# Knowledge Combination to Learn Rotated Detection Without Rotated Annotation Supplementary Material

Tianyu Zhu[1, 2,*] Bryce Ferenczi[2] Pulak Purkait[1] Tom Drummond[2]
Hamid Rezatofighi[2] Anton van den Hengel[1]
[1]Amazon IML. [2]Monash University
tianyuzhu52@gmail.com {alanzty, purkaitp, hengelah}@amazon.com
{tianyu.zhu, bryce.ferenczi, tom.drummond, hamid.rezatofighi}@monash.edu

## A. Semantic segmentation experiment

We furthermore verify the advantage of using rotated bounding boxes for downstream task such as semantic segmentation. An experiment was conducted comparing semantic segmentation using crops with rotated ground truth bounding box versus axis-aligned ground truth bounding box. We use a segmentation model deeplab [1]. The results in table C1 show that the model trained using rotated crops significantly outperforms the alternate in both IOU and pixel-wise accuracy. A potential contributing factor to this is that rotated bounding boxes have a tighter object representation.

Table C1. Comparisons of segmentation performance of banana using axis-aligned crops or rotated crops. The result shows the rotated cropping is superior to axis-aligned cropping.

| | Samples | mIoU↑ | Accuracy↑ |
|---|---|---|---|
| axis-aligned | 2000 | 80.83 | 89.4 |
| rotated | 2000 | **86.53** | **92.78** |

## B. Examples of banana images

We include some banana images in Figure B1 to show why rotated detection is superior to axis-aligned detection. It is also clear to see why rotated annotation is more expensive.

## C. KCR design insight

The first instinct we have to solve a weakly-supervised problem for rotated detection is to base the approach on the saliency image derived from the feature map combined with some heuristic extraction to compute the orientation of the box. But such heuristic-based method often suffers from generality. It might work well on one dataset with significant effort, but fails on a different dataset. We have also spent significant effort to use GrabCut [4] to generate reasonable rotated boxes but it is much less reliable and general compared to KCR.

We believe this problem should be solved by a learning process. Consider a human has learnt to draw a rotated bounding box of banana. Now if you tell the person what a cucumber is by simply pointing at it, the person should be able to draw a rotated bounding box enclosing the cucumber. We, as human, can effortless do this because we intrinsically separate task knowledge (skills) and domain knowledge (objects). Inspired by this, we think it is possible, not trivially, to design a learning framework enabling the rotated detector to learn task and domain knowledge from two separate datasets.

## D. Using Grabcut for Rotated Detection

In this section, we further elaborate how we tackle the weakly-supervised rotated detection problem via Grabcut as a heuristic baseline.

The axis-aligned boundary box for the detection is used as the initial ROI for the GrabCut algorithm. The mask is then post-processed with erosion and dilation operations to smooth the boundary of the generated mask. To determine the angle of the tightest boundary box to enclose a ship, we find the pixel coordinate that does not belong to this mask that is nearest the center of the ROI. From this, we can determine the width of the short side of the elongated object and the angle of rotation. Ideally the vector $V_{min}$ from the minimum non-masked coordinate to the center of the ROI is perpendicular to the side of the object $V_{perp}$. Hence, we can find the angle $R_\theta$ of our rotated boundary box $R$ from $V_{perp}$. We can then simply derive the width in equation 2 (where we subtract the extra dilation of the mask) and height in equation 1 with the estimated angle $R_{theta}$ and parameters from the original axis-aligned boundary box $B$, giving us all parameters needed for a tight rotated boundary box $R$.

As observed in figure D2a and D2c, there are scenarios

*Corresponding author.

