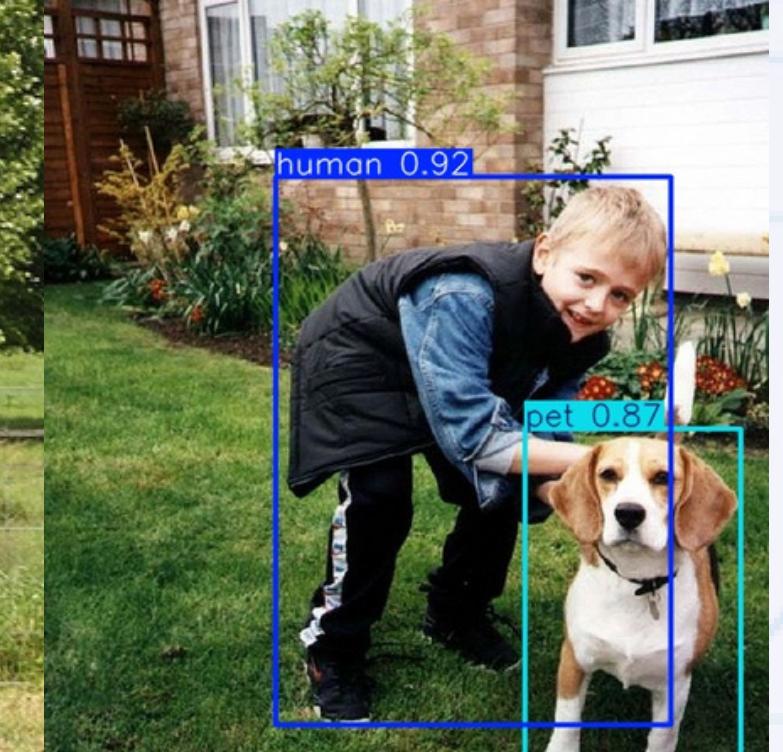
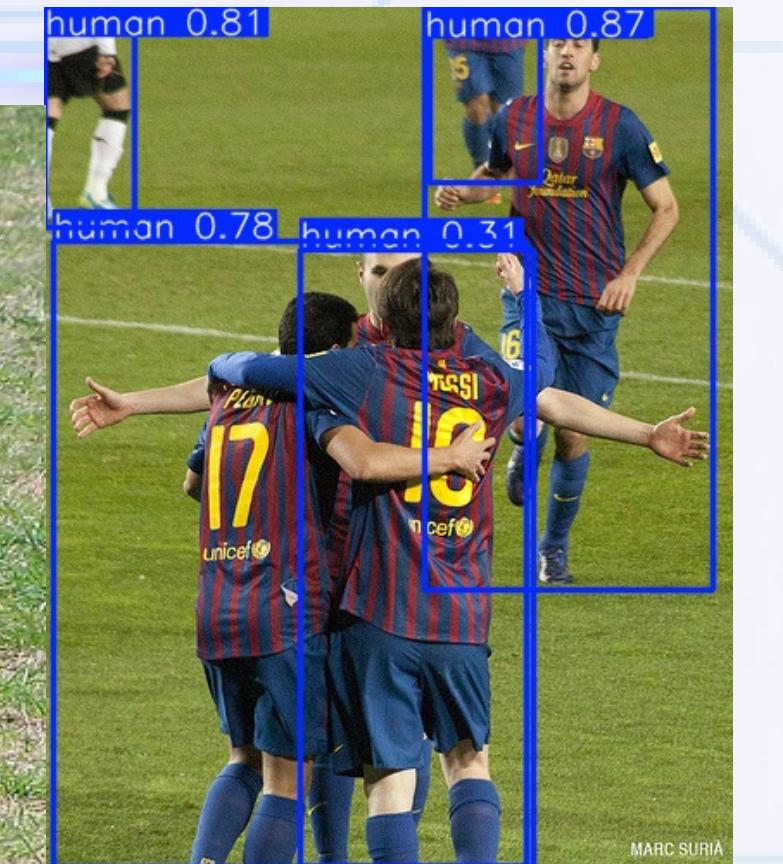
