Rethinking Boundary Discontinuity Problem for Oriented Object Detection

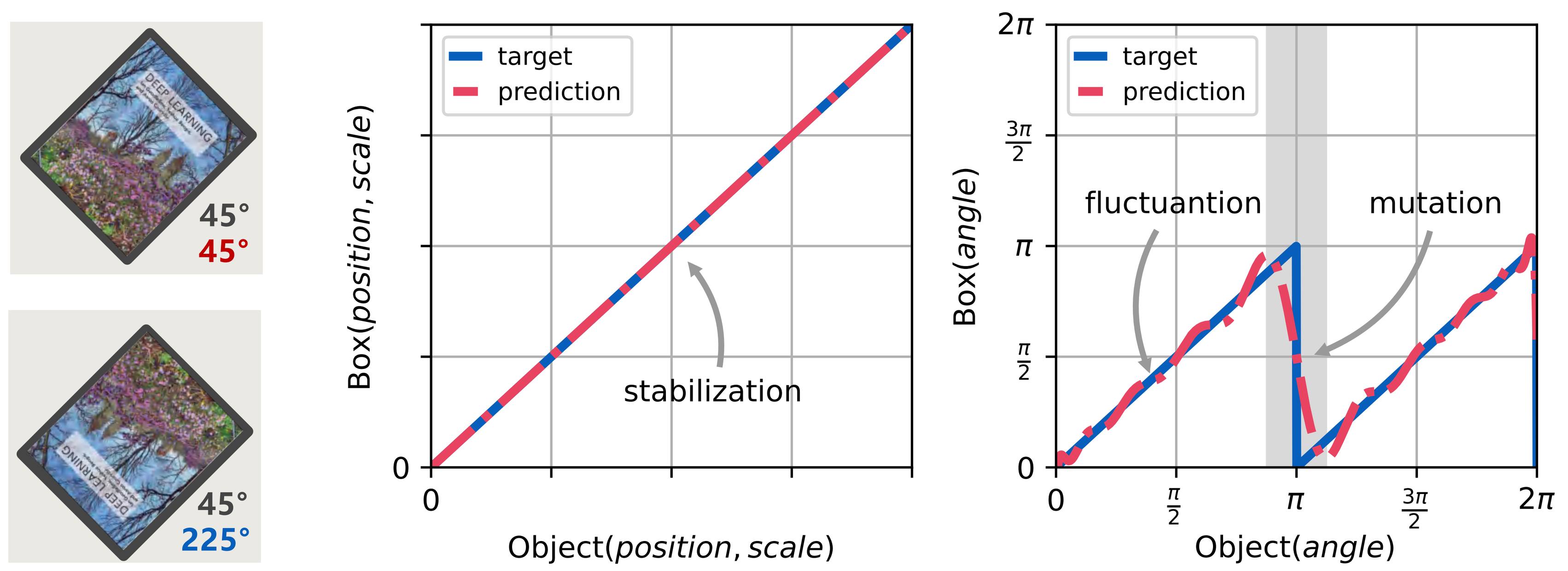
Hang Xu[1,2*] Xinyuan Liu[2*] Haonan Xu[2] Yike Ma[2] Zunjie Zhu[1] Chenggang Yan[1] Feng Dai[2†]

[1] Hangzhou Dianzi University [2] Institute of Computing Technology, Chinese Academy of Sciences * Equal contribution †Corresponding author

