ROBOT DREAM

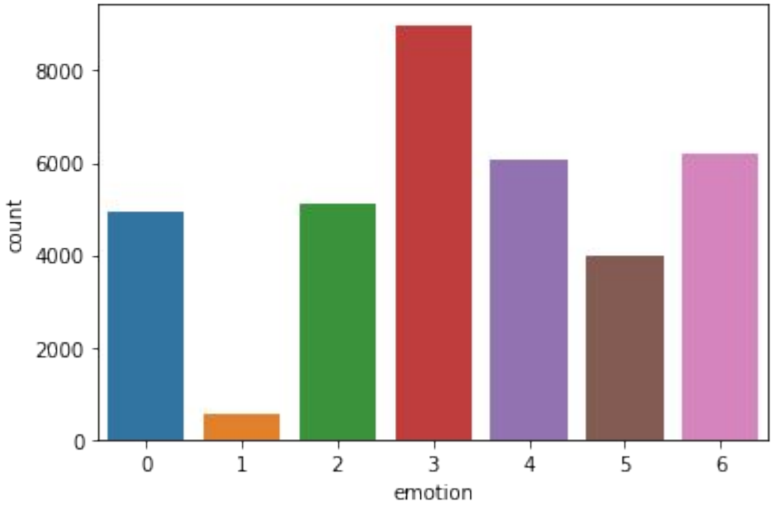
Zeyu
Nirvaan
Rashmi
Ekin
Quynh

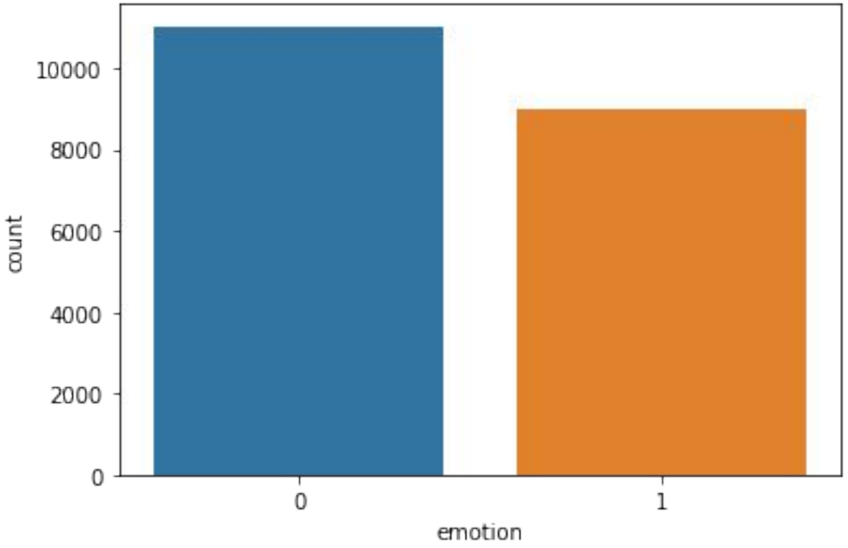
Agenda

- **Face Recognition + Emotion Detection**
- **Balloon Detection**
  - Evaluation Metrics
  - Modelling & Evaluation
    - DETR
    - Yolo V5

Emotion Detection - Dataset (Recap on 1st presentation)

Original (35887, 3) → Filtered (20019, 3)



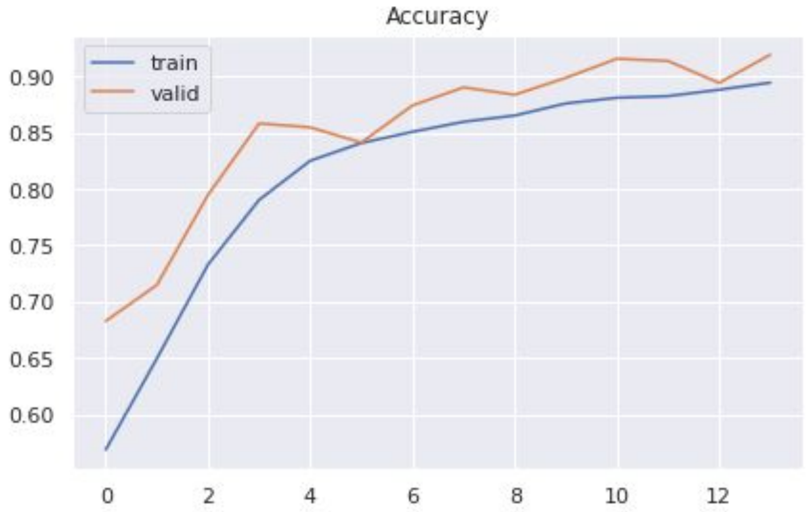


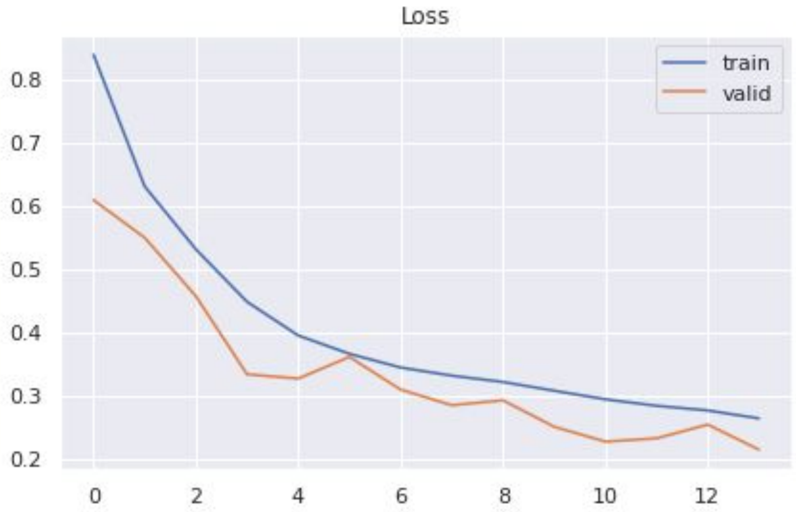
0 - Not Happy vs. 1 - Happy

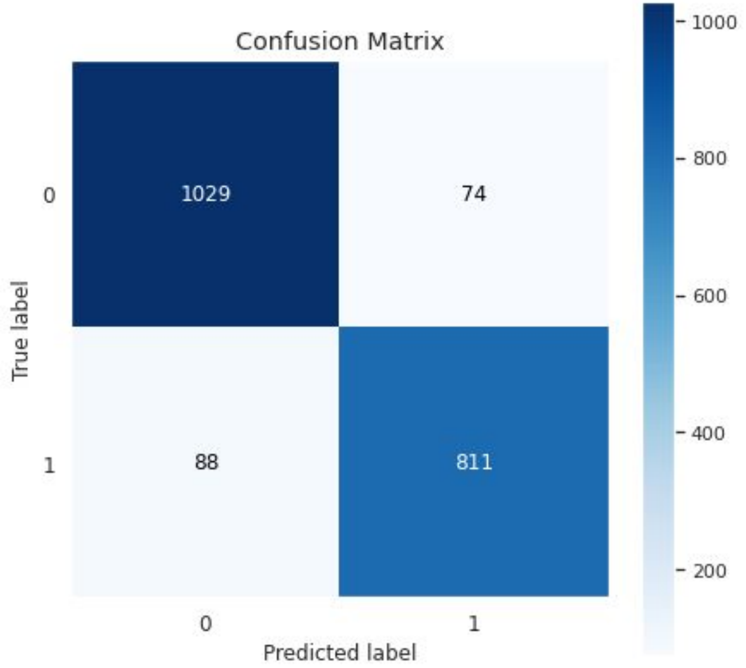
11030 vs. 8989

Then, we trained a DCNN model with <u>14 Epochs</u> and finally arrived at <u>91.91% accuracy</u> for the validation set.