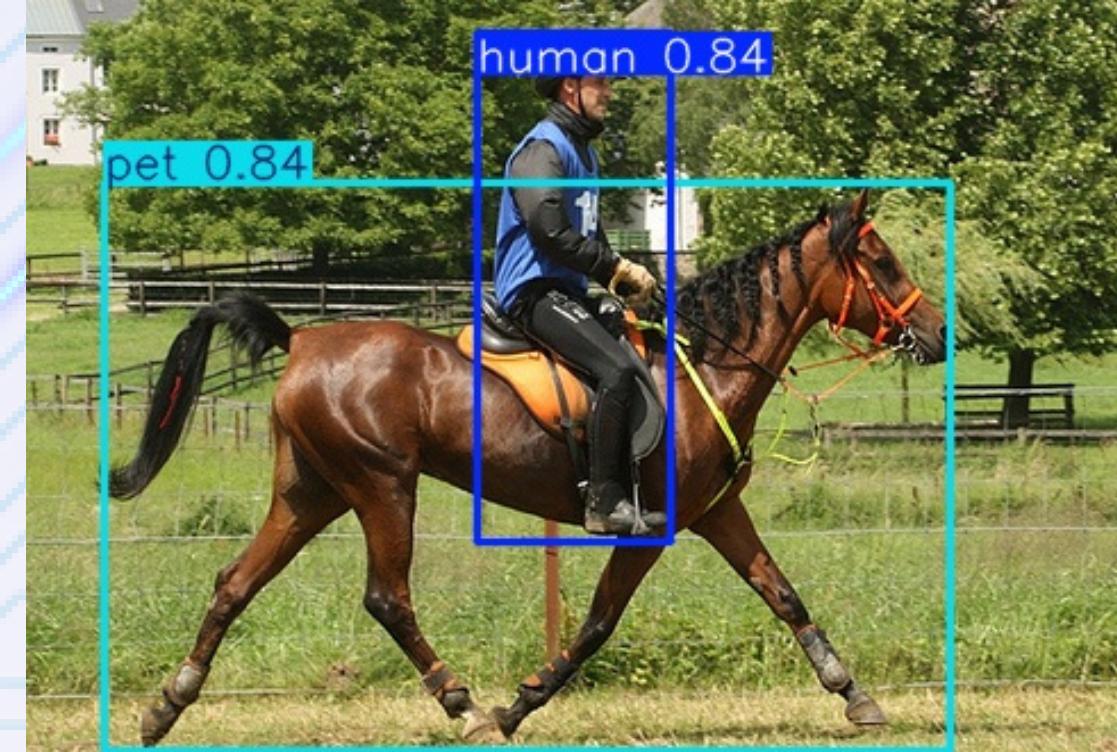
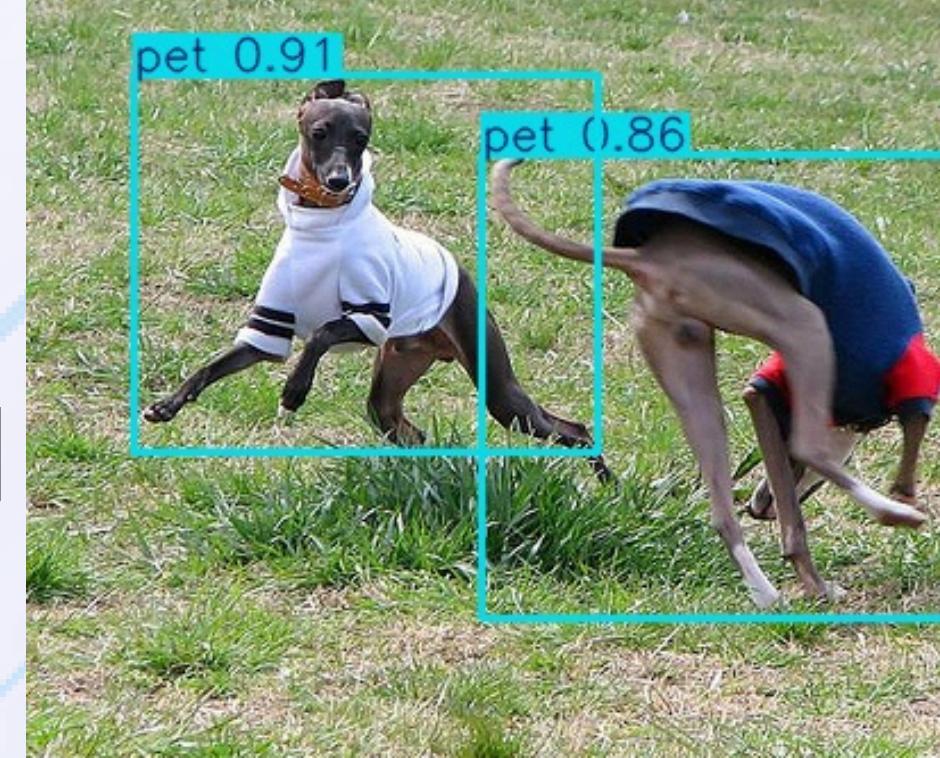




