, y=866.073486328125
Keypoint 4: x=1366.3092041015625, y=808.02197265625
Keypoint 5: x=1362.05810546875, y=815.5514526367188
Keypoint 6: x=1323.049072265625, y=824.59521484375
Keypoint 7: x=1343.263427734375, y=807.3923950195312
Keypoint 8: x=1322.3638916015625, y=829.5630493164062
Keypoint 9: x=1313.0184326171875, y=820.5009155273438
Keypoint 10: x=1330.812255859375, y=844.4539184570312
Keypoint 11: x=1322.53515625, y=842.51904296875

Process finished with exit code 0
```

<u>Results graphs:</u>



Confusion matrix:



Analysis:

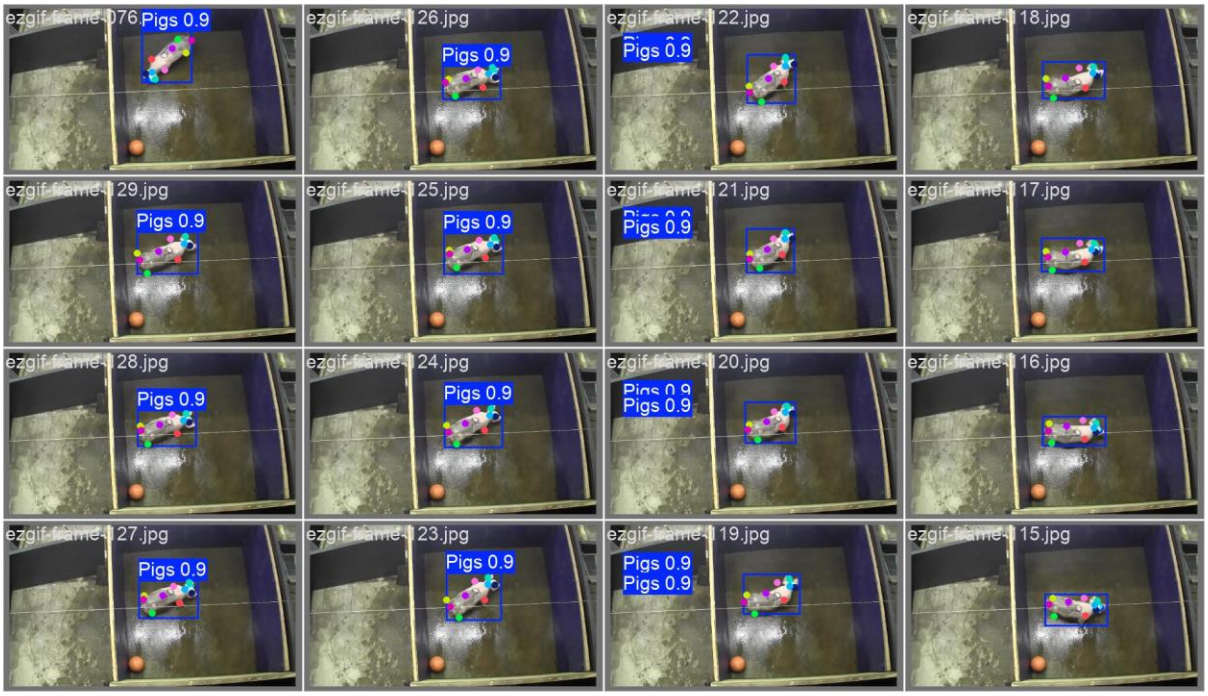
-Bad performance. Wrong key points, and bounding boxes imperfect.

-The confusion matrix shows that there are 61 True positives, 14 false negatives, 2 false positives. Could be much better.

Second test:

Increasing epochs to 100.

The same images as the first test, predicted by the model:



The key points predicted for three of these images with key points numbered for clarity:

Image 76:

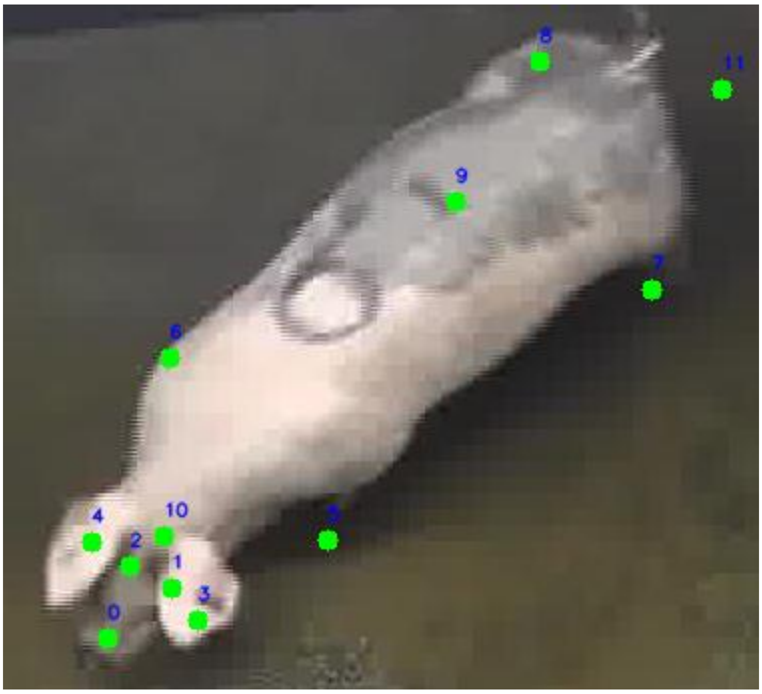


Image 99:

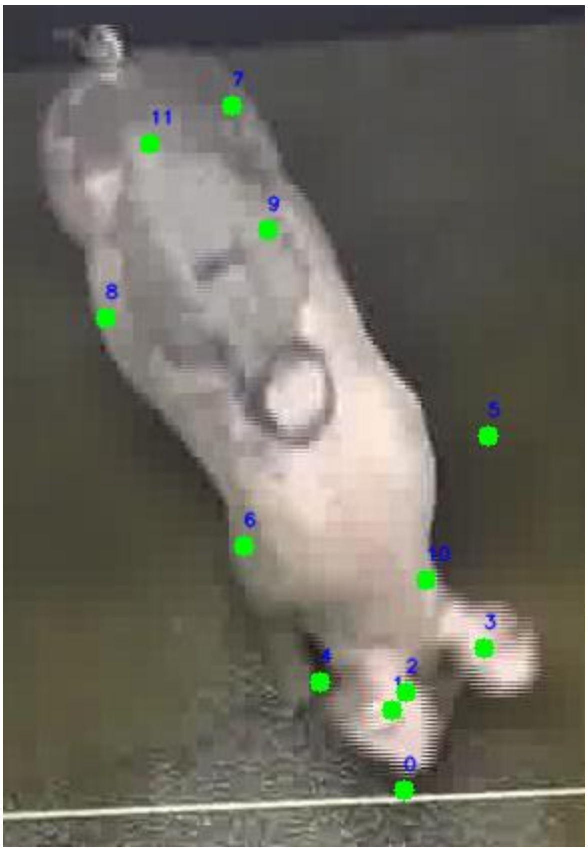
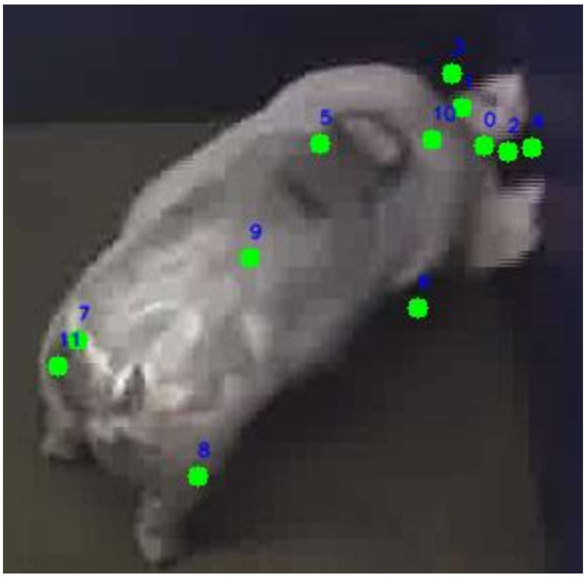
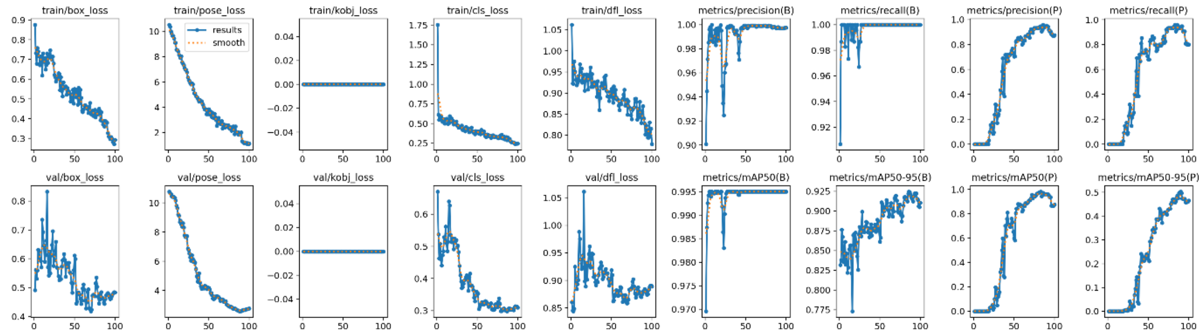


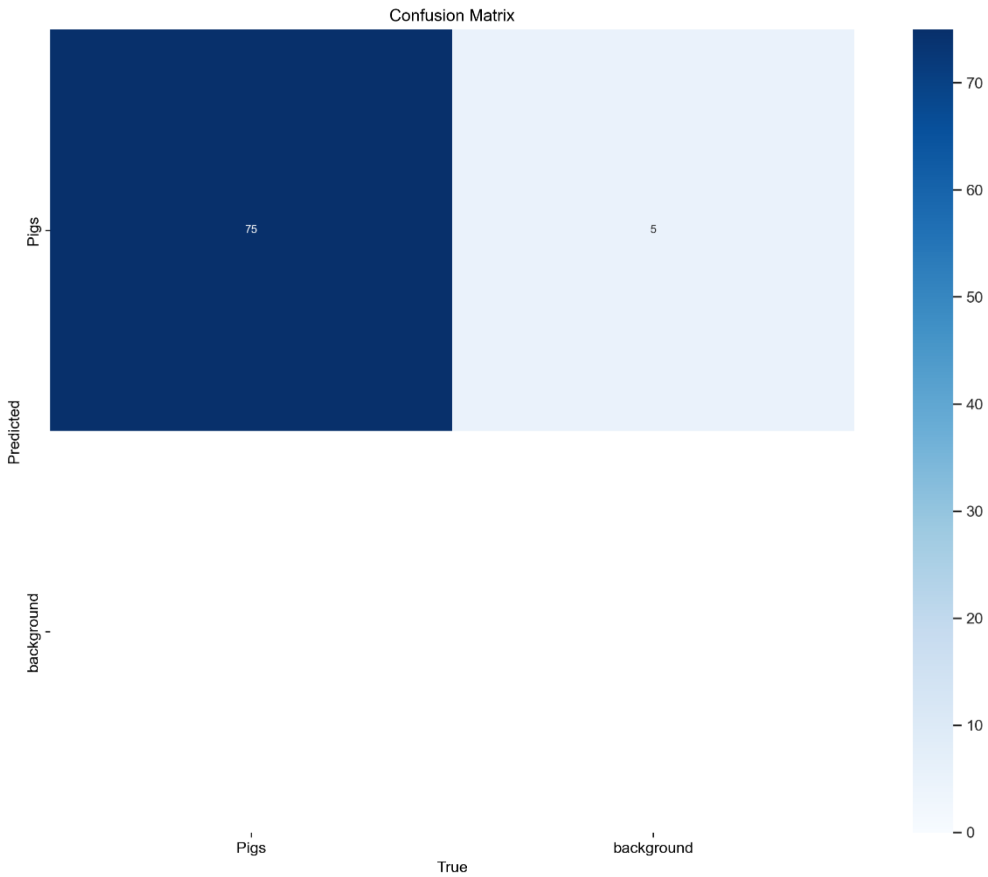
Image 149:



Results graphs:



Confidence graph matrix



Analysis:

-From the train/pose_loss and val/pose_loss graphs important data can be extracted. The train pose loss graph is good as loss is going down with higher epoch until 100 and may even be able to go to a lower loss with further epoch as the line has not yet levelled. The val pose loss graph however is showing increasing loss after about 80 epoch which is not optimal.

-Bounding box overall accurate. 0.8 minimum confidence score. Although some images don't have precision score. And some pictures have 2 bounding boxes/precision scores for a single image.

-The confusion matrix shows that there are 75 True positives, 5 false positives. Big improvement but could still be better.

