A REPORT TO IMAGE SEARCH APPLICATION

AYUSHI GARG
The National Institute Of Engineering,Mysuru

In this report I have put forward a model for a search engine where an image can be uploaded by the user to retreive information about it from the internet.Since an image is being used as a query it makes the search complicated as compared to searching of the keywords by inputing text.This approach is really beneficial for people who do not have an idea of an object.They simply upload the image of their query and get the relevant information from the internet.

SYSTEM ARCHITECTURE-

We see complicated sights with multiple overlapping objects and different backgrounds and we not only classify these different objects but also identify their boundaries,differences,and their relativeness to one another.

So when we are seeing an image with multiple things around it,we are detecting boundaries around it to classify.We tend to differentiate between one object and the other object and the relations between them.Other region proposal classification networks (fast RCNN) which perform detection on various region proposals and thus end up performing prediction multiple times for various regions in a image.

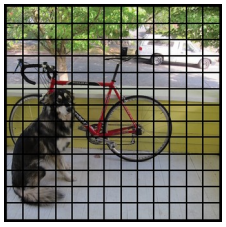
To further move on,we can use several image classifiers such a VGGNET or INCEPTION for image detection.

People are obsessed by the usage of conventional CNNs.In this fast growing technological world a new approach was put up which was called YOLO-YOU ONLY LOOK ONCE.

YOLO is a state of art,object detection system.Yolo architecture is more like FCNN (fully convolutional neural network) and passes the image (nxn) once through the FCNN and output is (mxm) prediction.

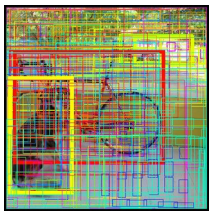
It actually looks at the image just once in a clever way.It divides the image into a s by s cells.

Each of these cells is responsible for predicting 2 bounding boxes.

A grid of 13 by 13 cells 

A bounding boxes describes the rectangle the image has been put in.Yolo also gives a confidence score that tells us the certainity about the class the bounding box is enclosing.The score does not say anything about what kind of object is in the box.

Bounding boxes with higher predictive scores are made thicker telling that we have something significant there.For each bounding boxes the cell also predicts the class.It works like a classfier.It gives a probability distribution over all the possible classes that the network has been trained of.

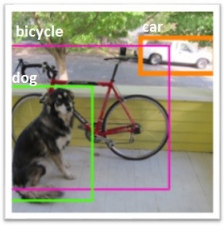


Higher predictive scores are made thicker

So now we combine the confidence score of the bounding box and the class prediction and combine them to tell about a specific bounding box and what class type object it contains.

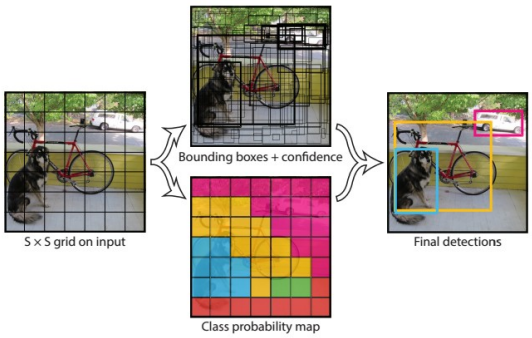
Since there are several bounding boxes and we cannot keep all of them,hence we set a limit .So the

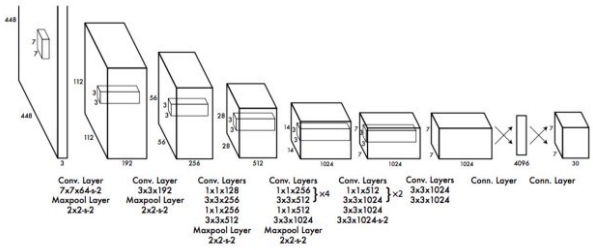
bounding boxes whose confidence scores are less than the set limit are discarded. Hence in the end result we are going to have only those boxes who are thicker and have a specific class.



The objects were identified

An illustration for better understanding is given as follows-
1. An image has been divided in a grid of 13 by 13 cells.
2.Boundary boxes and confidence boxes have been made around the classes.
3.A THRESHOLD for example of 30% has been set and only the boxes with more than 30% of the confidence scores are kept which in this image are 3.
4.The class identifier indentifies them as a dog,bicycle and a car.





ARCHITECTURE -DETECTION NETWORK HAS 24 CONVOLUTIONAL LAYERS FOLLOWED BY 2 FULLY CONNECTED LAYERS.ALTERNATING 1X1 CONVOLUTIONAL LAYERS REDUCE THE FEATURES SPACE FROM PRECEDING LAYERS.WE PRETRAIN THE CONVOLUTIONAL LAYERS ON THE IMAGE NET CLASSIFICATION TASK AT HALF THE RESOLUTION (224X224) INPUT IMAGE AND THEN DOUBLE THE RESOLUTION FOR DETECTION

DATASETS-

COCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features:

•Object segmentation
•Recognition in context
•Superpixel stuff segmentation
•330K images (>200K labeled)
•1.5 million object instances
•80 object categories
•91 stuff categories
•5 captions per image
•250,000 people with keypoints

I would not like to prefer making a dataset of my own as it will be time consuming,less accurate and data set would not be so large.

Now in the search engine-

1.The user will upload an image.
2.The image undergoes Yolo.
3.After the images are matched,the complete web links that hold a match with the uplaoded image are saved in a file and each of them are searched for the source code for a particular website.
4.After the image link is found the source code,the paragraph corresponding to the image is taken and segmented into words to find a certain word that frequentlly occur in sentences.Further that word is taken down in to more searches of those webpages that the image belongs.
Further page ranking techniques are used and finally the highest ranked web page along with that image is displayed.

ADVANTAGES-

1.This model is simple to construct.Specially it is trained on a loss function that directly corresponds to detection performance and the entiremodel gets trained jointly.
2.YOLO is a fast general-purpose object detector and it pushes the state-of-the-art in real-time object detection.YOLO also generalizes well to new domains making it ideal for applications that rely on fast, robust object detection.
3.Network understands generalized object representation (This allowed them to train the network on real world images and predictions on artwork was still fairly accurate).

RELEVANT EVALUATION CRITERIA FOR MODELS And SHORTCOMINGS-

1. Yolo imposes spatial constraints on bounding box predictions.Each cell predicts only 2 boxes and can have only one class.This contraint makes the model struggle with small objects present in image,for e.g – a flock of birds.
2. Keeping multiple objects in a same image might not give required results.

REFERENCES-

1.Original paper (CVPR 2016. OpenCV People's Choice Award) https://arxiv.org/pdf/1506.02640v5.pdf