

The KFIoU Loss for Rotated Object Detection

Xue Yang¹ Yue Zhou² Gefan Zhang¹ Jirui Yang³ Wentao Wang¹
 Junchi Yan¹ Xiaopeng Zhang⁴ Qi Tian⁴

Abstract

Differing from the well-developed horizontal object detection area whereby the computing-friendly IoU based loss is readily adopted and well fits with the detection metrics. In contrast, rotation detectors often involve a more complicated loss based on SkewIoU which is unfriendly to gradient-based training. In this paper, we argue that one effective alternative is to devise an approximate loss who can achieve trend-level alignment with SkewIoU loss instead of the strict value-level identity. Specifically, we model the objects as Gaussian distribution and adopt Kalman filter to inherently mimic the mechanism of SkewIoU by its definition, and show its alignment with the SkewIoU at trend-level. This is in contrast to recent Gaussian modeling based rotation detectors e.g. GWD, KLD that involves a human-specified distribution distance metric which requires additional hyperparameter tuning. The resulting new loss called KFIoU is easier to implement and works better compared with exact SkewIoU, thanks to its full differentiability and ability to handle the non-overlapping cases. We further extend our technique to the 3-D case which also suffers from the same issues as 2-D detection. Extensive results on various public datasets (2-D/3-D, aerial/text/face images) with different base detectors show the effectiveness of our approach.

1. Introduction

Rotated object detection is an relatively emerging but challenging area, due to the difficulties of locating the arbitrary-oriented objects and separating them effectively from the background, such as aerial images (Yang et al., 2018a; Ding et al., 2019; Yang et al., 2018b; 2020b; Ming et al., 2021b),

¹Department of Computer Science and Engineering, MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University

²Department of Electronic Engineering, Shanghai Jiao Tong University

³University of Chinese Academy of Sciences

⁴Huawei Inc.. Correspondence to: Junchi Yan <yanjunchi@sjtu.edu.cn>.

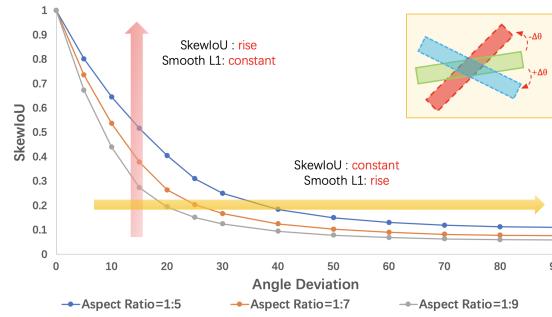


Figure 1. For rotation detection, there is a notable inconsistency between the final detection metric i.e. mAP (largely depending on SkewIoU) and regression-based loss e.g. the popular Smooth L1.

scene text (Jiang et al., 2017; Zhou et al., 2017; Ma et al., 2018). Though considerable progresses have been recently made, for practical settings, there still exist challenges for rotating objects with large aspect ratio, dense distribution.

The Skew Intersection over Union (SkewIoU) score between large aspect ratio objects is sensitive to the deviations of the object positions. This causes the negative impact of the inconsistency between metric (dominated by SkewIoU) and regression loss (e.g. l_n -norms), which is common in horizontal detection, and is further amplified in rotation detection. The red and orange arrows in Fig. 1 show the inconsistency between SkewIoU and Smooth L1 Loss. Specifically, when the angle deviation is fixed (red arrow), SkewIoU will decrease sharply as the aspect ratio increases, while the Smooth L1 loss is unchanged (mainly from the angle difference). Similarly, when SkewIoU does not change (orange arrow), Smooth L1 loss increases as the angle deviation increases. Solution for inconsistency between the metric and regression loss has been extensively discussed in horizontal detection by using IoU loss and related variants, such as GIoU (Rezatofighi et al., 2019) and DIoU (Zheng et al., 2020b). However, the applications of these solutions to rotation detection are blocked because the analytical solution of the SkewIoU calculation process is not easy to be provided due to the complexity of intersection between two rotated boxes (Zhou et al., 2019). Especially, there exist some custom operations (intersection of two edges and sorting the vertexes etc.) whose derivative functions have not been implemented in the existing deep learning frameworks (Abadi et al., 2016; Paszke et al., 2017). Based

on the above analysis, developing an easy-to-implement approximate SkewIoU loss is meaningful and several works (Chen et al., 2020; Zheng et al., 2020a; Yang et al., 2021d;e) have been proposed.

This paper aims to find an easy-to-implement and better-performing alternative. By calculating the error variance and comparing the final performance, we find the key to maintaining the consistency between metric and regression loss (including L1, KLD, GWD and KFIoU etc.) lies in the trend-level consistency between approximate and exact SkewIoU loss rather than value-level consistency. This finding greatly simplifies the difficulty of designing alternatives, and even better ones can be found. Inspired by the above finding, in this paper, we design a novel and effective alternative to SkewIoU loss based on Kalman filter, named KFIoU loss, which can be easily implemented by the existing operations of the deep learning framework without the need for additional acceleration (e.g. C++/CUDA). Specifically, we use convert the rotating bounding box into a Gaussian distribution, which can avoid the well-known boundary discontinuity and square-like problems (Yang & Yan, 2020; Qian et al., 2021a; Ming et al., 2021c; Yang et al., 2021d) in rotation detection. Then we use a center point loss to narrow the distance between the center of the two Gaussian distributions, follow by calculating the overlap area under the new position through Kalman filter. **The highlights of this paper are as follows:**

- 1) For rotation detection, instead of exactly computing the SkewIoU loss which is tedious ¹ and unfriendly to differentiable learning, we propose our new approximate loss – KFIoU loss. It follows the protocol of Gaussian modeling for objects (Yang et al., 2021d;e), yet innovatively uses Kalman filter to mimic SkewIoU’s computing mechanism.
- 2) Compared with plain SkewIoU loss, our KFIoU loss is easy-to-implement, and works better due to fully differentiable and able to handle the non-overlapping cases. Compared to Gaussian-based losses (GWD, KLD) that try to approximate SkewIoU by specifying a distance which requires extra hyperparameters tuning and metric selection, our mechanism level simulation to SkewIoU is more interpretable and natural, and free from hyperparameter tuning.
- 3) We show that KFIoU achieves a trend-level alignment with SkewIoU, where the trend deviation is measured by our devised error variance. While the error variance for GWD and KLD are much larger. The effectiveness of such a trend-level alignment strategy is verified by comparing KFIoU with ideal SkewIoU. On extensive benchmarks (aerial images, scene texts, face), our approach also outperforms other SOTA alternatives (GWD, KLD).

¹As said, the plain SkewIoU loss is difficult to implement and not readily supported by existing mainstream open source deep learning frameworks.

4) We further extend the Gaussian modeling and KFIoU loss from 2-D to 3-D rotation detection, with notable improvement compared with baselines. To our best knowledge, this is the first 3-D rotation detector based on Gaussian modeling which also verifies its effectiveness, which is in contrast to (Yang et al., 2021d;e) focusing on 2-D rotation detection. The source code is available at AlphaRotate ² and MMRotate ³.

2. Related Work

Rotated Object Detection. Rotated object detection is an emerging direction, which attempts to extend classical horizontal detectors (Girshick, 2015; Ren et al., 2015; Lin et al., 2017a;b) to the rotation case by adopting the rotated bounding boxes. Aerial images and scene text are popular application scenarios of rotation detector. For aerial images, objects are often arbitrary-oriented and dense-distributed with large aspect ratios. To this end, ICN (Azimi et al., 2018), ROI-Transformer (Ding et al., 2019), SCRDet (Yang et al., 2019), Mask OBB (Wang et al., 2019a), Gliding Vertex (Xu et al., 2020b), ReDet (Han et al., 2021b) are two-stage mainstreamed approaches whose pipeline is inherited from Faster RCNN (Ren et al., 2015), while DRN (Pan et al., 2020), DAL (Ming et al., 2021d), R³Det (Yang et al., 2021c), RSdet (Qian et al., 2021a;b) and S²A-Net (Han et al., 2021a) are based on single-stage methods for faster detection speed. For scene text detection, RRPn (Ma et al., 2018) employs rotated RPN to generate rotated proposals and further perform rotated bounding box regression. TextBoxes++ (Liao et al., 2018a) adopts vertex regression on SSD (Liu et al., 2016). RRD (Liao et al., 2018b) improves TextBoxes++ by decoupling classification and bounding box regression on rotation-invariant and rotation sensitive features, respectively. The regression loss of the above algorithms are all bounding box or point based or mask-based representation, and they are rarely SkewIoU loss due to the complexity of implementing SkewIoU (Zhou et al., 2019).

Variants of IoU-based Loss. The inconsistency between metric and regression loss is a common problem for both horizontal detection and rotation detection. Solution for this inconsistency has been extensively discussed in horizontal detection by using IoU loss and related variants. For instance, Unitbox (Yu et al., 2016) proposes an IoU loss which regresses the four bounds of a predicted box as a whole unit. More works (Rezatofighi et al., 2019; Zheng et al., 2020b) extend the idea of Unitbox by introducing GIoU (Rezatofighi et al., 2019) and DIoU (Zheng et al., 2020b) for bounding box regression. However, their applications to rotation detection are blocked due to the hard-to-implement

²<https://github.com/yangxue0827/RotationDetection>

³<https://github.com/open-mmlab/mmrotate>

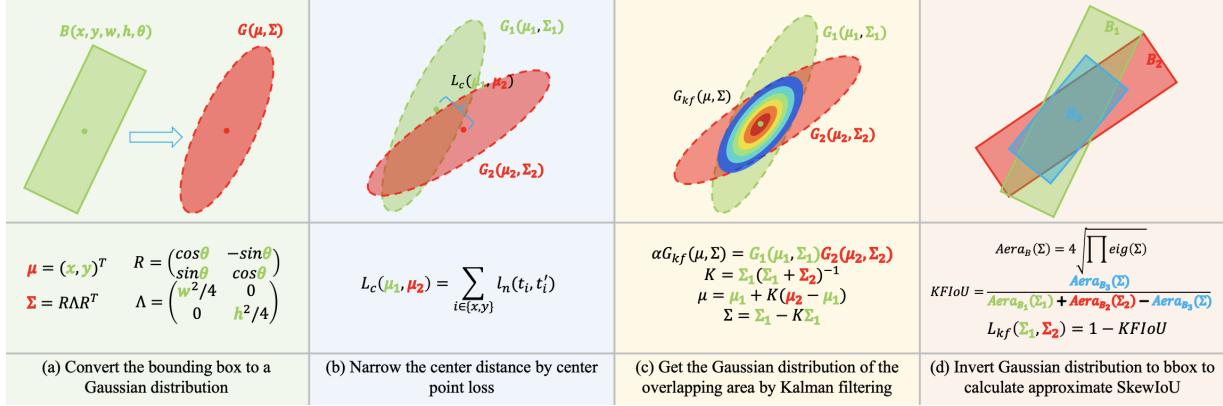


Figure 2. SkewIoU approximation process in two-dimensional space based on Kalman filter. Compared with GWD (Yang et al., 2021d) and KLD (Yang et al., 2021e), our approach follows the calculation process of SkewIoU without introducing additional hyperparameters. We believe such a design is more mathematically rigorous and more in line with SkewIoU approximation.

of the SkewIoU. Recently, some approximate methods for SkewIoU loss have been proposed. **Box/Polygon based:** SCRDet (Yang et al., 2019) propose IoU-Smooth L1, which partly circumvents the need for SkewIoU loss with gradient backpropagation by combining IoU and Smooth L1 loss. To tackle the uncertainty of convex caused by rotation, the work (Zheng et al., 2020a) proposes a projection operation to estimate the intersection area for both 2-D/3-D object detection. PolarMask (Xie et al., 2020) proposes Polar IoU loss that can largely ease the optimization and considerably improve the accuracy. CFA (Guo et al., 2021) proposes convex hull based CIoU loss for optimization of point based detectors. **Pixel based:** PIoU (Chen et al., 2020) calculates the SkewIoU directly by accumulating the contribution of interior overlapping pixels. **Gaussian based:** GWD (Yang et al., 2021d) and KLD (Yang et al., 2021e) simulate SkewIoU by Gaussian distance measurement.

3. Background on Gaussian Modeling

This section presents the preliminary according to (Yang et al., 2021d), for how to convert an arbitrary-oriented 2-D/3-D bounding box to a Gaussian distribution $\mathcal{G}(\mu, \Sigma)$.

$$\Sigma = \mathbf{R}\Lambda\mathbf{R}^\top, \mu = (x, y, (z))^\top \quad (1)$$

where \mathbf{R} represents the rotation matrix, and Λ represents the diagonal matrix of eigenvalues.

For 2-D object $B_{2d}(x, y, h, w, \theta)$,

$$\mathbf{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \Lambda = \begin{pmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{pmatrix} \quad (2)$$

and for 3-D object $B_{3d}(x, y, z, h, w, l, \theta)$,

$$\mathbf{R} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \Lambda = \begin{pmatrix} \frac{w^2}{4} & 0 & 0 \\ 0 & \frac{h^2}{4} & 0 \\ 0 & 0 & \frac{l^2}{4} \end{pmatrix} \quad (3)$$

and l, w, h represent the length, width, and height of the 3-D bounding box, respectively.

It is worth noting that the recent works GWD (Yang et al., 2021d) and KLD (Yang et al., 2021e) also belong to the Gaussian modeling based regression loss. Compared with our work, their difference is that they use the nonlinear transformation of distribution distance to approximate SkewIoU loss. In this process, additional hyperparameters are introduced. Since Gaussian modeling has the natural advantages of being immune to boundary discontinuity and square-like problems, in this paper, we will take another perspective to approximate the SkewIoU loss to better train the detector without any extra hyperparameter, which can be more in line with SkewIoU calculation. Tab. 1 shows the comparison of properties between different losses.