The key points analysed for image 76, 99, 149 respectively:

Nose:

- Correct
- Incorrect
- Close

Overall: Promising

Left eye

- Close
- Incorrect
- Close

Overall: Needs more work

Right eye

- Close
- Correct
- Incorrect

Overall: Promising

Left ear

- Good
- Good
- Close

Overall: Very promising

Right ear

- Correct
- Close
- Incorrect

Overall: Promising

Left front leg

- Close
- Incorrect
- Close

Overall: Needs more work

Right front leg

- Good
- Good
- Close

Overall: Very promising

Left back leg

- Close
- Correct
- Close

Overall: Promising

Right back leg

- Correct
- Incorrect
- Correct

Overall: Very promising

Back

- Correct
- Correct
- Correct

Overall: Correct

Neck

- Correct
- Correct
- Close

Overall: Very promising

Tail

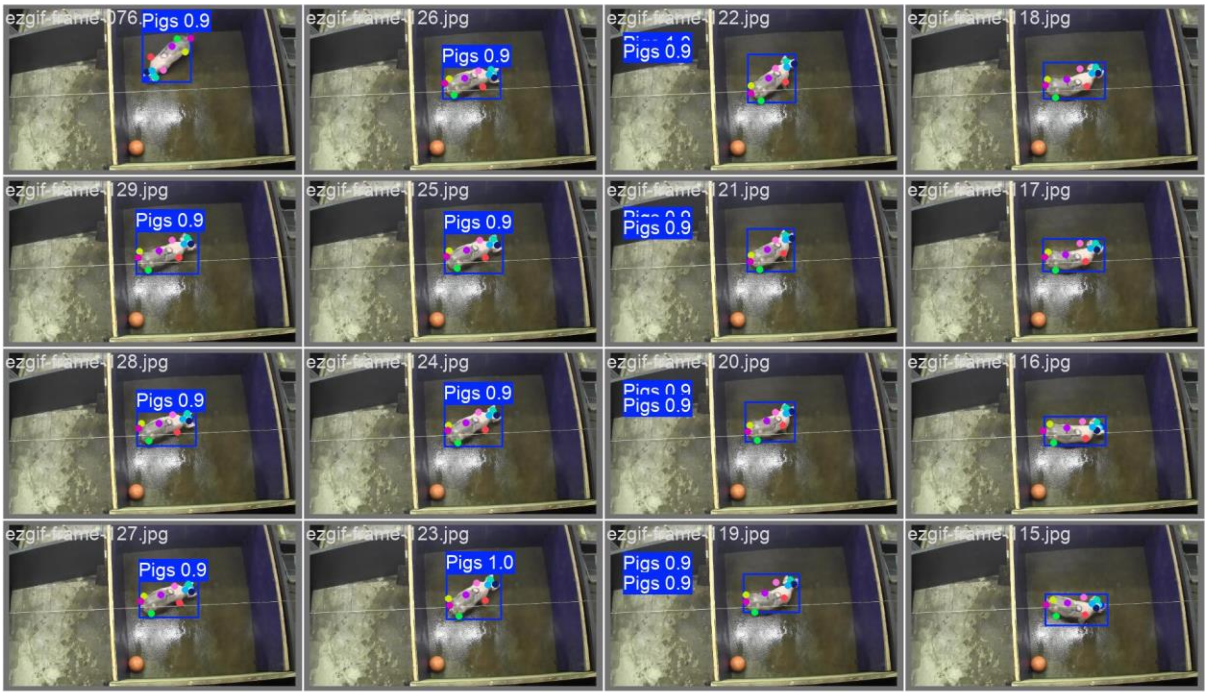
- Incorrect
- Incorrect
- Close

Overall: Bad

Third test:

Keeping epochs at 100 and changing batch from 16 to 32.

