Learning High-Precision Bounding Box for Rotated Object Detection via Kullback-Leibler Divergence

Xue Yang[1,2], Xiaojiang Yang[1,2], Jirui Yang[3], Qi Ming[4], Wentao Wang[1,2], Qi Tian[5], Junchi Yan[1,2]

[1]Department of CSE, SJTU [2]MoE Key Lab of Artificial Intelligence, AI Institute, SJTU

[3]University of Chinese Academy of Sciences [4]School of Automation, Beijing Institute of Technology [5]Huawei Inc.

yangxue-2019-sjtu@sjtu.edu.cn

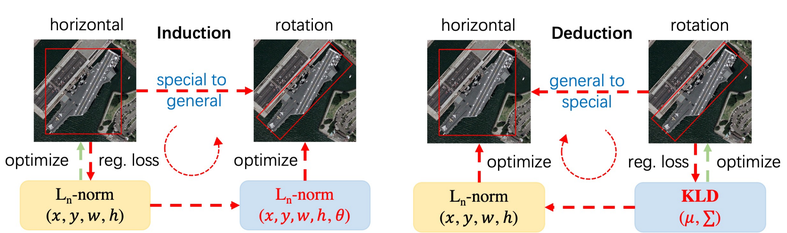
Introduction

➢ **Task:** Design a novel high-precision rotation detector for small, cluttered and rotated objects.

➢ **Contributions:**

- Differing from the dominant existing practices that build rotation detectors heavily upon the horizontal detectors, we develop new rotation detection loss from scratch and show that it is coherent with existing horizontal detection protocol in its degenerated case for horizontal detection.
- To achieve a more principled measurement between the prediction and ground truth, instead of computing the difference for each physically-meaningful parameter related to the bounding box which are in different scales and units, we innovatively convert the regression loss of rotation detection into the KLD of two 2-D Gaussian distributions, leading to a clean and coherent regression loss.
- Through the gradient analysis of each parameter in KLD, we further find that the self-modulated optimization mechanism of KLD greatly promotes the improvement of high-precision detection, which verify the advantage of our loss design. Importantly, we have theoretically shown that KLD is scale invariant for detection, which is crucial for the rotation cases.
- Extensive experimental results on seven public datasets and two popular detectors show the effectiveness of our approach, which achieves new state-of-the-art performance for rotation detection.

➢ **Code:** https://github.com/yangxue0827/RotationDetection



Left: previous methods follow the induction paradigm from special horizontal to general rotated detection.
Right: proposed method adopts a deduction methodology from general rotated to special horizontal detection.

Proposed Approach

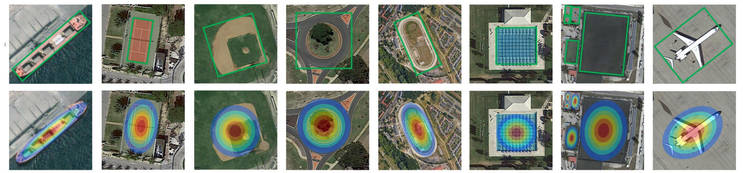
➢ **Gaussian Modeling:**

To break the original inductive design paradigm, we adopt deductive paradigm to construct more accurate rotation regression loss. We convert a arbitrary-oriented bounding box (x,y,h,w,θ) into a 2-D Gaussian (μ, Σ)

$$\boldsymbol{\mu} = (x, y)^{\top}$$

$$\boldsymbol{\Sigma}^{1/2} = \mathbf{R}\boldsymbol{\Lambda}\mathbf{R}^{\top} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \frac{w}{2} & 0 \\ 0 & \frac{h}{2} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} \frac{w}{2}\cos^2\theta + \frac{h}{2}\sin^2\theta & \frac{w-h}{2}\cos\theta\sin\theta \\ \frac{w-h}{2}\cos\theta\sin\theta & \frac{w}{2}\sin^2\theta + \frac{h}{2}\cos^2\theta \end{pmatrix}$$



