**DELHI TECHNOLOGICAL UNIVERSITY**

**Machine Learning Project**

**Video Content Analysis (Object Detection)**

**Under supervision of: Dr. Anil Singh Parihar**



**Submitted by:**

**PRAJJWAL CHITTORI (2K18/CO/249)**

**(Dept. of Computer Science and Engineering)**

**Fixes used to make object detection faster**

During implementation of Batch normalization calculating activation statistics from 4 different images on each layer.

Using Self-Adversarial Training (SAT), which operates in 2 forward backward stages. In the 1st stage the neural network alters the original image instead of the network weights. In this way the neural network executes an adversarial attack on itself, altering the original image to create the deception that there is no desired object on the image. In the 2nd stage, the neural network is trained to detect an object on this modified image in the normal way.

Fixing the model to low resolution scanning to keep the model relatively fast.

**Drawbacks and Contribution**

The first one is little object detection, for example, occurring in COCO dataset and in face detection task. To improve limitation accuracy on little objects under partial impediments, it is necessary to change network architectures from the accompanying angles.

• Multi-task joint enhancement and multi-modular information combination. Because of the correlations between different assignments inside and outside object detection, perform multiple tasks joint enhancement has already been concentrated by numerous researchers. However, apart from the errands referenced in Subs. III-A8, it is desirable to thoroughly consider the characteristics of different sub-assignments of object detection (for example superpixel semantic division in notable object detection) and stretch out perform various tasks advancement to other applications, for example, case division [66], multi-object tracking and multi-person present assessment . Additionally, given a particular application, the information from different modalities, for example, text , thermal information and pictures , can be combined to accomplish a more discriminant network.

• Scale adaption. Objects for the most part exist in different scales, which is more apparent in face detection and pedestrian detection. To increase the robustness to scale transforms, it is requested to train scale-invariant, multi-scale or scaleadaptive detectors. For scale-invariant detectors, more powerful spine architectures (for example ResNext ), negative example mining , reverse association and subcategory modeling are altogether useful. For multi-scale detectors, both the FPN which produces multi-scale feature maps and Generative Adversarial Network which narrows representation differences between little objects and the large ones with an ease architecture provide experiences into generating significant feature pyramid. For scale-versatile detectors, it is valuable to consolidate information graph, attentional component , course network and scale distribution assessment to distinguish objects adaptively.

• Spatial correlations and logical modeling. Spatial distribution assumes an important role in object detection. So, region proposal generation and grid regression are taken to get probable object areas. However, the correlations between numerous proposals and object categories are ignored. Plus, the worldwide structure information is deserted by the position-delicate score maps in R-FCN. To tackle these problems, we can refer to diverse subset choice and successive reasoning errands for potential arrangements. It is additionally important to veil notable parts and couple them with the worldwide structure in a joint-learning manner. The subsequent one is to release the burden on physical work and achieve real-time object detection, with the emergence of large-scale picture and video information. The accompanying three angles can be considered. • Cascade network. In a course network, a course of detectors are implicit different stages or layers. Furthermore, effectively recognizable models are rejected at shallow layers so that features and classifiers at latter stages can deal with more troublesome examples with the guide of the choices from previous stages. However, current falls are inherent a greedy manner, where previous stages in course are fixed when training another stage. So, the improvements of different CNNs are disconnected, which stresses the need of end-to end advancement for CNN course. Simultaneously, it is additionally a matter of concern to construct logical related course networks with existing layers.

• Unsupervised and pitifully supervised learning. It's very tedious to physically draw large amounts of bounding boxes. To release this burden, semantic prior, unsupervised object discovery, numerous occurrences learning and profound neural network prediction can be integrated to utilize picture level supervision to allot object category labels to corresponding object regions and refine object boundaries. Furthermore, pitifully comments (for example centre-click comments) are additionally useful for accomplishing great detectors with unassuming comment efforts, particularly supported by the portable platform.

• Network enhancement. Given explicit applications and platforms, it is critical to make an equilibrium among speed.

Further on we study the impact of different spine models on the detector accuracy. We notice that the model characterized with the best grouping accuracy isn't generally the best in terms of the detector accuracy. First, in spite of the fact that grouping accuracy of CSPResNeXt50 models trained with different features is higher compared to CSPDarknet53 models, the CSPDarknet53 model shows higher accuracy in terms of object detection. Second, utilizing BoF and Mish for the CSPResNeXt50 classifier training increases its grouping accuracy, however further utilization of these pre-trained weightings for detector training reduces the detector accuracy. However, utilizing BoF and Mish for the CSPDarknet53 classifier training increases the accuracy of both the classifier and the detector which utilizes this classifier pre-trained weightings. The performance of CSPDarknet53 is better in terms of computational load than CSPResNeXt50. We observe that the CSPDarknet53 model demonstrates a greater capacity to increase the detector accuracy inferable from various improvements.