Training your Model

Upload your dataset

We will be training your model on Google Colab.

To use your images obtained from the previous step there, we will need to upload your dataset to your Google Drive. But before you upload, let's first zip the `train` and `val` folders in the `assets/duckietown_object_detection_dataset` directory of this exercise.

Please go ahead and execute the following 2 cells' code (Please do NOT change).

In [ ]:
```
%load_ext autoreload
%autoreload 2
```

In [ ]:
```
import os
import tempfile
import shutil
from typing import List
from datetime import datetime


def zip_sub_dirs(abs_root_dir: str, lst_rel_subdirs: List[str], output_baser
    """Zip some sub-directories, return the zipped file's path"""

    # check no identical output file exists
    out_full = f"{output_basename}.zip"
    if os.path.exists(out_full):
        print(f"File already exists at: {out_full}")
        print("Rename/Move it to run.\nNo operations performed.")
        return ""

    # make temporary directory
    tmp_dir = tempfile.mkdtemp()
    print(f"[{datetime.now()}] Temporary directory created at: {tmp_dir}")

    # format subdir original and temporary paths
    original_paths = [os.path.join(abs_root_dir, _d) for _d in lst_rel_subdi
    tmp_paths = [os.path.join(tmp_dir, _d) for _d in lst_rel_subdirs]

    print(f"[{datetime.now()}] List of directories to include in the zip fil
    # ensure all specified subdirs exist
    for subdir in original_paths:
```

```
        assert os.path.exists(subdir), f"Specified path does not exist: {sub
        print(f" - {subdir}")

    print(f"[{datetime.now()}] Move subdirs to the temp root dir")
    # move subdirs to the tmp dir
    for ori, tmp in zip(original_paths, tmp_paths):
        shutil.move(ori, tmp)

    # create the zip archive
    print(f"[{datetime.now()}] Compressing and creating the archive...")
    ret = shutil.make_archive(output_basename, 'zip', tmp_dir)

    print(f"[{datetime.now()}] Move subdirs back to original location")
    # move directories back to original location
    for tmp, ori in zip(tmp_paths, original_paths):
        shutil.move(tmp, ori)

    print(f"[{datetime.now()}] Finished. Archive created at: {ret}")
    return ret


# NOTE: DO NOT change these
# zip file basename for our dataset
ZIPPED_DATASET_BASENAME_FILE = "duckietown_object_detection_dataset"
# file/dir location constants
DATASET_DIR = "/code/object-detection/assets/duckietown_object_detection_dat
# path and file name (without file extension)
ZIPPED_DATASET_BASENAME_FULL = os.path.join(DATASET_DIR, ZIPPED_DATASET_BASE
TRAIN_DIR = "train"
VALIDATION_DIR = "val"

_ = zip_sub_dirs(
    abs_root_dir=DATASET_DIR,
    lst_rel_subdirs=[TRAIN_DIR, VALIDATION_DIR],
    output_basename=ZIPPED_DATASET_BASENAME_FULL,
)
```

```
[2023-12-31 17:31:48.331990] Temporary directory created at: /tmp/tmpq8e_t4h
c
[2023-12-31 17:31:48.332112] List of directories to include in the zip file:
 - /code/object-detection/assets/duckietown_object_detection_dataset/train
 - /code/object-detection/assets/duckietown_object_detection_dataset/val
[2023-12-31 17:31:48.332705] Move subdirs to the temp root dir
[2023-12-31 17:31:50.641360] Compressing and creating the archive...
[2023-12-31 17:31:52.998908] Move subdirs back to original location
[2023-12-31 17:31:55.354511] Finished. Archive created at: /code/object-dete
ction/assets/duckietown_object_detection_dataset/duckietown_object_detection
_dataset.zip
```

If everything went well, you should see the following output now:

```
Finished. Archive created at: /code/object-
detection/assets/duckietown_object_detection_dataset/duckietown_obj
```

The **zip file** is located in the *assets/duckietown_object_detection_dataset* directory of this exercise. You can upload it by going on your Google Drive and dragging the file into your drive.