4. Proposed Method

In this section, we present our main approach. Fig. 2 shows the approximate process of SkewIoU loss in two-dimensional space based on Kalman filter. Briefly, we first convert the bounding box to a Gaussian distribution as discussed in Sec. 3, and move the center points of the two Gaussian distributions to make them coincide. Then, the distribution function of the overlapping area is obtained by Kalman filtering. Finally, the obtained distribution function is inverted into a rotating bounding box to calculate the overlapping area and the IoU.

4.1. SkewIoU based on Kalman Filtering

First of all, we can easily calculate the volume of the corresponding rotating box based on its covariance, when we obtain a new Gaussian distribution, where n denotes the number of dimensions.

$$V_B(\Sigma) = 2^n \sqrt{\prod eig(\Sigma)} = 2^n \cdot |\Sigma^{\frac{1}{2}}| = 2^n \cdot |\Sigma|^{\frac{1}{2}} \quad (4)$$

Table 1. Comparison of the properties and performance of different regression losses. Base model is RetinaNet.

LOSS	REGION FITTING	IMPLEMENT	CONTINUITY	SCALE INVARIANCE	CONSISTENCY	HYPERPARAMETER	EVAR [↓]	DOTA-v1.0	DOTA-v1.5	DOTA-v2.0
SMOOTH L1	BBOX	EASY	×	×	×	✓ (σ)	0.073201718	64.17	56.10	43.06
PLAIN SKEWIOU	BBOX	HARD	✓	✓	✓	✗	-	68.27	59.01	45.87
GWD	GAUSSIAN	EASY	✓	✗	✗	✓ (τ)	0.019041297	68.93	60.03	46.65
KLD	GAUSSIAN	EASY	✓	✓	✓	✓ (τ)	0.007653582	71.28	62.50	47.69
KFIoU (OURS)	GAUSSIAN	EASY	✓	✓	✓	✗	0.002348353	70.64	62.71	48.04

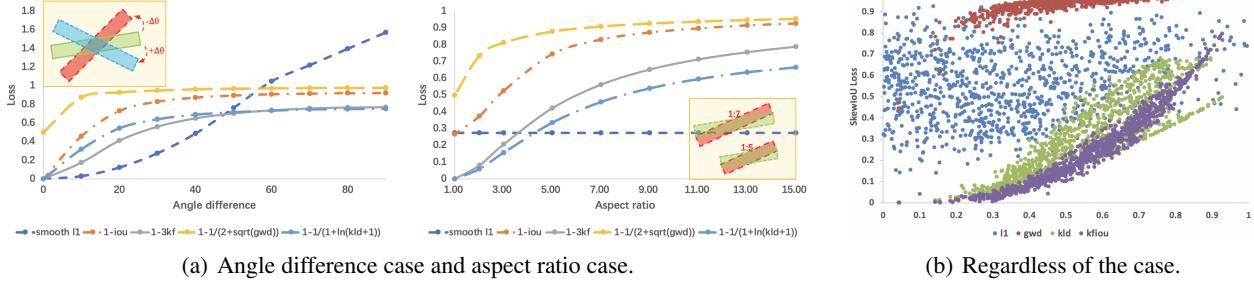


Figure 3. Behavior comparison of different losses in different cases for rotated object detection. (a) Left case depicts the relation between angle difference and loss functions. Right case shows the changes of the five loss under different aspect ratio condition. (b) Scatter plot between approximate losses and SkewIoU loss, 1,000 examples regardless of the case.

To obtain the final SkewIoU, calculating the area of overlap is critical. For two Gaussian distributions, we can use Kalman filter⁴ to get the distribution function of the overlapping area. Specifically:

$$\alpha \mathcal{G}_{kf}(\mu, \Sigma) = \mathcal{G}_1(\mu_1, \Sigma_1)\mathcal{G}_2(\mu_2, \Sigma_2) \quad (5)$$

Note here α is written by:

$$\begin{aligned} \alpha &= \mathcal{G}_\alpha(\mu_2, \Sigma_1 + \Sigma_2) \\ &= \frac{1}{\sqrt{\det(2\pi(\Sigma_1 + \Sigma_2))}} e^{-\frac{1}{2}(\mu_1 - \mu_2)^\top (\Sigma_1 + \Sigma_2)^{-1} (\mu_1 - \mu_2)} \end{aligned} \quad (6)$$

where $\mu = \mu_1 + \mathbf{K}(\mu_2 - \mu_1)$, $\Sigma = \Sigma_1 - \mathbf{K}\Sigma_1$, and \mathbf{K} is the Kalman gain, $\mathbf{K} = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}$.

We observe that Σ is only related to the covariance (Σ_1 and Σ_2) of the given two Gaussian distributions, which means that no matter how the two Gaussian distributions move, as long as the covariance is fixed, the area calculated by Eq. 4 will not change. This is obviously not in line with intuition: the overlapping area should be reduced when the two Gaussian distributions are far away. The main reason is $\alpha \mathcal{G}_{kf}(\mu, \Sigma)$ is not a standard Gaussian distribution (probability sum is not 1), we cannot directly use Σ to calculate the area of the current overlap by Eq. 4 without considering α . It can be found from Eq. 6 that α is related to the distance between the center points ($\mu_1 - \mu_2$) of the two Gaussian distributions. Based on the above findings, we can first use a center point loss $L_c(\mu_1, \mu_2)$ to narrow the distance between the center of the two Gaussian distributions. In this way, α can be approximated as a constant, and the introduction

⁴We model predicted boxes, ground truth boxes, and overlapping regions into predicted values, observed values, and uncertainties, respectively. It should be emphasized that we only borrow the technique of multiplying Gaussian distributions in Kalman filter, and the rest (e.g. iterative process) is not introduced in this paper.

of the $L_c(\mu_1, \mu_2)$ also allows the entire loss to continue to optimize the detector in non-overlapping cases. Then, calculate the overlap area under the new position by Eq. 4. According to Fig. 2, we can easily calculate the KFIoU loss $L_{kf}(\Sigma_1, \Sigma_2)$ when we get the overlap area.

$$\text{KFIoU} = \frac{\mathcal{V}_{\mathcal{B}_3}(\Sigma)}{\mathcal{V}_{\mathcal{B}_1}(\Sigma_1) + \mathcal{V}_{\mathcal{B}_2}(\Sigma_2) - \mathcal{V}_{\mathcal{B}_3}(\Sigma)} \quad (7)$$

where \mathcal{B}_1 , \mathcal{B}_2 and \mathcal{B}_3 refer to the three different bounding boxes shown in the right part of Fig. 2.

In the appendix, we prove that the upper bounds of KFIoU in n-dimensional space is $\frac{1}{2^{\frac{n}{2}} - 1}$. For 2-D/3-D detection, the upper bounds are $\frac{1}{3}$ and $\frac{1}{\sqrt{32} - 1}$ respectively when $n = 2$ and $n = 3$. We can easily stretch the range of KFIoU to $[0, 1]$ by linear transformation according to the upper bound, and then compare it with IoU for consistency. It should be noted that this linear transformation is unnecessary and does not necessarily improve the final performance, because we pay more attention to whether the changing trends of KFIoU and IoU are consistent rather than specific values.

Fig. 3(a) shows the curves of five loss forms for two bounding boxes with the same center in different cases. Note that we have expanded KFIoU by 3 times so that its value range is $[0, 1]$ like SkewIoU. Case 1 (left) depicts the relation between angle difference and loss functions. Though they all bear monotonicity, only Smooth L1 curve is convex while the others are not. Case 2 (right) shows the changes of the five loss under different aspect ratio conditions. It can be seen that the Smooth L1 loss of the two bounding boxes are constant (mainly from the angle difference), but other losses will change drastically as the aspect ratio varies. Regardless of the case in Fig. 3(b), KFIoU can maintain a more consistent trend to SkewIoU than others.

To further explore the behavior of different approximate

SkewIoU losses, we design the metrics of error mean (EMean) and error variance (EVar) as follows:

$$\begin{aligned} EMean &= \frac{1}{N} \sum_{i=1}^N (SkewIoU_{plain} - SkewIoU_{approx}) \\ EVar &= \frac{1}{N} \sum_{i=1}^N (SkewIoU_{approx} - EMean)^2 \end{aligned} \quad (8)$$

where EVar can better measure the trend-level consistency between the designed loss and the plain SkewIoU loss.

Tab. 1 calculates the EVar of different losses in Fig. 3(b). In general, $EVar_{kfiou} < EVar_{kld} < EVar_{gwd} < EVar_{l1}$. In our analysis, this is probably due to the fundamental inconsistency between the distribution distance as used in GWD/KLD and the definition of similarity in SkewIoU. Moreover, for GWD such inconsistency is more pronounced. Because it has no scale invariance and under the same IoU, and a case with a larger scale will get a larger loss value, it can greatly magnify its trend inconsistency with SkewIoU. The results in Tab. 1 also verifies our analysis. In contrast, the calculation process of KFIoU is essentially the calculation of the overlap rate, so it does not require additional hyperparameters and can maintain a high trend-level consistency with SkewIoU.

Combined with the corresponding performance on three datasets, smaller EVars tend to have better performance. Therefore, we come to the conclusion that the key to maintaining the consistency between metric and regression loss lies in the trend-level consistency between approximate and exact SkewIoU loss rather than value-level consistency. The reason why the Gaussian-based losses (e.g. KFIoU, KLD, GWD) outperform the plain SkewIoU loss is due to the advanced parameter optimization mechanism, effective measurement for non-overlapping cases, and complete derivation. However, the introduction of hyperparameters makes KLD and GWD less stable than KFIoU in terms of Evar and performance. Compared with GWD and KLD, which use the distribution distance to approximate SkewIoU, KFIoU is physically more reasonable (in line with the calculation process of SkewIoU) and simpler, as well as empirically more effective than GWD and KLD. In addition, KFIoU implementation is much simpler than plain SkewIoU and can be easily implemented by the existing operations of the deep learning framework.

4.2. The Proposed KFIoU Loss

We take 2-D object detection as the main example for notation brevity, though our experiments further cover the 3-D case (source code will also be released). We use the one-stage detector RetinaNet (Lin et al., 2017b) as the baseline. Rotated rectangle is represented by five parameters (x, y, w, h, θ) . First, we shall clarify that the network has not

changed the output of the original regression branch, that is, it is not directly predicting the parameters of the Gaussian distribution. The whole training process of detector is summarized as follows: i) predict offset $(t_x^*, t_y^*, t_w^*, t_h^*, t_\theta^*)$; ii) decode prediction box; iii) convert prediction box and target ground-truth into Gaussian distribution; iv) calculate L_c and L_{kf} of two Gaussian distributions. Therefore, the inference time remains unchanged.

The regression equation of (x, y, w, h) is as follows:

$$\begin{aligned} t_x &= (x - x_a)/w_a, t_y = (y - y_a)/h_a \\ t_w &= \log(w/w_a), t_h = \log(h/h_a) \\ t_x^* &= (x^* - x_a)/w_a, t_y^* = (y^* - y_a)/h_a \\ t_w^* &= \log(w^*/w_a), t_h^* = \log(h^*/h_a) \end{aligned} \quad (9)$$

where x, y, w, h denote the box's center coordinates, width, height and angle, respectively. x, x_a, x^* are for ground-truth box, anchor box, and predicted box (likewise for y, w, h).

For the regression of θ , we use two forms as the baselines:

i) Direct regression, marked as **Reg.** ($\Delta\theta$). The model directly predicts the angle offset t_θ^* :

$$t_\theta = (\theta - \theta_a) \cdot \pi/180, \quad t_\theta^* = (\theta^* - \theta_a) \cdot \pi/180 \quad (10)$$

ii) Indirect regression, marked as **Reg.*** ($\sin \theta, \cos \theta$). The model predicts two vectors $(t_{\sin \theta}^* \text{ and } t_{\cos \theta}^*)$ to match the two targets from the ground truth $(t_{\sin \theta} \text{ and } t_{\cos \theta})$:

$$\begin{aligned} t_{\sin \theta} &= \sin(\theta \cdot \pi/180), t_{\cos \theta} = \cos(\theta \cdot \pi/180) \\ t_{\sin \theta}^* &= \sin(\theta^* \cdot \pi/180), t_{\cos \theta}^* = \cos(\theta^* \cdot \pi/180) \end{aligned} \quad (11)$$

To ensure that $t_{\sin \theta}^{*2} + t_{\cos \theta}^{*2} = 1$ is satisfied, we will perform the following normalization processing:

$$t_{\sin \theta}^* = \frac{t_{\sin \theta}^*}{\sqrt{t_{\sin \theta}^{*2} + t_{\cos \theta}^{*2}}}, \quad t_{\cos \theta}^* = \frac{t_{\cos \theta}^*}{\sqrt{t_{\sin \theta}^{*2} + t_{\cos \theta}^{*2}}} \quad (12)$$

Indirect regression is a simpler way to avoid boundary discontinuity problem (Yang et al., 2019; Yang & Yan, 2020; Yang et al., 2021d). The multi-task loss is:

$$L_{total} = \lambda_1 \sum_{n=1}^{N_{pos}} L_{reg}(b_n, gt_n) + \frac{\lambda_2}{N} \sum_{n=1}^N L_{cls}(p_n, t_n) \quad (13)$$

where N and N_{pos} indicates the number of all anchors and that of positive anchors. b_n denotes the n -th predicted bounding box, gt_n is the n -th target ground-truth. t_n represents the label of the n -th object, p_n is the n -th probability distribution of classes calculated by sigmoid function. λ_1, λ_2 control the trade-off and are set to $\{0.01, 1\}$. The classification loss L_{cls} is set as the focal loss (Lin et al., 2017b). The regression loss is set by $L_{reg} = L_c + L_{kf}$, where

$$L_c(\mu_1, \mu_2) = \sum_{i \in (x, y)} l_n(t_i, t_i^*), \quad L_{kf}(\Sigma_1, \Sigma_2) = f(\text{KFIoU}) \quad (14)$$

Table 2. Ablation study of different KFIoU loss forms with different detectors on DOTA-v1.0.

