FCOS: A Simple and Strong Anchor-Free Object Detector – Critique

This is a critique of the above-mentioned paper done as a part of the assignment for CS 771. It follows the format described in the assignment writeup

Debarshi Deka [ddeka@wisc.edu](mailto:ddeka@wisc.edu)

Srinath <sgnamburi@wisc.edu>

# Summary

This paper proposes a new object detection model called FCOS (Fully Convolutional One-Stage Object Detection). Unlike the current state-of-the-art models such as Faster R-CNN, YOLOv3, and RetinaNet, which use a two-stage approach to solve the detection problem, FCOS uses a single-stage approach. The key idea is to eliminate the need for predetermined anchor boxes, which would otherwise complicate the computation of anchor boxes, thereby solving object detection in the neat per-pixel prediction fashion analogous to FCN (Fully Convolutional Networks) for semantic segmentation. Further, FCOS shows that simpler FCN-based detector achieves better performance than their anchor-based counterparts.

Diagram

Description automatically generated

# Experiments

FCOS uses focal loss for classification, IOU loss for regression, and binary cross-entropy loss for the center-ness of the image. It uses ResNet-50 architecture as the backbone, pretrained on ImageNet. The network is trained with SGD for 90k iterations with a batch size of 16 and a learning rate of 0.01. Finally, input images are resized to their shorter side being 800px and larger side being less than or equal to 1333px. For inference, the image is first passed through the forward pass of the network to obtain the bounding boxes and class scores, and then, the NMS threshold of 0.6 is used to filter out the overlapping boxes. The following experiments are performed by the authors as described in the paper.

1. They compared BPR (Best Possible Recall) between RetinaNet and FCOS. It was observed that FCOS and RetinaNet have similar BPR i.e. 98.95% vs 99.32% respectively. They also show that FCOS achieved better or similar AR when compared to RetinaNet upon using the same train and test settings.
2. They checked the impact of center sampling and FPN on FCOS to observe the effect of the number of ambiguous samples. It was seen that including center sampling in FCOS (without FPN) reduces the ambiguous samples ratio to only 3.48%. Further, applying FPN reduces this to 2.66%.
3. They showed that the center-ness had a significant impact in crowded scenarios and Center-ness increases the AP from 38% to 38.9%.
4. They performed an ablation study where they removed the group normalization which decreases the AP by 1% both in the case of classification and regression.
5. They used different methods to assign objects to the FPN levels. In one of the experiments, they assigned object proposals to FPNs which decreased the AP to 37.7%.
6. They conducted experiments on the COCO dataset where FCOS outperformed RetinaNet with 4.1% AP. It also performed better than the two-stage object detection approaches such as Faster-RCNN by 7% AP.
7. They compared the performance between FCOS and RetinaNet on the CrowdHuman dataset where it was observed that FCOS gives similar results as anchor-based RetinaNet on AP implying that anchor-based detectors no longer have advantages in high-crowded scenarios.
8. They tested FCOS in real-time scenarios. With ResNet-50, FCOS achieves 40.1% AP at 38 fps. In another experiment, they replaced ResNet-50 with DLA-34 which resulted in a better speed/accuracy tradeoff (40.3% AP at 46 fps). and it surpassed CenterNet by 1.7% AP at the same speed.

# Major Contributions

* FCOS is a single-stage detector that performs object detection in a proposal-free and anchor-free manner. This improves the efficiency and simplicity compared to the previous object detection methods making it a good choice for deployment in resource-constrained environments. Further, FCOS has 9x lesser network parameters as compared to the famous anchor-based detectors.
* Since FCOS uses FPNs (Feature Pyramid Network) to apply scale-invariant properties, it allows the model to effectively detect objects of varying sizes within an image. This improves the overall accuracy of the model. Moreover, sharing heads between different feature levels when using FPN in FCO makes the object detector parameter-efficient as well as improve the overall performance of the model.
* The authors also propose an effective strategy to decrease the low-level detections produced by the pixel locations distant from the center of the object. They mention that a single-layer branch is added which predicts the center-ness of the location. Here, center-ness depicts the normalized distance from the center of the object to that location. This decreases the weights of bounding boxes for the locations far from the center of the object which can then be filtered out in the post-processing using NMS(non-maximum suppression).

# Strengths and Weaknesses

The following points describe the strengths of FCOS and the corresponding tradeoff i.e. weakness which results upon addressing it.

* Simplification, efficiency, and performance:
  + **(Strength)** FCOS performs object detection in an anchor-free manner, which simplifies the computations related to the anchor boxes, such as IOU, and matching with ground truth boxes during training. This leads to faster training and testing compared to anchor-based detectors. This also leads to FCOS having significantly lesser design parameters.
  + **(Weakness)** FCOS doesn’t work well with low-resolution images. Further, FCOS has limited accuracy. Recent works such as Deformable DETR (an end-to-end object detector) proved to be more efficient and fast-converging and their model also achieved higher AP compared to FCOS.
  + **(Weakness)** FCOS predicts the center points of objects and then regresses to the object’s bounding box. Therefore, it is sensitive to the starting predictions made by the model thereby requiring diligent initialization or pre-training to achieve good performance.
* Two-stage detection and multiple tasks
  + **(Strength)** Compared to the other state-of-the-art approaches such as RetinaNet, YOLOv3, and Faster R-CNN, FCOS is very customizable. For example, FCOS can be easily modified to perform tasks like instance segmentation, text spotting, and keypoint detection.
  + **(Strength)** FCN-related tasks, such as semantic segmentation, can be unified with detection, making it easier to apply ideas from those tasks to detection. For instance, there is a technique called structured knowledge distillation, developed for dense prediction tasks which can be used directly for object detection using the FCOS network.
  + **(Weakness)** FCOS lacks refinement when compared to two-stage detectors such as Faster R-CNN and YOLOv3. Therefore, FCOS is not reliable when dealing with complex datasets.

# Improvement Suggestions and Comments

* To deal with low-resolution images, input images of varying scales and sizes should be used during training and inference.
* Additional context can be incorporated by using additional context branches or increased input image sizes to help understand the relationship between objects and surrounding environments.
* Experiments should be performed with ensemble or cascade methods for object detection which may improve model performance.