Please be aware that

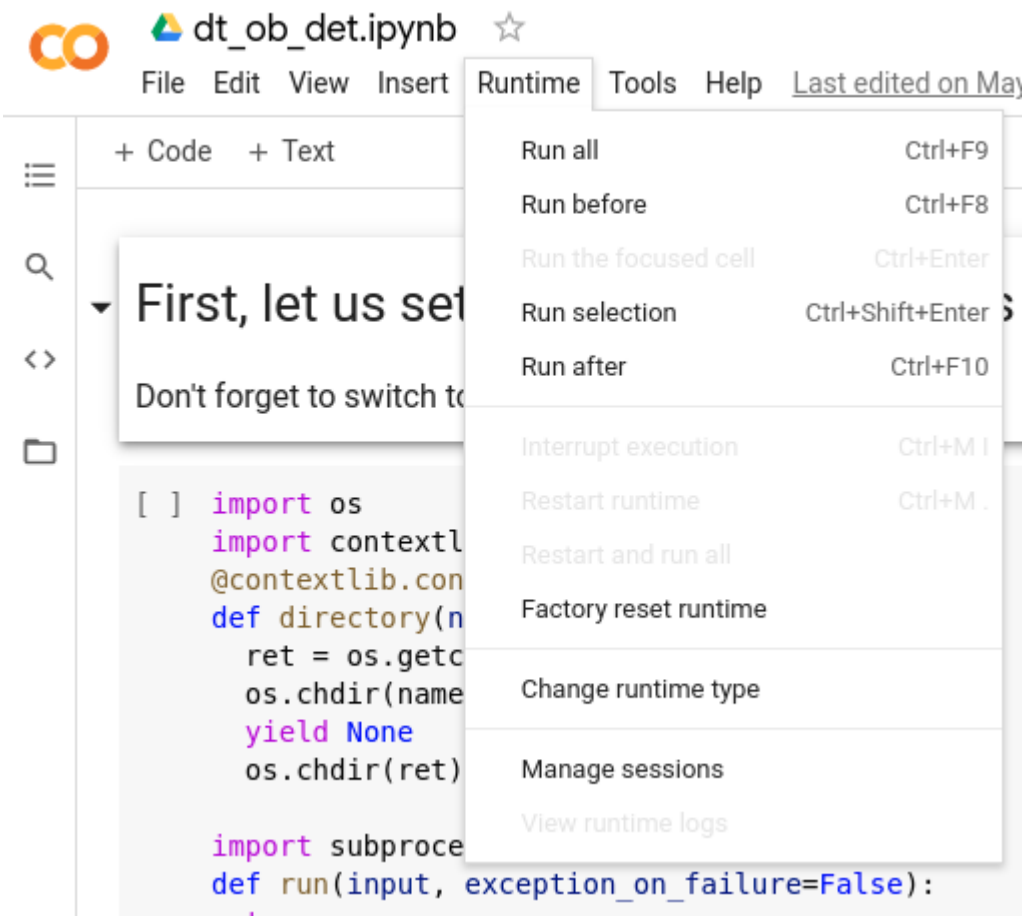
- you should **not** rename the dataset zip file
- the file should be uploaded to the out-most ***"My Drive"*** area

Training with Google Colab

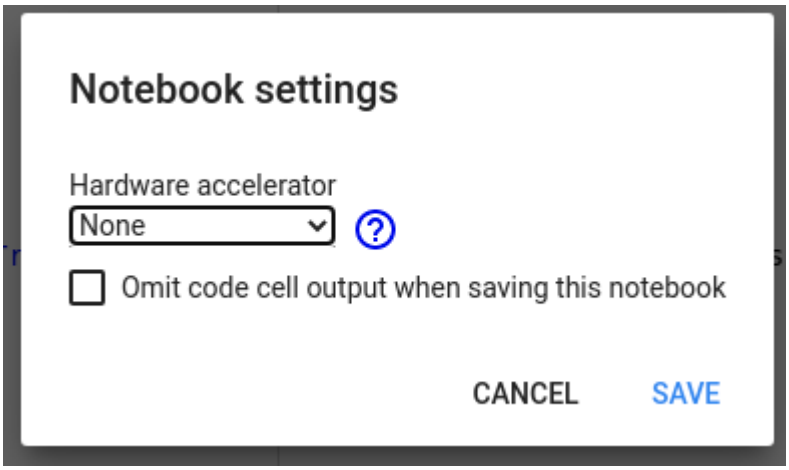
IMPORTANT: Make sure you carefully read the rest of this section **BEFORE** running anything in Colab.

- Use this Google Colab Notebook to train your model.
- Make sure the runtime type to GPU-accelerated!

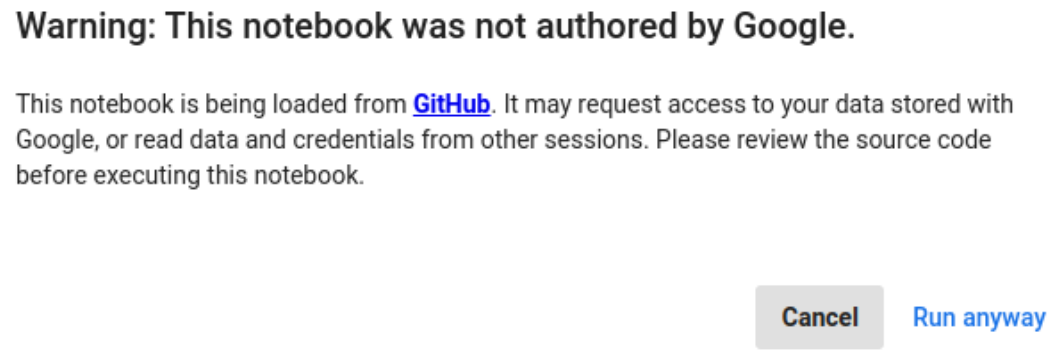
  Click on Runtime > Change Runtime Type.



