2023 IEEE International Conference on Image Processing (ICIP) | 978-1-7281-9835-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIP49359.2023.10222610

THE ELLIPTIC ENERGY LOSS FOR ROTATED OBJECT DETECTION IN AERIAL IMAGES

Cong Zhang[⋆] *Kunming Luo*[†] *Fanman Meng*[⋆‡] *Qingbo Wu*[⋆]

[⋆] University of Electronic Science and Technology of China, Chengdu, China
[†]The Hong Kong University of Science and Technology, Hong Kong, China

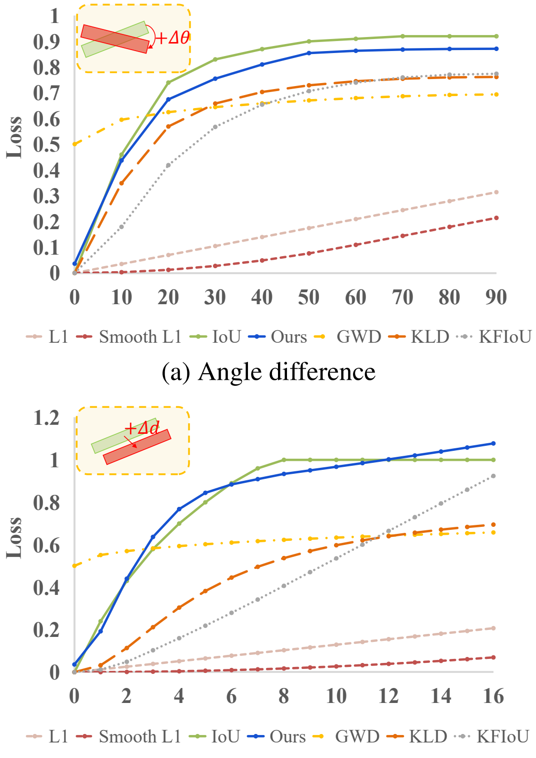
ABSTRACT

Rotated object detection is a promising yet challenging task in computer vision. Existing algorithms mainly train the rotated detector by the L_n-norm loss, which is inconsistent with the evaluation metric of Intersection over Union (IoU). However, the concise and efficient solution using the loss based on IoU between oriented boxes is hindered by its non-differentiability. In this paper, we propose Elliptic Energy Loss, a differentiable loss based on curve energy, to fit with the evaluation metric IoU. Specifically, given a pair of predicted and ground truth boxes, we first convert them to curve representations using the elliptic transformation. Then, the curve energy is calculated to measure the similarity between the predicted and ground truth curves. Finally, the curve energy is used as regression loss to optimize rotated detectors. We conduct experiments with different detectors on DOTA and HRSC2016 datasets, which demonstrate that the performance is significantly improved by our proposed loss. The code is available at https://github.com/zhangc-uestc/EEL.

Index Terms— Object detection, rotated object detection, regression loss

1. INTRODUCTION

Rotated object detection is a fundamental task in computer vision, with various applications in environmental change monitoring [1] and intelligent transportation [2]. Compared with horizontal object detection [3–5], rotated object detection regresses oriented bounding boxes (OBBs) that locate objects more tightly, which exhibits apparent advantages when detecting objects with arbitrary orientation and dense arrangements [6–8]. To train rotated detectors, existing algorithms mainly use the L_n-norm loss as the regression loss. However, the L_n-norm loss fails to accurately reflect the change in the evaluation metric Intersection over Union (IoU) [9, 10], which limits the performance of rotated detectors. As is demonstrated in horizontal detection, using IoU loss or its variants [11–13] can effectively solve this inconsistency problem. Unfortunately, these solutions cannot be applied to rotated detection since the calculation of IoU between OBBs (called SkewIoU [14]) is non-differentiable.



(a) Angle difference

(b) Center offset

Fig. 1. Comparison of our Elliptic Energy Loss with ideal IoU loss and existing losses such as L1, Smooth L1 [3], GWD [15], KLD [16], KFIoU [10]. Given predicted and ground truth OBBs, we compute the loss values of (a) angle difference and (b) center offset.