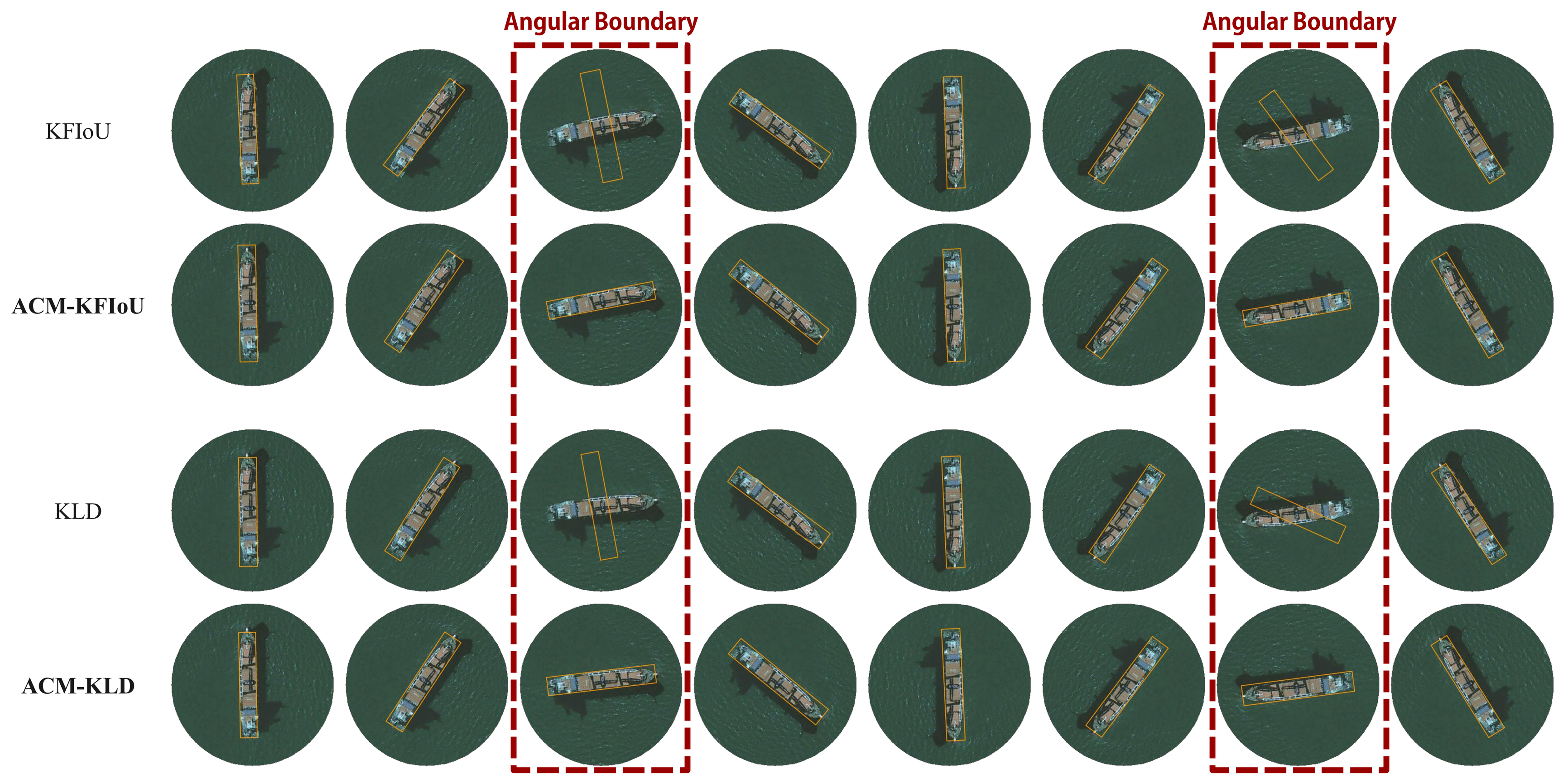
CVPR JUNE 17-21, 2024 SEATTLE, WA

1 Motivation

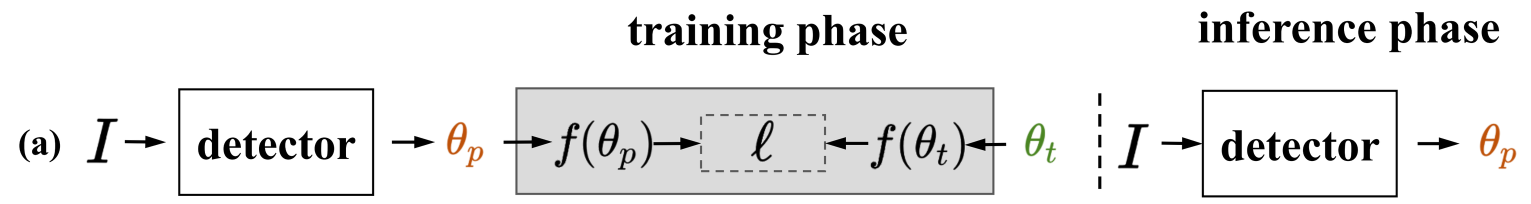
1.1 The Root of All Evil is "Box ≠ Object"



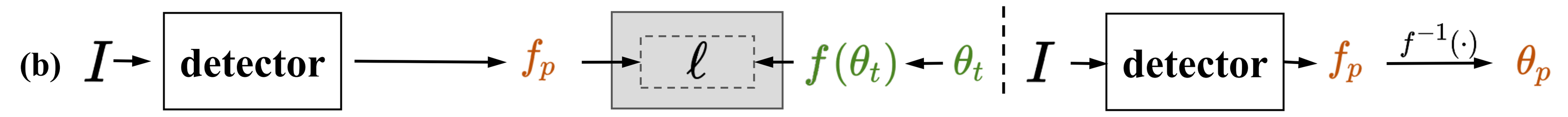
When objects rotate near the boundary angle, SOTA IoU-like methods (e.g., KFIoU, KLD) actually suffer from severe mutation in angular prediction.