METHOD	SMOOTH L1	$-\ln(\text{KFIoU} + \epsilon)$	$1 - \text{KFIoU}$	$e^{1-\text{KFIoU}} - 1$	$e^{1-3\text{KFIoU}} - 1$	$-\ln(3\text{KFIoU} + \epsilon)$
RETINANET	65.73	69.80 (+4.07)	70.19 (+4.46)	70.64 (+4.91)	69.64 (+3.91)	—
R ³ DET	70.66	72.28 (+1.62)	71.09 (+0.43)	71.58 (+0.92)	—	71.77 (+1.11)

 Table 3. Results on KITTI val split 3D detection. [†] indicates our implementation, others are quoted from (Chen et al., 2019).

METHOD	MAP	CAR - 3D DETECTION			PED. - 3D DETECTION			CYC. - 3D DETECTION		
		MOD.	EASY	MOD.	HARD	EASY	MOD.	HARD	EASY	MOD.
POINTPILLARS	59.50	85.90	73.88	67.98	50.17	45.11	41.09	78.66	59.51	56.02
POINTPILLARS [†] + KFIoU	61.34 64.98	85.66 86.45	75.48 76.49	68.39 74.41	55.46 58.11	48.69 54.22	43.71 49.53	79.37 82.68	59.84 64.23	55.92 60.07

 Table 4. Results on KITTI val BEV detection. [†] indicates our implementation, others are quoted from (Chen et al., 2019).

METHOD	MAP	CAR - BEV DETECTION			PED. - BEV DETECTION			CYC. - BEV DETECTION		
		MOD.	EASY	MOD.	HARD	EASY	MOD.	HARD	EASY	MOD.
POINTPILLARS	66.97	90.14	85.27	79.81	57.80	52.53	47.50	79.96	63.10	59.35
POINTPILLARS [†] + KFIoU	68.16 70.91	89.89 89.59	86.97 86.81	79.64 83.21	61.04 63.34	54.94 58.43	49.26 54.80	81.76 84.61	62.56 67.50	60.54 64.52

and $f(\cdot)$ represents the loss concerning KFIoU, such as $-\ln(\text{KFIoU} + \epsilon)$, $1 - \text{KFIoU}$, $e^{1-\text{KFIoU}} - 1$.

5. Experiments

5.1. 2-D Datasets and Implementation Details

Aerial image dataset: DOTA (Xia et al., 2018) is one of the largest datasets for oriented object detection in aerial images with three released versions: DOTA-v1.0, DOTA-v1.5 and DOTA-v2.0. DOTA-v1.0 contains 15 common categories, 2,806 images and 188,282 instances. DOTA-v1.5 uses the same images as DOTA-v1.0, but extremely small instances (less than 10 pixels) are also annotated. Moreover, a new category (CC-container crane), containing 402,089 instances in total is added in this version. While DOTA-v2.0 contains 18 common categories (two new categories: AP-airport and HP-helipad), 11,268 images and 1,793,658 instances. Compared to DOTA-v1.5, it further includes the new categories. The 11,268 images in DOTA-v2.0 are split into training, validation, test-dev, and test-challenge sets. We divide the images into 600×600 subimages with an overlap of 150 pixels and scale it to 800×800 , in line with the popular cropping protocol in literature. **HRSC2016** (Liu et al., 2017) contains images from two scenarios with ships on sea and close inshore. The training, validation and test set include 436,181 and 444 images.

Scene text dataset: ICDAR2015 (Karatzas et al., 2015) includes 1,000 training images and 500 testing images. **MSRA-TD500** (Yao et al., 2012) has 300 training images and 200 testing images. They are popular for oriented scene text detection and spotting.

Face dataset: Fddb (Jain & Learned-Miller, 2010) is a dataset designed for unconstrained face detection, in which faces have a wide variability of face scales, poses, and appearance. This dataset contains annotations for 5,171 faces

in a set of 2,845 images taken from the faces in the Wild dataset (Berg et al., 2005). We manually use 70% as the training set and the rest as the validation set.

We use AlphaRotate (Yang et al., 2021f) for implementation, where many advanced rotation detectors are integrated. Experiments are performed on a server with GeForce RTX 3090 Ti and 24G memory. Experiments are initialized by ResNet50 (He et al., 2016) by default unless otherwise specified. We perform experiments on two aerial benchmarks, two scene text benchmarks and one face benchmark to verify the generality of our techniques. Weight decay and momentum are set 0.0001 and 0.9, respectively. We employ MomentumOptimizer over 4 GPUs with a total of 4 images per mini-batch (1 image per GPU). All the used datasets are trained by 20 epochs, and learning rate is reduced tenfold at 12 epochs and 16 epochs, respectively. The initial learning rate is 1e-3. The number of image iterations per epoch for DOTA-v1.0, DOTA-v1.5, DOTA-v2.0, HRSC2016, IC-DAR2015, MSRA-TD500 and FDDB are 54k, 64k, 80k, 10k, 10k, 5k and 4k respectively, and increases exponentially if data augmentation (i.e. random graying, flipping and rotation) and multi-scale training are enabled.

5.2. 3-D Datasets and Implementation Details

KITTI (Geiger et al., 2012) contains 7,481 training and 7,518 testing samples for 3-D object detection. The training samples are generally divided into the train split (3,712 samples) and the val split (3,769 samples). The evaluation is classified into Easy, Moderate or Hard according to the object size, occlusion and truncation. All results are evaluated by the mean average precision with a rotated IoU threshold 0.7 for cars and 0.5 for pedestrian and cyclists. To evaluate the model’s performance on KITTI val split, we train our model on the train set and report the results on the val set.

We use third-party tools, MMDetection3D (Chen et al.,

Table 5. High-precision detection under different regression losses. ‘R’, ‘F’ and ‘G’ indicate random rotation, flipping, and graying.

METHOD	DATASET	DATA AUG.	REG. LOSS	HMEAN/AP ₅₀	HMEAN/AP ₆₀	HMEAN/AP ₇₅	HMEAN/AP ₈₅	HMEAN/AP _{50:95}
RETINANET	HRSC2016	R+F+G	SMOOTH L1 KFIoU	84.28 84.41 (+0.13)	74.74 82.23 (+7.49)	48.42 58.32 (+9.90)	12.56 18.34 (+5.78)	47.76 51.29 (+3.53)
	MSRA-TD500	R+F	SMOOTH L1 KFIoU	70.98 76.30 (+5.32)	62.42 69.84 (+7.42)	36.73 47.58 (+10.85)	12.56 19.21 (+6.65)	37.89 44.96 (+7.07)
	ICDAR2015	F	SMOOTH L1 KFIoU	69.78 75.90 (+6.12)	64.15 69.28 (+5.13)	36.97 40.03 (+3.06)	8.71 9.18 (+0.47)	37.73 41.17 (+3.44)
	FDDB		SMOOTH L1 KFIoU	95.92 97.25 (+1.33)	87.50 94.89 (+7.39)	55.81 77.38 (+21.57)	12.67 25.62 (+12.93)	52.77 63.25 (+10.48)
	DOTA-v1.0		SMOOTH L1 KFIoU	65.00 67.68 (+2.68)	57.84 62.18 (+4.34)	33.68 37.30 (+3.62)	11.39 14.21 (+2.82)	35.16 38.51 (+3.35)

Table 6. Accuracy (%) comparison on DOTA. The bold red and blue indicate the top two performances. D_{oc} and D_{le} denotes OpenCV Definition ($\theta \in [-90^\circ, 0^\circ]$) and Long Edge Definition ($\theta \in [-90^\circ, 90^\circ]$) of RBox. ‘H’ and ‘R’ denotes the horizontal and rotating anchors, respectively.

METHOD	BOX DEF.	v1.0	v1.5	v2.0
RETINANET-H (REG.) (2017B)	D_{oc}	65.73	58.87	44.16
RETINANET-H (REG.) (2017B)	D_{le}	64.17	56.10	43.06
RETINANET-H (REG.*) (2017B)	D_{le}	65.78	57.17	43.92
RETINANET-R (REG.) (2017B)	D_{oc}	67.25	56.50	42.04
PIOU (2020)	D_{oc}	65.85	57.65	45.23
IOU-SMOOTH L1 (2019)	D_{oc}	66.99	59.16	46.31
MODULATED LOSS (2021A)	D_{oc}	66.05	57.75	45.17
MODULATED LOSS (2021A)	QUAD.	67.20	61.42	46.71
RIL (2021C)	QUAD.	66.06	58.91	45.35
CSL (2020)	D_{le}	67.38	58.55	43.34
DCL (BCL) (2021B)	D_{le}	67.39	59.38	45.46
PLAIN SKEWIoU (2019)	D_{oc}	68.27	59.01	45.87
GWD (2021D)	D_{oc}	68.93	60.03	46.65
KLD (2021E)	D_{oc}	71.28	62.50	47.69
KFIoU (OURS)	D_{oc}	70.64	62.71	48.04

Table 7. Ablation study of training strategies and tricks on DOTA-v1.0. ‘MS’: multi-scale training and testing. ‘R’: random rotation.

METHOD	KFIoU	BACKBONE	SCHED.	MS	R	MAP ₅₀
RETINANET	✓	R-50	12E			65.19
		R-50	12E			69.86
ROI TRANS.	✓	R-50	12E			73.89
		R-50	12E			74.27
	✓	R-50	12E			75.99
		SWIN-TINY	12E			76.17
		SWIN-TINY	12E			77.18
		R-50	24E	✓	✓	80.22
	✓	SWIN-TINY	12E	✓	✓	80.93
R ³ DET	✓	R-50	12E			70.41
		R-50	12E			72.04
	✓	R-50	12E	✓		78.50
		SWIN-TINY	12E	✓		79.35
	✓	R-50	12E	✓	✓	79.73
		R-101	12E	✓	✓	79.83
	✓	SWIN-TINY	12E	✓	✓	80.06
	✓	SWIN-TINY	24E	✓	✓	80.90

2019) and use PointPillar (Lang et al., 2019) as the baseline, and the training schedule inherited from SECOND (Yan et al., 2018): ADAM optimizer with a cosine-shaped cyclic learning rate scheduler that spans 160 epochs. The learning rate starts from 1e-4 and reaches its peak value 1e-3 at the 60th epoch, and then goes down gradually to 1e-7 finally. In the development phase, the experiments are conducted with a single model for 3-class joint detection.

5.3. Ablation Study and Further Comparison

Ablation study of three forms of KFIoU loss on two detectors. We use two different detectors and three different KFIoU based loss functions to verify its effectiveness, as shown in Tab. 2. RetinaNet-based detector will have a large number of low-SkewIoU prediction bounding box in the early stage of training, and will produce very large loss after the log function, which weakens the improvement of the model. Compared with the linear function, the derivative of the exp-based function will pay more attention to the training of difficult samples, so it has a higher performance, at **70.64%**. In contrast, R³Det-based detector can generate high-quality prediction box at the beginning of training by adding refinement stages, so it will not suffer the same troubles as RetinaNet. Due to the same mechanism of focusing on difficult samples, log and exp-based functions are both better than linear functions, and the best performance is achieved on the log-based function, about **72.28%**. We also expand KFIoU by 3 times to make its range truly consistent with the IoU loss, at [0, 1]. However, this consistency do not bring any additional gains, so the following experiments are all use the KFIoU before non-expansion.

Ablation study of KFIoU loss on 3-D detection. We generalize the KFIoU loss from 2-D to 3-D detection, with results in Tab. 3 and Tab. 4. It involves 3-D detection and BEV detection on KITTI val split, and significant performance improvements are also achieved. On the moderate level of 3-D detection, KFIoU loss improves PointPillars[†] by **3.64%**. On the moderate level of BEV detection, KFIoU loss achieves gains of **2.75%**, at **70.91%**.

High-precision detection experiment. Tab. 5 compares Smooth L1 loss and KFIoU loss by high-precision detection indicators. For HRSC2016 containing a large number of ships with large aspect ratios, KFIoU has a **9.90%** improvement over Smooth L1 on AP₇₅. For the scene text datasets MSRA-TD500 and ICDAR2015, KFIoU achieves **7.07%** and **3.44%** improvements on Hmean_{50:95}, reaching **44.96%** and **41.17%** respectively. For face dataset FDDB and aerial dataset DOTA-v1.0, KFIoU still maintains an advantage.

Comparison with peer methods. Methods in Tab. 6 are based on the same baseline RetinaNet, and initialized by

Table 8. AP of different objects on DOTA-v1.0. R-101 denotes ResNet-101 (likewise for R-50, R-152). RX-101 and H-104 denotes ResNeXt101 (Xie et al., 2017) and Hourglass-104 (Newell et al., 2016). **Red** and **blue**: top two performances.