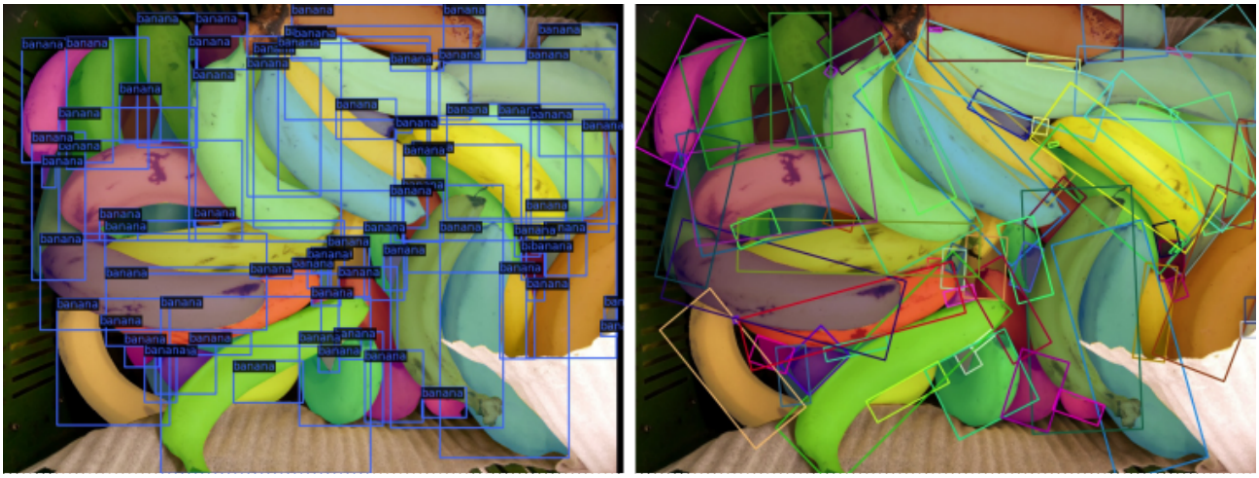

Figure B1. An example of banana dataset. Rotated detection is superior to axis-aligned detection in terms of clarity. We can hardly tell which banana a bounding box is for with axis-aligned detection.

where this method works effectively. It is also noted that in figure D2c, other methods such as using minimum enclosing rectangle on the mask itself would not work due to the over-masking, we would end up with the same axis-aligned bounding box. However, this method is obviously not infallible, if the ROI is not centered on the target properly as observed in D2b, or the mask does not smoothly enclose the object on at least one side, there is a significant error in the sensitive angle calculation. Heuristically, this algorithm produced superior results on HRSC [3] compared with minimum area rectangle that encloses the mask, which is why it was chosen as the alternate baseline for this paper. Furthermore, we demonstrated that using this algorithm as a post-processing step on model predictions results in higher AP than compared to the native axis-aligned detection.

$$R_{height} = 2 * (|V_{min}| - C_{dilation}) \tag{1}$$

$$R_{width} = \frac{max(B_{height}, B_{width})}{max(|\cos(R_\theta)|, |\sin(R_\theta)|)} \tag{2}$$

## E. More Qualitative Visualizations

In this section, we show more visualizations of KCR on test set in Figure E3 and some failure cases in Figure E4.

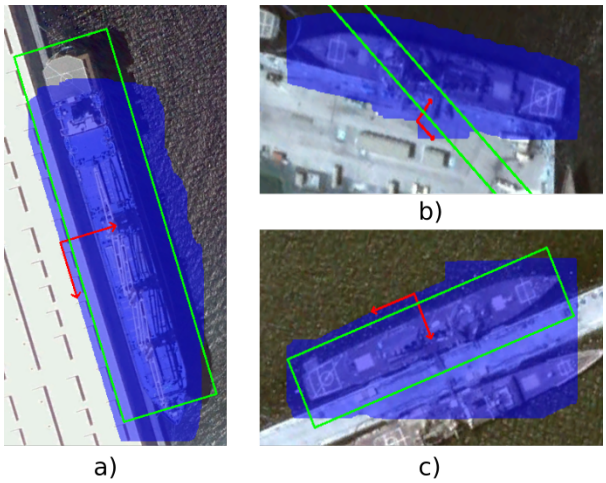


Figure D2. Examples of the GrabCut mask and vectors used to calculate the rotated boundary box. The origin of the vectors show the location of the minimum distance pixel. One vector points to the middle of the ROI and its length determines the short side of the box, and the other is rotated 90 degrees and used for the boundary box angle. The green rotated boundary box is calculated from various parameters.