1.2 The Devil is in Encoding Mode



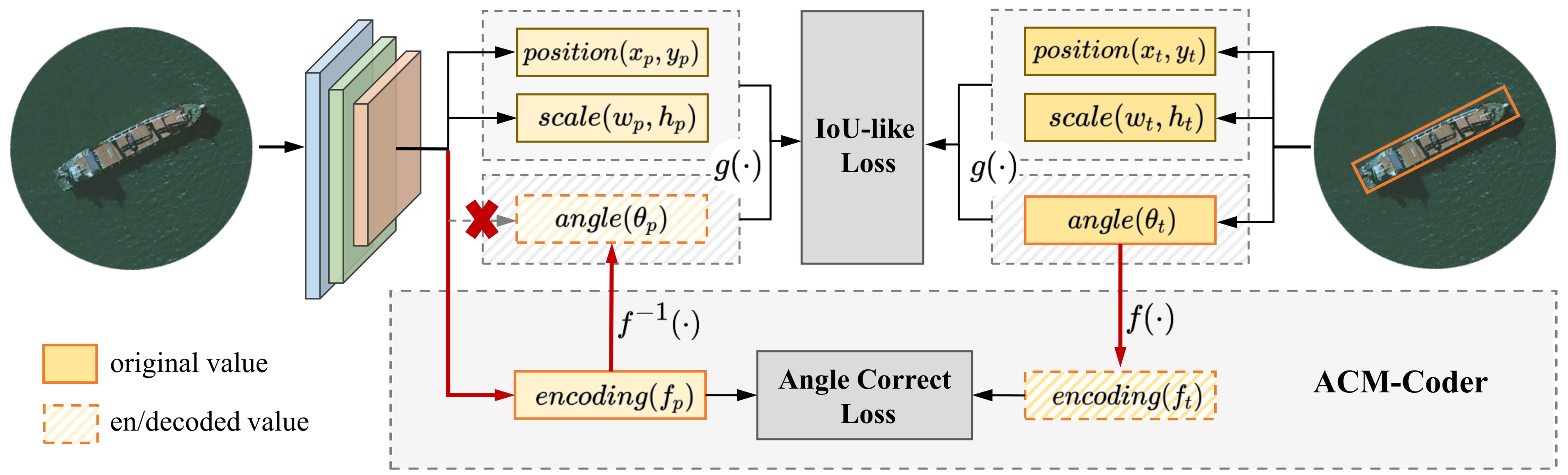
$$\theta = \arg\min_{\theta_p} \ell\big(f(\theta_p); f(\theta_t)\big) \rightarrow \textbf{model fits discontinuous } \theta_p$$



$$\theta = f^{-1}\Big(\arg\min_{f_p} \ell\big(f_p; f(\theta_t)\big)\Big) \rightarrow \textbf{model fits continous } f_p$$

2 Method

2.1 Dual-Optimization Paradigm



2.2 ACM-Coder

To participate in joint-optimization loss calculation, both f and its inverse f^{-1} must be continuous, differentiable, and reversible. We use the simple yet classic complex-exponential transformation for reversible transformation.

$$z = f(\theta) = e^{j\omega\theta} = \cos(\omega\theta) + j\sin(\omega\theta)$$

$$\theta = f^{-1}(z) = -\frac{j}{\omega}\ln z = \frac{1}{\omega}((\operatorname{arctan2}(z_{im}, z_{re}) + 2\pi) \bmod 2\pi)$$

To determine the appropriate ω, we discuss the relationship of $f_{box} \sim f_{obj}$ as following:

$$f_{box} = e^{i\omega\theta_{box}} = e^{i\omega(\theta_{obj} \bmod \pi)} = \begin{cases} e^{i\omega\theta_{obj}}, & \theta_{obj} \in [0, \pi) \\ e^{i\omega\theta_{obj}} \cdot e^{-i\omega\pi}, & \theta_{obj} \in [\pi, 2\pi) \end{cases}$$

1) When $\omega = 1$, $e^{-i\omega\pi} = -1$, then

$$f_{box} = e^{i\omega\theta_{box}} = \begin{cases} e^{i\omega\theta_{obj}}, \theta_{obj} \in [0, \pi) \\ -e^{i\omega\theta_{obj}}, \theta_{obj} \in [\pi, 2\pi) \end{cases} = \begin{cases} f_{obj}, \theta_{obj} \in [0, \pi) \\ -f_{obj}, \theta_{obj} \in [\pi, 2\pi) \end{cases} = f_{obj} \cdot \mathrm{sign}(\pi - \theta_{obj})$$

2) When $\omega = 2$, $e^{-i\omega\pi} = 1$, then

$$f_{box} = e^{i\omega\theta_{box}} = \begin{cases} e^{i\omega\theta_{obj}}, \theta_{obj} \in [0, \pi) \\ e^{i\omega\theta_{obj}}, \theta_{obj} \in [\pi, 2\pi) \end{cases} = f_{obj}$$

Deriving the formula reveals: **2)** At $\omega = 1$, $f_{box} = f_{obj} \cdot \mathrm{sign}(\pi - \theta_{obj}) \neq f_{obj}$; **1)** At $\omega = 2$, $f_{box} = f_{obj}$, aligning with our goals. Thus, we select $\omega = 2$ for ACM. It should be noted that when the object is in the shape of a square, $\omega = 4$.