	METHOD	BACKBONE	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	MAP ₅₀
SINGLE-STAGE	PIoU (2020)	DLA-34	80.90	69.70	24.10	60.20	38.30	64.40	64.80	90.90	77.20	70.40	46.50	37.10	57.10	61.90	64.00	60.50
	O ² -DNET (2020)	H-104	89.31	82.14	47.33	61.21	71.32	74.03	78.62	90.76	82.23	81.36	60.93	60.17	58.21	66.98	61.03	71.04
	DAL (2021d)	R-101	88.61	79.69	46.27	70.37	65.89	76.10	78.53	90.84	79.98	78.41	58.71	62.02	69.23	71.32	60.65	71.78
	P-RSDET (2020)	R-101	88.58	77.83	50.44	69.29	71.10	75.79	78.66	90.88	80.10	81.71	57.92	63.03	66.30	69.77	63.13	72.30
	BBAVECTORS (2021)	R-101	88.35	79.96	50.69	62.18	78.43	78.95	87.94	90.85	83.58	84.35	54.13	60.24	65.22	64.28	55.70	72.32
	DRN (2020)	H-104	89.71	82.34	47.22	64.10	76.22	74.43	85.84	90.57	86.18	84.89	57.65	61.93	69.30	69.63	58.48	73.23
	DCL (2021b)	R-152	89.10	84.13	50.15	73.57	71.48	58.13	78.00	90.89	86.64	86.78	67.97	67.25	65.63	74.06	67.05	74.06
	POLARDET (2021)	R-101	89.65	87.07	48.14	70.97	78.53	80.34	87.45	90.76	85.63	86.87	61.64	70.32	71.92	73.09	67.15	76.64
	GWD (2021d)	R-152	86.96	83.88	54.36	77.53	74.41	68.48	80.34	86.62	83.41	85.55	73.47	67.77	72.57	75.76	73.40	76.30
	KFIoU (OURS)	R-152	89.46	85.72	54.94	80.37	77.16	69.23	80.90	90.79	87.79	86.13	73.32	68.11	75.23	71.61	69.49	77.35
REFINE-STAGE	CFC-NET (2021a)	R-101	89.08	80.41	52.41	70.02	76.28	78.11	87.21	90.89	84.47	85.64	60.51	61.52	67.82	68.02	50.09	73.50
	R ³ DET (2021c)	R-152	89.80	83.77	48.11	66.77	78.76	83.27	87.84	90.82	85.38	85.51	65.67	62.68	67.53	78.56	72.62	76.47
	CFA (GUO ET AL., 2021)	R-152	89.08	83.20	54.37	66.87	81.23	80.90	87.17	90.21	84.32	86.09	52.34	69.94	75.52	80.76	67.96	76.67
	DAL (2021d)	R-50	89.69	83.11	55.03	71.00	78.30	81.90	88.46	90.89	84.97	87.46	64.41	65.65	76.86	72.09	64.35	76.95
	DCL (2021b)	R-152	89.26	83.60	53.54	72.76	79.04	82.56	87.31	90.67	86.59	86.98	67.49	66.88	73.29	70.56	69.99	77.37
	RIDET (2021c)	R-50	89.31	80.77	54.07	76.38	79.81	81.99	89.13	90.72	83.58	87.22	64.42	67.56	78.08	79.17	62.07	77.62
	S ² A-Net (2021a)	R-50	88.89	83.60	57.74	81.95	79.94	83.19	89.11	90.78	84.87	87.81	70.30	68.25	78.30	77.01	69.58	79.42
	GWD (2021d)	R-152	89.66	84.99	59.26	82.19	78.97	84.83	87.70	90.21	86.54	86.85	73.47	67.77	76.92	79.22	74.92	80.23
	KLD (2021e)	R-152	89.92	85.13	59.19	81.33	78.82	84.38	87.50	89.80	87.33	87.00	72.57	71.35	77.12	79.34	78.68	80.63
	R ³ DET-KFIoU (OURS)	SWIN-TINY	89.50	84.26	59.90	81.06	81.74	85.45	88.77	90.85	87.03	87.79	70.68	74.31	78.17	81.67	72.37	80.90
TWO-STAGE	ICN (2018)	R-101	81.40	74.30	47.70	70.30	64.90	67.80	70.00	90.80	79.10	78.20	53.60	62.90	67.00	64.20	50.20	68.20
	ROI-TRANS. (2019)	R-101	88.64	78.52	43.44	75.92	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
	SCRDETR (2019)	R-101	89.98	80.65	52.09	68.36	68.36	60.32	72.41	90.85	87.94	86.86	65.02	66.68	66.25	68.24	65.21	72.61
	GLIDING VERTEX (2020b)	R-101	89.64	85.00	52.26	77.34	73.01	73.14	86.82	90.74	79.02	86.81	59.55	70.91	72.94	70.86	57.32	75.02
	MASK OBB (2019a)	RX-101	89.56	85.95	54.21	72.90	76.52	74.16	85.63	89.85	83.81	86.48	54.89	69.64	73.94	69.06	63.32	75.33
	CENTERMAP (2020b)	R-101	89.83	84.41	54.60	70.25	77.66	78.32	87.19	90.66	84.89	85.27	56.46	69.23	74.13	71.56	66.06	76.03
	CSL (2020)	R-152	90.25	85.53	54.64	75.31	70.44	73.51	77.62	90.84	86.15	86.69	69.60	68.04	73.83	71.10	68.93	76.17
	RSDET-II (2021a)	R-152	89.93	84.45	53.77	74.35	71.52	78.31	78.12	91.14	87.35	86.93	65.64	65.17	75.35	79.74	63.31	76.34
	SCRDETR++ (2020c)	R-101	90.05	84.39	55.44	73.99	77.54	71.11	86.05	90.67	87.32	87.08	69.62	68.90	73.74	71.29	65.08	76.81
	REDET (2021b)	RER-50	88.81	82.48	60.83	80.82	78.34	86.06	88.31	90.87	88.77	87.03	68.65	66.90	79.26	79.71	74.67	80.10
ROI-TRANS.-KFIoU (OURS)	ORIENTED R-CNN (2021)	R-50	89.84	85.43	61.09	79.82	79.71	85.35	88.82	90.88	86.68	87.73	72.21	70.80	82.42	78.18	74.11	80.87
	SWIN-TINY	89.44	84.41	62.22	82.51	80.10	86.07	88.68	90.90	87.32	88.38	72.80	71.95	78.96	74.95	75.27	80.93	

ResNet50 (He et al., 2016) without using data augmentation and multi-scale training/testing. They are trained/tested under the same environment and hyperparameters. These methods are all published solutions to the boundary discontinuity in rotation detection.

First, we conduct ablation experiments on anchor form (horizontal and rotating anchors), rotated bounding box definition form (OpenCV definition and Long Edge definition), and angle regression form (direct regression and indirect regression) based on RetinaNet. Rotating anchors provides accurate prior, which makes the model show strong performance in large aspect ratio objects (e.g. SV, LV, SH). However, the large number of anchors makes it time-consuming. Therefore, we use horizontal anchors by default to balance accuracy and speed. OpenCV definition (D_{oc}) (Yang et al., 2019) and Long Edge definition (D_{le}) (Ma et al., 2018) are two popular methods for defining bounding boxes with different angles. Experiments show that D_{oc} is slightly better than D_{le} on the three versions of DOTA. Angle direct regression (Reg.) always suffers from the standing boundary discontinuity problem as widely studied recently (Yang & Yan, 2020). In contrast, angle indirect regression (Reg^{*}) is a simpler way to avoid above problems and brings performance boost according to Tab. 6.

PIoU calculates the SkewIoU directly by accumulating the contribution of interior overlapping pixels but the effect is not significant. IoU-Smooth L1 partly circumvents the need for SkewIoU loss with gradient backpropagation by combining IoU and Smooth L1 loss. Although IoU-Smooth L1 has achieved an improvement of 1.26%/0.29%/2.15%

on DOTA-v1.0/v1.5/v2.0, the gradient is still dominated by Smooth L1 but still worse than plain SkewIoU loss. Modulated Loss and RIL implement ordered and disordered quadrilateral detection respectively, and the more accurate representation makes them both have a considerable performance improvement. In particular, Modulated Loss achieves the third highest performance on DOTA-v1.5/v2.0. CSL and DCL convert the angle prediction from regression to classification, cleverly eliminating the boundary discontinuity problem caused by the angle periodicity. GWD, KLD and KFIoU loss are three different regression losses based on Gaussian distribution. In contrast, KFIoU loss has a more obvious performance increase due to its scale invariance and a more consistent calculation process with SkewIoU loss.

Training strategies and tricks. In line with many advanced methods implemented with MMDetection (Chen et al., 2019), we also reimplement KFIoU based on them. For ResNet (He et al., 2016), SGD optimizer is adopted with an initial learning rate of 0.0025. The momentum and weight decay are 0.9 and 0.0001, respectively. For Swin Transformer (Liu et al., 2021d), AdamW (Kingma & Ba, 2014; Loshchilov & Hutter, 2018) optimizer is adopted with an initial learning rate of 0.0001. The weight decay is 0.05. In addition, we adopt learning rate warmup for 500 iterations, and the learning rate is divided by 10 at each decay step. Tab. 7 performs ablation experiments on four detectors: RetinaNet (Lin et al., 2017b), S²A-Net (Han et al., 2021a), R³Det (Yang et al., 2021c), and ROI Transformer (Ding et al., 2019). The experimental results prove that KFIoU can stably enhance the performance of the detector. In order to further improve the performance of the model

on DOTA, we verified many training strategies and tricks, including backbone, training schedule, data augmentation and multi-scale training and testing, as shown in Tab. 7.

5.4. Comparison with the State-of-the-Art

Tab. 8 compares recent detectors on DOTA-v1.0, as categorized by single-, refine-, and two-stage based methods. Since different methods use different image resolution, network structure, training strategies and various tricks, we cannot make absolutely fair comparisons. In terms of overall performance, our method has achieved the best performance so far, at around **77.35%/80.90%/80.93%**.

6. Conclusion

We have presented a trend-level consistent approximate to the ideal but gradient-training unfriendly SkewIoU loss for rotation detection, and we call it KFIoU loss as the Kalman filter is adopted to directly mimic the computing mechanism of SkewIoU by definition. This design differs from the distribution distance based losses including GWD and KLD which in our analysis have fundamental difficulty in achieving trend-level alignment with SkewIoU loss. Moreover, KFIoU is easier to implement and works better than plain SkewIoU due to the effective measurement for non-overlapping cases and complete derivation. Experimental results on both 2D and 3D cases, on various datasets, show the effectiveness of our approach.

References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pp. 265–283, 2016.
- Azimi, S. M., Vig, E., Bahmanyar, R., Körner, M., and Reinartz, P. Towards multi-class object detection in unconstrained remote sensing imagery. In *Asian Conference on Computer Vision*, pp. 150–165. Springer, 2018.
- Berg, T. L., Berg, A. C., Edwards, J., and Forsyth, D. A. Who’s in the picture. In *Advances in neural information processing systems*, pp. 137–144, 2005.
- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.
- Cao, L., Zhang, X., Wang, Z., and Ding, G. Multi angle rotation object detection for remote sensing image based on modified feature pyramid networks. *International Journal of Remote Sensing*, 42(14):5253–5276, 2021.
- Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., Dai, J., Wang, J., Shi, J., Ouyang, W., Loy, C. C., and Lin, D. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.
- Chen, L., Liu, C., Chang, F., Li, S., and Nie, Z. Adaptive multi-level feature fusion and attention-based network for arbitrary-oriented object detection in remote sensing imagery. *Neurocomputing*, 451:67–80, 2021a.
- Chen, S.-B., Dai, B.-M., Tang, J., Luo, B., Wang, W.-Q., and Lv, K. A refined single-stage detector with feature enhancement and alignment for oriented objects. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:8898–8908, 2021b.
- Chen, Y., Li, J., Xiao, H., Jin, X., Yan, S., and Feng, J. Dual path networks. In *Advances in Neural Information Processing Systems*, pp. 4467–4475, 2017.
- Chen, Z., Chen, K., Lin, W., See, J., Yu, H., Ke, Y., and Yang, C. Piou loss: Towards accurate oriented object detection in complex environments. In *European Conference on Computer Vision*, pp. 195–211. Springer, 2020.
- Cheng, G., Wang, J., Li, K., Xie, X., Lang, C., Yao, Y., and Han, J. Anchor-free oriented proposal generator for object detection. *arXiv preprint arXiv:2110.01931*, 2021.
- Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., and Wei, Y. Deformable convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 764–773, 2017.
- Ding, J., Xue, N., Long, Y., Xia, G.-S., and Lu, Q. Learning roi transformer for oriented object detection in aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2849–2858, 2019.
- Feng, K., Li, W., Han, J., Pan, F., and Zheng, D. Ts4net: Two-stage sample selective strategy for rotating object detection. *arXiv preprint arXiv:2108.03116*, 2021.
- Feng, P., Lin, Y., Guan, J., He, G., Shi, H., and Chambers, J. Toso: Student’s t distribution aided one-stage orientation target detection in remote sensing images. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4057–4061. IEEE, 2020.
- Fu, K., Chang, Z., Zhang, Y., and Sun, X. Point-based estimator for arbitrary-oriented object detection in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 2020a.