Emotion Detection - Modelling







The epochs history shows that accuracy gradually increases and achieved +90% accuracy on both training and validation set. We didn't notice any overfitting of our model.

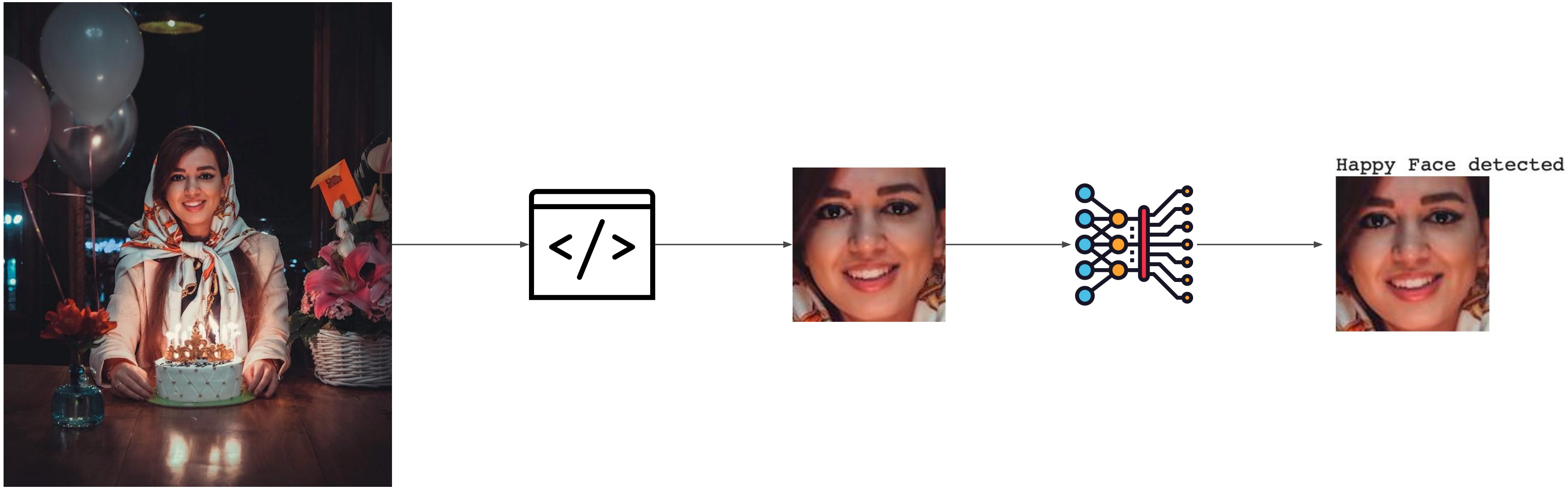
The confusion matrix clearly shows that our model is doing an excellent job on both classes. But for **Happy**, the false negative is higher and overall F1 score is slightly lower than **Not Happy**. One of the reason for this could be the fact that class **Happy** has less data.

	precision	recall	f1-score	support
0	0.92	0.93	0.93	1103
1	0.92	0.90	0.91	899

Face Recognition + Emotion Detection

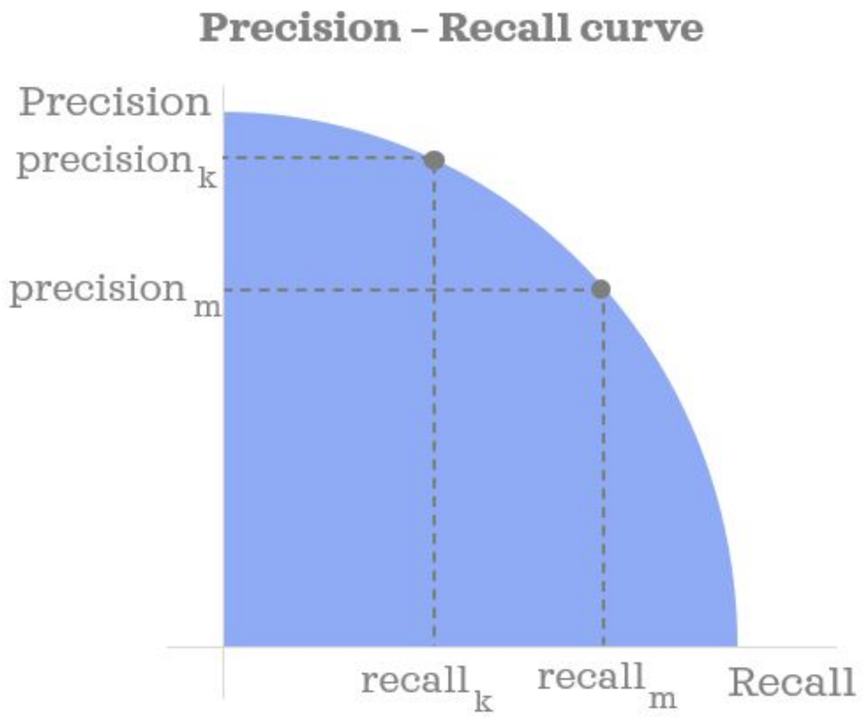
We managed to upload one photo to the notebook and start recognizing the face. Then the pixels will be analyzed using our trained model.

Limitation so far, we haven’t tested on images with multiple faces.



Balloon Detection - Evaluation Metrics

- AP(Average Precision)
- mAP(Mean Average Precision)
- PR Curve

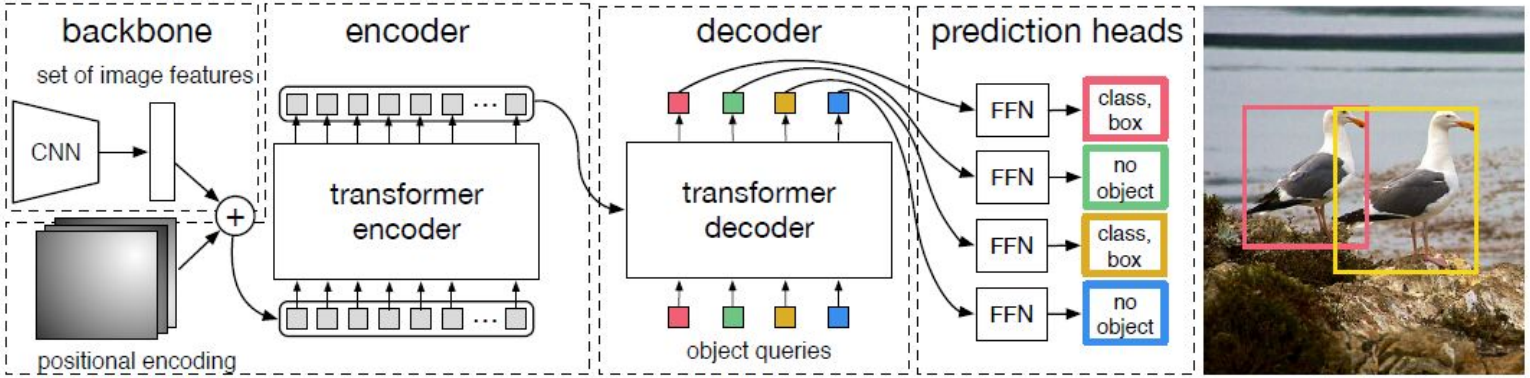


$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

Balloon Detection - DETR (End-to-End Object Detection with Transformers)