2.3 Loss Functions

Given a batch of images, the detector outputs the classification c_p, position (x_p, y_p), scale (w_p, h_p), and angular encoding f_p, and the corresponding ground truth is c_t, (x_t, y_t), (w_t, h_t), and θ_t. The total loss is as follows ($\lambda_{box}, \lambda_{acm}$ are coefficients to balance each parts of loss):

$$\mathcal{L}_{tal} = \mathcal{L}_{cls} + \lambda_{box}\mathcal{L}_{box} + \lambda_{acm}\mathcal{L}_{acm},$$

where $\mathrm{L}_{cls} = \ell_{focal}\big(c_p, c_t\big), \mathcal{L}_{box} = \ell\big(B(xywh_p, \theta_p), B(xywh_t, \theta_t)\big), \mathcal{L}_{acm} = \ell_{smooth\_l1}\big(f_p, f_t\big)$

3 Experiment

3.1 Comparison with the State-of-the-Art

Method	MS	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	AP_{50}
PIoU		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
RoI-Trans.	✓	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
O^2-DNet	✓	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	✓	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	✓	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	✓	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
DRN	✓	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
CFC-Net	✓	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
Gliding Vertex		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	✓	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	✓	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	✓	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
R^3Det	✓	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
GWD	✓	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
SCRDet++	✓	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
KFIoU	✓	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
DCL	✓	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	✓	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
PSC	✓	89.86	86.02	54.94	62.02	81.90	85.48	88.39	90.73	86.90	88.82	63.94	69.19	76.84	82.75	63.24	78.07
KLD	✓	88.91	85.23	53.64	81.23	78.20	76.99	84.58	89.50	86.84	86.38	71.69	68.06	75.95	72.23	75.42	78.32
CenterNet-ACM	✓	89.84	85.50	53.84	74.78	80.77	82.81	88.92	90.82	87.18	86.53	64.09	66.27	77.51	79.62	69.57	**78.53**
RoI-Trans.-ACM	✓	85.55	80.53	61.21	75.40	80.35	85.60	88.32	89.88	87.13	87.10	68.15	67.94	78.75	79.82	75.96	**79.45**

3.2 Ablation Study

Typical IoU-like methods are improved to the same level, indicating that the primary distinction between them lies in their optimization capabilities for the angular boundary.

Method	HRSC2016 (Ship) AP_{50}	HRSC2016 (Ship) AP_{75}	UCAS-AOD (Car) AP_{50}	UCAS-AOD (Car) AP_{75}	UCAS-AOD (Plane) AP_{50}	UCAS-AOD (Plane) AP_{75}	DOTA-v1.0 AP_{50}	DOTA-v1.0 AP_{75}
GWD	84.94	61.87	87.25	28.46	90.34	38.22	73.12	34.98
ACM-GWD	90.63 (+5.69)	86.71 (+24.84)	88.69 (+1.44)	29.15 (+0.69)	90.35 (+0.01)	76.00 (+37.78)	73.71 (+0.59)	41.97 (+6.99)
KLD	90.01	79.29	87.54	29.99	90.33	29.19	73.41	35.25
ACM-KLD	90.55 (+0.54)	87.45 (+8.16)	88.76 (+1.22)	30.40 (+0.41)	90.39 (+0.06)	75.65 (+46.46)	73.95 (+0.54)	42.97 (+7.72)
KFIoU	88.26	62.95	85.74	24.44	90.34	16.81	71.97	26.11
ACM-KFIoU	90.55 (+2.29)	87.77 (+24.82)	88.31 (+2.57)	34.81 (+10.37)	90.40 (+0.06)	74.48 (+57.67)	74.51 (+2.54)	40.49 (+14.38)
SkewIoU	89.39	76.43	87.73	27.59	90.34	63.64	73.62	38.01
ACM-SkewIoU	90.47 (+1.08)	88.33 (+11.09)	88.27 (+0.54)	29.13 (+1.74)	90.37 (+0.03)	75.13 (+11.49)	74.21 (+0.59)	42.83 (+4.37)

3.3 Visualized Results

Our ACM greatly eliminates the angular prediction errors in the original KFIoU.