To narrow the gap between the regression loss and the evaluation metric, IoU-Smooth L1 loss [17] takes the negative logarithm of SkewIoU as the magnitude of the gradient directly, PIoU [9] estimates the intersection of two OBBs in a pixel-wise manner. These methods improved the accuracy of localization, but the inconsistency problem is only par-

[‡]Corresponding author. Email: fmmeng@uestc.edu.cn

This work was supported in part by the National Natural Science Foundation of China under Grant 62271119, the Natural Science Foundation of Sichuan Province under Grant 2023NSFSC1972.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on October 14,2023 at 15:31:56 UTC from IEEE Xplore. Restrictions apply.

tially solved. Recently, Gaussian-based losses (GWD [15], KLD [16], KFIoU [10]) are proposed to mimic the mechanism of SkewIoU using Gaussian distribution, which achieves superior performance. Despite their success, as illustrated in Fig. 1, Gaussian-based losses do not completely fit the IoU loss. In the center offset case, GWD and KLD tend to be constant as the distance increases. It is also the flaw of IoU loss, which is suboptimal for initial learning of detectors.

In this paper, we propose a new regression loss called Elliptic Energy loss, which approximates the non-differentiable SkewIoU using a differentiable similarity metric called elliptic energy. Specifically, we first convert the predicted OBB and the ground truth OBB to their inscribed ellipse representations. Then we use the energy function inspired by the active contour model [18] to measure the similarity between the predicted curve and the ground truth curve. Since the curve transformation and the energy calculation are differentiable, they can be applied as the regression loss. As shown in Fig. 1, our loss is the closest approximation to SkewIoU compared with other losses. We conduct various experiments on two popular datasets, DOTA and HRSC2016, to validate the effectiveness of our method, achieving obvious gains over the baseline and state-of-the-art performance.

2. PROPOSED METHOD

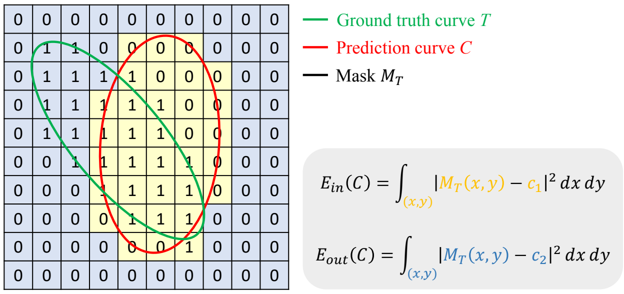
Rotated detection aims at locating objects in images using OBBs that are typically represented by 5 parameters (x_c, y_c, w, h, θ), which stand for the center of the object, width (long side), height (short side), and the angle from the positive direction of the x-axis to the direction parallel to w in the Long Edge definition [14]. To train the rotated detector, regression loss is required to measure the difference between the predicted and ground truth OBBs. In this paper, we propose Elliptic Energy loss, in which we first convert OBBs into elliptic curve representations (Sec 2.1), and then calculate the energy of curves (Sec 2.2), which can finally be used as the regression loss (Sec 2.3) that fits the SkewIoU well.

2.1. Curve representation of the object

Given two OBBs, we transform them into their curve representations, which should be continuous and differentiable. Thus, we propose to describe the object using the inscribed elliptic curve of OBB, which can be formulated as follows:

$$
\begin{aligned}
f_{rt}(x, y) &= \frac{[(x - x_c)\cos\theta + (y - y_c)\sin\theta]^2}{(w/2)^2} \\
&\quad + \frac{[(y - y_c)\cos\theta - (x - x_c)\sin\theta]^2}{(h/2)^2} \qquad (1) \\
&= 1,
\end{aligned}
$$

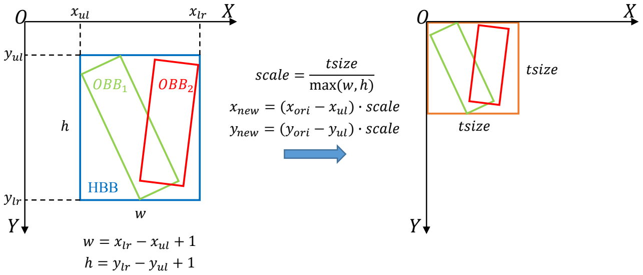
where x_c, y_c, w, h, θ are parameters of OBB, and (x, y) is the point on the ellipse if $f_{rt}(x, y) = 1$. Similarly, the relative position from any point (x, y) to the ellipse can be determined by $f_{rt}(x, y)$. In this way, object representation is transformed from the bounding box to the elliptic curve, which facilitates the similarity measure between curves later.