- DETR model is an End-to-End Object Detection with Transformers.
  - A new technique introduced by the Facebook research team in 2020
  - The DETR architecture has 3 layers:
    - The CNN layers are used to extract features from the image (Backbone)
    - Encoder-decoder structure in Transformer
    - To a set loss function which performs bipartite matching between predicted and ground-truth objects



- DETR requires a very intensive training schedule. In the original paper, the authors train their model using
  - dataset with at least 10K photos
  - 16 Nvidia V100 GPUs
  - 300+ epochs
  - 3+ days of training time (about 4,000 USD at AWS)

Balloon Detection - DETR (End-to-End Object Detection with Transformers)

- Fine-tuning a pre-trained DETR model on a our balloon dataset
  - DETR RC 50

	name	backbone	schedule	inf_time	box AP	url	size
0	DETR	R50	500	0.036	42.0	model | logs	159Mb
1	DETR-DC5	R50	500	0.083	43.3	model | logs	159Mb
2	DETR	R101	500	0.050	43.5	model | logs	232Mb
3	DETR-DC5	R101	500	0.097	44.9	model | logs	232Mb

  - change the model-building structure to enable any amount of classes (in our case, num_classes = 1)
  - replace the classification head
  - transform the images to have maximum dimension of 800 x 1333 pixels

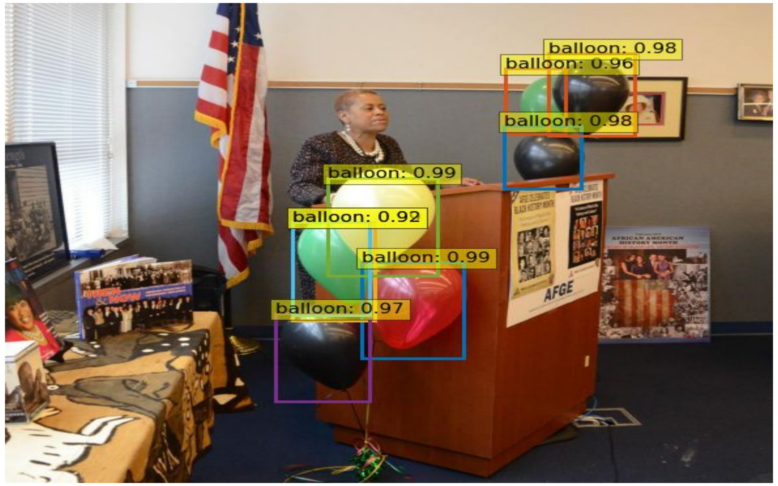
- Our model training with
  - dataset of 2645 photos
  - 6 epochs (stop in middle of epoch 7)
  - free GPU on Google Collab
  - 1 hour training time

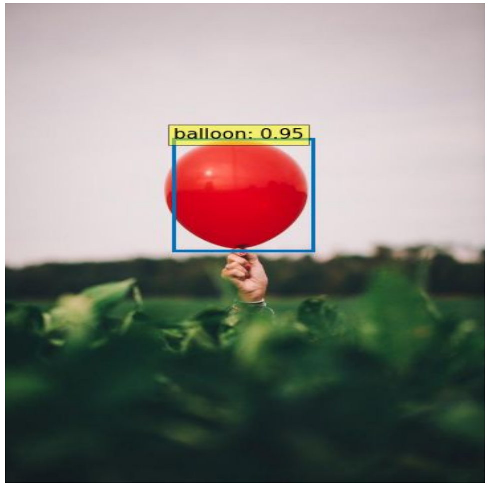
Balloon Detection - DETR (End-to-End Object Detection with Transformers)

- Give **good result** in
  - avoid non-balloon object



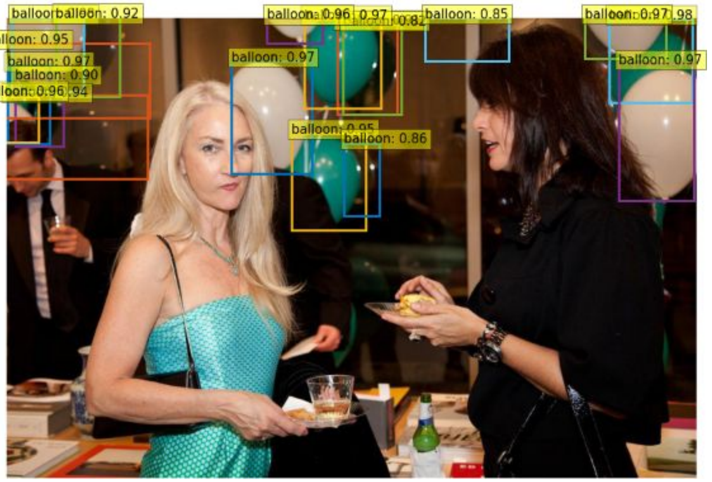

  - detect balloons if the balloons are clearly seen and in small group (2-7 balloons)

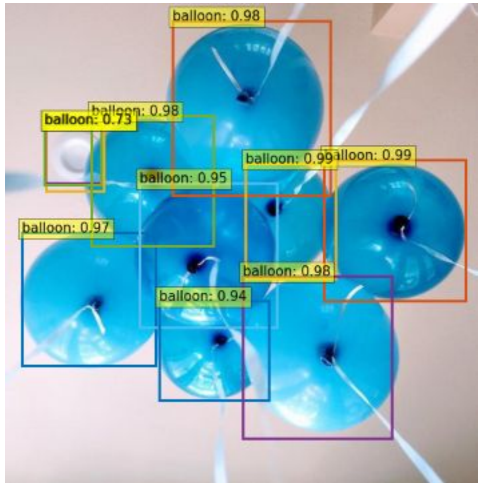
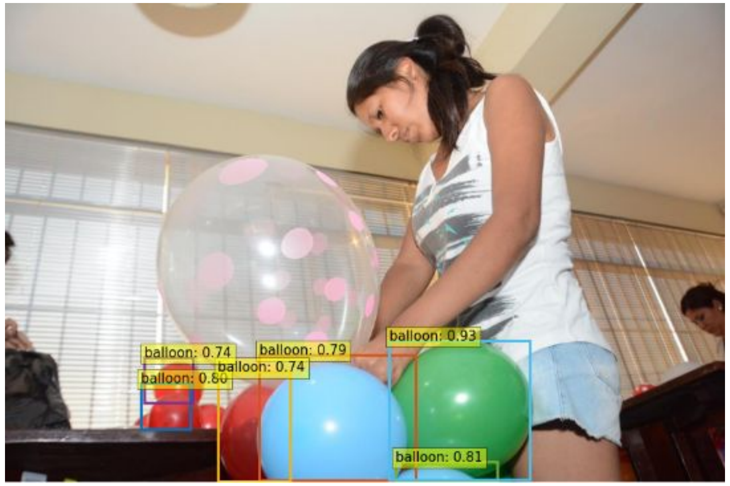
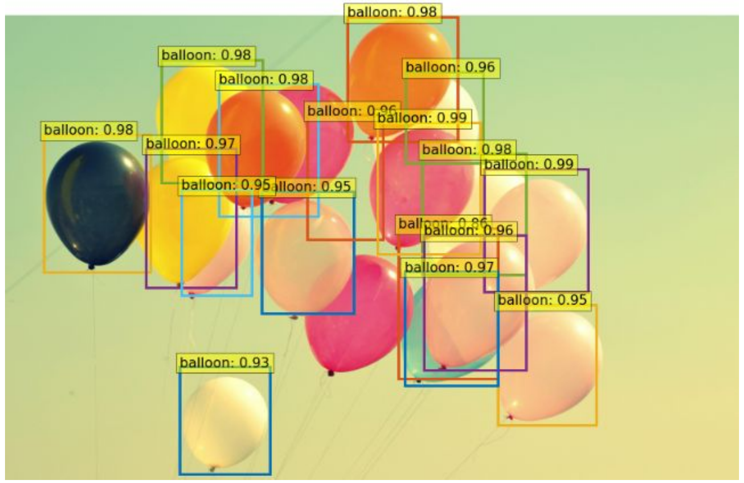
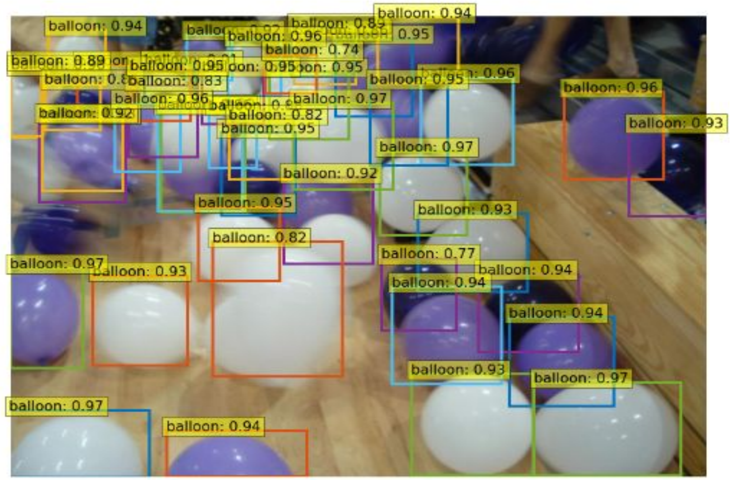
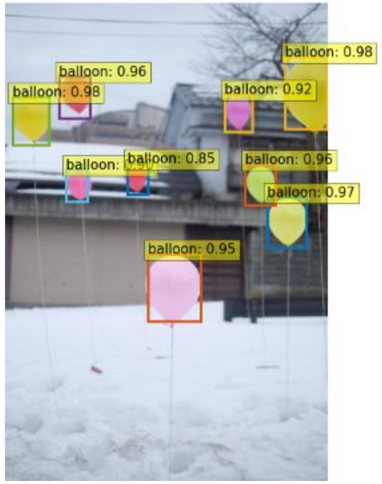




