1. What foundation did you use for your project? What did you modify and what did you keep identical?
2. Describe your processing pipeline and pre-processing steps, if any

   I started initially with an example from *pyimagesearch* using keras with tensorflow background. In fact such examples can easily be found on the internet. I extended the example to process videos. I spent quite some time in getting my installation on windows working since I often ended up with conflicting versions of keras and tensorflow or opencv and the code would crash. Most of the community follows keras (with tensorflow background) on ubunto. In the beginning I started with vgg16 model and later extended it to using numerous other models to check difference in accuracy of prediction. To speed up processing I tried using multithreading but this was not completely successful so I implemented a model where 9 out of 10 frames are dropped. This gave acceptable performance (time-wise). All models were pre-trained on 'imagenet' dataset. Also extended this to include webcam and finally I made a generic code where the model can be selected (using variable 'modeltype') . From my initial results: Inceptionv3 was worst in correctly identifying objects while vgg16,19 and resnet 50 gave good results. A short demo on a nature video (from BBC) is stored in the results folder.

   Then I extended the code to process multiple regions in the image separately by drawing boxes. Performance was very bad. Then I found that there is model called SSD (Single Shot multibox Detector) that speeds this up. I found an implementation of this in tensorflow examples that was trained on COCO dataset (common objects in context). This gave much better performance. An example result of this with webcam is stored in the results folder.

3. How much time did you spend on this exercise?

   I did not track this explicitly as my time was fragmented. But in total around 24 to 28 hours.

4. What shortcuts did you have to take to deliver the exercise in the short period of time allotted?

   In order to speed up development, I used pycharm as it significantly improves productivity. I lost quite some time in dealing with installation issues. Multi-threading was not working well especially in combination with offline videos so I implemented frame-dropping to speedup.

5. If you had more time, how would you expand on this submission? What is your wishlist for the app?

- Cleanup code, create a reproducible build environment, installer.
- Build a website where one good enter a youtube URL and get objects identified
- Use NLP and concepts from another DL technique called YOLO to get scene descriptions. This allows to search in videos using the video content interfacing with text.