W251 HW 8

Thomas Drage <draget@berkeley.edu>, June 2019.

PART 1

1. In the time allowed, how many images did you annotate?

   380 in 71 mins. This is ~11s per image.

2. Home many instances of the Millennium Falcon did you annotate? How many TIE Fighters?

   308x Millenium Falcon
   272x TIE Fighter

3. Based on this experience, how would you handle the annotation of large image data set?

   I think annotation of a large set should be distributed amongst multiple people or crowd sourced. It is a tedious and time consuming task. Such approaches are taken commercially (e.g. Mechanical Turk or even Google's new image CAPTCHA).

4. Think about image augmentation? How would augmentations such as flip, rotation, scale, cropping, and translation effect the annotations?

   Such augmentations would allow firstly the generation of a greater number of training samples for the same labelling effort, but secondly allow a greater generalisation of the model developed from the training. This is because the model would become less sensitive to position and scale factors which are difficult to control when in real-world (inference) use.

PART 2

1. Describe the following augmentations in your own words

- Flip
  The image mirrored about either the vertical or horizontal axis.

- Rotation

  The image is rotated through an angle about the centre.

- Scale

  While maintaining the same dimensions, the image is scaled in each axis to create a distortion in aspect or enlargement of an area of the image (or both).

- Crop

  In a crop the sample image is reduced and only a sub-rectangle is retained.

- Translation

  The position of the sample within the frame is altered by moving it in either axis.

- Noise

  A random signal is added to the pixels, resulting in a distortion of the image. The image retains visual similarity overall, but takes on difference at pixel level.

PART 3

1. Image annotations require the coordinates of the objects and their classes; in your option, what is needed for an audio annotation?

   In an audio annotation the most simplistic view is that a particular class of sound is a temporal slice from an audio track, thus the dataset would consist of a source trace and start/stop coordinates corresponding to a certain class. A more complex annotation could take on the spectral density and indicate the "coordinates" of a slice in frequency space across a time span, allowing a particular audio source to be isolated depending on pitch (consider annotating a bird singing at the same time as a bass guitar playing). However, obtaining a quality annotation is most dependant upon obtaining clean source media.