- Fu, K., Chang, Z., Zhang, Y., Xu, G., Zhang, K., and Sun, X. Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161:294–308, 2020b.
- Geiger, A., Lenz, P., and Urtasun, R. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354–3361. IEEE, 2012.
- Girshick, R. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, 2015.
- Guan, Q., Qu, Z., Zeng, M., Shen, J., and Du, J. Cgp box: An effective direction representation strategy for oriented object detection in remote sensing images. *International Journal of Remote Sensing*, 42(17):6666–6687, 2021a.
- Guan, Q., Qu, Z., Zhao, P., Zeng, M., and Liu, J. Dual-det: a fast detector for oriented object detection in aerial images. *International Journal of Remote Sensing*, 42(24): 9542–9564, 2021b.
- Guo, Z., Liu, C., Zhang, X., Jiao, J., Ji, X., and Ye, Q. Beyond bounding-box: Convex-hull feature adaptation for oriented and densely packed object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8792–8801, 2021.
- Han, J., Ding, J., Li, J., and Xia, G.-S. Align deep features for oriented object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2021a.
- Han, J., Ding, J., Xue, N., and Xia, G.-S. Redet: A rotation-equivariant detector for aerial object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2786–2795, 2021b.
- He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
- He, X., Ma, S., He, L., and Ru, L. High-resolution polar network for object detection in remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 2020.
- He, X., Ma, S., He, L., Ru, L., and Wang, C. Learning rotated inscribed ellipse for oriented object detection in remote sensing images. *Remote Sensing*, 13(18):3622, 2021a.
- He, X., Ma, S., He, L., Zhang, F., Liu, X., and Ru, L. Aroa: Attention refinement one-stage anchor-free detector for objects in remote sensing imagery. In *International Conference on Image and Graphics*, pp. 269–279. Springer, 2021b.
- He, Z., Ren, Z., Yang, X., Yang, Y., and Zhang, W. Mead: a mask-guided anchor-free detector for oriented aerial object detection. *Applied Intelligence*, pp. 1–16, 2021c.
- Hou, J.-B., Zhu, X., and Yin, X.-C. Self-adaptive aspect ratio anchor for oriented object detection in remote sensing images. *Remote Sensing*, 13(7):1318, 2021.
- Hou, L., Lu, K., Xue, J., and Hao, L. Cascade detector with feature fusion for arbitrary-oriented objects in remote sensing images. In *2020 IEEE International Conference on Multimedia and Expo*, pp. 1–6. IEEE, 2020.
- Huang, Z., Li, W., Xia, X.-G., and Tao, R. A general gaussian heatmap labeling for arbitrary-oriented object detection. *arXiv preprint arXiv:2109.12848*, 2021.
- Jain, V. and Learned-Miller, E. Fddb: A benchmark for face detection in unconstrained settings. 2010.
- Jiang, Y., Zhu, X., Wang, X., Yang, S., Li, W., Wang, H., Fu, P., and Luo, Z. R2cnn: rotational region cnn for orientation robust scene text detection. *arXiv preprint arXiv:1706.09579*, 2017.
- Karatzas, D., Gomez-Bigorda, L., Nicolaou, A., Ghosh, S., Bagdanov, A., Iwamura, M., Matas, J., Neumann, L., Chandrasekhar, V. R., Lu, S., et al. Icdar 2015 competition on robust reading. In *2015 13th International Conference on Document Analysis and Recognition*, pp. 1156–1160. IEEE, 2015.
- Kim, B., Lee, J., Lee, S., Kim, D., and Kim, J. Tricubenet: 2d kernel-based object representation for weakly-occluded oriented object detection. *arXiv preprint arXiv:2104.11435*, 2021.
- Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., and Beijbom, O. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12697–12705, 2019.
- Lang, S., Ventola, F., and Kersting, K. Dafne: A one-stage anchor-free deep model for oriented object detection. *arXiv preprint arXiv:2109.06148*, 2021.
- Li, C., Xu, C., Cui, Z., Wang, D., Jie, Z., Zhang, T., and Yang, J. Learning object-wise semantic representation for detection in remote sensing imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 20–27, 2019a.
- Li, C., Xu, C., Cui, Z., Wang, D., Zhang, T., and Yang, J. Feature-attentioned object detection in remote sensing

- imagery. In *2019 IEEE International Conference on Image Processing*, pp. 3886–3890. IEEE, 2019b.
- Li, K., Wan, G., Cheng, G., Meng, L., and Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159:296–307, 2020a.
- Li, W. and Zhu, J. Oriented repoints for aerial object detection. *arXiv preprint arXiv:2105.11111*, 2021.
- Li, W., Wei, W., and Zhang, L. Gsdet: Object detection in aerial images based on scale reasoning. *IEEE Transactions on Image Processing*, 30:4599–4609, 2021a.
- Li, W. T., Li, L. W., Li, S. Y., Mou, J. C., and Hei, Y. Q. Efficient vertex coordinate prediction based csp-hourglass net for object obb detection in remote sensing. *IEEE Geoscience and Remote Sensing Letters*, 2021b.
- Li, Y., Huang, Q., Pei, X., Jiao, L., and Shang, R. Radet: Refine feature pyramid network and multi-layer attention network for arbitrary-oriented object detection of remote sensing images. *Remote Sensing*, 12(3):389, 2020b.
- Li, Y., Kong, C., Dai, L., and Chen, X. Single-stage detector with dual feature alignment for remote sensing object detection. *IEEE Geoscience and Remote Sensing Letters*, 2021c.
- Li, Y., Wang, Z., Wang, J., and Wang, P. Sdcdet: Robust remote sensing object detection based on instance segmentation direction correction. In *2021 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, pp. 385–389. IEEE, 2021d.
- Li, Z., Hou, B., Wu, Z., Jiao, L., Ren, B., and Yang, C. Fcosr: A simple anchor-free rotated detector for aerial object detection. *arXiv preprint arXiv:2111.10780*, 2021e.
- Li, Z., Yuan, Y., and Ma, D. One-stage detector from coarse to fine for rotating object of remote sensing. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pp. 5307–5310. IEEE, 2021f.
- Liang, D., Geng, Q., Wei, Z., Vorontsov, D. A., Kim, E. L., Wei, M., and Zhou, H. Anchor retouching via model interaction for robust object detection in aerial images. *arXiv preprint arXiv:2112.06701*, 2021a.
- Liang, D., Wei, Z., Zhang, D., Geng, Q., Zhang, L., Sun, H., Zhou, H., Wei, M., and Gao, P. Learning calibrated-guidance for object detection in aerial images. *arXiv preprint arXiv:2103.11399*, 2021b.
- Liao, M., Shi, B., and Bai, X. Textboxes++: A single-shot oriented scene text detector. *IEEE Transactions on Image Processing*, 27(8):3676–3690, 2018a.
- Liao, M., Zhu, Z., Shi, B., Xia, G.-s., and Bai, X. Rotation-sensitive regression for oriented scene text detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5909–5918, 2018b.
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, 2017a.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, 2017b.
- Lin, Y., Feng, P., and Guan, J. Ienet: Interacting embranchment one stage anchor free detector for orientation aerial object detection. *arXiv preprint arXiv:1912.00969*, 2019.
- Liu, S., Zhang, L., Hao, S., Lu, H., and He, Y. Polar ray: A single-stage angle-free detector for oriented object detection in aerial images. In *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 3124–3132, 2021a.
- Liu, S., Zhang, L., Lu, H., and He, Y. Center-boundary dual attention for oriented object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2021b.
- Liu, T., Wang, Y., Hou, S., Li, W., Guan, J., Zhou, S., and Qin, R. Rba-centernet: Feature enhancement by rotated border alignment for oriented object detection. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, 2021c.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. Ssd: Single shot multibox detector. In *European Conference on Computer Vision*, pp. 21–37. Springer, 2016.
- Liu, Z., Yuan, L., Weng, L., and Yang, Y. A high resolution optical satellite image dataset for ship recognition and some new baselines. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, volume 2, pp. 324–331, 2017.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE International Conference on Computer Vision*, 2021d.
- Llerena, J. M., Zeni, L. F., Kristen, L. N., and Jung, C. Gaussian bounding boxes and probabilistic intersection-over-union for object detection. *arXiv preprint arXiv:2106.06072*, 2021.

- Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2018.
- Lu, D. Oskdet: Towards orientation-sensitive keypoint localization for rotated object detection. *arXiv preprint arXiv:2104.08697*, 2021.
- Lu, J., Li, T., Ma, J., Li, Z., and Jia, H. Sar: Single-stage anchor-free rotating object detection. *IEEE Access*, 8: 205902–205912, 2020.
- Ma, J., Shao, W., Ye, H., Wang, L., Wang, H., Zheng, Y., and Xue, X. Arbitrary-oriented scene text detection via rotation proposals. *IEEE Transactions on Multimedia*, 20 (11):3111–3122, 2018.
- Ma, T., Mao, M., Zheng, H., Gao, P., Wang, X., Han, S., Ding, E., Zhang, B., and Doermann, D. Oriented object detection with transformer. *arXiv preprint arXiv:2106.03146*, 2021.
- Ming, Q., Zhou, Z., Miao, L., Zhang, H., and Li, L. Dynamic anchor learning for arbitrary-oriented object detection. *arXiv preprint arXiv:2012.04150*, 1(2):6, 2020.
- Ming, Q., Miao, L., Zhou, Z., and Dong, Y. Cfc-net: A critical feature capturing network for arbitrary-oriented object detection in remote sensing images. *arXiv preprint arXiv:2101.06849*, 2021a.
- Ming, Q., Miao, L., Zhou, Z., Song, J., and Yang, X. Sparse label assignment for oriented object detection in aerial images. *Remote Sensing*, 13(14):2664, 2021b.
- Ming, Q., Miao, L., Zhou, Z., Yang, X., and Dong, Y. Optimization for arbitrary-oriented object detection via representation invariance loss. *IEEE Geoscience and Remote Sensing Letters*, 2021c.
- Ming, Q., Zhou, Z., Miao, L., Zhang, H., and Li, L. Dynamic anchor learning for arbitrary-oriented object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 2355–2363, 2021d.
- Newell, A., Yang, K., and Deng, J. Stacked hourglass networks for human pose estimation. In *European Conference on Computer Vision*, pp. 483–499. Springer, 2016.
- Pan, X., Ren, Y., Sheng, K., Dong, W., Yuan, H., Guo, X., Ma, C., and Xu, C. Dynamic refinement network for oriented and densely packed object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11207–11216, 2020.
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. Automatic differentiation in pytorch. 2017.
- Pei, L., Cheng, G., Sun, X., Li, Q., Zhang, M., and Miao, S. Multi-scale bidirectional feature fusion for one-stage oriented object detection in aerial images. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pp. 2592–2595. IEEE, 2021.
- Qian, W., Yang, X., Peng, S., Yan, J., and Guo, Y. Learning modulated loss for rotated object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 2458–2466, 2021a.
- Qian, W., Yang, X., Peng, S., Yan, J., and Zhang, X. Rsdet++: Point-based modulated loss for more accurate rotated object detection. *arXiv preprint arXiv:2109.11906*, 2021b.
- Qin, R., Liu, Q., Gao, G., Huang, D., and Wang, Y. Mrdet: A multi-head network for accurate oriented object detection in aerial images. *arXiv preprint arXiv:2012.13135*, 2020.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016.
- Ren, S., He, K., Girshick, R., and Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pp. 91–99, 2015.
- Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 658–666, 2019.
- Song, Q., Yang, F., Yang, L., Liu, C., Hu, M., and Xia, L. Learning point-guided localization for detection in remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 1084–1094, 2020.
- Sun, P., Zheng, Y., Zhou, Z., Xu, W., and Ren, Q. R4det: Refined single-stage detector with feature recursion and refinement for rotating object detection in aerial images. *Image and Vision Computing*, 103:104036, 2020.
- Tang, T., Liu, Y., Zheng, Y., Zhu, X., and Zhao, Y. Rotating objects detection in aerial images via attention denoising and angle loss refining. *DEStech Transactions on Computer Science and Engineering*, (cisnr), 2020.
- Tian, J., Zhang, J., Li, W., Li, J., and Zhuo, L. Structurally re-parameterized rotation detector for arbitrary-oriented objects in high-resolution remote sensing images. *International Journal of Remote Sensing*, 43(1):241–269, 2022.

- Wang, J., Ding, J., Guo, H., Cheng, W., Pan, T., and Yang, W. Mask obb: A semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images. *Remote Sensing*, 11(24):2930, 2019a.
- Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al. Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020a.
- Wang, J., Yang, W., Li, H.-C., Zhang, H., and Xia, G.-S. Learning center probability map for detecting objects in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5):4307–4323, 2020b.
- Wang, J., Yang, L., and Li, F. Predicting arbitrary-oriented objects as points in remote sensing images. *Remote Sensing*, 13(18):3731, 2021a.
- Wang, M., Zhang, X., Yu, C., Guo, T., Gu, J., and Peng, B. Oriented object detection with fine-grained enhancement and angle constraint. In *2021 16th International Conference on Computer Science & Education (ICCSE)*, pp. 752–757. IEEE, 2021b.
- Wang, Y., Zhang, Y., Zhang, Y., Zhao, L., Sun, X., and Guo, Z. Sard: Towards scale-aware rotated object detection in aerial imagery. *IEEE Access*, 7:173855–173865, 2019b.
- Wang, Y., Wang, L., Lu, H., and He, Y. Segmentation based rotated bounding boxes prediction and image synthesizing for object detection of high resolution aerial images. *Neurocomputing*, 388:202–211, 2020c.
- Wang, Y., Zou, X., Zhang, Z., Xu, W., Chen, L., Zhong, S., Yan, L., and Wang, G. Learning oriented remote sensing object detection via naive geometric computing. *arXiv preprint arXiv:2112.00504*, 2021c.
- Wei, H., Zhang, Y., Chang, Z., Li, H., Wang, H., and Sun, X. Oriented objects as pairs of middle lines. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169:268–279, 2020.
- Wu, Q., Xiang, W., Tang, R., and Zhu, J. Bounding box projection for regression uncertainty in oriented object detection. *IEEE Access*, 9:58768–58779, 2021.
- Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., and Zhang, L. Dota: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3974–3983, 2018.
- Xiao, Z., Qian, L., Shao, W., Tan, X., and Wang, K. Axis learning for orientated objects detection in aerial images. *Remote Sensing*, 12(6):908, 2020.
- Xiao, Z., Wang, K., Wan, Q., Tan, X., Xu, C., and Xia, F. A2s-det: Efficiency anchor matching in aerial image oriented object detection. *Remote Sensing*, 13(1):73, 2021.
- Xie, E., Sun, P., Song, X., Wang, W., Liu, X., Liang, D., Shen, C., and Luo, P. Polarmask: Single shot instance segmentation with polar representation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 12193–12202, 2020.
- Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1492–1500, 2017.
- Xie, X., Cheng, G., Wang, J., Yao, X., and Han, J. Oriented r-cnn for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3520–3529, 2021.
- Xu, C., Li, C., Cui, Z., Zhang, T., and Yang, J. Hierarchical semantic propagation for object detection in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6):4353–4364, 2020a.
- Xu, Y., Fu, M., Wang, Q., Wang, Y., Chen, K., Xia, G.-S., and Bai, X. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020b.
- Yan, Y., Mao, Y., and Li, B. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10):3337, 2018.
- Yang, F., Li, W., Hu, H., Li, W., and Wang, P. Multi-scale feature integrated attention-based rotation network for object detection in vhr aerial images. *Sensors*, 20(6):1686, 2020a.
- Yang, G.-Y., Li, X.-L., Martin, R. R., and Hu, S.-M. Sampling equivariant self-attention networks for object detection in aerial images. *arXiv preprint arXiv:2111.03420*, 2021a.
- Yang, X. and Yan, J. Arbitrary-oriented object detection with circular smooth label. In *European Conference on Computer Vision*, pp. 677–694. Springer, 2020.
- Yang, X., Sun, H., Fu, K., Yang, J., Sun, X., Yan, M., and Guo, Z. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. *Remote Sensing*, 10(1):132, 2018a.
- Yang, X., Sun, H., Sun, X., Yan, M., Guo, Z., and Fu, K. Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network. *IEEE Access*, 6:50839–50849, 2018b.

