1.In my opinion, modern methods focus on location and segmentation. To extract more information from a video we may think from other aspects, like how to distinguish individuals in a video. See for details in part 4.

2.Training a model can be divided in 4 steps:

a. Configure model and parameters with protobuf library

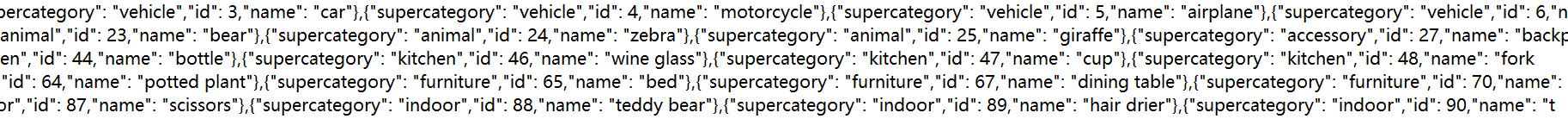
b. Prepare the data set (COCO here)

c. Train the model

d. Export model

Above all are brief description of training a model, more tests and environment configuration are needed if practise.

The model I tested was trained base on COCO and some other models use Kitti, Open Images etc.

* At first I didn't find a precise description about how many classes these models can detect on neither GitHub of TensorFlow Detection Model Zoo website nor COCO website. *I do find a 'pipeline.config' file under folder of this model and line 3 'num\_classes: 90' indicates it can detect 90 class but I'm not sure.* (The file has been uploaded to ‘config file’ on ***https://github.com/applepiiie/Human-body-detection***)
* Considering the data set downloaded from COCO was unique (no classified packages), I suppose these models trained with COCO data set can detect all classes that COCO provide, though the number of class was not informed on COCO's official website. Through the 'tensorflow-human-detection.py' file line 78 '# Class 1 represents human' and line 79 'if classes[i] == 1 and scores[i] > threshold:' I know by changing the number which classes[i] equals to can these models output other detected classes with highlighted boxes.
* After some searching I notice by downloading the COCO data set file and looking through the 'instances\_train2017.json' file under 'annotations\_trainval2017' folder may find the description of classes. Eventually I saw the class name and their number at the end of the .json file (part of the description shown below as screenshot).

3.90 in total objects can be detected like giraffe, teddy bear etc. by merely changing the code I mentioned in Answer 2. I can make a demo video if needed. TensorFlow Detection Model Zoo provides 32 models trained with 5 data platform and I believe all 25 COCO-trained models can detect the 90 objects, while other models need to be confirmed after looking into the four left data platform.

4.If wants the model to count number of people, then it needs to distinguish each individual in case of people reappear in the camera. I think facial recognition or gait recognition will increase technical difficulty while won't achieve satisfying result under UAV circumstance. I believe identifying individuals by calculating the length of body parts (like arm, leg, waist), various proportions and using math algorithm to generate different parameters worth trying.