The same images as the first test, predicted by the model:



The key points predicted for three of these images with key points numbered for clarity:

Image 76:

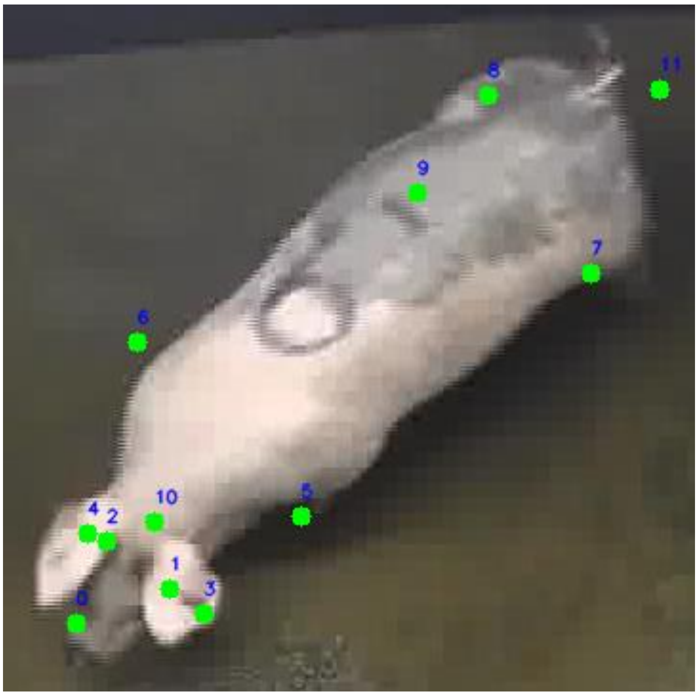


Image 99:

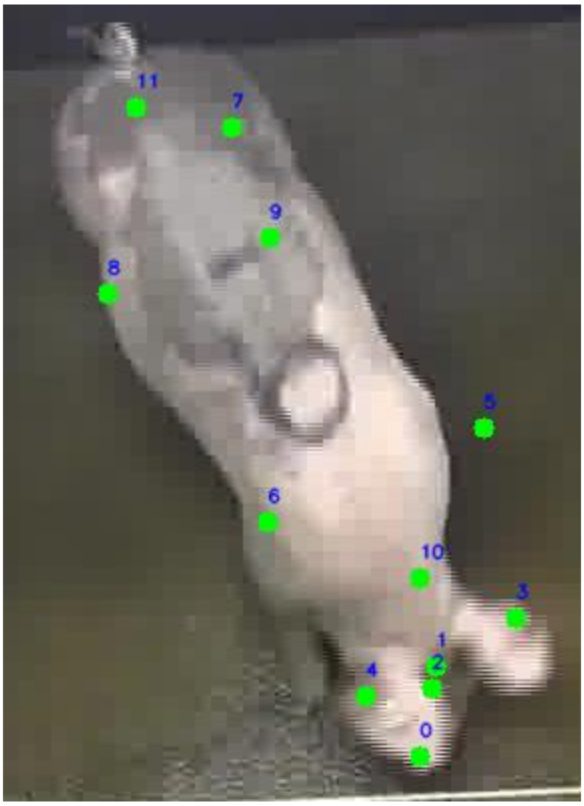
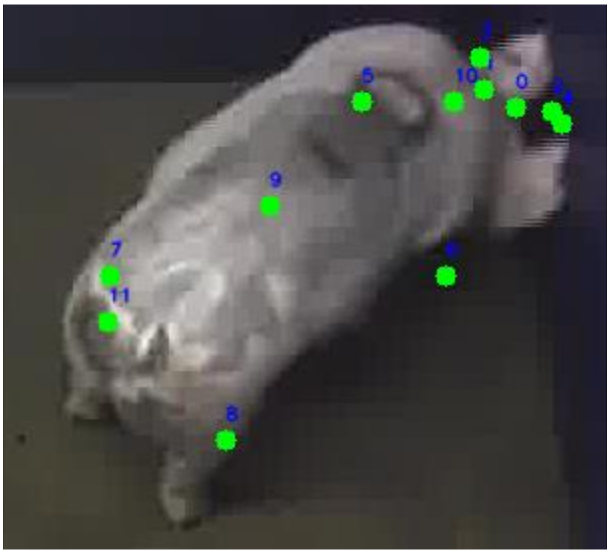
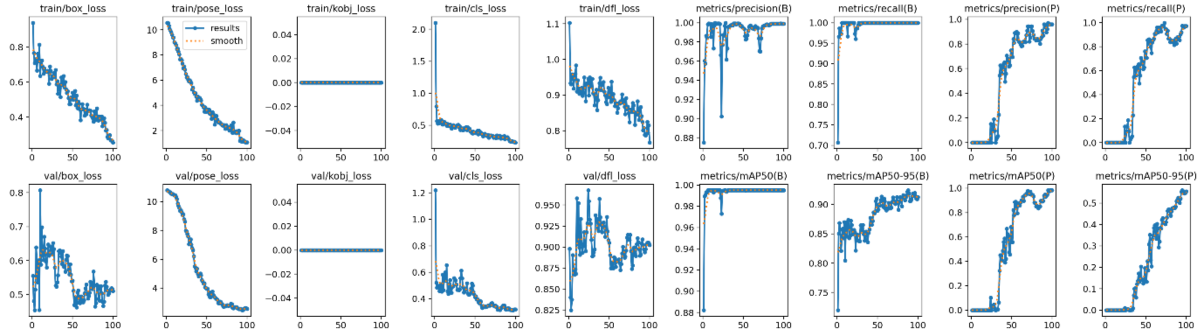


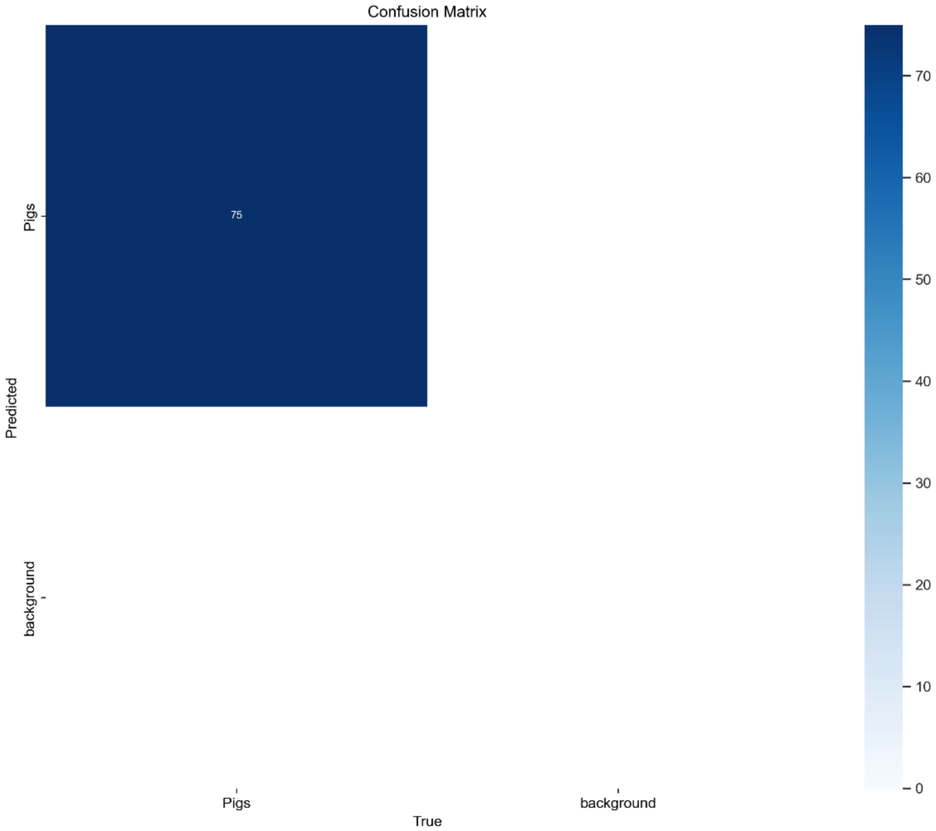
Image 149:



Results graphs:



Confidence graph matrix



Analysis:

- The train pose loss graph is good as loss is going down with higher epoch until 100 and may even be able to go to a lower loss with further epoch as the line has not yet levelled. The val pose loss graph is better than test 2 since the loss is not rising. It is starting to level at 100 epochs.

-Bounding box similar performance as test 2.

-The confusion matrix shows that there are 75 True positives only. Optimal performance. A good confusion matrix has high true positives and true negatives while keeping the false positives and negatives ones low (Suresh, 2021).

- Overall performance of key points predicted is similar to test 2 even if batch changes. Some points better and some worse than test 2. But since the loss is lower and the confusion matrix is better, it can be understood that increasing batch improved performance.

Further work:

- Increase dataset images number
- Can further tweak hyperparameters/tuning/optimisation.

References:

Suresh, A. (2021) *What is a confusion matrix? Analytics Vidhya*. 22 June 2021 [online]. Available from: https://medium.com/analytics-vidhya/what-is-a-confusion-matrix-d1c0f8feda5#:~:text=A%20Confusion%20matrix%20is%20an [Accessed 30 July 2024].

Logbook 2

Logbook 1 tested a 150-image dataset, all of one single angle of a single pig. It showed promising results but needed deeper work. This logbook shows all the new updates.