Balloon Detection - DETR (End-to-End Object Detection with Transformers)

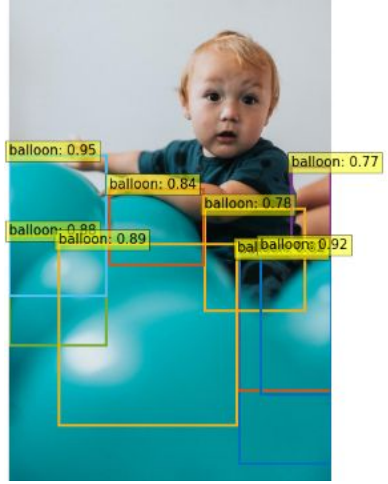
- Give **good overall result with some mistakes** in detecting multiple balloons in big group

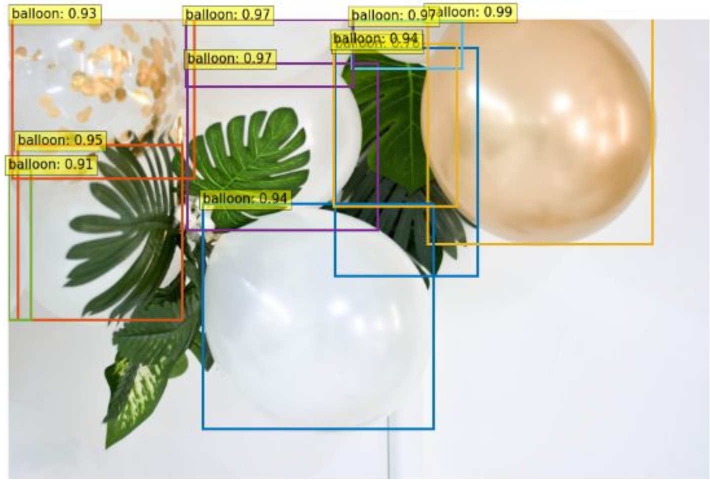


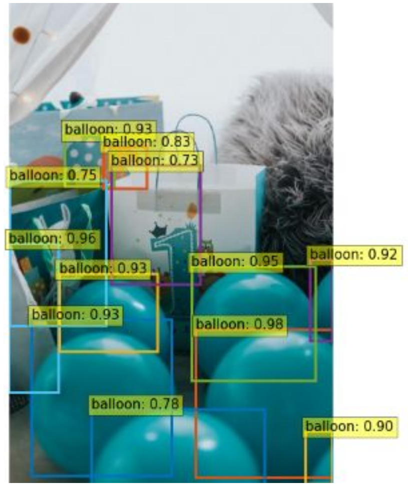







Balloon Detection - DETR (End-to-End Object Detection with Transformers)

- Give **not so good result** when balloons are in less common position or blurry or far away or color too close to the background







Balloon Detection - DETR (End-to-End Object Detection with Transformers)

- Count number of balloons

Balloon Detection - DETR (End-to-End Object Detection with Transformers)

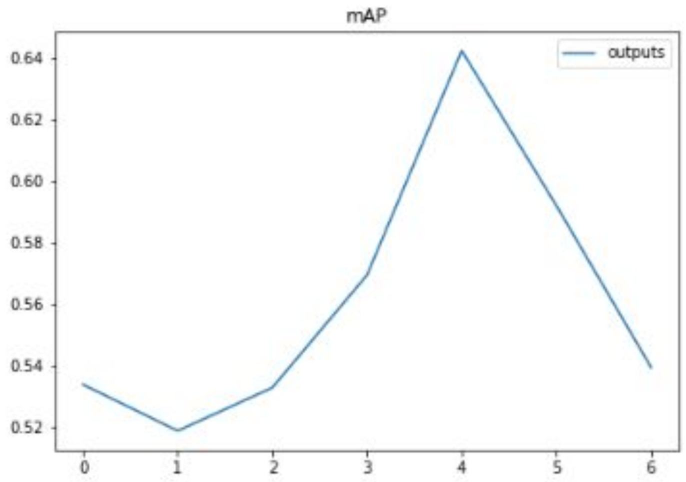
- Encountered problems:
  - Run out of free GPU in Google Collab (in middle of training for epoch 7)
  - Dataset is small with only 2,645 photos → difficult to train the model to detect balloons in “tricky” situations
    - balloons in different positions on the edge, far away, on the ground, etc
    - balloons with reflection on the mirror or shadows on the wal, etc.
    - balloons with color matching with the background, etc.
    - balloons too blurry and far away in the background, etc.