0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0
0	1	1	1	1	0	0	0	0	0
0	1	1	1	1	1	0	0	0	0
0	1	1	1	1	1	0	0	0	0
0	0	1	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	0	1	1	1	0	0	0
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0

Ground truth curve T
Prediction curve C
Mask M_T
$E_{in}(C) = \int_{(x,y)} |M_T(x,y) - c_1|^2 \, dx \, dy$
$E_{out}(C) = \int_{(x,y)} |M_T(x,y) - c_2|^2 \, dx \, dy$

Fig. 2. Illustration of the curve energy. Mask is constructed in terms of the ground truth curve T. The internal energy $E_{in}(C)$ in Eq. (2) comes from the yellow region, where c_1 is the average of the yellow region. The external energy $E_{out}(C)$ comes from the blue region, where c_2 is the average of the blue region.



O
x_{ul}
x_{lr}
X
y_{ul}
OBB_1
OBB_2
h
HBB
y_{lr}
Y
w
$w = x_{lr} - x_{ul} + 1$
$h = y_{lr} - y_{ul} + 1$
$scale = \frac{tsize}{\max(w,h)}$
$x_{new} = (x_{ori} - x_{ul}) \cdot scale$
$y_{new} = (y_{ori} - y_{ul}) \cdot scale$
O
X
$tsize$
$tsize$
Y

Fig. 3. Diagram of normalization. For given two boxes, OBB$_1$ and OBB$_2$, we first find their external horizontal box HBB (the smallest enclosing box covering two boxes) and its relevant parameters including the upper left corner (x_{ul}, y_{ul}), the lower right corner (x_{lr}, y_{lr}), the width w and the height h. Then we calculate the scale factor $scale$, and the vertex coordinates (x_{new}, y_{new}) of the new boxes using $scale$ and (x_{ul}, y_{ul}), so that the new boxes are limited to $tsize \times tsize$.

2.2. Similarity measure between curves

After elliptic transformation of OBBs, we then measure the similarity between the predicted curve C and the ground truth curve T. Inspired by the active contour model [18], we weigh the similarity with customized energy, which contains two fitting items as Eq. (2).

$$
\begin{aligned}
E(C) &= E_{in}(C) + E_{out}(C) \\
&= \int_{inside(C)} |M_T(x, y) - c_1|^2 dxdy \qquad (2) \\
&\quad + \int_{outside(C)} |M_T(x, y) - c_2|^2 dxdy,
\end{aligned}
$$

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on October 14,2023 at 15:31:56 UTC from IEEE Xplore. Restrictions apply.

in which

$$M_T(x,y) = \begin{cases} 1, & \text{if } f_{rt}^T(x,y) \le 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

where $E_{in}(C)$ and $E_{out}(C)$ represent the internal and external energies of C on M_T, respectively. M_T is a special image we construct with T that corresponds to the input image in the active contour model, where pixel values inside and including T are 1 and pixel values outside T are 0, as shown in Fig. 2. The constants c_1, c_2 represent the averages of M_T inside and outside C, respectively. Eq. (2) shows that if C is identical to T or completely inside T, $E_{in}(C) = 0$, otherwise $E_{in}(C) > 0$, which indicates that minimizing the internal energy tightens the curve C inwards. And if C is identical to T or completely outside T, $E_{out}(C) = 0$, otherwise $E_{out}(C) > 0$, which means that minimizing the external energy stops the curve C when C shrinks to T. Finally, fitting energy is minimized when $C = T$.

However, we observe that Eq. (3) is not a continuous and differentiable function. It will break the back propagation of the gradient if we use Eq. (3) directly. Inspired by PIoU [9], we update Eq. (3) based on a kernel function:

$$M(x,y) \approx K(f_{rt}(x,y), 1) \tag{4}$$

in which

$$K(d,s) = 1 - \frac{1}{1 + e^{-k(d-s)}} \tag{5}$$

where $K(d,s)$ is the kernel function, $k(> 0)$ is an adjustable factor that controls the sensitivity. $K(d,s)$ tends to 0 when $d > s$, and tends to 1 when $d < s$. Thus, $K(f_{rt}(x,y), 1)$ can be used as a continuous and differentiable alternative to Eq. (3).