1) New key points set (23)

Snout centre

Left snout nostril

Right snout nostril

Left eye centre

Left eye corner

Right eye centre

Right eye corner

Middle point of upper lip

Middle point of lower lip

Left corner of pig mouth

Right corner of pig mouth

Left cheek

Right cheek

Left ear hole

Left ear top

Right ear hole

Right ear top

Left front leg

Left back leg

Right front leg

Right back leg

Tail centre

Tail tip

These key points were picked based on input from lecturer Disi Chen as well as doing an in-depth literature review.

(Neethirajan, Reimert and Kemp, 2021) summarised the body changes depending on the emotions of common farm animals. Figure 1 shows that the tail, ears, snout, and eyes can be valuable for assessing the emotions in pigs.

Pigs	High duration lateral tail movement—Positive emotions or play behavior	[42]
Pigs	Tail raised and forming a loop—Positive emotion Ears forward—Alert and neutral emotion Ears backward—Negative emotion Hanging ears flipping in the direction of eyes—Normal state (Neutral emotion Standing upright ears—Normal neutral state	[43,44]
Pigs	Smaller snout ration and ears forward oriented—Aggression or negative emotion state Ears backward and less open eyes—Retreat from aggression or transition to neutral state	[45]
Pigs	Tail hang loose—Negative or neutral emotion state	[46]
Pigs	Curled up tails and ears directed forward—Positive emotion state Tucked under tails—Negative emotion	[47]

Figure 1- How the body of a pig changes with emotion

Note: Another valuable sign they mentioned that was not assessed through the current project due to insufficient relevant images is the visibility of the white sclera in pigs.

