Anchor Boxes

Tuesday, January 8, 2019   7:11 PM

In the algorithms like YOLO and SSD to predict multiple objects in a picture, the network is making thousands of predictions and only shows the ones that it decides as an object.

So the multiple predictions are output with the following format:

Prediction 1: (X, Y, Height, Width), Class
….
Prediction ~80,000: (X, Y, Height, Width), Class

The bounding box is defined by (X, Y, Height, Width).
While labeling these quantities are hand encoded by the humans.

So for example take the picture below.

In this example consider both of the fruits fall under the same grid and each grid is able to two bounding boxes.

So here we have a problem of assigning which information to which bounding box?

Prediction 1: Pear
Prediction 2: Apple

Or should it be:

Prediction 1: Apple
Prediction 2: Pear

What if the network predicts as shown below.

Prediction 1: Apple
Prediction 2: Apple

So we need to make these predictors responsible for particular scenarios.
This can be done by assigning the first predictor with the parameters of the objects that fall in the left side from the center of the grid box.
Similarly the right side objects will be assigned to the second predictor.

This method of selecting the predictor can also be done for the size and as well the aspect ratio of the objects in the grid cell.

Then the objects will be assigned based on the location of the center of the boxes of the objects inside the grid cell, the IOU of the predictor box (anchor box) with the actual box.

Reference:
C4W3L08 Anchor Boxes