- Yang, X., Yang, J., Yan, J., Zhang, Y., Zhang, T., Guo, Z., Sun, X., and Fu, K. Scrdet: Towards more robust detection for small, cluttered and rotated objects. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 8232–8241, 2019.
- Yang, X., Yan, J., and He, T. On the arbitrary-oriented object detection: Classification based approaches revisited. *arXiv preprint arXiv:2003.05597*, 2020b.
- Yang, X., Yan, J., Yang, X., Tang, J., Liao, W., and He, T. Scrdet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing. *arXiv preprint arXiv:2004.13316*, 2020c.
- Yang, X., Hou, L., Zhou, Y., Wang, W., and Yan, J. Dense label encoding for boundary discontinuity free rotation detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 15819–15829, 2021b.
- Yang, X., Yan, J., Feng, Z., and He, T. R3det: Refined single-stage detector with feature refinement for rotating object. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 3163–3171, 2021c.
- Yang, X., Yan, J., Qi, M., Wang, W., Xiaopeng, Z., and Qi, T. Rethinking rotated object detection with gaussian wasserstein distance loss. In *International Conference on Machine Learning*, 2021d.
- Yang, X., Yang, X., Yang, J., Ming, Q., Wang, W., Tian, Q., and Yan, J. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence. In *Advances in Neural Information Processing Systems*, 2021e.
- Yang, X., Zhou, Y., and Yan, J. Alpharotate: A rotation detection benchmark using tensorflow. *arXiv preprint arXiv:2111.06677*, 2021f.
- Yang, Z., He, K., Zou, F., Cao, W., Jia, X., Li, K., and Jiang, C. Ropdet: real-time anchor-free detector based on point set representation for rotating object. *Journal of Real-Time Image Processing*, 17(6):2127–2138, 2020d.
- Yao, C., Bai, X., Liu, W., Ma, Y., and Tu, Z. Detecting texts of arbitrary orientations in natural images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1083–1090. IEEE, 2012.
- Ye, X., Xiong, F., Lu, J., Zhou, J., and Qian, Y. F3-net: Feature fusion and filtration network for object detection in optical remote sensing images. *Remote Sensing*, 12(24):4027, 2020.
- Yi, J., Wu, P., Liu, B., Huang, Q., Qu, H., and Metaxas, D. Oriented object detection in aerial images with box boundary-aware vectors. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pp. 2150–2159, 2021.
- Yu, F., Wang, D., Shelhamer, E., and Darrell, T. Deep layer aggregation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2403–2412, 2018.
- Yu, J., Jiang, Y., Wang, Z., Cao, Z., and Huang, T. Unitbox: An advanced object detection network. In *Proceedings of the 24th ACM international conference on Multimedia*, pp. 516–520, 2016.
- Yu, X., Lin, M., Lu, J., and Ou, L. Oriented object detection in aerial images based on area ratio of parallelogram. *arXiv preprint arXiv:2109.10187*, 2021.
- Zand, M., Etemad, A., and Greenspan, M. Oriented bounding boxes for small and freely rotated objects. *IEEE Transactions on Geoscience and Remote Sensing*, 2021.
- Zhang, C., Xiong, B., Li, X., Zhang, J., and Kuang, G. Learning higher quality rotation invariance features for multioriented object detection in remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:5842–5853, 2021a.
- Zhang, F., Wang, X., Zhou, S., and Wang, Y. Dardet: A dense anchor-free rotated object detector in aerial images. *IEEE Geoscience and Remote Sensing Letters*, 2021b.
- Zhang, G., Lu, S., and Zhang, W. Cad-net: A context-aware detection network for objects in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):10015–10024, 2019.
- Zhang, K., Zeng, Q., and Yu, X. Rosd: Refined oriented staged detector for object detection in aerial image. *IEEE Access*, 9:66560–66569, 2021c.
- Zhang, S., Shao, Z., Huang, X., Bai, L., and Wang, J. An internal-external optimized convolutional neural network for arbitrary orientated object detection from optical remote sensing images. *Geo-spatial Information Science*, pp. 1–12, 2021d.
- Zhang, T., Zhang, X., Zhu, P., Chen, P., Tang, X., Li, C., and Jiao, L. Foreground refinement network for rotated object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2021e.
- Zhang, Y., Li, H., Wang, R., Zhang, M., and Hu, X. Constrained-siou: A metric for horizontal candidates in multi-oriented object detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021f.

- Zhao, P., Qu, Z., Bu, Y., Tan, W., and Guan, Q. Polardet: A fast, more precise detector for rotated target in aerial images. *International Journal of Remote Sensing*, 42(15): 5821–5851, 2021.
- Zheng, X., Zhang, W., Huan, L., Gong, J., and Zhang, H. Apronet: Detecting objects with precise orientation from aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181:99–112, 2021a.
- Zheng, Y., Zhang, D., Xie, S., Lu, J., and Zhou, J. Rotation-robust intersection over union for 3d object detection. In *European Conference on Computer Vision*, pp. 464–480. Springer, 2020a.
- Zheng, Y., Sun, P., Zhou, Z., Xu, W., and Ren, Q. Adt-det: Adaptive dynamic refined single-stage transformer detector for arbitrary-oriented object detection in satellite optical imagery. *Remote Sensing*, 13(13):2623, 2021b.
- Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., and Ren, D. Distance-iou loss: Faster and better learning for bounding box regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 12993–13000, 2020b.
- Zhong, B. and Ao, K. Single-stage rotation-decoupled detector for oriented object. *Remote Sensing*, 12(19):3262, 2020.
- Zhou, D., Fang, J., Song, X., Guan, C., Yin, J., Dai, Y., and Yang, R. IoU loss for 2d/3d object detection. In *2019 International Conference on 3D Vision*, pp. 85–94. IEEE, 2019.
- Zhou, L., Wei, H., Li, H., Zhao, W., Zhang, Y., and Zhang, Y. Arbitrary-oriented object detection in remote sensing images based on polar coordinates. *IEEE Access*, 8: 223373–223384, 2020.
- Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., and Liang, J. East: an efficient and accurate scene text detector. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5551–5560, 2017.
- Zhu, Y., Du, J., and Wu, X. Adaptive period embedding for representing oriented objects in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10): 7247–7257, 2020.
- Zou, F., Xiao, W., Ji, W., He, K., Yang, Z., Song, J., Zhou, H., and Li, K. Arbitrary-oriented object detection via dense feature fusion and attention model for remote sensing super-resolution image. *Neural Computing and Applications*, pp. 1–14, 2020.
- Zou, M., Hu, Z., Guan, Y., Gan, Z., Guan, C., and Leng, S. Feature adaption with predicted boxes for oriented object detection in aerial images. In *Pacific Rim International Conference on Artificial Intelligence*, pp. 366–378. Springer, 2021.

A. Proof of KFIoU Upper Bound

For an n-dimensional Gaussian distribution, its volume is:

$$\mathcal{V} = 2^n \cdot |\Sigma|^{\frac{1}{2}} = 2^n \cdot |\Sigma|^{\frac{1}{2}} \quad (15)$$

For Σ_{kf} , we have

$$\begin{aligned} |\Sigma_{kf}| &= |\Sigma_1 - \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}\Sigma_1| \\ &= |\Sigma_1(\Sigma_1 + \Sigma_2)^{-1}\Sigma_2| \\ &= \frac{|\Sigma_1| \cdot |\Sigma_2|}{|\Sigma_1 + \Sigma_2|} \end{aligned} \quad (16)$$

According to Minkowski's inequality:

$$|\Sigma_1 + \Sigma_2|^{\frac{1}{n}} \geq |\Sigma_1|^{\frac{1}{n}} + |\Sigma_2|^{\frac{1}{n}} \quad (17)$$

Simultaneous mean inequalities:

$$|\Sigma_1 + \Sigma_2|^{\frac{1}{n}} \geq |\Sigma_1|^{\frac{1}{n}} + |\Sigma_2|^{\frac{1}{n}} \geq 2 \cdot |\Sigma_1|^{\frac{1}{2n}} \cdot |\Sigma_2|^{\frac{1}{2n}} \quad (18)$$

Thus:

$$\begin{aligned} \frac{|\Sigma_1|^{\frac{1}{2n}} \cdot |\Sigma_2|^{\frac{1}{2n}}}{|\Sigma_1 + \Sigma_2|^{\frac{1}{n}}} &\leq \frac{1}{2} \\ \frac{|\Sigma_1|^{\frac{1}{2}} \cdot |\Sigma_2|^{\frac{1}{2}}}{|\Sigma_1 + \Sigma_2|} &\leq \frac{1}{2^n} \end{aligned} \quad (19)$$

and

$$\begin{aligned} |\Sigma_{kf}| &= \frac{|\Sigma_1| \cdot |\Sigma_2|}{|\Sigma_1 + \Sigma_2|} \leq \frac{|\Sigma_1|^{\frac{1}{2}} \cdot |\Sigma_2|^{\frac{1}{2}}}{2^n} \\ |\Sigma_{kf}|^{\frac{1}{2}} &\leq \frac{|\Sigma_1|^{\frac{1}{4}} \cdot |\Sigma_2|^{\frac{1}{4}}}{2^{\frac{n}{2}}} \end{aligned} \quad (20)$$

Combine the mean inequalities again:

$$|\Sigma_{kf}|^{\frac{1}{2}} \leq \frac{|\Sigma_1|^{\frac{1}{4}} \cdot |\Sigma_2|^{\frac{1}{4}}}{2^{\frac{n}{2}}} \leq \frac{|\Sigma_1|^{\frac{1}{2}} + |\Sigma_2|^{\frac{1}{2}}}{2^{\frac{n}{2}+1}} \quad (21)$$

According to Eq. 15, we have

$$\mathcal{V}_{kf} \leq \frac{\mathcal{V}_1 + \mathcal{V}_2}{2^{\frac{n}{2}+1}} \quad (22)$$

Therefore, the upper bound of KFIoU is

$$\text{KFIoU} = \frac{\mathcal{V}_{kf}}{\mathcal{V}_1 + \mathcal{V}_2 - \mathcal{V}_{kf}} \leq \frac{1}{2^{\frac{n}{2}+1} - 1} \quad (23)$$

When $n = 2$ and $n = 3$, the upper bounds are $\frac{1}{3}$ and $\frac{1}{\sqrt{32}-1}$ respectively.

B. Supplementary Experiment

In this section, we reimplement KFIoU based on the more powerful benchmark, MMDetection (Chen et al., 2019). We use a single GeForce RTX 3090 Ti with a total batch size of 2 for training. For ResNet (He et al., 2016), SGD optimizer is adopted with an initial learning rate of 0.0025. The momentum and weight decay are 0.9 and 0.0001, respectively. For Swin Transformer (Liu et al., 2021d), AdamW (Kingma & Ba, 2014; Loshchilov & Hutter, 2018) optimizer is adopted with an initial learning rate of 0.0001. The weight decay is 0.05. In addition, we adopt learning rate warmup for 500 iterations, and the learning rate is divided by 10 at each decay step.

Tab. 9 performs ablation experiments on four detectors: RetinaNet (Lin et al., 2017b), S²A-Net (Han et al., 2021a), R³Det (Yang et al., 2021c), and RoI Transformer (Ding et al., 2019). The experimental results prove that KFIoU can stably enhance the performance of the detector. In order to further improve the performance of the model on DOTA, we verified many

The KFIoU Loss for Rotated Object Detection

Table 9. Ablation study of training strategies and tricks. Rotate and MS indicate rotation augmentation and multi-scale training and testing.

