

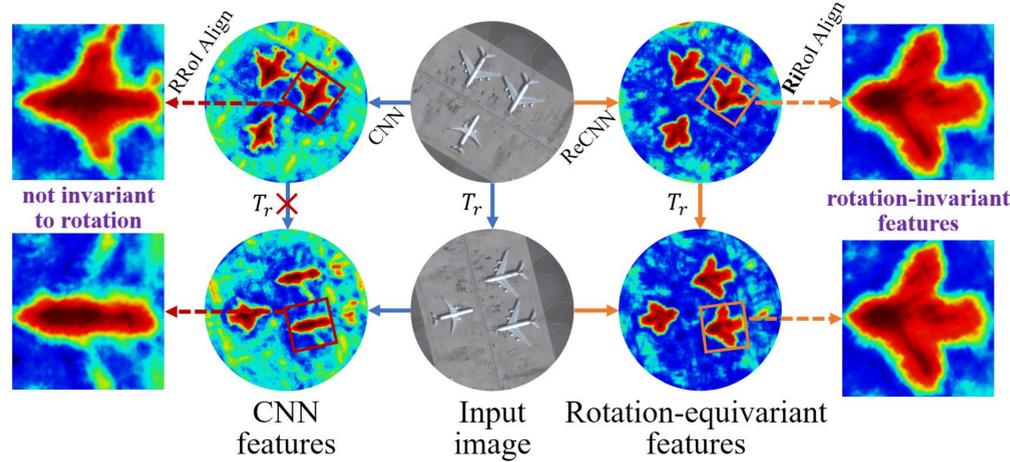


ReDet: A Rotation-equivariant Detector for Aerial Object Detection

Jiaming Han*, Jian Ding*, Nan Xue, Gui-Song Xia†
{hanjiaming, jian.ding, xuenan, guisong.xia}@whu.edu.cn
Wuhan University; *Equal contribution; † Corresponding author



1 Introduction



Background:

- Objects in aerial images are often distributed with arbitrary orientations.
- Aerial object detection is usually formulated as an oriented object detection task, which not only regresses the box size, but also the orientation of an object.
- CNN features are not equivariant to rotation. Therefore, region features warped by RRoI Align are usually unstable and delicate as the orientation changes.
- To encode the orientation information, the detector requires more parameters which are often highly redundant and inefficient.

Contributions:

- propose a Rotation-equivariant Detector (ReDet) for high-quality aerial object detection, which encodes both rotation equivariance and rotation invariance.
- design a novel RiRoI Align to extract rotation-invariant features from rotation-equivariant features..
- achieves the state-of-the-art 80.10, 76.80 and 90.46 mAP on DOTA-v1.0, DOTA-v1.5 and HRSC2016 while reducing the number of parameters by 60%.



Paper

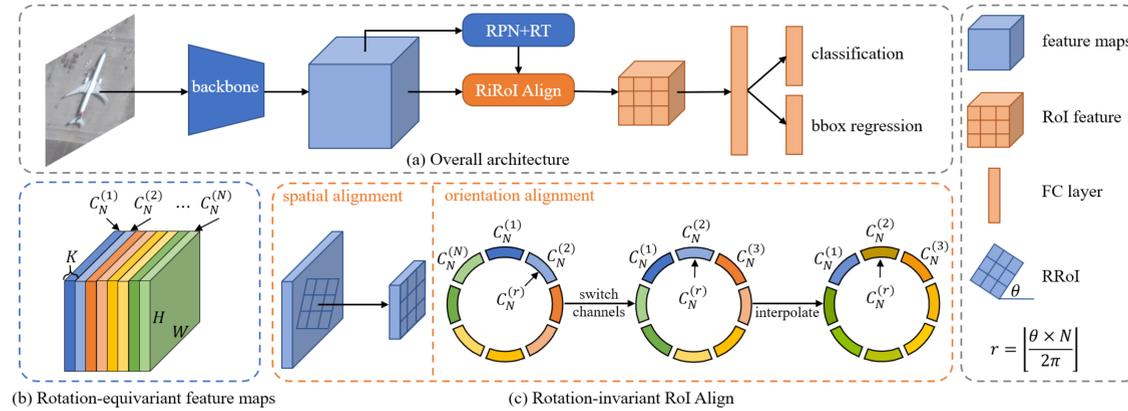


Code



Website

2 Rotation-equivariant Detector



Preliminaries:

For a rotation-equivariant network $\Phi = \{L_i \mid i \in \{1, 2, \dots, M\}\}$, the rotation T_r will be preserved by the whole network:

$$\left[\prod_{i=1}^M L_i \right] (T_r I) = T_r \left[\prod_{i=1}^M L_i \right] (I)$$

For an image region $I_R \in I$ and its rotated version $T_r I_R$, we have:

$$\Phi(T_r I_R) = T_r \Phi(I_R).$$

By applying an inverse rotation T_r' , we get the rotation-invariant representation:

$$\Phi(I_R) = T_r' \Phi(T_r I_R).$$

Rotation-equivariant backbone:

We implement a rotation-equivariant backbone (ReResNet + ReFPN), which is equivariant to discrete rotation group $(\mathbb{R}^2, +) \rtimes C_N$. The produced feature maps f have N orientation channels: $f = \{f^{(i)} \mid i \in \{1, 2, \dots, N\}\}$, and feature maps from each orientation is corresponding to an element in C_N .

