More about Modern History of Object Recognition Infographic

Yuan An

May 5, 2018

In the last article, six key technologies in CV were introduced. There are also other knowledges to be presented.

Something about CNN were referred in my previous article. So only some important object recognition concepts are introduced there.

Important Object Recognition Concepts [1]

Bounding box proposal

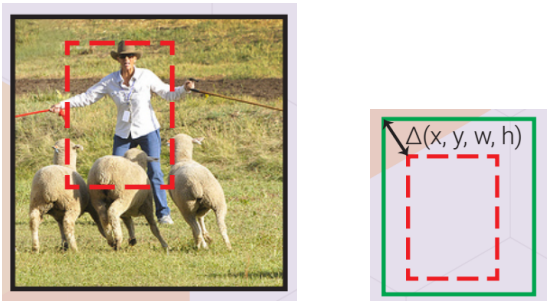

Figure 1: Bounding box proposal

Bounding box proposal are also named region of interest, region proposal or box proposal. As is shown in Figure 1, a rectangular region of the input image that potentially contains an object inside. A bounding box can be represented as a 4-elements vector, either storing its two corner coordinates (x0,y0,x1,y1), or storing its center location and its width and height (x,y,w,h), as presentation in right figure in Figure 1. And a bounding box is usually accompanied by a confidence score of how likely the box contains an object.

Intersection over Union

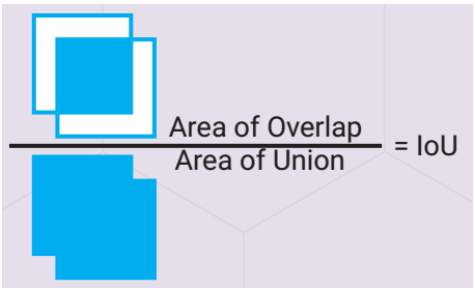

Figure 2: Intersection over Union

Intersection over Union (IoU, or called Jaccard similarity) is a metric that measures the similarity bewteen two bounding boxes.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

Non Maxium Suppression

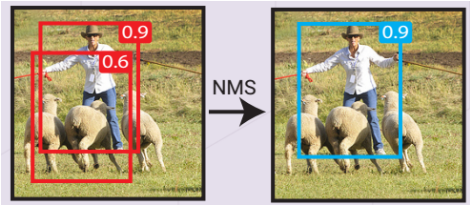

Figure 3: Non Maxium Suppression

Non Maxium Suppression (NMS) is a common algorithm to merge overlapping bounding boxes (proposals or detections). As displayed in the Figure 3, a bounding box that significantly overlaps (IoU > IoU_threshold) with a higher-confident bounding box is suppressed or removed.

Bounding box regression

Figure 4: Bounding box regression

Bounding box regression (also named bounding box refinement) is that by looking at an input region, we can infer the bounding box that better fit the object inside, even if the object is only partly visible. As shown in the Figure 4, it illustrates the possibility of inferring the ground truth box only by looking at part of object. So one regressor can be trained to look at an input region and predict the $offset_{\Delta}(x, y, w, h)$ between the input region box and the ground truth box.

Prior box

Figure 5: Prior box

Instead of using the input region as the only prior box, we can train multiple bounding box regressors, each look at the same input region but has a different prior box and learns to predict the offset between its own prior box and the ground the ground truth box. In this way, regressors with different prior boxs can learn to predict bounding with different properties (such as aspect ratio, scale, locations).

Box Matching Strategy

Figure 6: Box Matching Strategy

In general, a bounding box is not able to predict a bounding box of an object that is far away from its input region or its prior box. Therefore a box matching strategy is needed to decide which prior box is matching with a ground truth.

Hard negative example mining

For each prior box, there is a bounding box classifier that estimates the likelihood of having an object inside. And after box matching, all matched prior boxes are positive examples for the classifier while other prior boxes are negatives. When used all of these hard negative examples, the positives and negatives are significant imbalance.

References

[1] Đặng Hà Thế Hiển. Modern history of object recognition. https://github.com/Nikasa1889/HistoryObjectRecognition/blob/master/HistoryOfObjectRecognition.pdf/.