- Cons:
  - Require intensive training and powerful computing power
  - Time consuming
  - Expensive

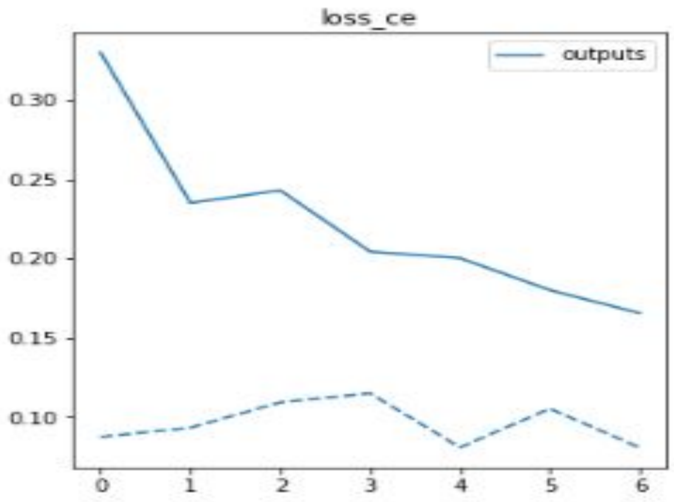
- Pros:
  - Give decent result right off the bat even with training on only 6 epochs

Balloon Detection - Evaluating DETR

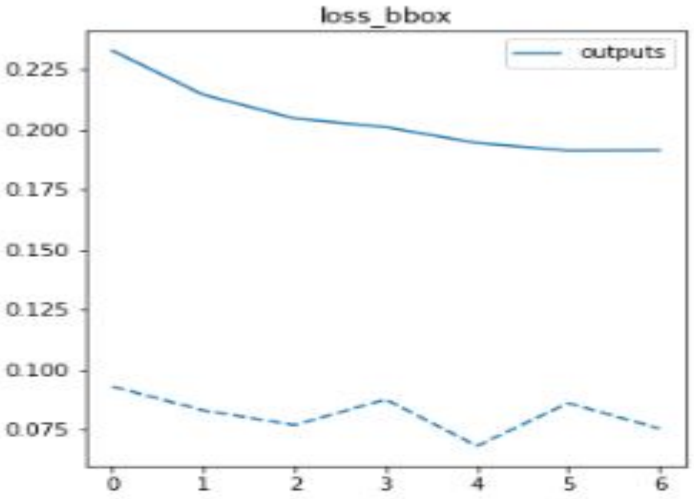
Mean Average Precision



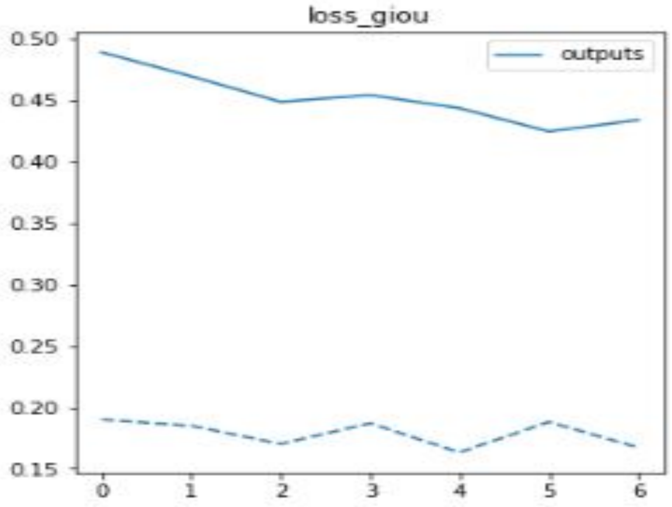
Cross-Entropy Loss



Bounding Box Prediction Loss



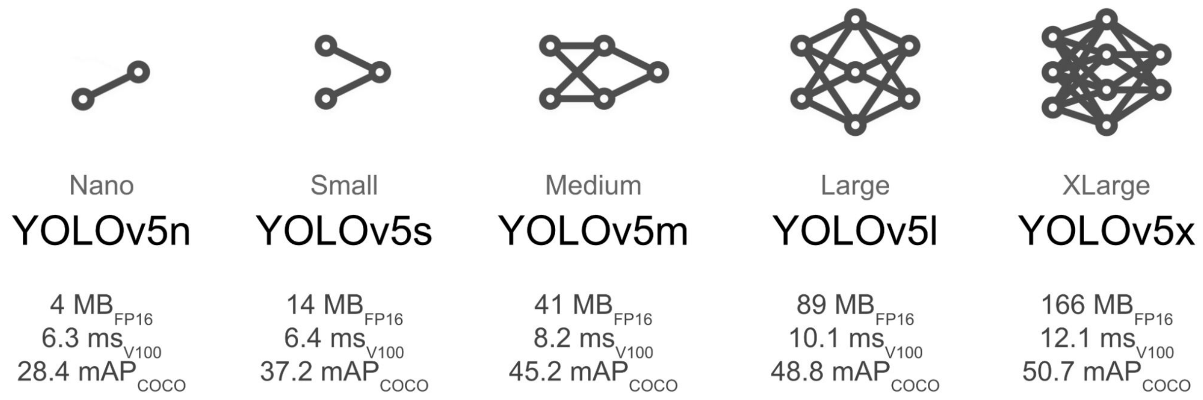
Generalised IoU Loss



- Max mAP reached - 0.64
- The model was quite accurate in obtaining true positives compared to False positives
- The dotted line which is the validation results is much lower than the training results in terms of loss
- Very good accuracy in detecting balloons for the validation set

Balloon Detection - Yolo v5

- Different Yolov5 Models



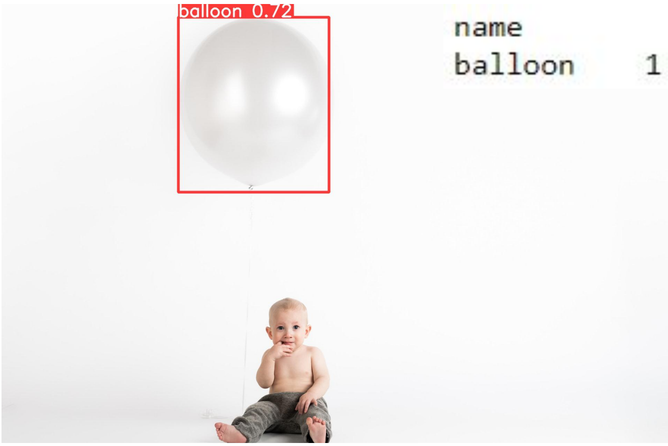
- Fine-tuning a pre-trained Yolov5 model on a our balloon dataset
  - Model used - YOLOv5s
  - structured the data as per the model requirements

- Our training with
  - dataset of 2245 training images
  - dataset of 414 validation images
  - 50 epochs
  - trained on local machine
  - 5.25 hours of training time

Balloon Detection - Yolo v5

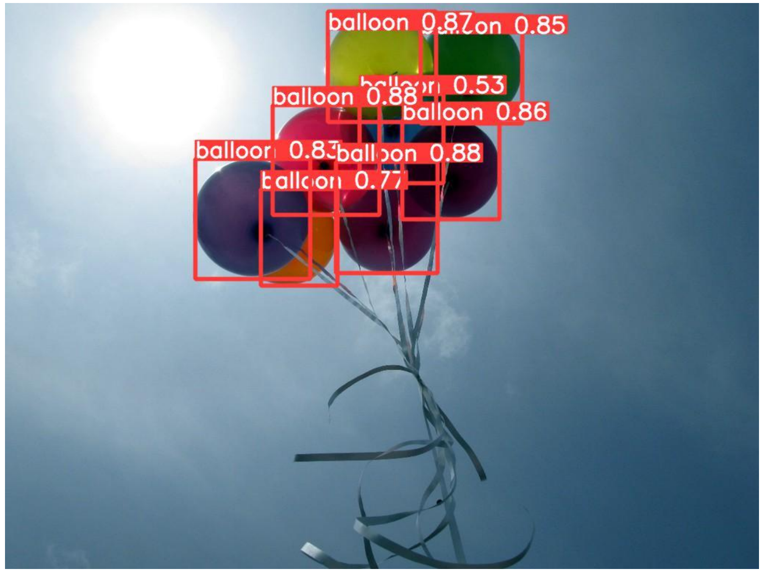
- Give **good result** in
  - avoid non-balloon object