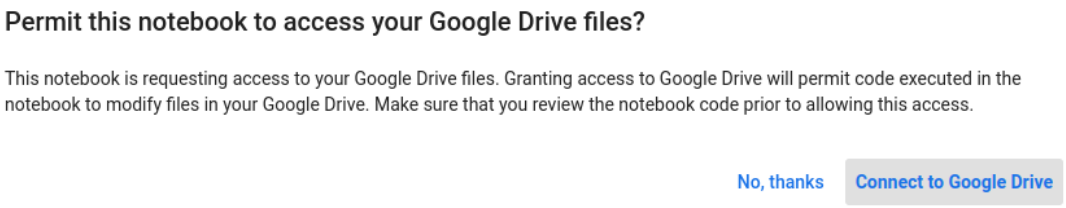
Then, in the drop-down menu, select "GPU"



- When prompted the warning below, click "*Run anyway*"



- When prompted the notification below, click "*Connect to Google Drive*"



NOTE: If you want to view the training statistics and artefacts, do NOT close the Colab notebook after training. Follow the instructions in the following **Debugging and Model Inspection** section to examine these contents.

Now follow the instructions in the Colab notebook. For reference, on google colab, training with the default settings takes a few minutes.

Optional (**not officially supported**) - local training

This is only recommended for experienced Machine Learning enthusiasts. Training the model requires a GPU. If you have one, you can run the command,

```
git clone -b dt-obj-det
https://github.com/duckietown/yolov5.git
```

You must now install all dependencies required by yolov5, and then call

```
python3 train.py --img 416 --batch 16 --epochs 100 --data
duckietown.yaml --weights yolov5n.pt
```

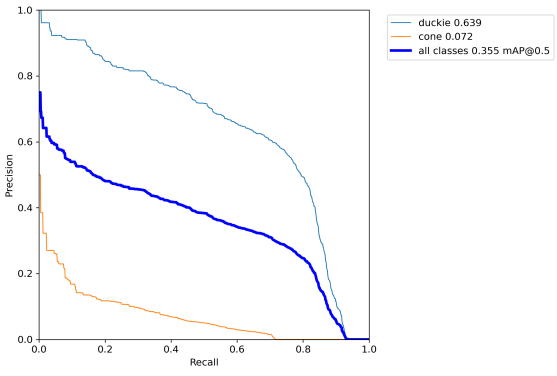
For reference, on a computer with `Nvidia GTX 1080TI` GPU and `Ryzen 3700x` CPU, training took about 20 minutes.

Debugging and Model Inspection

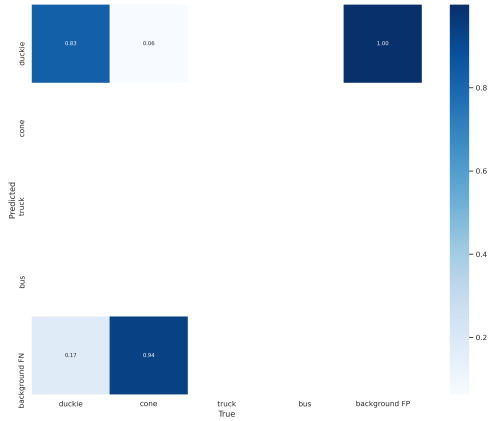
One you have finished training on Colab, there are a bunch of interesting outputs that will get generated during the training process that can be helpful for you to look at.

- During the Colab notebook execution, a session temporary workspace directory has been created. The path is shown in a cell output, similar as `Session workspace created at: /tmp/tmpxe3g50sz`
- Navigate to the workspace folder in the left Navigation Menu (on Colab), in the Files tree. (You might need to click the `..` to go up to `/`)
- After locating the `/tmp/...` workspace folder, go into the `yolov5` directory inside.
- Then navigate to `runs/train/expX/` where `X` is incremented each time you train.
- If you want to save these results, using the left Navigation Menu's Files tree, drag the folders/files you want to keep, to the `/content/drive/MyDrive` directory. Then you will be able to find them in your Google Drive even after closing the Colab session.

In here you can see things like your PR curve, e.g.:



Your confusion matrix:



And sample training outputs:

Next step

Onto the Integration notebook!