➢ **Kullback-Leibler Divergence:**

To explore the more appropriate regression loss, we adopt the Kullback-Leibler divergence (KLD). The KLD between two 2-D Gaussian is:

$$\mathbf{D}_{kl}(\mathcal{N}_p||\mathcal{N}_t) = \underbrace{\frac{1}{2}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_t)^{\top}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_t)}_{\text{term about } x_p \text{ and } y_p} + \underbrace{\frac{1}{2}\mathbf{Tr}(\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\Sigma}_p) + \frac{1}{2}\ln\frac{|\boldsymbol{\Sigma}_t|}{|\boldsymbol{\Sigma}_p|} - 1}_{\text{coupling terms about } h_p,\ w_p \text{ and } \theta_p}$$

or

$$\mathbf{D}_{kl}(\mathcal{N}_t||\mathcal{N}_p) = \underbrace{\frac{1}{2}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_t)^{\top}\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_t) + \frac{1}{2}\mathbf{Tr}(\boldsymbol{\Sigma}_p^{-1}\boldsymbol{\Sigma}_t) + \frac{1}{2}\ln\frac{|\boldsymbol{\Sigma}_p|}{|\boldsymbol{\Sigma}_t|} - 1}_{\text{chain coupling of all parameters}}$$

It can be seen that each item in $D_{kl}(N_p||N_t)$ is composed of partial parameter coupling, which makes all parameters form a chain coupling relationship. In the optimization process of the KLD-based detector, the parameters influence each other and are jointly optimized which make optimization mechanism of the model is self-modulated. In contrast, $D_{kl}(N_t||N_p)$ and GWD are both semi-coupled, but $D_{kl}(N_t||N_p)$ has a better central point optimization mechanism.

Experiments

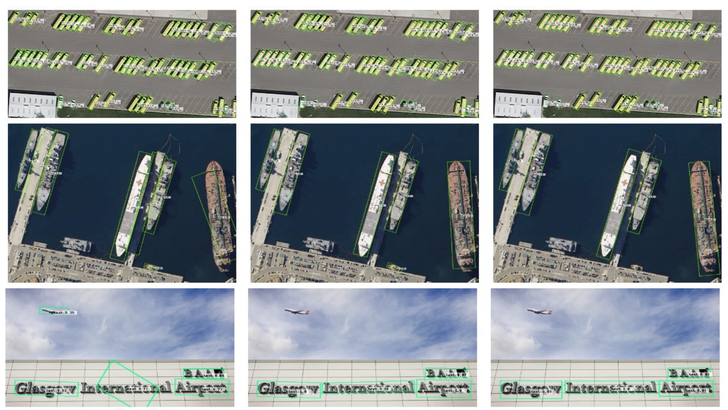
➢ **High-Precision Detection Experiment:**

Method	Dataset	Data Aug.	Reg. Loss	$Hmean_{50}/AP_{50}$	$Hmean_{60}/AP_{60}$	$Hmean_{75}/AP_{75}$	$Hmean_{85}/AP_{85}$	$Hmean_{50:95}/AP_{50:95}$
RetinaNet	HRSC2016	R+F+G	Smooth L1	84.28	74.74	48.42	12.56	47.76
			GWD	85.56 (+1.28)	84.04 (+9.30)	60.31 (+11.89)	17.14 (+4.58)	52.89 (+5.13)
			KLD	**87.45** (+3.17)	**86.72** (+11.98)	**72.39** (+23.97)	**27.68** (+15.12)	**57.80** (+10.04)
R^3Det			Smooth L1	88.52	79.01	43.42	4.58	46.18
			GWD	89.43 (+0.91)	88.89 (+9.88)	65.88 (+22.46)	15.02 (+10.44)	56.07 (+9.89)
			KLD	**89.97** (+1.45)	**89.73** (+10.72)	**77.38** (+33.96)	**25.12** (+20.54)	**61.40** (+15.22)
RetinaNet	MSRA-TD500	R+F	Smooth L1	70.98	62.42	36.73	12.56	37.89
			GWD	76.76 (+5.78)	68.58 (+6.16)	44.21 (+7.48)	17.75 (+5.19)	43.62 (+5.73)
			KLD	**76.96** (+5.98)	**70.08** (+7.66)	**46.95** (+10.22)	**19.59** (+7.03)	**45.24** (+7.35)
RetinaNet	ICDAR2015	F	Smooth L1	69.78	64.15	36.97	8.71	37.73
			GWD	74.29 (+4.51)	68.34 (+4.19)	43.39 (+6.42)	10.50 (+1.79)	41.68 (+3.95)
			KLD	**75.32** (+5.54)	**69.94** (+5.79)	**44.46** (+7.49)	**10.70** (+1.99)	**42.68** (+4.95)
		R+F	Smooth L1	74.83	69.46	42.02	11.59	41.98
			GWD	76.15 (+1.32)	71.26 (+1.80)	**45.59** (+3.57)	**11.65** (+0.06)	43.58 (+1.60)
			KLD	**77.92** (+3.09)	**72.77** (+3.31)	43.27 (+1.25)	11.09 (-0.50)	**43.65** (+1.67)
R^3Det		F	Smooth L1	74.28	68.12	35.73	8.01	39.10
			GWD	75.59 (+1.31)	68.36 (+0.24)	40.24 (+4.51)	9.15 (+1.14)	40.80 (+1.70)
			KLD	**77.72** (+2.43)	**71.99** (+3.87)	**43.95** (+8.22)	**10.43** (+2.42)	**43.29** (+4.19)
		R+F	Smooth L1	75.53	69.69	37.69	9.03	40.56
			GWD	77.09 (+1.56)	71.52 (+1.83)	41.08 (+3.39)	10.10 (+1.07)	42.17 (+1.61)
			KLD	**79.63** (+4.63)	**73.30** (+3.61)	**43.51** (+5.82)	**10.61** (+1.58)	**43.61** (+3.05)