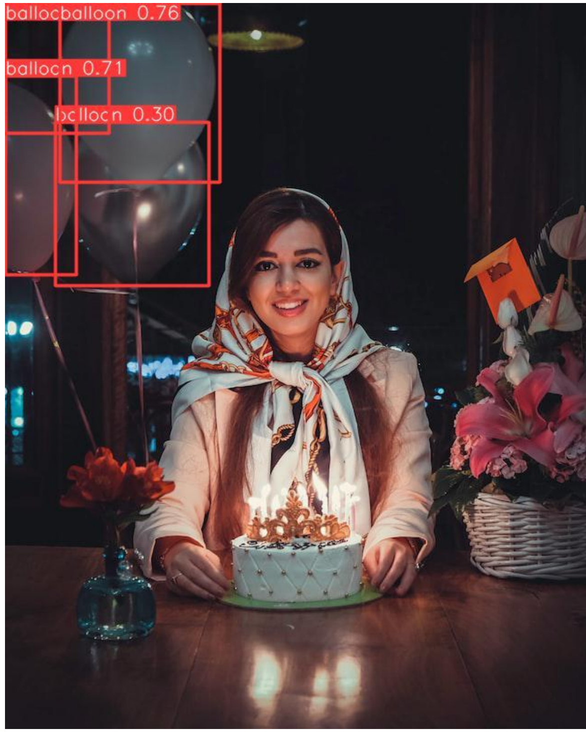
  - detect balloons if the balloons are clearly seen and in groups

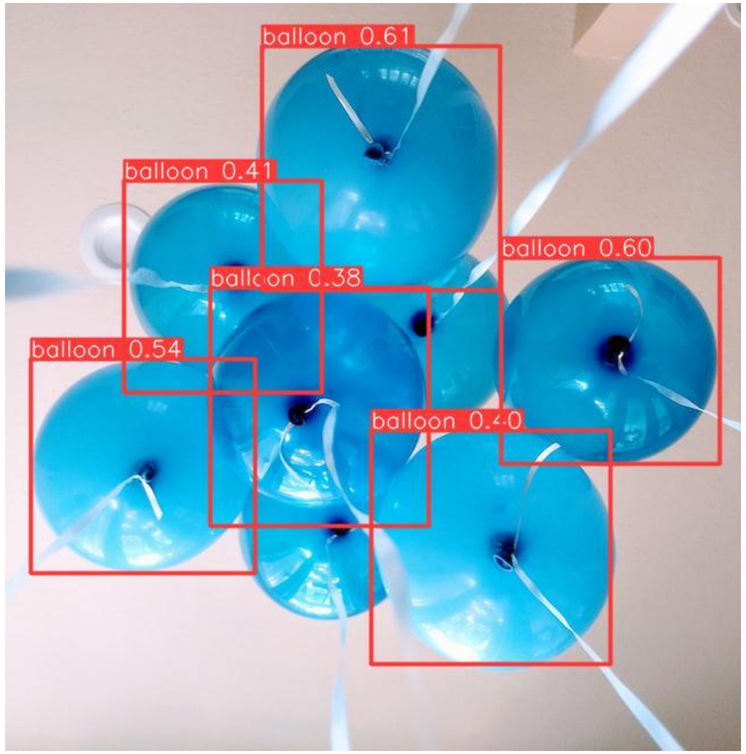


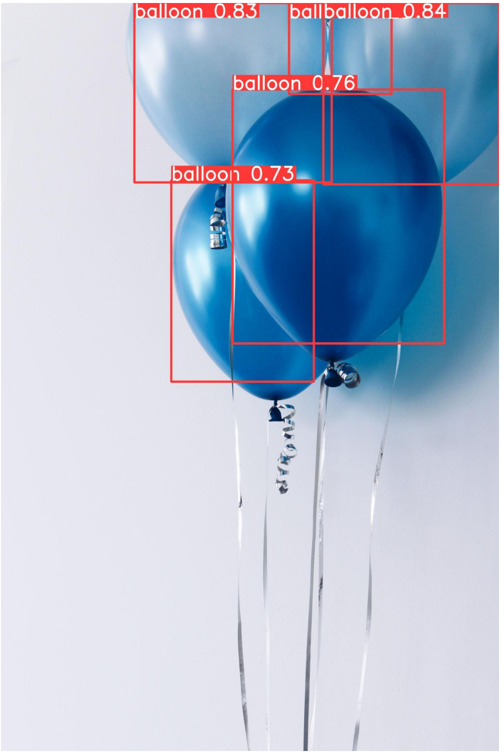
Balloon Detection - Yolo v5

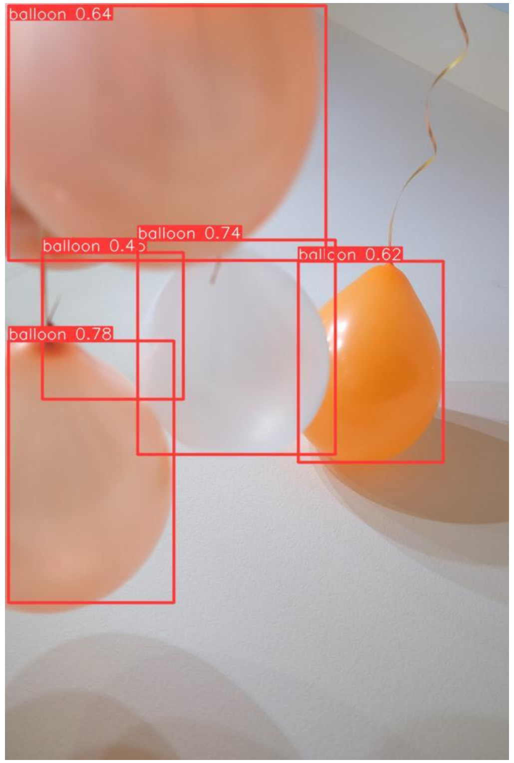
- Give **mix result** in
  - detect multiple balloons in big group or with less “visual”





