MIDS W251 Spring 2020

Section: Monday 6:30pm Suhas Gupta

HW8 – Data set annotation and augmentation

1. Image Annotation

- I annotated 384 images in one hour
- I annotated 299 instances of the Millennium FALCON
- I annotated 270 instances of TIE fighters
- There are a couple of good options for performing the annotations:
- Hire professionals to do the annotations manually
- Develop an unsupervised generative neural network that can perform annotations
- The augmentations will change the coordinates of the annotated objects in the XML files. Both human annotators and generative neural network should be able to perform well on these augmentations, however I expect the human annotators to perform slightly better.

2. Image Augmentation

- *Flip*: Flip is mirroring transformation of the image along the vertical or horizontal axis of passing through the image's center.
    - Hflip: Mirror across vertical axis
    - Vflip: Mirror across horizontal axis
- *Rotation*: Rotation is a more general case of the "flip" augmentation. In the rotation transformation, the image can be rotated at an arbitrary angle along an axis passing through the center of the image.
- *Scale:* Scale is same as stretching of an image. In this augmentation, the image pixels are multiplied with a scalar value resulting in a scaled/stretched version of the original image (This is also called resizing of the image).
- *Crop:* Crop augmentation is the separation of a part of the original image to generate a new image. The new image only contains the "cropped" region of the original image and all other original image's features are removed.
- *Translation:* Translation is the shifting of the image pixels from their original location. In this augmentation, the image is moved either along the x-axis or y-axis.
- *Noise:* In this augmentation, random noise is added to the image. This helps reduce overfitting to certain image features by reducing their dominance over other features in the image.

3. Audio Augmentation

- For audio augmentation, the time domain and frequency domain waveforms of the audio signals are needed.
    - Augmentations such as stretching, speed up, slow down can be efficiently applied to the time domain waveforms of the audio signal.
    - Augmentations such as noise injection can be efficiently applied to the frequency domain waveform of the audio signal.

- The spectrogram of the audio signal effectively lends the treatment of the audio signal as a visual one. In this domain, we can mask blocks of consecutive frequency channels, and, mask blocks of utterances in time [1]

References :

Google AI Blog: SpecAugment: A New Data Augmentation Method for Automatic Speech Recognition