OBJECT DETECTION

by:
tomer harel 207207317
yuval gandler 326686359



■ Why do we need object detection?

■ Object detection is essential because it allows machines to not only recognize what is in an image but also where those objects are located. This ability is crucial in many real-world applications such as autonomous driving, security surveillance, and healthcare diagnostics.



The Dataset

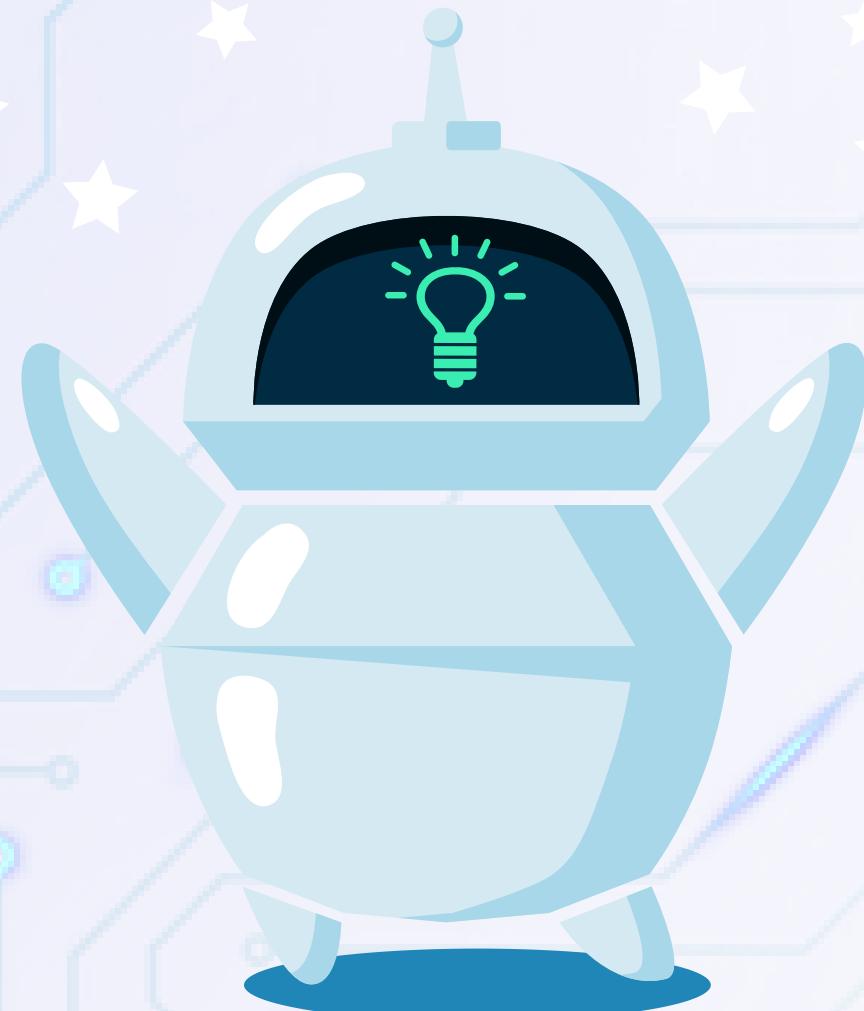
- Dataset Source: Flickr image collection
- Target Classes: Human and Pet (dog, cat, horse)
- Annotation Method: Used Florence-2 object detection model
- Label Categorization: Grouped by model-predicted keywords



The Dataset

We faced several challenges during the project.