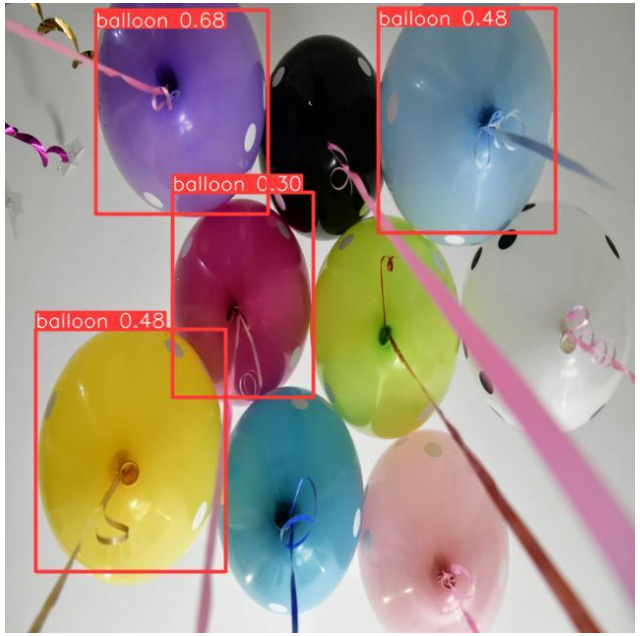






Balloon Detection - Yolo v5

- Give **bad result** when
  - balloon with less common position



  - balloons are blurry or far away or color too close to the background

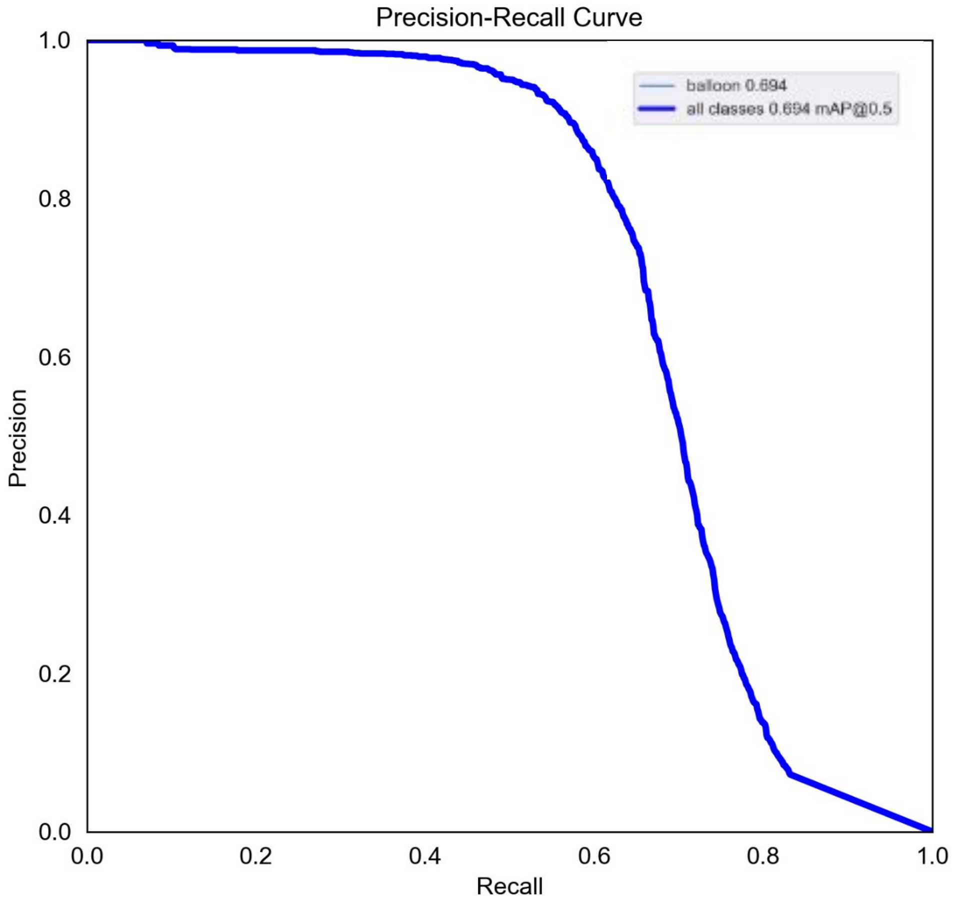
Balloon Detection - Yolo v5

- Cons:
  - Needs a larger Dataset
  - Is not a transformer bases model
  - Larger models need higher computing power

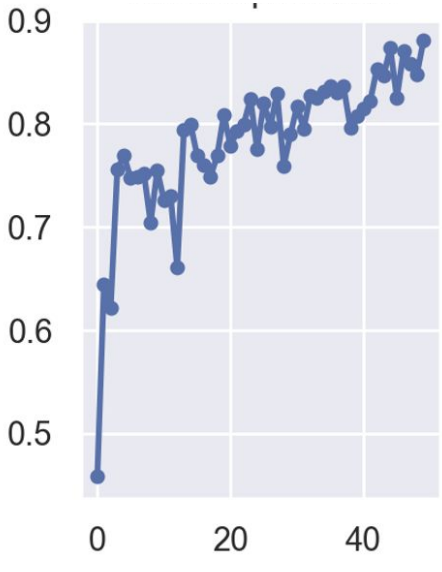
- Pros:
  - Smaller model size
  - Easy to set up
  - Smaller models can be trained without GPU

Balloon Detection - Evaluating Yolo v5

Results from the best epoch -
Epoch no - 43
Mean Average Precision (MAP) - 69.369%
Recall - 60.45
Precision - 84.64