➢ **Comparison of Peer Methods:**

Baseline	Method	Box Def.	v1.0 tranval/test								v1.0 train/val			v1.5	v2.0
			BR†	SV†	LV†	SH†	HA†	ST‡	RA‡	7-AP_{50}	AP_{50}	AP_{50}	AP_{75}	$AP_{50:95}$	AP_{50}	AP_{50}
RetinaNet	-	D_{oc}	42.17	65.93	51.11	72.61	53.24	78.38	62.00	60.78	65.73	64.70	32.31	34.50	58.87	44.16
	-	D_{le}	38.31	60.48	49.77	68.29	51.28	78.60	60.02	58.11	64.17	62.21	26.06	31.49	56.10	43.06
	IoU-Smooth L1 [3]	D_{oc}	44.32	63.03	51.25	72.78	56.21	77.98	63.22	61.26	66.99	64.61	34.17	36.23	59.16	46.31
	Modulated Loss [44]	D_{oc}	42.92	67.92	52.91	72.67	53.64	**80.22**	58.21	61.21	66.05	63.50	33.32	34.61	57.75	45.17
	Modulated Loss [44]	Quad.	43.21	70.78	54.70	72.68	**60.99**	79.72	62.08	63.45	67.20	65.15	40.59	**39.12**	**61.42**	**46.71**
	RIL [33]	Quad.	40.81	67.63	55.45	72.42	55.49	78.09	**64.75**	62.09	66.06	64.07	**40.98**	39.05	58.91	45.35
	CSL [4]	D_{le}	42.25	68.28	54.51	72.85	53.10	75.59	58.99	60.80	67.38	64.40	32.58	35.04	58.55	43.34
	DCL (BCL) [45]	D_{le}	41.40	65.82	56.27	73.80	54.30	79.02	60.25	61.55	67.39	**65.93**	35.66	36.71	59.38	45.46
	GWD [5]	D_{oc}	**44.07**	**71.92**	**62.56**	**77.94**	60.25	79.64	63.52	**65.70**	**68.93**	65.44	38.68	38.71	60.03	46.65
	KLD	D_{oc}	**44.00**	**74.45**	**72.48**	**84.30**	**65.54**	**80.03**	**65.05**	**69.41**	**71.28**	**68.14**	**44.48**	**42.15**	**62.50**	**47.69**
R^3Det [27]	-	D_{oc}	44.15	75.09	72.88	86.04	56.49	82.53	61.01	68.31	70.66	67.18	38.41	38.46	62.91	48.43
	DCL (BCL) [45]	D_{le}	**46.84**	74.87	74.96	85.70	57.72	**84.06**	**63.77**	69.70	71.21	67.45	35.44	37.54	61.98	48.71
	GWD [5]	D_{oc}	46.73	**75.84**	**78.00**	**86.71**	**62.69**	**83.09**	61.12	**70.60**	**71.56**	**69.28**	**43.35**	**41.56**	**63.22**	**49.25**
	KLD	D_{oc}	**48.34**	**75.09**	**78.88**	**86.52**	**65.48**	82.08	**61.51**	**71.13**	**71.73**	**68.87**	**44.48**	**42.11**	**65.18**	**50.90**

➢ **Visual Comparison:**



ArXiv Code Homepage Lab