METHOD	KFIoU	BACKBONE	SCHED.	MS	ROTATE	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	MAP ₅₀
RETINANET (2017B)	✓	R-50	12E			87.76	72.61	43.86	66.61	69.70	56.61	74.15	90.86	75.27	79.09	47.81	64.60	58.93	63.37	26.58	65.19
		R-50	12E			88.90	80.68	47.12	70.40	72.20	62.49	74.84	90.91	79.63	79.73	58.54	66.40	63.67	67.13	45.33	69.86
S ² A-NET (2021A)	✓	R-50	12E			89.18	79.35	49.11	72.97	79.08	78.03	86.67	90.91	85.90	85.04	64.06	65.64	66.71	67.60	48.08	73.89
		R-50	12E			89.24	83.46	51.44	70.88	78.70	76.31	86.90	90.90	82.22	84.81	61.67	66.93	65.62	67.99	57.05	74.27
ROI TRANS. (2019)	✓	R-50	12E			89.02	81.71	53.84	71.65	79.00	77.76	87.85	90.90	87.04	85.70	61.73	64.55	75.06	71.71	62.38	75.99
		R-50	12E			89.08	82.62	53.90	71.78	78.73	77.91	87.97	90.90	86.68	85.37	63.17	67.65	74.30	71.19	61.35	76.17
	✓	SWIN-TINY	12E			88.96	82.81	53.34	76.55	78.66	83.54	88.00	90.90	86.95	86.47	41.94	64.17	76.29	72.87	63.95	77.18
	✓	SWIN-TINY	12E			88.9	83.77	53.98	77.63	78.83	84.22	88.15	90.91	87.21	86.14	67.79	65.73	75.80	73.68	63.30	77.74
	✓	R-50	24E	✓	✓	89.12	84.54	60.73	78.86	79.65	85.79	88.45	90.90	87.03	88.28	69.15	70.28	78.88	81.54	70.05	80.22
	✓	SWIN-TINY	12E	✓	✓	89.44	84.41	62.22	82.51	80.10	86.07	88.68	90.90	87.32	88.38	72.80	71.95	78.96	74.95	75.27	80.93
R ³ DET (2021C)	✓	R-50	12E			89.02	74.52	47.93	69.64	77.02	74.07	82.56	90.90	79.39	83.67	59.02	62.51	63.56	65.06	37.22	70.41
	✓	R-50	12E			89.06	73.89	49.82	68.39	78.13	75.35	86.65	90.89	82.57	83.84	59.63	62.03	66.16	66.22	47.98	72.04
	✓	R-50	12E	✓		89.06	82.49	55.91	81.04	80.14	83.24	88.56	90.90	84.61	86.83	66.25	71.50	75.60	77.64	63.66	78.50
	✓	SWIN-TINY	12E	✓		89.41	83.66	56.92	79.76	80.45	84.34	88.71	90.91	85.69	87.64	67.69	72.88	76.34	73.63	72.21	79.35
	✓	R-50	12E	✓	✓	89.33	84.19	58.78	81.30	80.48	84.49	88.85	90.84	85.56	87.57	69.14	70.79	77.33	80.82	66.51	79.73
	✓	R-101	12E	✓	✓	89.28	83.32	59.40	80.29	80.43	84.70	88.85	90.87	84.51	87.95	71.86	71.60	78.31	79.42	66.60	79.83
	✓	SWIN-TINY	12E	✓	✓	89.24	83.75	59.77	79.40	80.95	84.61	88.84	90.84	86.86	87.93	71.71	71.17	76.79	77.42	71.59	80.06
	✓	SWIN-TINY	24E	✓	✓	89.50	84.26	59.90	81.06	81.74	85.45	88.77	90.85	87.03	87.79	70.68	74.31	78.17	81.67	72.37	80.90

Table 10. AP of different objects on DOTA-v1.0 test-dev (instance-level based methods). R-101 denotes ResNet-101 (likewise for R-18, R-50, R-152), RX-101, H-104, D-53 and CSPD-53 stand for ResNeXt101 (Xie et al., 2017), Hourglass-104 (Newell et al., 2016), DarkNet53 (Redmon et al., 2016) and CSPDarkNet53 (Bochkovskiy et al., 2020). Other backbone include DPN-92 (Chen et al., 2017), DLA-34 (Yu et al., 2018), DCN (Dai et al., 2017), HRNet-W48 (Wang et al., 2020a).

METHOD	BACKBONE	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	MAP ₅₀
FR-O (2018)	R-101	79.09	69.12	17.17	63.49	34.20	37.16	36.20	89.19	69.60	58.96	49.4	52.52	46.69	44.80	46.30	52.93
MARNET (2021)	R-101	88.91	77.91	39.88	71.17	62.79	58.96	66.25	90.87	73.73	79.04	57.57	64.33	62.47	61.64	51.80	67.15
ICN (2018)	R-101	81.40	74.30	47.70	70.30	64.90	67.70	70.00	90.80	79.10	78.20	53.60	62.90	67.00	64.20	50.20	68.20
GSDET (2021A)	R-101	81.12	76.78	40.78	75.89	64.50	58.37	74.21	89.92	79.40	78.83	64.54	63.67	66.04	58.01	52.13	68.28
KARNET (2020)	R-50	89.33	83.55	44.79	71.61	63.05	67.06	69.53	90.47	79.46	77.84	51.04	60.97	65.38	69.46	49.53	68.87
RADET (2020b)	RX-101	79.45	76.99	48.05	65.83	65.46	74.40	68.86	89.70	78.14	74.97	49.92	64.63	66.14	71.58	62.16	69.09
ROI-TRANS. (2019)	R-101	88.64	78.52	43.44	75.92	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
CAD-NET (2019)	R-101	87.8	82.4	49.4	73.5	71.1	63.5	76.7	90.9	79.2	73.3	48.4	60.9	62.0	67.0	62.2	69.9
AOOD (2020)	DPN-92	89.99	81.25	44.50	73.20	69.80	60.33	66.86	89.89	80.99	86.23	64.98	63.88	65.24	68.36	62.13	71.18
CASCADE-FF (2020)	R-152	89.9	80.4	51.7	77.4	68.2	75.2	75.6	90.8	78.8	84.4	62.3	64.6	57.7	69.4	50.1	71.8
SCRDET (2019)	R-101	89.98	80.65	52.09	68.36	68.36	60.32	72.41	90.85	87.94	86.86	65.02	66.68	66.25	68.24	65.21	72.61
SARD (2019B)	R-101	89.93	84.11	54.19	72.04	68.41	61.18	66.00	90.82	87.79	86.59	65.65	64.04	66.68	68.84	68.03	72.95
GLS-NET (2020A)	R-101	88.65	77.40	51.20	71.03	73.30	72.16	84.68	90.87	80.43	85.38	58.33	62.27	67.58	70.69	60.42	72.96
FADET (2019B)	R-101	90.21	79.58	45.49	76.41	73.18	68.27	79.56	90.83	83.40	84.68	53.40	65.42	74.17	69.69	64.86	73.28
MFIAR-NET (2020A)	R-152	89.62	84.03	52.41	70.30	70.13	67.64	77.81	90.85	85.40	86.22	63.21	64.14	68.31	70.21	62.11	73.49
SDCDET (2021D)	R-101	90.22	82.67	43.76	70.53	75.36	78.23	70.53	90.68	84.66	86.21	68.82	70.53	68.26	62.52	65.43	73.89
SEGMRDET (2020c)	R-50	89.91	80.31	52.85	70.72	76.72	79.43	86.33	90.65	82.91	87.1	48.41	69.29	74.68	72.36	50.48	74.14
GLIDING VERTEX (2020B)	R-101	89.64	85.00	52.26	77.34	73.01	73.14	86.82	90.74	79.02	86.81	59.55	70.91	72.94	70.86	57.32	75.02
SAR (2020)	R-152	89.67	79.78	54.17	68.29	71.70	77.90	84.63	90.91	88.22	87.07	60.49	66.95	75.13	75.28	64.29	75.28
MASK OBB (2019A)	RX-101	89.56	85.95	54.21	72.90	76.52	74.16	85.63	89.85	83.81	86.48	54.89	69.64	73.94	69.06	63.32	75.33
FFA (2020B)	R-101	90.1	82.7	54.2	75.2	71.0	79.9	83.5	90.7	83.9	84.6	61.2	68.0	70.7	76.0	63.7	75.7
APE (2020)	RX-101	89.96	83.62	53.42	76.03	74.01	77.16	79.45	90.83	87.15	84.51	67.72	60.33	74.61	71.84	65.55	75.75
F ³ -NET (2020)	R-152	88.89	78.48	54.62	74.43	72.80	77.52	87.54	90.78	87.64	85.63	63.80	64.53	78.06	72.36	63.19	76.02
CENTERMAP (2020B)	R-101	89.83	84.41	54.60	70.25	77.66	78.32	87.19	90.66	84.89	85.27	56.46	69.23	74.13	71.56	66.06	76.03
CSL (2020)	R-152	90.25	85.53	54.64	75.31	70.44	73.51	77.62	90.84	86.15	86.69	69.60	68.04	73.83	71.10	68.93	76.17
MRDET (2020)	R-101	89.49	84.29	55.40	66.68	76.27	82.13	87.86	90.81	86.92	85.00	52.34	65.98	76.22	76.78	67.49	76.24
AFC-NET (2021A)	R-101	89.65	84.20	53.06	77.62	74.66	75.45	87.08	90.84	86.84	85.33	63.76	65.52	72.26	70.64	67.13	76.27
RSDET (2021A; 2021B)	R-152	89.93	84.45	53.77	74.35	71.52	78.31	82.12	91.14	87.35	86.93	65.64	65.17	75.35	79.74	63.31	76.34
OWSR (2019A)	R-152	90.41	85.21	55.00	78.27	76.19	72.19	82.14	90.70	87.22	86.87	66.62	68.43	75.43	72.70	57.99	76.36
OPLD (2020)	R-101	89.37	85.82	54.10	79.58	75.00	75.13	86.92	90.88	86.42	86.62	62.46	68.41	73.98	68.11	63.69	76.43
SiOU (2021F)	R-50	89.39	84.00	54.76	73.71	76.22	78.31	83.73	90.87	84.96	87.37	62.38	67.15	73.86	69.44	71.98	76.54
SCRDETT++ (2020C)	R-101	90.05	84.39	55.44	73.99	77.54	71.11	86.05	90.67	87.32	87.08	69.62	68.90	73.74	71.29	65.08	76.81
MSFF (2021A)	R-50	89.60	85.51	54.59	77.28	73.31	79.44	85.94	90.82	87.68	86.34	65.05	70.38	77.67	72.0	62.27	77.46
OSKDET (2021)	R-101	90.04	87.25	54.41	79.48	72.66	80.29	88.20	90.84	83.91	86.90	63.39	71.76	75.63	72.59	69.75	77.81
HSP (2020A)	R-101	90.39	86.23	56.12	80.59	77.52	73.26	83.78	90.8								

The KFIoU Loss for Rotated Object Detection

Table 11. AP of different objects on DOTA-v1.0 test-dev (image-level based methods).