To simplify the calculation, we also construct the mask M_C corresponding to the predicted curve C, so that Eq. (2) can be computed as follows:

$$\begin{aligned} E(C) = &\sum_{x,y} \left[M_C(x,y) \times (M_T(x,y) - c_1)\right]^2 \\ &+ \sum_{x,y} \left[(1 - M_C(x,y)) \times (M_T(x,y) - c_2)\right]^2 \end{aligned} \tag{6}$$

in which

$$c_1 = \frac{\sum_{x,y} M_C(x,y) \times M_T(x,y)}{\sum_{x,y} M_C(x,y)} \tag{7}$$

$$c_2 = \frac{\sum_{x,y} (1 - M_C(x,y)) \times M_T(x,y)}{W \times H - \sum_{x,y} M_C(x,y)} \tag{8}$$

where W (H) represents the width (height) of the M_C, which is also the width (height) of the input image and M_T. $x \in [0, W), y \in [0, H)$. By this way, the similarity between the predicted and the ground truth curve can be measured using defined energy, i.e., Eq. (6)-(8).

2.3. Regression loss

Note that the calculation above is done pixel by pixel. To lower the consumption of computational resources, we suggest normalizing oriented boxes first. The specific steps are depicted in Fig. 3, where $tsize$ is the size after normalizing and is generally set to 50 or 100. It further ensures scale invariance of the loss with rescaling operations. Following normalization, all pixel-wise calculations are restricted to the range of $tsize \times tsize$, thus only $tsize \times tsize$ is needed to construct the mask, that is, $x \in [0, tsize), y \in [0, tsize)$. But the normalization method shown in Fig. 3 introduces a new problem. Assuming that the predicted box is far away from the ground truth box, their external horizontal box will be large, resulting in the predicted and ground truth boxes becoming small after rescaling. The slight difference between M_T and M_C will result in a small loss. This means that the current method is not suitable for measuring predicted and ground truth boxes that are far apart, as the ideal loss should increase with their distance. To fix this problem, we add the distance between the centers of the predicted and ground truth boxes to the loss as center loss, so that the center loss takes the lead when the two boxes are far away, and the energy takes the lead otherwise.

We select Smooth L1 loss as the center loss, and condense energy using a linear function as the energy loss, which are indicated by Eq. (9) and Eq. (10).

$$L_{ct}(t, t^*) = \frac{1}{2} \sum_{j \in \{x,y\}} \text{Smooth L1}(t_j - t_j^*) \tag{9}$$

$$L_e(b, gt) = \lambda E(C, T) \tag{10}$$

where t (t^*) denotes the offset from the center of the predicted box (the ground truth box) to the center of the anchor box. $t_x = \frac{x_c - x_c'}{w'}, t_y = \frac{y_c - y_c'}{h'}, t_x^* = \frac{x_c^* - x_c'}{w'}, t_y^* = \frac{y_c^* - y_c'}{h'}$. x_c, y_c, w, h denote the center coordinates, width, and height of the box, respectively. Variables x_c, x_c', x_c^* are for the predicted box, anchor box, and ground-truth box (likewise for y_c, w, h). b and gt represent the predicted box and ground truth box, respectively. λ is hyper-parameter determined by follow-up experiments.

The entire regression loss is as follows:

$$L = \frac{1}{N_{pos}} \sum_{i=1}^{N_{pos}} L_{ct}\left(t_i, t_i^*\right) + L_e\left(b_i, gt_i\right) \tag{11}$$

where N_{pos} indicates the number of positive samples.

3. EXPERIMENTS

3.1. Datasets and Implementation Details

We conduct experiments on DOTA [19] and HRSC2016 [2], which are the most common datasets for oriented object detection. All experiments are implemented by MMRotate [20]

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on October 14,2023 at 15:31:56 UTC from IEEE Xplore. Restrictions apply.

Table 1. Ablation study of hyperparameters with the RetinaNet-HBB detector on DOTA. The **bold** and underlined fonts indicate the top two performances, respectively. The metric is mAP_{50}.

Parameter	$\lambda = 1/500$	$\lambda = 1/1000$	$\lambda = 1/2000$
$k = 10$	68.85	**70.31**	70.05
$k = 100$	68.19	69.50	69.76
$k = 1000$	66.89	68.57	68.63

Table 2. Ablation study of hyperparameters with the RetinaNet-HBB detector on HRSC2016. The metrics used are AP_{50} / AP_{75} / $\mathrm{mAP}_{50:95}$.

Parameter	$\lambda = 1/500$	$\lambda = 1/1000$	$\lambda = 1/2000$
$k = 10$	86.8 / 60.3 / 54.49	**87.4** / 68.9 / 57.75	86.7 / **70.9** / **59.60**
$k = 100$	85.6 / 60.3 / 54.02	87.0 / 69.1 / 56.22	86.6 / 69.1 / 58.69
$k = 1000$	84.9 / 53.5 / 48.58	86.3 / 59.7 / 53.73	85.7 / 68.9 / 56.99

Table 3. Comparisons with peer losses under different detectors on DOTA. The definition methods of oriented boxes on both detectors are OpenCV Definition ($\theta \in (0, \frac{\pi}{2}]$). The values in brackets show the gain or reduction of each loss relative to the Smooth L1 loss.

