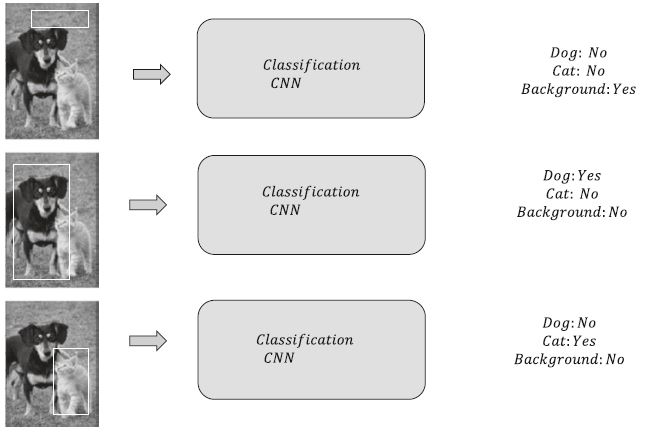
CNN

The task of detecting objects in images is a classical problem in computer vision, CNN is a very useful method to deal with this problem. However CNN usually detect only one imagine patch when it was used, so if use it to detect multiple objects, it could be computationally expensive.

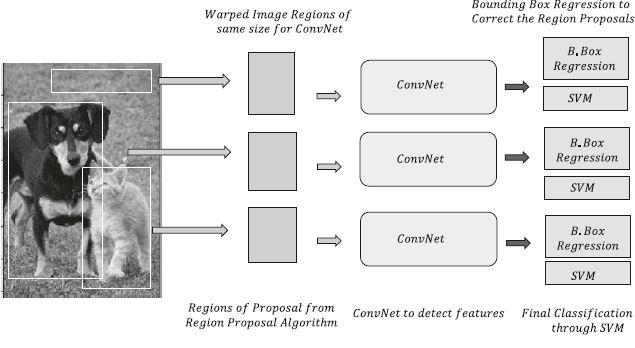


But of cause finally it could work.

R-CNN/Fast RCNN

In R-CNN the letter R means region proposals. The region proposals are usually derived in through an algorithm called selective search. A selective search on an image generally provides around 2000 region proposals of interest. Selective search usually use traditional imagine processing technology to find the object.

These 2000 regions of interest are then fed to the classification and localization network to predict the class of the object along with associated bounding boxes. The classification network is a convolutional neural network followed by a support-vector machine for the final classification. It just like to make a preprocessing before use classification CNN.



R-CNN perform better than CNN but it also have some problem. First it has about 2000 proposals which make the net very slow. Besides it use separate models to predict the bounding and localization so we are not learning anything specific to localization of the objects based on the training data.

Fast R-CNN use a common convolution path to make the network run faster, but it has less accuracy.

[1] Navin Kumar, Manaswi ; Deep Learning with Applications Using Python.

[2] Santanu Pattanayak; Pro Deep Learning with TensorFlow