SCRDet: Towards More Robust Detection for Small, Cluttered and Rotated Objects

**[Introduction]**

Object detection에는 세 가지 challenging task가 있다.

1) Small object

2) Cluttered arrangement : Object들이 뭉쳐 있는 경우가 많다.

3) Arbitrary orientation: Aerial image에서는 다양한 방향을 가진 물체들이 존재한다.

이를 해결하기 위해 저자는 다음 세가지 Method를 제안한다.

1) sampling fusion network(SFnet)을 이용하여 small object에 대한 issue를 해결한다.

2) multi-dimensional attention network를 이용하여 noisy back ground를 억제한다.

**[Proposed method]**

**3.1. Finer Sampling and Feature Fusion Network**

Small object detection에는 feature information의 부족과 부적절한 anchor sample이 문제점으로 존재한다. 이유는 pooling layer를 이용할 경우 small object feature information이 사라지며 large sampling stride는 smaller object를 skip할 가능성이 높다.

**Feature fusion**을 위해서 location information을 담고 있는 low-level feature maps과 high semantic feature를 담고 있는 high-level feature maps을 이용한다. 이 때 이용되는 network는 FPN, TDM, RON등이 있다.

**Finer sampling**: insufficient training sample & imbalance는 network의 성능에 안 좋은 영향을 준다. EMO를 이용하여 object와 anchor사이의 expected IoU를 계산한다. EMO의 저자는 smaller stride of anchor가 높은 EMO score를 달성한다는 것을 확인했다.

위의 두 가지 논의를 토대로 SF-net을 제안하였다. S\_A는 Original image의 feature map의 reduction factor와 동일하다. 즉, S\_A의 사이즈는 2의 배수가 될 수 밖에 없다. 이러한 문제점을 SF-net은 feature map의 사이즈를 변경하는 것으로 해결하였다. 이는 더욱 적합한 sampling에 도움이 되었다. 또한 network의 파라미터를 줄이기 위해서 resnet에서 C3, C4 layer만 이용한다. 이는 semantic information과 location information을 혼합하고 상관도가 낮은 information을 배재하는 효과를 가진다.

**3.2. Multi-Dimensional Attention Network**

Excessive noise는 object information을 압도할 수 있다. 이는 그 결과 잘못된 detection 또는 false alarms을 증가시킬 수 있다. 이 때문에 object information cues를 증가시키고 non-objective information을 감소시키는 것이 중요하다. 기존의 많은 연구들은 occlusion, noise 그리고 blurring의 문제점을 해결하였지만 대부분의 method는 비지도방식 이므로 정확한 목적을 가지고 network learning을 진행시킬 수 없다는 문제점이 있다.

저자는 small object against complex background 를 더욱 잘 capture하기 위해서 Multi dimensional attention network를 제안한다. MDAnet은 F3를 inception module을 통과시킨 다음 two channel saliency map을 얻는다. 이 saliency map은 각각 background score와 foreground score을 나타낸다. 그리고 softmax layer를 적용시키고 F3의 channel 중 하나를 선택하여 곱해준다. 이러한 과정은 object information을 강화 시켜주고 noise를 줄여주며 동시에 이러한 값들은 연속적이므로 특정 상황 정보를 유지시켜주고 더욱 안정적인 network를 만들 수 있다.

또한, 저자는 supervised learning을 위하여 label을 binary map으로 표현하여 saliency map이랑 pixel wise sigmoid CE loss를 구해서 이를 attention loss로 이용한다. 거기다 SENet을 이용하여 channel attention을 적용한다.

**3.3. Rotation Branch**

기존의 RPN network들은 NMS를 사용하여 연산량을 줄여서 연산속도의 증가를 가져왔습니다. Training stage 에서는 12000 🡪 2000 proposal 그리고 test stage에서는 10000 🡪 300 proposal 입니다. [-pi/2, 0)의 영역에서 회전을 합니다.

**3.4. Loss Function**

Loss function은 다음과 같습니다. 기존의 smooth l1 loss(smooth l1 loss는 l1 loss보다 loss가 작기 때문에 낮은 loss 값에서는 빠르게 수렴한다는 장점이 있습니다.)는 비슷한 boundary value를 가지고 있더라도 angle difference가 크면 loss가 커져서 norm을취하고 log(IoU)값을 곱해줘서 훨씬 안정적으로 학습이 될 수 있게 수정을 해주었습니다. 또한 t’\_n은 foreground or background value인데 background인 경우 t’\_n=0이 되고 이 loss의 경우 학습이 진행되지 않습니다.