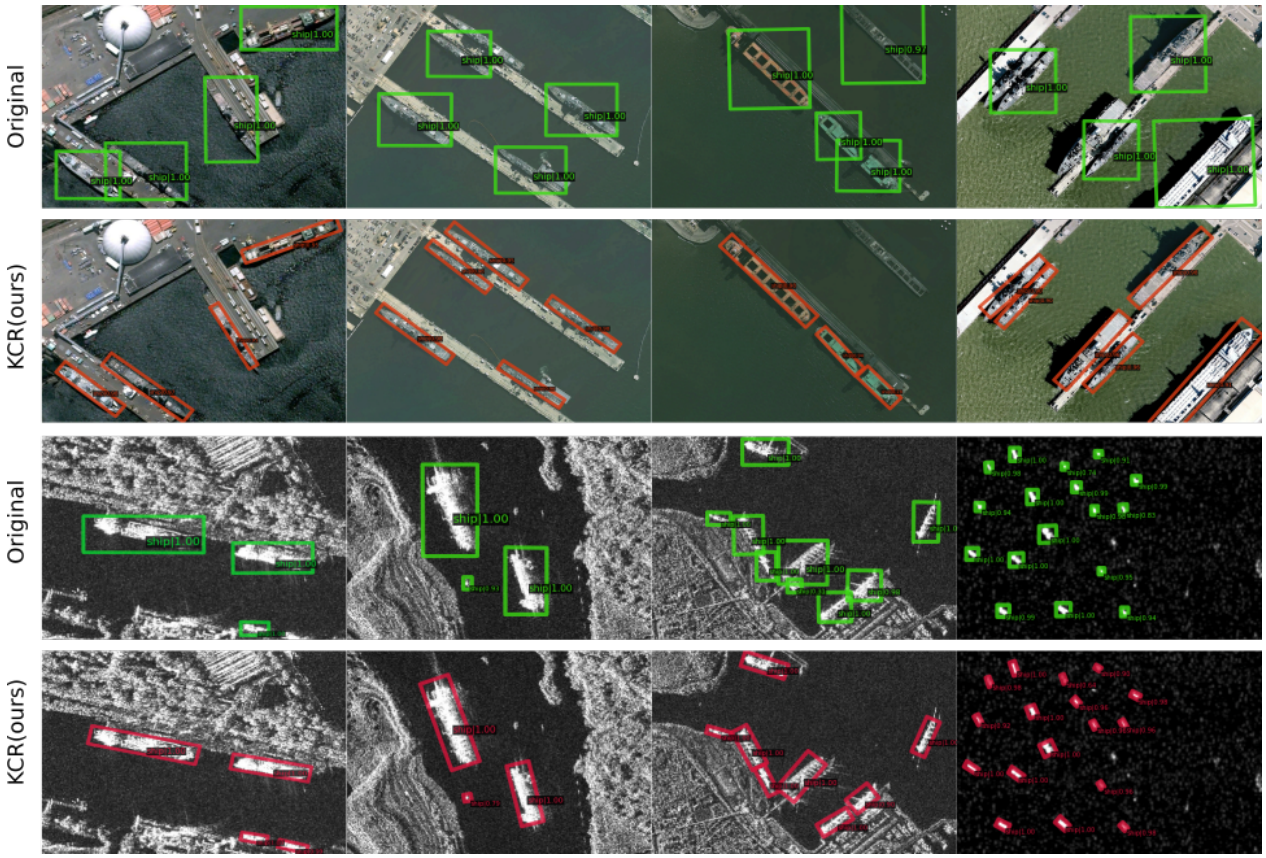


Figure E3. More comparison between original learning and KCR both trained with axis-aligned ground truth of the target dataset. For KCR, we use rotation augmented coco as source datasets. The first two rows are from HRSC [2] test dataset. The last two rows are from SSDD [5] test dataset.


Figure E4. Failure cases of KCR. For the first two cases, some KCR rotated boxes are not tight. But they are still much tighter than axis-aligned boxes. For last two cases, KCR either detects two ships as one or detects the ship trace. Original detector also fails in those two cases.

## References


[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 1

[2] Zikun Liu, Hongzhen Wang, Lubin Weng, and Yiping Yang. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds. *IEEE Geoscience and Remote Sensing Letters*, 13(8):1074–1078, 2016. 3

[3] Zikun Liu, Liu Yuan, Lubin Weng, and Yiping Yang. A high resolution optical satellite image dataset for ship recognition and some new baselines. In *International conference on pattern recognition applications and methods*, volume 2, pages 324–331. SciTePress, 2017. 2

[4] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. " grabcut" interactive foreground extraction using iterated graph cuts. *ACM transactions on graphics (TOG)*, 23(3):309–314, 2004. 1

[5] J. Li T. Zhang, X. Zhang and X. Xu. Sar ship detection dataset (ssdd): Official release and comprehensive data analysis. *Remote Senseing*, 13(18):3690, Sep. 2021. 3