(Contributors, 2021) also mentioned how the cheeks can be used to assess emotions in pigs, which is why they were included in the key points as well.

2) Dataset size

The plan was to use 1000 images, but due to various issues encountered as well as the unexpected great difficulty of annotating multiple pigs in a single image accurately, not all 1000 were annotated. The final number of images annotated (for test 4 in section 5) was 100.

3) Questions before training the model:

1) Many of the images have pigs slightly out of frame. i.e. rather than a body part being obscured by something INSIDE the image (which is not a problem) the body part is cropped OUTSIDE of the image entirely. What should be done?

Tests were carried out and datasets like the Tiger Dataset (Ultralytics, 2023) were analysed. The solution was to just ignore the pigs that are partly cropped out of frame. This is because if the parts showing were annotated with a bounding box, and then not with key points (because not all 23 key points are present) this may have caused the model confusion. Therefore, it was just better to completely ignore them and focus on the pigs fully inside the frame.

2) Should images with multiple pigs just be cropped into multiple images containing a single pig or should the images be used as they were (with multiple pigs in each image). Perhaps it is more accurate and easier for the machine to crop the images. However, realistically videos from farms where a model like this may be used to analyse emotion are not going to have just one pig but rather multiple. Furthermore, most datasets online have multiple pigs in each image. Hence, no cropping was done.

3) Use the 2-stage method consisting of 2 neural networks- one for the prediction of the bounding boxes and one for the key points? Or use a single network that predicts both bounding boxes and key points at the same time? The latter was used as this method has already shown promising results in previous tests done for this project.

4) What neural network can detect multiple objects/animals in an image? Yolov8 was picked. Two variants were tested. Yolov8n (nano) and the larger model Yolov8m (medium). Yolov8m was used in the end as it provided more accurate results, as expected.

4) Split into train and val:

75% of images were used for training and 25% for validation.

5) Process:

Test 1) Testing with a few images in to evaluate whether the code and model are working effectively.

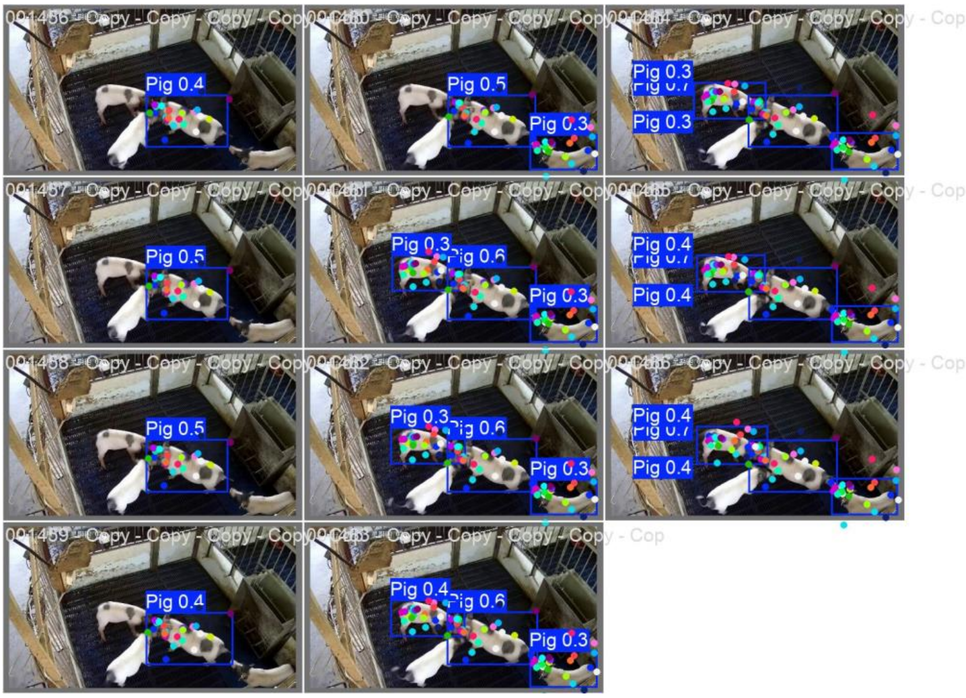


Figure 2-The test 1 predicted points and boxes

The key points were not accurate, but that is because the dataset needed more images at this stage. Some pigs were missing bounding boxes as well, which needed fixing.

<u>Test 2)</u> Now that the project is dealing with multiple pigs in each image, the annotation stage got tricky. For test 1, the pigs in the annotation stage on CVAT were labelled with bounding boxes all labelled with the same label 'Pig'. For test 2, this was changed. Each bounding box for the pigs in the image was labelled differently: 'Pig1', 'Pig2', 'Pig3', and 'Pig4'. This proved to be the better choice. All the pigs' bounding boxes were now predicted and drawn. Note, the key points were wrong because a lower number of epochs were used this time as the only goal of this test was to check if the bounding boxes prediction would improve which does not need a high epoch number.