Model	Reg.Loss	mAP_{50}
RetinaNet-HBB [5]	Smooth L1	65.08
	GWD [15]	68.70(+3.62)
	KLD [16]	69.39(+4.31)
	KFIoU [10]	69.48(+4.40)
	EEL (Ours)	**70.31(+5.23)**
R^{3}Det [8]	Smooth L1	69.01
	GWD [15]	72.42(+3.41)
	KLD [16]	72.73(+3.72)
	KFIoU [10]	72.25(+3.24)
	EEL (Ours)	**73.15(+4.14)**

Table 4. Comparisons with peer losses under different detectors on HRSC2016.

Model	Reg.Loss	AP_{50}	AP_{75}	$\mathrm{mAP}_{50:95}$
RetinaNet-HBB [5]	Smooth L1	81.3	54.5	48.18
	GWD [15]	84.9(+3.6)	67.1(+12.6)	55.98(+7.8)
	KLD [16]	**86.8(+5.5)**	69.6(+15.1)	58.48(+10.3)
	KFIoU [10]	84.7(+3.4)	60.4(+5.9)	54.61(+6.43)
	EEL (Ours)	86.7(+5.4)	**70.9(+16.4)**	**59.60(+11.42)**
R^{3}Det [8]	Smooth L1	86.2	57.6	52.33
	GWD [15]	87.0(+0.8)	**69.7(+12.1)**	**58.55(+6.22)**
	KLD [16]	87.5(+1.3)	68.2(+10.6)	55.92(+3.59)
	KFIoU [10]	86.5(+0.3)	68.8(+11.2)	57.51(+5.18)
	EEL (Ours)	**88.5(+2.3)**	68.2(+10.6)	55.77(+3.44)

using one TITAN Xp GPU. We set batch size to 2 and train models on DOTA and HRSC2016 for 12 and 72 epochs, respectively. For fair comparison, we use the same configuration to train the models, except with different losses.

3.2. Ablation Studies

In Table 1 and Table 2, we conduct ablation experiments on the hyperparameter settings of k and λ in our elliptic energy loss on DOTA and HRSC2016 datasets. As can be seen, performance drops rapidly when k increases. The reason may be that a larger k will cause a larger gradient of the loss, making the loss more sensitive to prediction errors during training. Also, choosing a smaller λ can improve performance when given a larger k. As suggested by Table 1 and Table 2, we set $k = 10, \lambda = 1/1000$ on DOTA, and $k = 10, \lambda = 1/2000$ on HRSC2016 for subsequent experiments.

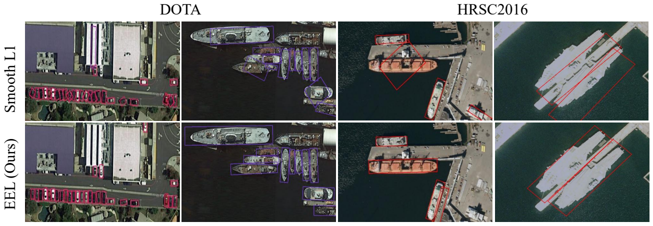


Fig. 4. Qualitative comparison between the Smooth L1 loss and the proposed Elliptic Energy loss with same RetinaNet-HBB detector on DOTA and HRSC2016.

3.3. Comparisons with previous methods

In Table 3 and Table4, we compare our method with previous regression losses, such as Smooth L1 [3], GWD [15], KLD [16] and KFIoU [10]. We train RetinaNet-HBB [5] and R^{3}Det [8] detectors on DOTA and HRSC2016 datasets using different regression losses. Smooth L1 loss is used as the baseline. As the results show, on DOTA dataset, our method achieves $\mathrm{mAP}_{50} = 70.31\%$ and $\mathrm{mAP}_{50} = 73.15\%$ using RetinaNet-HBB and R^{3}Det, which is the best compared with previous methods. On HRSC2016 dataset, our method can also produce superior results of $\mathrm{AP}_{75} = 70.9\%$ and $\mathrm{mAP}_{50:95} = 59.60\%$ using RetinaNet-HBB and $\mathrm{AP}_{50} = 88.5\%$ using $\mathrm{R}^3\mathrm{Det}$. The qualitative comparison between the Smooth L1 loss and our proposed loss is shown in Fig. 4. As a result, the model trained by our loss can locate objects more precisely than the model trained by the Smooth L1 loss.