Rotation-invariant RoI Align:

We extend the RoI Align to produce completely rotation-invariant features, which includes two parts: (1) spatial alignment, same as RoI Align. (2) orientation alignment: align features from different orientation channels.

$$\hat{f}_R = \text{Int}(SC(f_R, r), \theta), r = \lfloor \theta N / 2\pi \rfloor$$

3 Experiments

➤ Rotation-equivariant backbone

backbone	group	cls. (%)	det. (%)	size (Mb)
R50-FPN	-	76.55	65.03	103
ReR50-ReFPN	C_4	72.81	65.43	24
ReR50-ReFPN	C_8	71.20	66.86	12
ReR50-ReFPN	C_{16}	61.60	64.36	6

method	backbone	mAP (%)	size (Mb)
FR-O	R50-FPN	62.00	158
	ReR50-ReFPN	62.36	68
RetinaNet-O	R50-FPN	58.74	140
	ReR50-ReFPN	59.64	34

➤ Rotation-invariant RoI Align

method	#interpolate	mAP (%)
RRoI Align	-	65.99
RRoI Align+MP.	-	64.60 (-1.39)
RiRoI Align	1	66.44 (+0.45)
RiRoI Align	2	66.86 (+0.87)
RiRoI Align	4	66.32 (+0.33)

➤ vs. rotation augmentation

method	rot.	schd.	mAP (%)	training (h)
ReDet	×	1x	62.62	8
baseline	✓	1x	64.07	11
ReDet*	×	1x	66.66	13
baseline	✓	2x	67.34	22

➤ vs. SOTA on DOTA-v1.0

method	backbone	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
single-scale:																	
FR-O [35]	R101	79.42	77.13	17.70	64.05	35.30	38.02	37.16	89.41	69.64	59.28	50.30	52.91	47.89	47.40	46.30	54.13
ICN [11]	R101-FPN	81.36	74.30	47.70	70.32	64.89	67.82	69.98	90.76	79.06	78.20	53.64	62.90	67.02	64.17	50.23	68.16
CADNet [42]	R101-FPN	87.80	82.40	49.40	73.50	71.10	63.50	76.60	90.90	79.20	73.30	48.40	60.90	62.00	67.00	62.20	69.90
DRN [24]	H-104	88.91	80.22	43.52	63.35	73.48	70.69	84.94	90.14	83.85	84.11	50.12	58.41	67.62	68.60	52.50	70.70
CenterMap [30]	R50-FPN	88.88	81.24	53.15	60.65	78.62	66.55	78.10	88.83	77.80	83.61	49.36	66.19	72.10	72.36	58.70	71.74
SCRDet [40]	R101-FPN	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
R ³ Det [37]	R152-FPN	89.49	81.17	50.53	66.10	70.92	78.66	78.21	90.81	85.26	84.23	61.81	63.77	68.16	69.83	67.17	73.74
S ² A-Net [10]	R50-FPN	89.11	82.84	48.37	71.11	78.11	78.39	87.25	90.83	84.90	85.64	60.36	62.60	65.26	69.13	57.94	74.12
ReDet (Ours)	ReR50-ReFPN	88.79	82.64	53.97	74.00	78.13	84.06	88.04	90.89	87.78	85.75	61.76	60.39	75.96	68.07	63.59	76.25
multi-scale:																	
RoI Trans. [7]	R101-FPN	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* [31]	H104	89.30	83.30	50.10	72.10	71.10	75.60	78.70	90.90	79.90	82.90	60.20	60.00	64.60	68.90	65.70	72.80
DRN* [24]	H104	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
Gliding Vertex* [36]	R101-FPN	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
BBAVectors* [41]	R101	88.63	84.06	52.13	69.56	78.26	80.40	88.06	90.87	87.23	86.39	56.11	65.62	67.10	72.08	63.96	75.36
CenterMap* [30]	R101-FPN	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* [38]	R152-FPN	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
SCRDet++* [39]	R152-FPN	88.68	85.22	54.70	73.71	71.92	84.14	79.39	90.82	87.04	86.02	67.90	60.86	74.52	70.76	72.66	76.56
S ² A-Net* [10]	R50-FPN	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
ReDet* (Ours)	ReR50-ReFPN	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

➤ vs. SOTA on HRSC2016

method	RC2 [19]	RRPN [22]	R ² PN [43]	RRD [16]	RoI Trans. [7]	Gliding Vertex [36]
mAP	75.7	79.08	79.6	84.3	86.2	88.2
method	R ³ Det [37]	DRN [24]	CenterMap [30]	CSL [38]	S ² A-Net [10]	ReDet (Ours)
mAP	89.26	92.7*	92.8*	89.62	90.17 / 95.01*	90.46 / 97.63*

➤ Visualization

