"R 3Det: Refined Single–Stage Detector with Feature Refinement for Rotating Object".

Link: https://arxiv.org/pdf/1908.05612.pdf

A single shot rotation detector. This paper mainly talks about detecting object in large scale aerial images, containing many different types of objects in different scales.
They mention 3 challenges:

1. Large aspect ratio: The Intersection over Union score is sensitive to change in angle.
2. Densely arranged objects: Aerial images usually contain many densely arranged objects.
3. Category unbalance: Many aerial image datasets categories are extremely unbalanced.

They found that rotating anchors can perform better in dense scenes, while horizontal anchors can achieve higher recalls in fewer quantities. So, they adopt a combination strategy of the two types of anchors. First, they use horizontal anchors to for the faster speed and higher recall, then they use rotating anchors for the refinement stages. In addition, they found that existing refined single–stage detectors suffer from feature misalignment, which limits the reliability of classification and regression during the refinement stages. R 3Det is based on RetinaNet, the refining stage, that can be repeated many times, is added to refine the bounding box, and the FRM is used during the refinement stage for feature map reconstruction.

Feature Refinement Module:

To solve the feature misalignment problem, FRN is introduced. Using feature interpolation to obtain the position information of the refined anchor and reconstructing the feature map to achieve feature alignment. FRM can also reduce the number of refined bounding box after the first stage, thus speeding up the model. FRM contributed in all 3 of the challenges mentioned above. Only the bounding box with the highest score of each feature point is preserved in the refinement stage to increase the speed.

The bounding box:

Every bounding box is represented using 5 parameters (x, y, w, h, θ). (x, y) is the middle point of the bounding box, w is width, h is height. Θ is denotes the acute angle to the x–axis. Therefore, it calls for predicting an additional angular offset in the regression subnet.

To improve the IoU score for large scale object, multiple refinement stages are joined with different IoU thresholds. In addition to using the foreground IoU threshold 0.5 and background IoU threshold 0.4 in the first stage, the thresholds of first refinement stage are set 0.6 and 0.5, respectively. If there are multiple refinement stages, the remaining thresholds are 0.7 and 0.6.

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

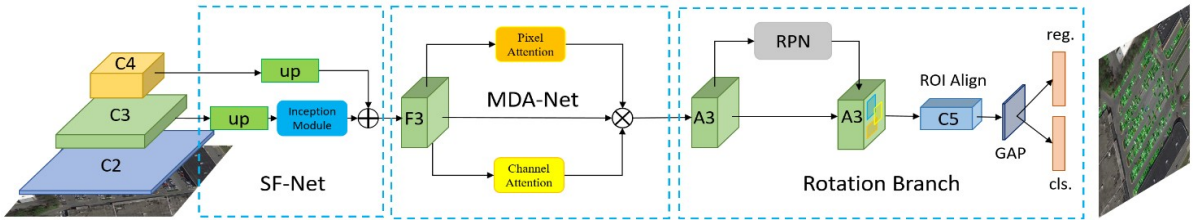
Link: https://arxiv.org/pdf/1811.07126.pdf

This paper presents a novel multi-category rotation detector for small, cluttered and rotated objects, namely SCRDet. The main topic of this article- their approach in the context of remote sensing aerial images, problems in general and, tested with various datasets beyond aerial images.

They mention 3 challenges:

1. Small objects: Aerial images often contain small objects overwhelmed by complex surrounding scenes.
2. Cluttered arrangement: Objects for detection are often densely arranged, such as vehicles and ships.
3. Arbitrary orientations: Objects in aerial images can appear in various orientations.

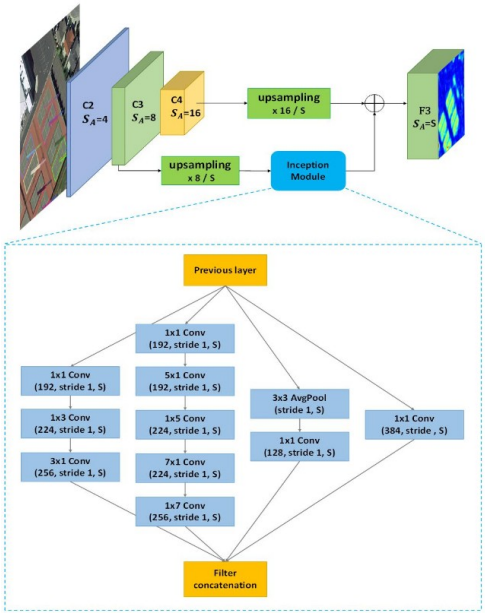
The structure of the network:



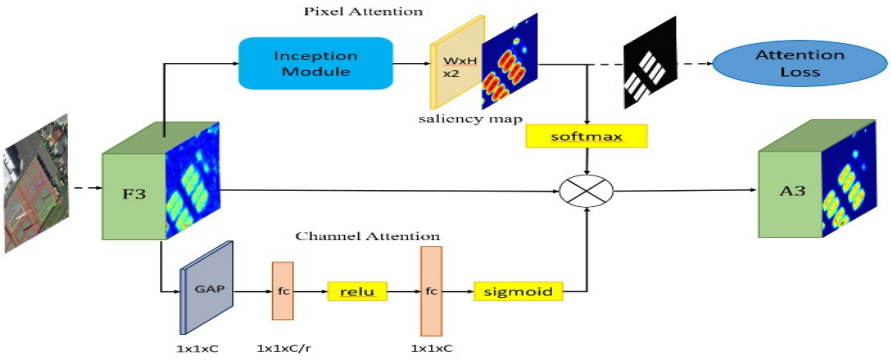
The proposed method include two stage. In the first stage, the feature map is expected to contain more feature information and less noise by adding SF-Net and MDA-Net. In the second stage, can obtain the final detection results under arbitrary rotations by the improved five-parameters regression and the rotation nonmaximum-suppression (R-NMS) operation for each proposal.

finer sam–pling and feature fusion network (SF–Net)

In order to reduce network parameters, SF–Net only uses C3 and C4 in ResNet. The first channel of SF–Net upsample C4, The second channel also upsample the C3 to the same size. Then, C3 pass through an inception structure to expand its receptive field and increase semantic information. Finally, anew feature map F3 is obtained by element–wise addition of the two channels.



Multi–Dimensional Attention Network (MDA–Net)



The goal is to enhance the object cues and weaken the non–object information. The feature map F3 passes through an inception structure with different ratio convolution kernels, and then a two–channels aliency map is learned through a convolution operation. The saliency map represents the scores of the foreground and background, respectively. Then, Softmax operation is performed on the saliency map and one of the channels is selected to multiply with F3. Finally, a new information feature map A3 is obtained.

Rotation Branch

The RPN network provides coarse proposals for the second stage. In order to improve the calculation speed of RPN, they take the highest score of 12,000 regression boxes for NMS operation in the training stage and get 2,000 as proposals. In the test stage, 300 proposals are taken from 10,000 regression boxes by NMS. Every bounding box is represented using 5 parameters (x, y, w, h, θ). (x, y) is the middle point of the bounding box, w is width, h is height. Θ is denotes the acute angle to the x−axis. IoU computation on axis−aligned bounding box may lead to an inaccurate IoU of the skew interactive bounding box and further ruin the bounding box prediction. In order to deal with this, the writers use rotation nonmaximum−suppression (R−NMS) as a post−processing operation based on skew IoU computation.