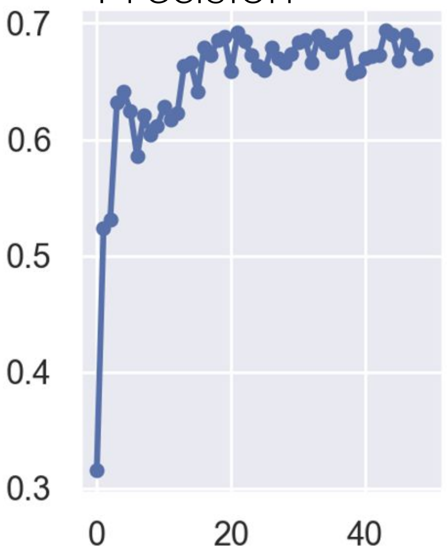


Precision



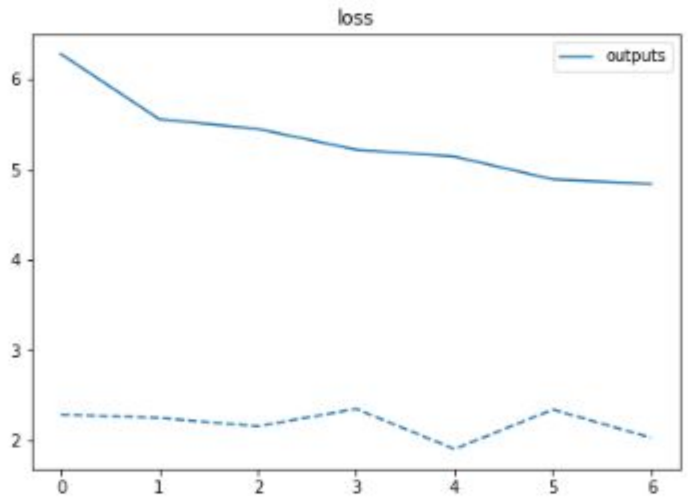
Recall

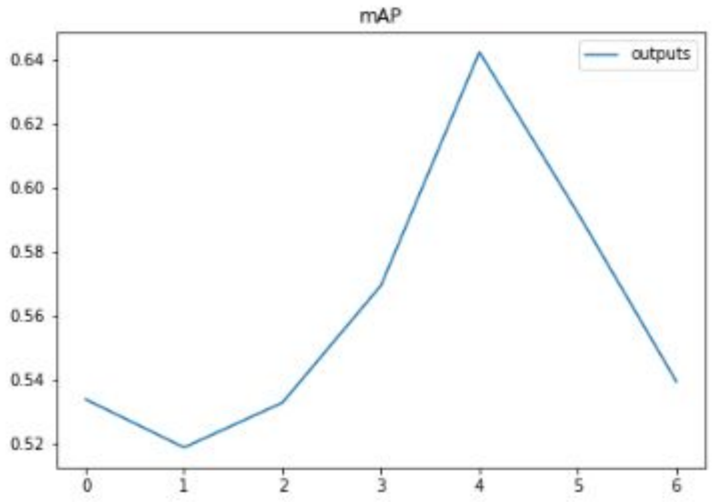
Mean Average Precision

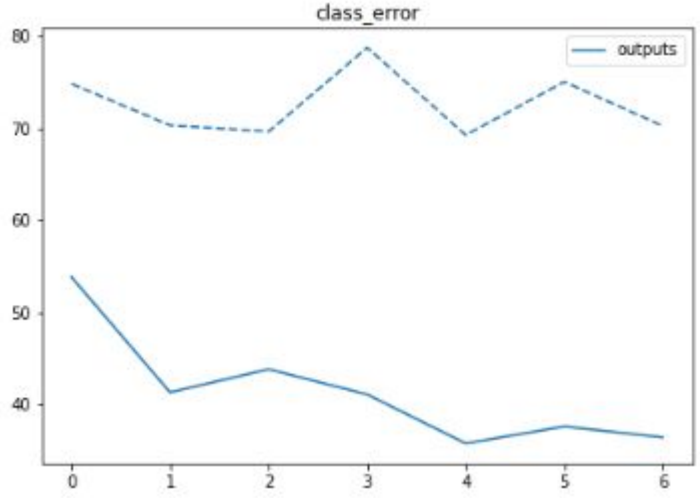


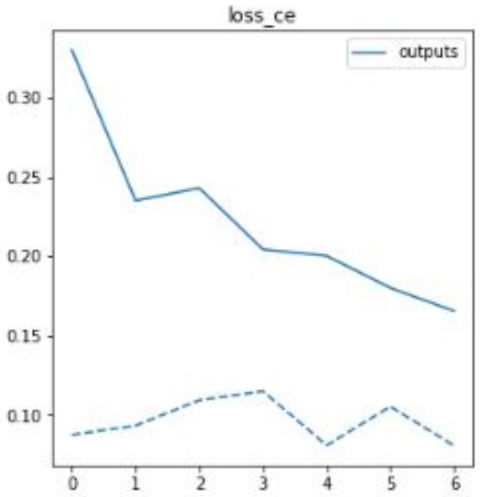
Q & A

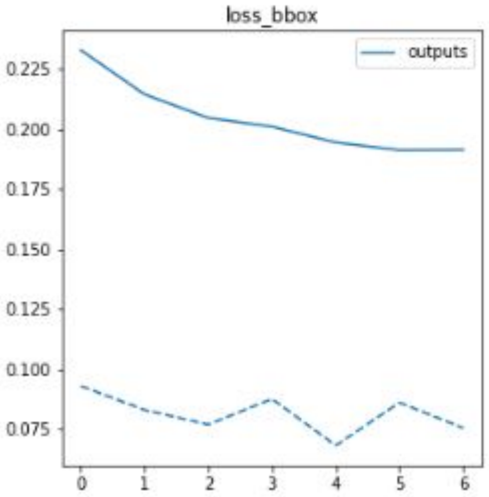
Balloon Detection - Evaluating DETR

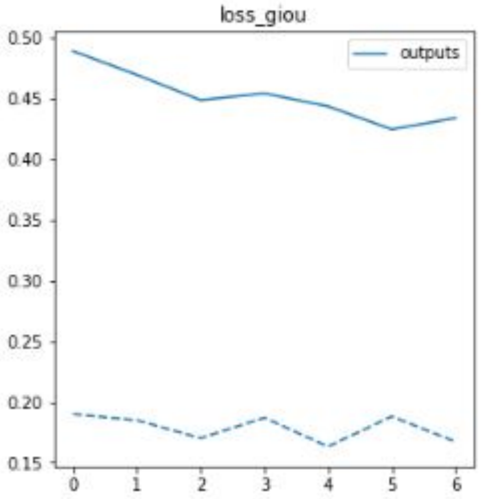


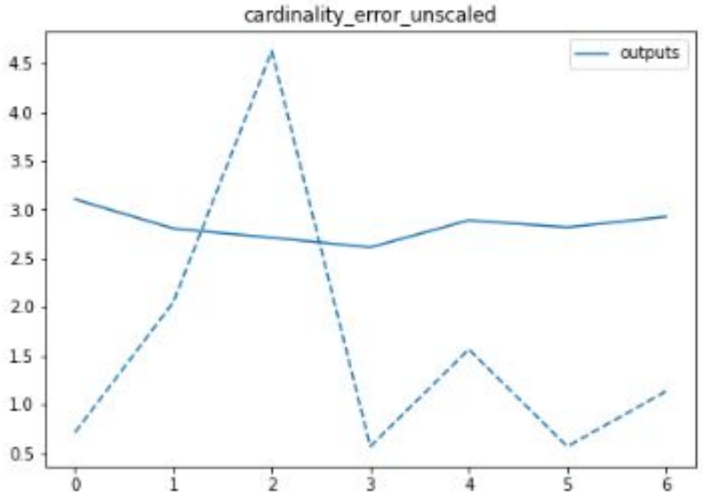






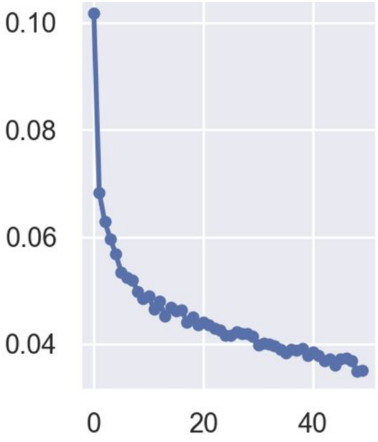




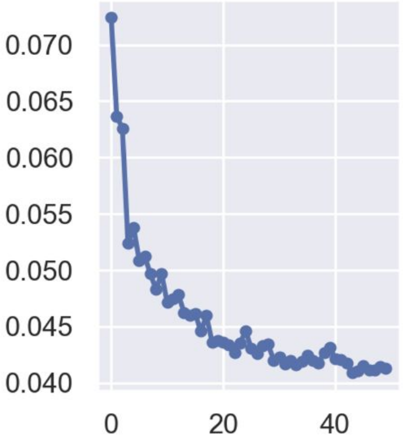


Balloon Detection - Evaluating Yolo v5

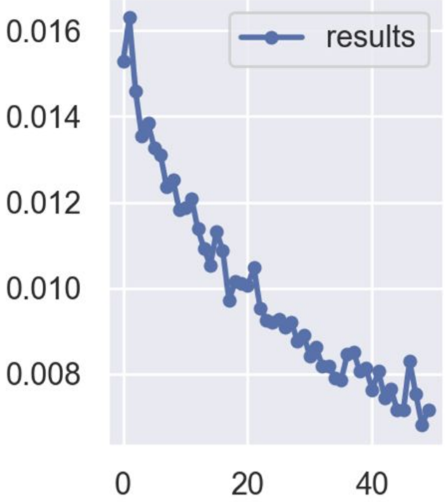
Train Box Loss



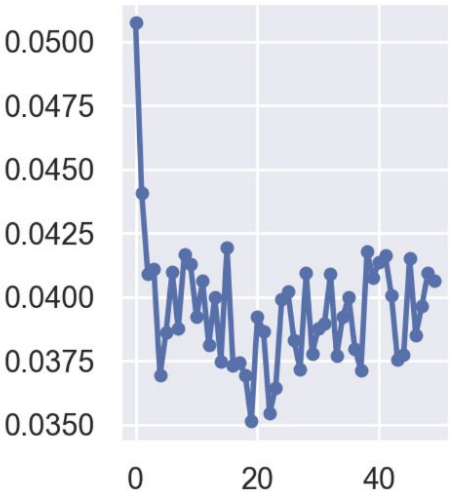
Validation Box Loss



Train Objectness



Validation Objectness



- Box: loss due to a box prediction not exactly covering an object.
- Objectness: loss due to a wrong box-object IoU **[1]** prediction