METHOD	BACKBONE	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	MAP ₅₀
IENET (2019)	R-101	80.20	64.54	39.82	32.07	49.71	65.01	52.58	81.45	44.66	78.51	46.54	56.73	64.40	64.24	36.75	57.14
TOSO (2020)	R-101	80.17	65.59	39.82	39.95	49.71	65.01	53.58	81.45	44.66	78.51	48.85	56.73	64.40	64.24	36.75	57.92
PIOU (2020)	DLA-34	80.9	69.7	24.1	60.2	38.3	64.4	64.8	90.9	77.2	70.4	46.5	37.1	57.1	61.9	64.0	60.5
AXIS LEARNING (2020)	R-101	79.53	77.15	38.59	61.15	67.53	70.49	76.30	89.66	79.07	83.53	47.27	61.01	56.28	66.06	36.05	65.98
Li <i>et al.</i> (2021F)	R-50	-	-	-	-	71.94	59.80	75.47	-	-	-	-	-	-	-	-	66.01
PROBIOU (2021)	R-50	89.09	72.15	46.92	62.22	75.78	74.70	86.62	89.59	78.35	83.15	55.83	64.01	65.50	65.46	46.23	70.04
A ² S-DET (2021)	R-101	89.59	77.89	46.37	56.47	75.86	74.83	86.07	90.58	81.09	83.71	50.21	60.94	65.29	69.77	50.93	70.64
O ² -DNET (2020)	H-104	89.31	82.14	47.33	61.21	71.32	74.03	78.62	90.76	82.23	81.36	60.93	60.17	58.21	66.98	61.03	71.04
CGP BOX (2021A)	R-18	89.00	82.83	34.12	63.91	74.04	74.70	79.35	90.81	80.19	84.74	55.77	67.51	64.72	73.15	55.47	71.35
PEI <i>et al.</i> (2021)	R-101	91.06	79.42	47.36	71.38	68.99	76.89	82.85	91.88	79.29	82.34	60.45	60.86	70.52	59.52	53.63	71.76
P-RSDET (2020)	R-101	88.58	77.83	50.44	69.29	71.10	75.79	78.66	90.88	80.10	81.71	57.92	63.03	66.30	69.77	63.13	72.30
BBAVECTORS (2021)	R-101	88.35	79.96	50.69	62.18	78.43	78.98	87.94	90.85	83.58	84.35	54.13	60.24	65.22	64.28	55.70	72.32
ZHANG <i>et al.</i> (2021D)	R-101	88.67	79.51	52.44	70.01	75.01	74.00	84.03	90.85	80.49	85.35	55.64	66.08	66.17	66.25	51.02	72.37
ROPDET (2020D)	R-101-DCN	90.01	82.82	54.47	69.65	69.23	70.78	75.78	90.84	86.13	84.76	66.52	63.71	67.13	68.38	46.09	72.42
HRP-NET (2020)	HRN-Net-W48	89.33	81.64	48.33	75.21	71.39	74.82	77.62	90.86	81.23	81.96	62.93	62.17	66.27	66.98	62.13	72.83
PROJBB (2021)	R-101	88.96	79.32	53.98	70.21	60.67	76.20	89.71	90.22	78.94	76.82	60.49	63.62	73.12	71.43	61.96	73.03
DRN (2020)	H-104	89.71	82.34	47.22	64.10	76.22	74.43	85.84	90.57	86.18	84.89	57.65	61.93	69.30	69.63	58.48	73.23
RBA-CENTERNET (2021C)	R-101	89.13	82.60	49.47	65.53	77.42	76.48	86.04	90.87	88.28	86.85	48.65	62.74	65.45	70.08	61.73	73.41
CFC-NET (2021A)	R-101	89.08	80.41	52.41	70.02	76.28	78.11	87.21	90.89	84.47	85.64	60.51	61.52	67.82	68.02	50.09	73.50
DUAL-DET (2021B)	R-18	89.91	85.33	44.25	63.68	78.65	76.00	86.48	90.86	80.04	85.79	56.41	62.63	67.22	74.30	62.76	73.62
Li <i>et al.</i> (2021B)	CSP-HOURGLASS	93.90	80.90	50.10	67.40	77.10	81.50	77.10	89.60	84.60	87.70	58.20	58.60	71.60	69.20	59.70	73.70
CENTERROT (2021A)	R-152	89.69	81.42	51.16	68.82	78.77	81.45	87.23	90.82	80.31	84.27	56.13	64.24	75.80	74.68	56.51	74.75
MEAD (2021C)	R-101	88.42	79.00	49.29	68.76	77.41	77.68	86.60	90.78	85.55	84.54	62.10	66.57	72.59	72.84	59.83	74.80
FEDET (2021B)	RX-101	89.09	79.87	51.13	70.20	78.42	80.54	87.84	90.86	83.91	85.31	58.33	66.90	67.74	69.74	63.43	74.89
SREP-RDET (2022)	REPVGG-B1G2	89.26	78.52	45.70	71.06	78.95	80.63	87.09	90.87	85.98	85.13	60.56	60.58	66.78	76.79	65.45	74.89
OSSDET (2021B)	RX-101	89.49	81.10	51.23	71.30	76.80	76.97	87.27	90.79	83.43	84.71	60.55	64.92	71.21	70.44	66.00	75.08
TRICUBE NET (2021)	H-104	88.75	82.12	49.24	72.98	77.64	74.53	84.65	90.81	86.02	85.38	58.69	63.59	73.82	69.67	71.08	75.26
SAOA (2021B)	R-101	88.33	82.73	56.02	71.58	72.98	77.59	78.29	88.63	83.33	86.61	65.93	63.52	76.03	78.43	61.33	75.41
ZAND <i>et al.</i> (2021)	D-53	89.0	80.4	50.5	76.0	78.2	77.8	87.4	90.0	83.2	86.93	63.3	68.7	67.6	70.8	63.8	75.5
TS ⁴ NET (2021)	R-101	89.59	84.07	53.41	73.23	78.76	77.11	87.04	90.88	85.94	85.48	63.04	64.98	68.05	72.34	60.58	75.63
CBDA-NET (2021B)	DLA-34-DCN	89.17	85.92	50.28	65.02	77.72	82.32	87.89	90.48	86.47	85.90	66.85	66.88	67.41	71.33	62.89	75.74
R ⁴ DET (2020)	R-152	88.96	85.42	52.91	73.84	74.86	81.52	80.29	90.79	86.95	85.25	64.05	60.93	69.00	70.55	67.76	75.84
RIE (2021A)	R-101	89.85	85.68	58.81	70.56	76.66	82.47	88.09	90.56	80.89	82.27	60.46	63.67	76.63	71.56	60.89	75.94
DA-NET (2021C)	R-101	89.70	82.41	53.28	69.55	78.24	79.54	89.04	90.68	84.76	86.33	65.03	65.70	76.16	73.37	58.86	76.11
SLA (2021B)	R-50	88.33	84.67	48.78	73.34	77.47	77.82	86.53	90.72	86.98	86.43	58.86	68.27	74.10	73.09	69.30	76.36
R ³ DET (2021C)	R-152	89.80	83.77	48.11	66.77	78.76	83.27	87.84	90.82	85.38	85.51	65.67	62.68	67.53	78.56	72.62	76.47
POLAR RAY (2021A)	R-101	88.78	85.51	51.89	73.42	74.36	77.98	87.78	90.86	86.50	86.84	60.69	69.84	75.29	72.22	67.22	76.50
POLARDET (2021)	R-101	89.65	87.07	48.14	70.97	78.53	80.34	87.45	90.76	85.63	86.87	61.64	70.32	71.92	73.09	67.15	76.64
CFA (2021)	R-152	89.08	83.20	54.37	66.87	81.23	80.96	87.17	90.21	84.32	86.09	52.34	69.94	75.52	80.76	67.96	76.67
DAL (2020)	R-50	89.69	83.11	55.03	71.00	78.30	81.90	88.46	90.89	87.46	84.41	65.65	76.86	72.09	64.35	76.95	
DAFNE (2021)	R-101	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	76.95
GGHL (2021)	D-53	89.74	85.63	44.50	77.48	76.72	80.45	86.16	90.83	88.18	86.25	67.07	69.40	73.38	68.45	70.14	76.95
DCL (2021B)	R-152	89.26	83.60	53.54	72.76	79.04	82.56	87.31	90.67	86.59	86.98	67.49	66.88	73.29	70.56	69.99	77.37
RIDET (2021C)	R-50	89.31	80.77	54.07	76.38	79.81	81.99	89.13	90.72	83.58	87.22	64.42	67.56	78.08	79.17	62.07	77.62
RDD (2020)	R-101	89.15	83.92	52.51	73.06	77.81	79.00	87.08	90.62	86.72	87.15	63.96	70.29	76.98	75.79	72.15	77.75
CG-NET (2021B)	R-101	88.76	85.36	60.81	71.88	74.04	83.43	88.29	90.89	86.00	85.55	63.51	67.02	78.78	80.86	68.70	78.26
ORIENTED REPPONTS (2021)	R-101	88.72	80.56	55.69	75.07	81.84	82.40	87.97	90.80	84.33	87.64	62.80	67.91	77.69	82.94	65.46	78.12
FORDET (2021E)	VGG16	89.62	85.88	47.55	81.45	80.63	81.84	88.08	90.87	88.27	86.41	72.42	67.69	73.91	72.67	64.63	78.13
DARDET (2021B)	R-50	89.08	84.30	56.64	77.83	81.10	83.39	88.46	90.88	85.44	87.56	62.77	66.23	77.97	82.03	67.40	78.74
FCOSR (2021E)	RX-101	88.78	85.38	54.29	76.81	81.52	82.76	88.38	90.80	86.61	87.25	67.58	67.03	76.86	73.22	74.68	78.80
S ² A-NET (2021A)	R-50	88.89	83.60	57.74	81.95	79.94	83.19	89.11	90.78	84.87	87.81	70.30	68.25	78.30	77.01	69.58	79.42
OFA-NET (2021)	R-101	88.75	82.75	57.14	81.37	78.83	85.45	88.63	90.90	86.60	87.97	66.46	70.66	77.35	78.82	71.55	79.52
SES-NET (2021A)	R-50	88.96	82.99	56.99	82.84	80.75	85.89	89.35	90.58	86.67	87.88	67.96	69.23	78.75	73.04	75.19	79.80
SARA (2021)	R-50	89.40	84.29	56.72	82.29	80.49	83.01	89.37	90.67	86.20	87.44	71.34	69.06	78.49	80.98	68.98	79.91
ARP (2021)	CSPD-53	89.32	85.28	58.12	81.12	78.54	84.11	89.43	90.91	87.42	87.88	68.76	67.01	77.87	81.87	71.32	79.93
ADT-DET (2021B)	R-152	89.71	84.71	59.63	80.94	80.30	83.53	88.94	90.86	87.06	87.81	70.72	70.92	78.66	79.40	65.99	79.95
GWD (2021D)	R-152	89.66	84.99	59.26	82.19	78.97	84.83	87.70	90.21	86.54	86.85	73.47	67.77	76.92	79.22	74.92	80.23
KLD (2021E)	R-152	89.92	85.13	59.19	81.33	78.82	84.38	87.50	89.80	87.33	87.00	72.57	71.35	77.12	79.34	78.68	80.63
KFIoU (OURS)	SWIN-TINY	89.50	84.26	59.90	81.06	81.74	85.45	88.77	90.85	87.03	87.79</						

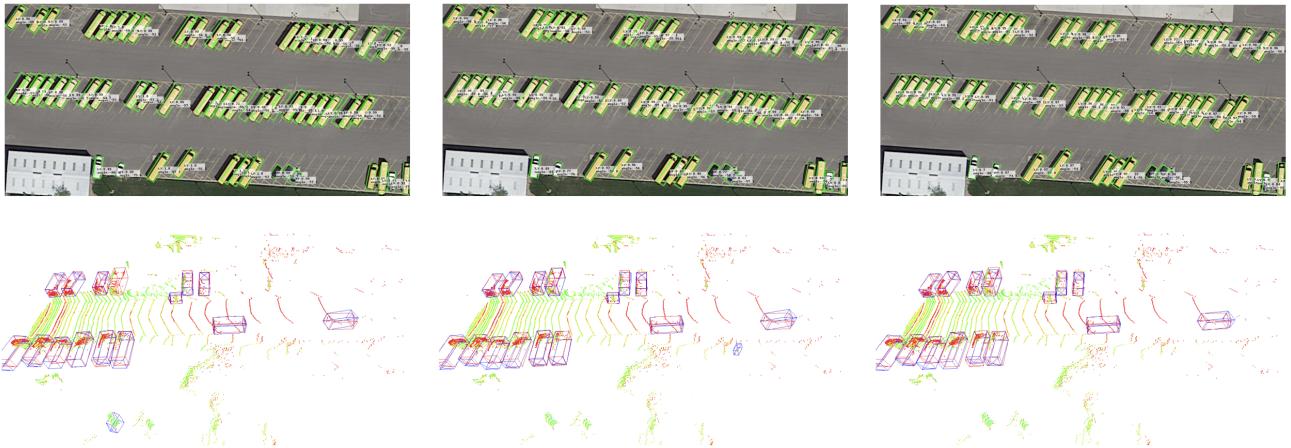


Figure 4. Visual comparison between Smooth L1 loss-based (left), GWD-based (middle) and the KFIoU-based (right) detectors on DOTA (2-D) and KITTI (3-D). For 3-D object detection, red and blue box denotes ground-truth and predict bounding box, respectively.

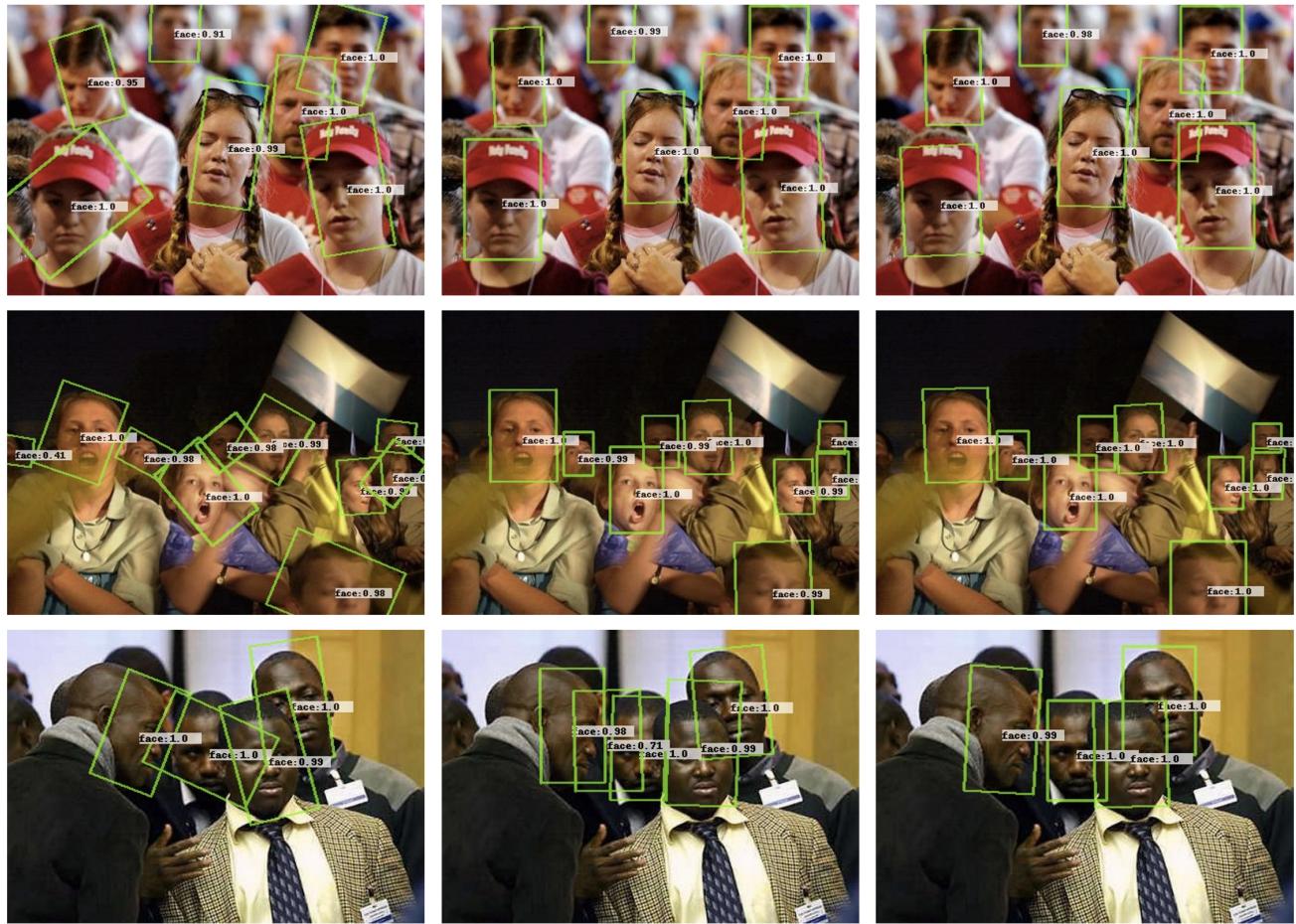


Figure 5. Visual comparison between Smooth L1 loss-based (left), GWD-based (middle) and the KFIoU-based (right) detectors on Fddb.