Learning RoI Transformer for Detecting

Oriented Objects in Aerial Images

# 0. Abstract

Bird-view object detection에서 densely and oriented object는 기존의 horizontal proposal로는 RoI랑 실제 object와의 괴리로 인한 문제가 존재했다. 이를 해결하기 위해 RRoI가 제안되었으나 RRoI는 computational complexity를 증가시키는 문제점이 존재했다. 이를 해결하기위해 저자는 RoI Transformer를 제안한다. RRoI learner는 HRoI를 RRoI로 바꾸는 model로 고안되었다. 이러한 RRoI learner를 기반으로 rotation invariant feature를 추출하는 Rotated Position Sensitive RoI Align module를 만들어서 classification 과 regression에서 성능을 올렸다.

저자가 제시한 RoI Transformer는 deformable Position Sensitive RoI pooling의 성능을 압도했다.

# 1. Introduction

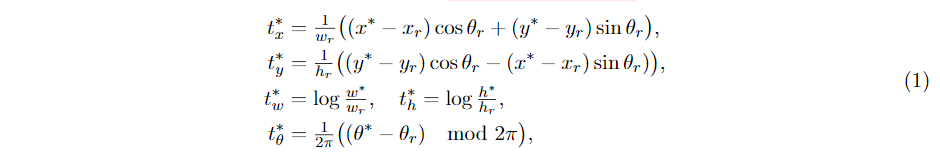
Bird-view object detection에서는 object들이 무작위로 방향이 정렬되어 있다. 또한, background 또한 복잡하고 object의 다양한 형태는 object detection을 힘들게 한다. 이러한 문제를 해결하는 task를 저자는 “*oriented and densely packed detection task”*라고 부른다. 기존의 HRoI technique의 문제점은 위에서 기술하였고 RRoI technique의 문제점은 high recall을 달성하기 위해서는 RRoI를 상당히 많은 양을 generation을 해야한다.

# 2. Related Work

# 3. RoI Transformer

## 3.1 RRoI Learner

이상적인 상황(?)에서는 HRoI는 RRoI의 바깥 상자 이므로 fully connected layer를 이용하여 feature map 로부터 RRoI의 geometry를 추론하려고 합니다. Regression target은 다음과 같습니다.

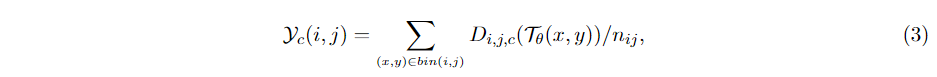


은 RRoI를 나타내는 vector이며 는 oriented bounding box의 ground truth parameters입니다. RRoI learner의 full connected layer 는 모든 feature map 를 입력으로 받아서 ***t***를 예측을 합니다. 를 training 하는 동안 계산의 효율성을 위해 HRoI와 bbox를 matching 시킵니다. HRoI가 matching이 되면 를 (1)의 식을 통해 구해줍니다. Loss function은 smooth L1 loss를 사용합니다. Forward processing을 통해 가 구해지면 우리는 offset으로부터 RRoI를 decoding합니다.

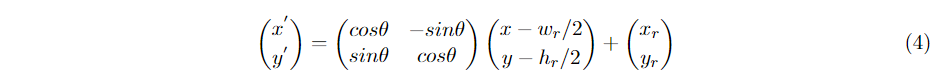
## 3.2 Rotated Position Sensitive RoI Align

해당 module은 간단하게 설명을 하면 R-FCN의 PS RoI pooling과 RoI Align을 합친 module이다.

RPS RoI pooling은 RRoI를 개의 bin으로 나누고 ()를 output으로 가집니다. 전체적인 식은 다음과 같습니다.

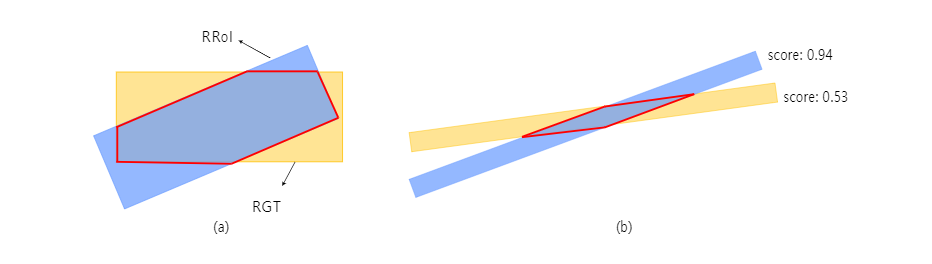


는 feature map의 결과 feature map입니다. Channel mapping은 기존의 PS RoI pooling이랑 동일합니다. 는 bin에서 sampling한 location의 개수입니다. 마지막으로 각 는 에 의해 로 변환됩니다. 식은 다음과 같습니다.

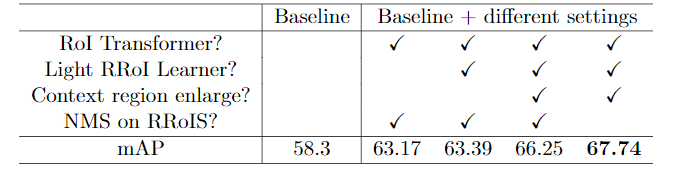


## 3.3 RoI Transformer for Oriented Object Detection

**IoU between OBBs.** True Positive sample을 고를 때와 NMS에 사용된다. 다만 long and thin bounding box에서 약간의 흔들림으로 IoU에 큰 변화를 줌으로 NMS 사용이 쉽지 않을 수가 있다.



# 4 Experiments and Analysis



**Light RRoI Learner.** fc layer에서의 정보손실로 인해 성능의 저하가 발생하였다.

**Contextual RRoI.** HRoI는 너무 많은 background를 포함하여 성능 저하를 일으키지만 반대로 highly enlarged RoI는 contexture information이 너무 없어서 성능을 저하시킨다. 적절한 enlargement는 성능을 향상시킨다.

**NMS on RRoIs**. 구해진 RRoI에 대하여 NMS를 적용시킬지에 대한 논의이다. 더 높은 Recall을 위해서는 추가적인 NMS가 없는 것이 성능이 더 좋게 나왔다.