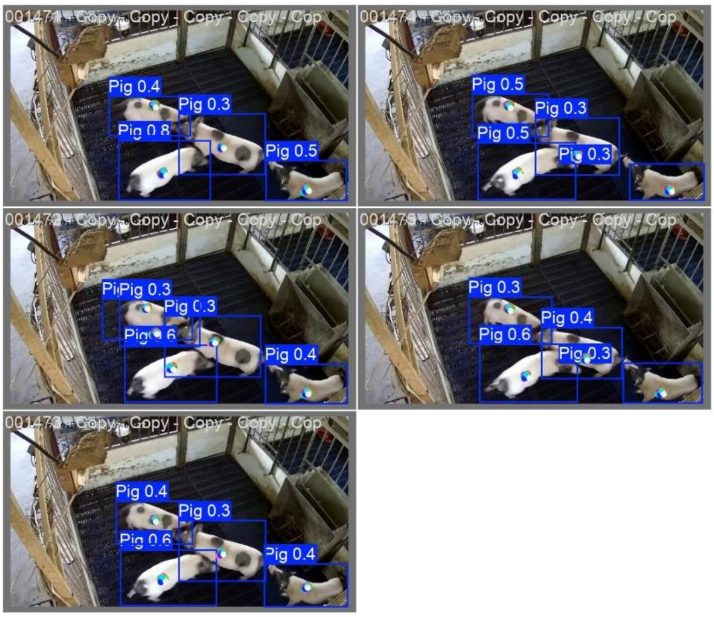


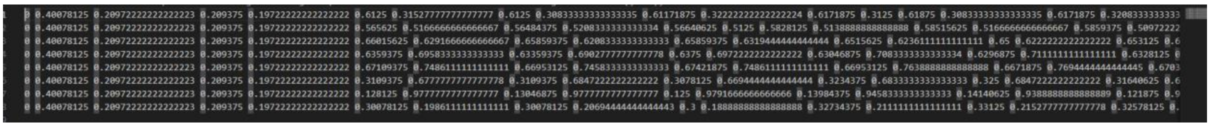
Figure 3- The test 2 predicted points and bounding boxes

Test 3) Note that for the previous test, only images of the pigs from one angle were used. Now pigs from different angles were added to the dataset to test if the system gets confused or works as expected. It was during this test that another significant problem was noticed. When images with 8 pigs were used, even if the change from test 2 (annotating each bounding box with label Pig1, Pig2…. Pig8) was implemented, not all the pig bounding boxes were identified in the prediction stage. This was an issue with the code that converts the CVAT annotated points to COCO format.

Figure 4-Manual annotations for the image with 8 pigs

An image showing the manual annotations of the points was extracted (figure 4). The code used, inspired by (Computer vision engineer, 2023), was assuming that only one object is present in every image. Therefore, it was only taking one bounding box per annotated image and ignoring the rest. This obviously then means the model is trained on one box per image ignoring the rest of the pigs hence the predictions will be wrong.

Another sign is that the text files in the COCO format looked like this upon further analysis:



```
0 0.40078125 0.20972222222222223 0.209375 0.19722222222222222 0.6125 0.31527777777777777 0.6125 0.30833333333333335 0.61171875 0.32222222222222224 0.6171875 0.3125 0.61875 0.30833333333333335 0.6171875 0.320833333333333
0 0.40078125 0.20972222222222223 0.209375 0.19722222222222222 0.565625 0.5166666666666667 0.56484375 0.5208333333333334 0.56640625 0.5125 0.5828125 0.5138888888888888 0.58515625 0.5166666666666667 0.5859375 0.50972222
0 0.40078125 0.20972222222222223 0.209375 0.19722222222222222 0.66015625 0.6291666666666667 0.65859375 0.6208333333333333 0.65859375 0.6319444444444444 0.6515625 0.6236111111111111 0.65 0.6222222222222222 0.653125 0.6
0 0.40078125 0.20972222222222223 0.209375 0.19722222222222222 0.6359375 0.6958333333333333 0.63359375 0.6902777777777778 0.6375 0.6972222222222222 0.63046875 0.7083333333333334 0.6296875 0.7111111111111111 0.6328125 0
0 0.40078125 0.20972222222222223 0.209375 0.19722222222222222 0.67109375 0.7486111111111111 0.66953125 0.7458333333333333 0.67421875 0.7486111111111111 0.66953125 0.7638888888888888 0.6671875 0.7694444444444445 0.6703
0 0.40078125 0.20972222222222223 0.209375 0.19722222222222222 0.3109375 0.6777777777777778 0.3109375 0.6847222222222222 0.3078125 0.6694444444444444 0.3234375 0.6833333333333333 0.325 0.6847222222222222 0.31640625 0.6
0 0.40078125 0.20972222222222223 0.209375 0.19722222222222222 0.128125 0.9777777777777777 0.13046875 0.9777777777777777 0.125 0.9791666666666666 0.13984375 0.9458333333333333 0.14140625 0.9388888888888889 0.121875 0.9
0 0.40078125 0.20972222222222223 0.209375 0.19722222222222222 0.30078125 0.1986111111111111 0.30078125 0.2069444444444443 0.3 0.1888888888888888 0.32734375 0.2111111111111111 0.33125 0.2152777777777778 0.32578125 0.
```

Figure 5-Text file for the annotations

Each line represents the annotated points for each of the 8 pigs. The first number of each line represents the classes. Then, numbers 2-5 are the points of the bounding box for each pig. Each line had the same second to fifth number, further indicating that the code is only using one bounding box for all of the pigs due to assuming only one pig per image.

Note: This problem was not noticed during test 2, which seemed to correctly draw the bounding boxes, because every pig in figure 3 had at least one manually annotated image in the training dataset where that pig was the sole pig with the bounding box. Therefore, in the prediction stage the model had a reference bounding box to base its prediction on for every one of the 4 pigs.

Plenty of tweaking and testing was needed to fix this code as there was no sources on the internet to fix this exact problem. The code was changed so that it loops through all bounding boxes in the xml file and then convert them all to COCO format.