- we had to use only 30% of the data (9534 images)
- we encountered an issue when trying to use the Florence model's output function inside another function.
- non equally number of object in each class



The Dataset

an addition to the dataset we could have made:

- Test-Time Augmentation (TTA): Apply different image transformations during inference to uncover missed or weak annotations in order to Helps detect edge cases and improves annotation coverage.
- Model Ensembles: Use predictions from multiple models to identify inconsistencies or missing labels. in order Increases confidence in annotations through consensus.



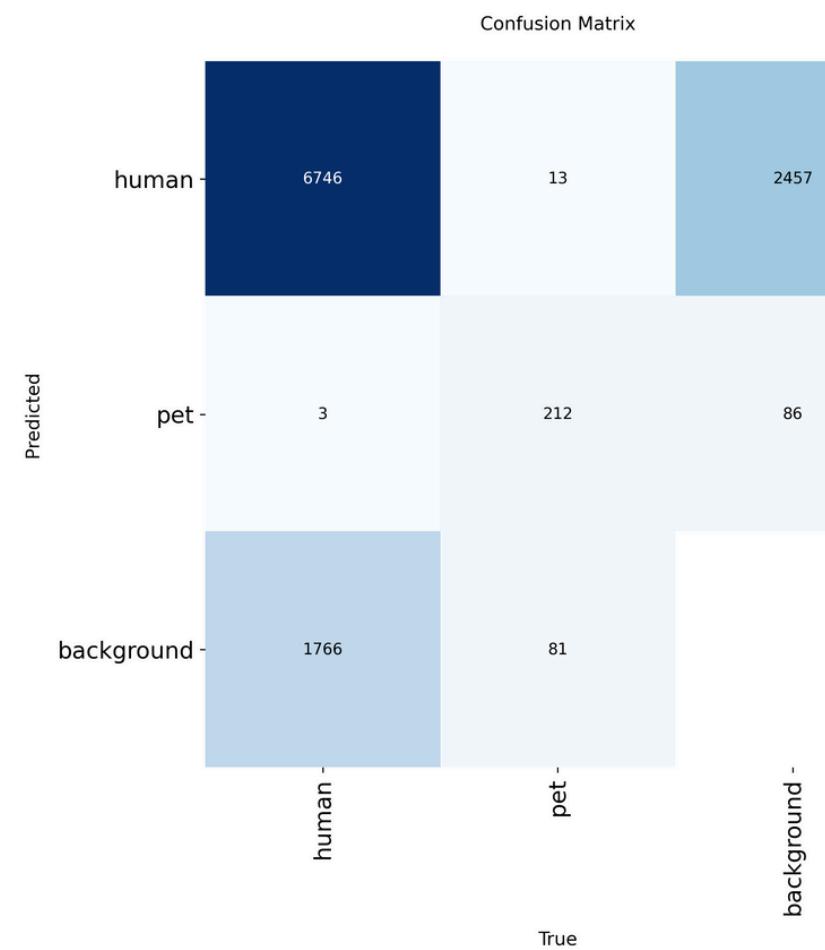
The Model

- we chose YOLO (You Only Look Once) because it offers an excellent balance between speed and accuracy by predicting where objects are and what they are in one look.
- we used YOLOv11 as it the best version of the today model.
- we also use ultralytics framework because simplifies training, evaluation, and deployment