4. CONCLUTIONS

In order to alleviate the inconsistency between the regression loss and the evaluation metric, this paper proposes the Elliptic Energy loss. The loss first converts the paired predicted box and the ground truth box into their corresponding curve representations, and then utilizes the curve energy to assess how similar the predicted curve and the ground truth curve are. Additionally, by transforming it appropriately and increasing the center loss, it can be used for training. Numerous experiments on DOTA and HRSC2016 demonstrate the effectiveness of our loss.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on October 14,2023 at 15:31:56 UTC from IEEE Xplore. Restrictions apply.

5. REFERENCES

[1] Courage Kamusoko, “Importance of remote sensing and land change modeling for urbanization studies,” in *Urban development in Asia and Africa*, pp. 3–10. Springer, 2017.

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

[3] Ross Girshick, “Fast r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[4] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *Advances in neural information processing systems*, vol. 28, 2015.

[5] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[6] Jian Ding, Nan Xue, Yang Long, Gui-Song Xia, and Qikai Lu, “Learning roi transformer for oriented object detection in aerial images,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2849–2858.

[7] Jiaming Han, Jian Ding, Jie Li, and Gui-Song Xia, “Align deep features for oriented object detection,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

[8] Xue Yang, Junchi Yan, Ziming Feng, and Tao He, “R3det: Refined single-stage detector with feature refinement for rotating object,” in *Proceedings of the AAAI conference on artificial intelligence*, 2021, vol. 35, pp. 3163–3171.

[9] Zhiming Chen, Kean Chen, Weiyao Lin, John See, Hui Yu, Yan Ke, and Cong Yang, “Piou loss: Towards accurate oriented object detection in complex environments,” in *European conference on computer vision*. Springer, 2020, pp. 195–211.

[10] Xue Yang, Yue Zhou, Gefan Zhang, Jitui Yang, Wentao Wang, Junchi Yan, Xiaopeng Zhang, and Qi Tian, “The kfiou loss for rotated object detection,” *arXiv preprint arXiv:2201.12558*, 2022.

[11] Jiahui Yu, Yuning Jiang, Zhangyang Wang, Zhimin Cao, and Thomas Huang, “Unitbox: An advanced object detection network,” in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 516–520.

[12] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese, “Generalized intersection over union: A metric and a loss for bounding box regression,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 658–666.

[13] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, and Dongwei Ren, “Distance-iou loss: Faster and better learning for bounding box regression,” in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, pp. 12993–13000.

[14] Jianqi Ma, Weiyuan Shao, Hao Ye, Li Wang, Hong Wang, Yingbin Zheng, and Xiangyang Xue, “Arbitrary-oriented scene text detection via rotation proposals,” *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 3111–3122, 2018.

[15] Xue Yang, Junchi Yan, Qi Ming, Wentao Wang, Xiaopeng Zhang, and Qi Tian, “Rethinking rotated object detection with gaussian wasserstein distance loss,” in *International Conference on Machine Learning*. PMLR, 2021, pp. 11830–11841.

[16] Xue Yang, Xiaojiang Yang, Jirui Yang, Qi Ming, Wentao Wang, Qi Tian, and Junchi Yan, “Learning high-precision bounding box for rotated object detection via kullback-leibler divergence,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 18381–18394, 2021.

[17] Xue Yang, Jirui Yang, Junchi Yan, Yue Zhang, Tengfei Zhang, Zhi Guo, Xian Sun, and Kun Fu, “Scrdet: Towards more robust detection for small, cluttered and rotated objects,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 8232–8241.

[18] Tony F Chan and Luminita A Vese, “Active contours without edges,” *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.

[19] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang, “Dota: A large-scale dataset for object detection in aerial images,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3974–3983.

[20] Yue Zhou, Xue Yang, Gefan Zhang, Jiabao Wang, Yanyi Liu, Liping Hou, Xue Jiang, Xingzhao Liu, Junchi Yan, Chengqi Lyu, et al., “Mmrotate: A rotated object detection benchmark using pytorch,” *arXiv preprint arXiv:2204.13317*, 2022.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on October 14,2023 at 15:31:56 UTC from IEEE Xplore. Restrictions apply.