Test 4) The final test included 100 images of pigs, 100 epochs, and 32 batch. The images had variety in the number of pigs, backgrounds, and angles. The improvements from test 3 proved to be instrumental. The bounding boxes of all pigs were predicted. However, the key points were still not perfect due to the small number of images in the dataset, although improved.

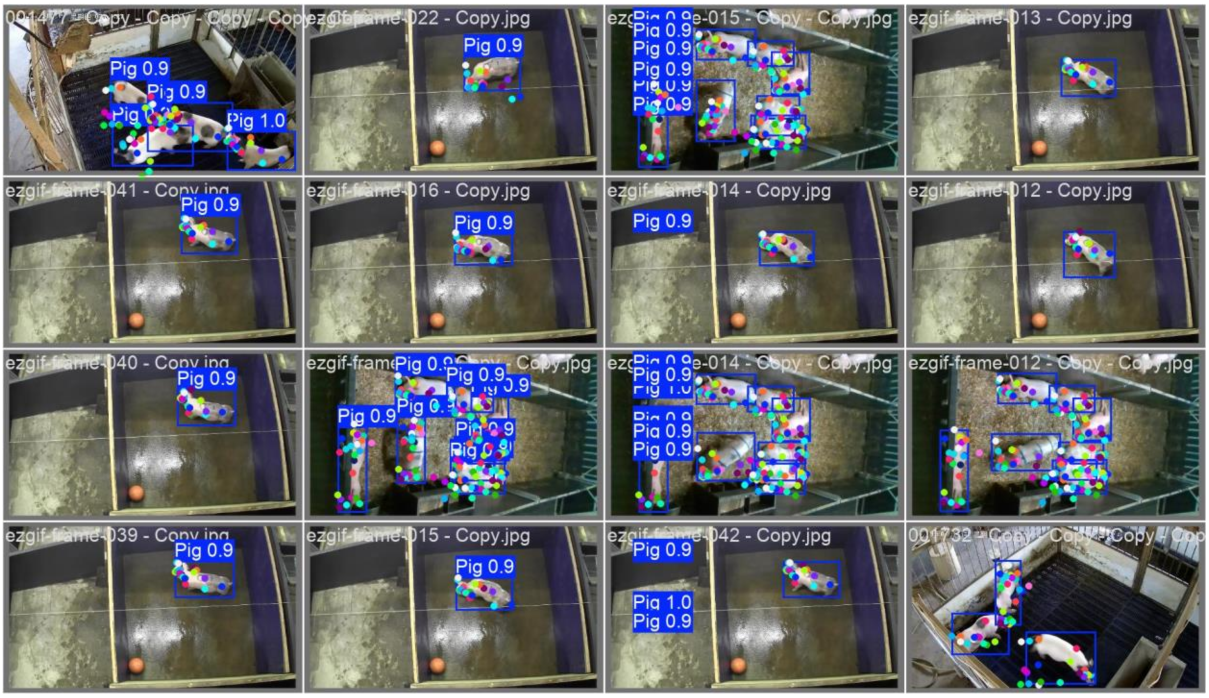


Figure 6-The final predicted points

6) Difficulties:

1) In images with a high number of pigs, there are a myriad of points of key points and bounding boxes to annotate in each image. Some bounding boxes can be annotated automatically, but for this project it was more convenient and accurate to do it manually. However, getting each point correctly annotated was challenging, considering the fact that there was a high number of key points. It was easy to mistake the left nostril for the right nostril for example, especially if the pigs in the same image are facing opposite directions and the left and right side are flipped. Another aspect of this difficulty is that maintaining consistency with the location of each manually annotated key point in consecutive images was strenuous and very time consuming. For example, if care is not taken, a point like the left front leg may vary a little in the manual annotation from each image even if the pig has not moved because it is very hard to maintain consistency.

2) Fixing the code to detect multiple pigs was time consuming.

7) Further work and recommendations:

1) Get clearer images of the pigs so that more detailed analysis can be done, like analysing the visibility of the white sclera in the eyes, eye lid movement, etc.

2) Use a larger dataset

3) Once the key points are a little bit more accurate through using a larger dataset, measuring the distance between the labelled position and the predicted position of the points could be useful to analyse accuracy.

4) A smaller number of key points should be used. This will make the annotation quicker and more images can then be annotated. Furthermore, it will be easier for the model to then predict these key points. Less key points also mean less of a chance that there will be mistakes in the manual annotation stage.

5) Plan the annotation process in advance. What does 'middle of upper lip' or 'Left ear top' mean exactly? An how will this vary when the angle of the image changes? This is to ensure consistency in the location of the annotated key points.

9) References:

Computer vision engineer (2023) *Train pose detection Yolov8 on custom data | Keypoint detection | Computer vision tutorial YouTube*. 26 April 2023 [online]. Available from: https://www.youtube.com/watch?v=gA5N54IO1ko [Accessed 17 September 2024].

Contributors (2021) *Measuring pig emotions and why it matters - Pig Progress Pig Progress*. 17 March 2021 [online]. Available from: https://www.pigprogress.net/health-nutrition/measuring-pig-emotions-and-why-it-matters/ [Accessed 9 September 2024].

Neethirajan, S., Reimert, I. and Kemp, B. (2021) Measuring Farm Animal Emotions—Sensor-Based Approaches. *Sensors* [online]. 21 (2), p. 553.

Ultralytics (2023) *Tiger-pose Ultralytics.com*. 2023 [online]. Available from: https://docs.ultralytics.com/datasets/pose/tiger-pose/ [Accessed 17 September 2024].