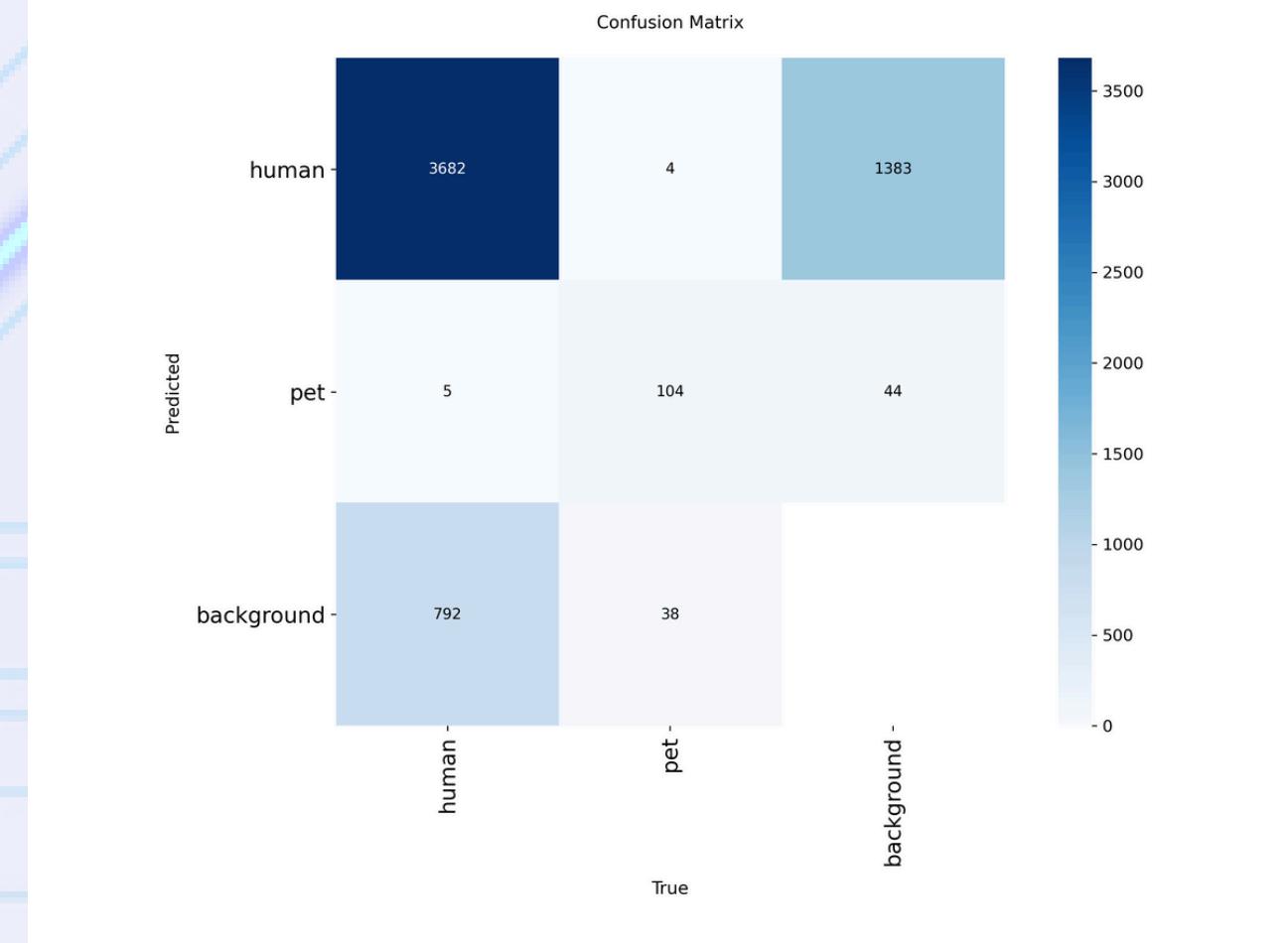
The Model

our first model is a basic
with no augmentations



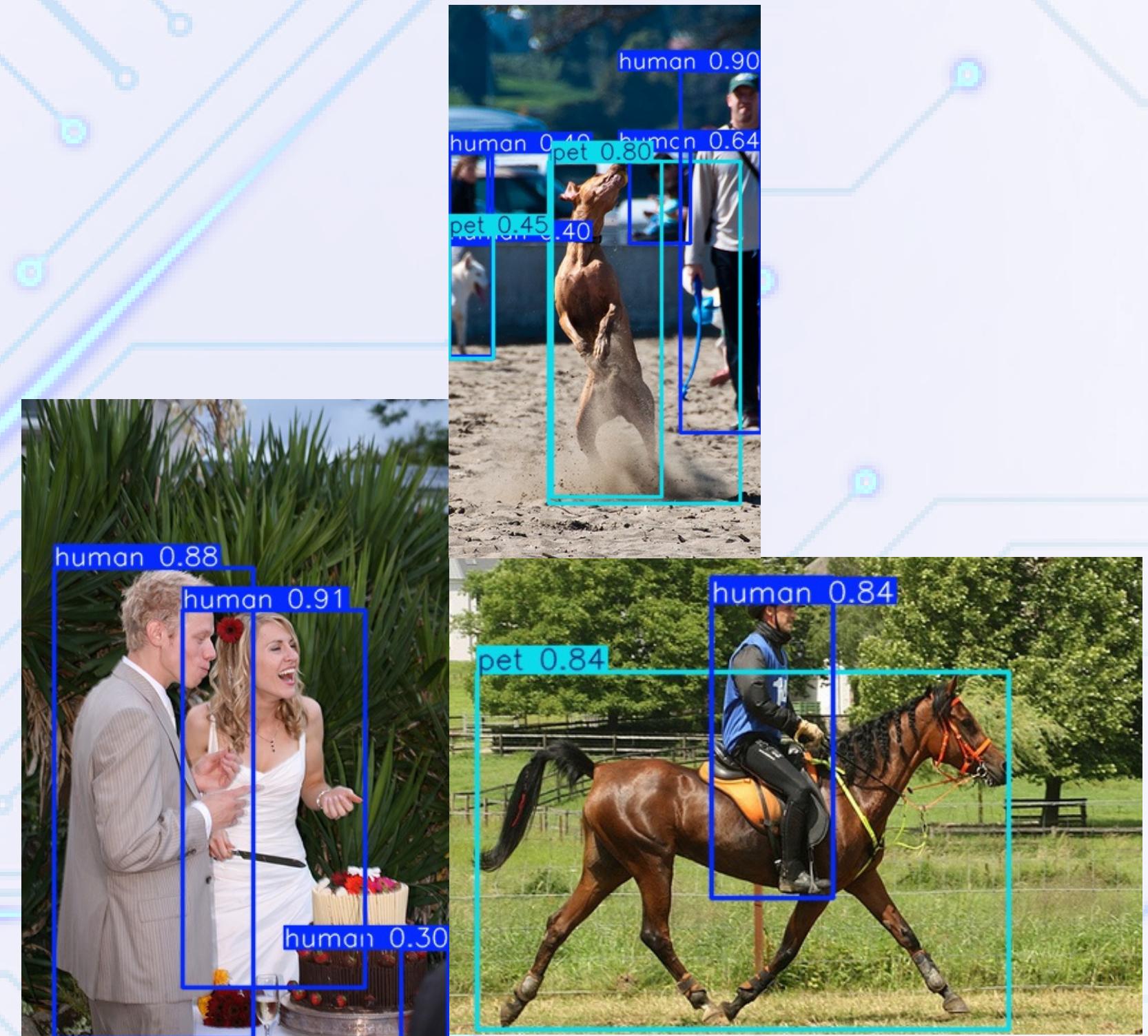
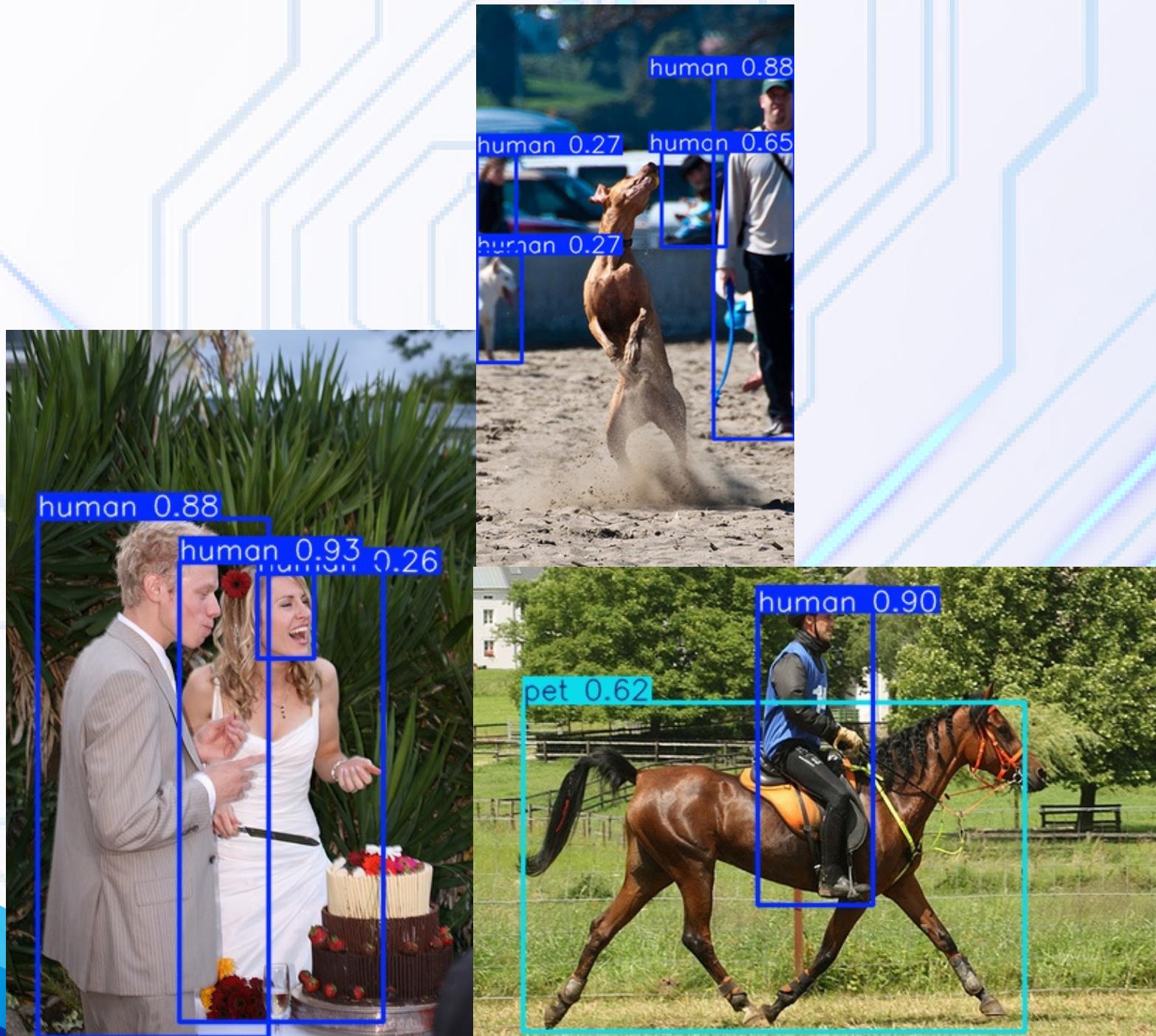
with mAP50: 0.77718
mAP50-95: 0.59856

our fourth model with
augmentations on it and split to
train, val and test



with mAP50: 0.79139
mAP50-95: 0.59176

The Models predictions



If we had more time...

1. We would add more "heavy" augmentations and train the model on them to improve it (such as Gaussian Blur, Gray scaling, Salt and Pepper noise, rain etc)
2. We would work on all pictures (On the notebook we took 30% of the pictures: ~9000)
3. We would probably run the model several times not only 4 times because every run took about 2 hours

■ Reflections

conclusions:

data augmentations contributed to better generalization and slightly improved mAP in training..

What we learned:

Creating and curating a dataset took more time and effort than expected, but was essential for training success.

What We Would Have Done Differently?

We could've used smaller images or fewer epochs early on to speed things up and avoid waiting hours just to see if something worked.



THANK YOU

