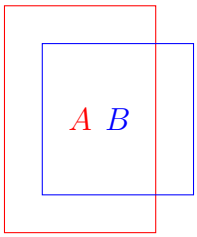
CO316 - Computer Vision (60006)

Lecture 1 - Introduction

Computer vision tries to build a system that can understand the world in a similar way to a human. At a higher level, the pipeline for vision consists of sensing an image or video, processing it, and then understanding it. For a human, the sensor is the eyes, and the processor is done by the primary visual cortex. On the other hand, a sensor can be a camera, or some form of medical imaging device, and the processor is the computer itself (and more importantly, the algorithm).

A **classification** problem has the goal of determining the **label** of what is in the picture. Classification is considered to be successful if one of the labels the algorithm predicts matches the true label. On the other hand, object **detection** attempts to draw a bounding box around an object (where are objects in the picture). We can quantify the success of detection based on the following. Consider the following, where the region in red is drawn by a human, and the region in blue is predicted by the algorithm;



We consider the detection of the intersection over union (IoU) is above 0.5;

$$\text{IoU} = \frac{A \cap B}{A \cup B} > 0.5$$

Another more complex piece of information we can extract is to perform **image segmentation**, allowing us to draw contours for each object.

Applications

Computer vision is used in our lives daily;

- **face detection**

  This can be noticed in most camera applications on modern smartphones, when a small box is drawn around faces. The algorithm first extracts **Haar** features from an image, and then determines (with these features) whether a region is a face or not.

  One example of these features is checking the contrast between the eyes and nose (horizontally); as the eyes tend to be quite dark in comparison. Another contrast is checked, this time between your eyes, as the nose tends to be brighter.

- **automatic number plate recognition**

  Automated barriers in parking lots can read number plates in order to calculate how long a car stays. Similarly, this can also be used to recognise building numbers, which is overlaid onto *Google Maps*, allowing for a large database of street numbers to be built in an automated fashion.

- **autonomous driving**
- **image style transfer**

  Choi et al. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation - used for changing features on inputs. Related to face motion capture (see *Face2Face*). Also see *DeepFake*.

- **Kinect**

  Works by taking a depth image, segmenting it into body parts, locating key points and building a skeleton.

- **design**

  See *OpenAI*'s *DALL-E*, combining NLP and computer vision by generating images based on the concepts of words in a sentence.

- **healthcare**

  Medical image analysis can be used for disease diagnosis. For example, identifying breast cancer lesions from mammograms.