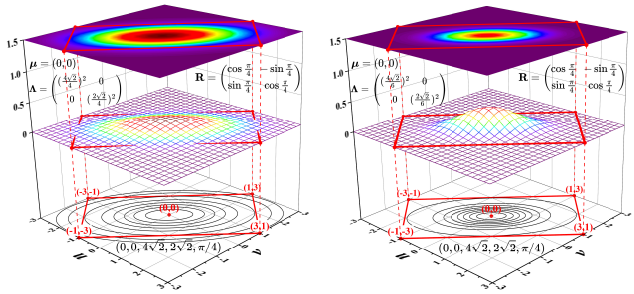
ReDiffDet: Rotation-equivariant Diffusion Model for Oriented Object Detection

Supplementary Material

1. Oriented box→2D Gaussian

Different scaling factors m will transform an oriented box into different scaling distributions, as shown in Fig 1. For example, if approximately 99.7% of the samples of a Gaussian distribution lie within a box, then six standard deviations correspond to the widths or heights according to the 68-95-99.7 rule, making $m = 6$.



(a) m=4 (b) m=6

Figure 1. Illustration of converting an oriented box to a Gaussian distribution under m=4 and m=6. Here the example oriented box is $(cx, cy, w, h, a) = (0, 0, 4\sqrt{2}, 2\sqrt{2}, \frac{\pi}{4})$.

2. Reverse Process

Time 1<t<T. Inspired by [4], DDIM sampling strategy [34] is adopted, which allows for much faster sampling. When $1 < t < T$, the reverse process is defined as:

$$p_\theta(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t) := \mathcal{N}(\boldsymbol{z}_{t-1}; \boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t), \boldsymbol{\Sigma}_\theta(\boldsymbol{z}_t, t)) \tag{1}$$

where

$$\begin{aligned}\boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t) &= \sqrt{\bar{\alpha}_{t-1}}\left(\frac{\boldsymbol{z}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\boldsymbol{z}_t, t)}{\sqrt{\bar{\alpha}_t}}\right) \\ &\quad + \sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}\epsilon_\theta(\boldsymbol{z}_t, t) \qquad (2)\\ \boldsymbol{\Sigma}_\theta(\boldsymbol{z}_t, t) &= \sigma_t^2\mathbf{I} \qquad (3)\end{aligned}$$

Formally, let $\mathbf{R}$ be the rotation transformation where $\mathbf{R} \in SO(2)$, we have:

Proposition 2. *During $1 < t < T$ in the reverse process p_θ, the function $p_\theta(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t)$ is rotation equivariant, i.e., $p_\theta(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t) = p_\theta(\mathbf{R}(\boldsymbol{z}_{t-1})|\mathbf{R}(\boldsymbol{z}_t))$ under $SO(2)$, as long as $\mathbf{R}(\boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t)) = \boldsymbol{\mu}_\theta(\mathbf{R}(\boldsymbol{z}_t), t)$. Proof:*

$$\begin{aligned}&p_\theta(\mathbf{R}(\boldsymbol{z}_{t-1})|\mathbf{R}(\boldsymbol{z}_t)) \\ &= \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}_\theta(\mathbf{R}(\boldsymbol{z}_t), t)|}}\exp\{-\frac{1}{2}(\mathbf{R}(\boldsymbol{z}_{t-1}) - \boldsymbol{\mu}_\theta(\mathbf{R}(\boldsymbol{z}_t), t))^\top \\ &\quad \cdot \boldsymbol{\Sigma}_\theta(\mathbf{R}(\boldsymbol{z}_t), t)^{-1} \cdot (\mathbf{R}(\boldsymbol{z}_{t-1}) - \boldsymbol{\mu}_\theta(\mathbf{R}(\boldsymbol{z}_t), t))\} \\ &= (2\pi\sqrt{|\sigma_t^2\mathbf{I}|})^{-1}\exp\{-\frac{1}{2}(\mathbf{R}(\boldsymbol{z}_{t-1}) - \mathbf{R}(\boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t)))^\top \\ &\quad \cdot (1/\sigma_t^2)\mathbf{I} \cdot (\mathbf{R}(\boldsymbol{z}_{t-1}) - \mathbf{R}(\boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t)))\} \\ &= (2\pi\sqrt{|\sigma_t^2\mathbf{I}|})^{-1}\exp\{-\frac{1}{2}(\mathbf{R}(\boldsymbol{z}_{t-1} - \boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t)))^\top \\ &\quad \cdot (1/\sigma_t^2)\mathbf{I} \cdot \mathbf{R}(\boldsymbol{z}_{t-1} - \boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t))\} \\ &= (2\pi\sqrt{|\sigma_t^2\mathbf{I}|})^{-1}\exp\{-\frac{1}{2}(\boldsymbol{z}_{t-1} - \boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t))^\top\mathbf{R}^\top \\ &\quad \cdot (1/\sigma_t^2)\mathbf{I} \cdot \mathbf{R}(\boldsymbol{z}_{t-1} - \boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t))\} \\ &= (2\pi\sqrt{|\sigma_t^2\mathbf{I}|})^{-1}\exp\{-\frac{1}{2}(\boldsymbol{z}_{t-1} - \boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t))^\top \\ &\quad \cdot (1/\sigma_t^2)\mathbf{I} \cdot (\boldsymbol{z}_{t-1} - \boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t))\} \\ &= p_\theta(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t) \qquad (4)\end{aligned}$$

Time t=1. The $p(\boldsymbol{z}_0)$ can be derived from the previous state. The $p(\boldsymbol{z}_0)$ is calculated as:

$$p_\theta(\boldsymbol{z}_0) = \int p_\theta(\boldsymbol{z}_{0:T})\mathrm{d}\boldsymbol{z}_{1:T} = \int p(\boldsymbol{z}_T)\prod_{t=1}^{T} p_\theta(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t)\mathrm{d}\boldsymbol{z}_{1:T} \tag{5}$$

Formally, let $\mathbf{R}$ be the rotation transformation where $\mathbf{R} \in SO(2)$, we have:

Proposition 3. *At time $t = 1$ in the reverse process p_θ, given the rotation invariant $p(\boldsymbol{z}_T)$, i.e., $p(\boldsymbol{z}_T) = p(\mathbf{R}(\boldsymbol{z}_T))$, and the rotation equivariant $p_\theta(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t)$, i.e., $p_\theta(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t) = p_\theta(\mathbf{R}(\boldsymbol{z}_{t-1})|\mathbf{R}(\boldsymbol{z}_t))$, then the $p_\theta(\boldsymbol{z}_0)$ is also rotation invariant, i.e., $p_\theta(\boldsymbol{z}_0) = p_\theta(\mathbf{R}(\boldsymbol{z}_0))$ under $SO(2)$.* Because:

$$\begin{aligned}p_\theta(\mathbf{R}(\boldsymbol{z}_0)) &= \int p(\mathbf{R}(\boldsymbol{z}_T))\textstyle\prod_{t=1}^{T} p_\theta(\mathbf{R}(\boldsymbol{z}_{t-1})|\mathbf{R}(\boldsymbol{z}_t))\mathrm{d}\boldsymbol{z}_{1:T} \\ &= \int p(\boldsymbol{z}_T)\textstyle\prod_{t=1}^{T} p_\theta(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t)\mathrm{d}\boldsymbol{z}_{1:T} \\ &= p_\theta(\boldsymbol{z}_0) \qquad (6)\end{aligned}$$

3. Training loss

In oriented object detection, there are two subtasks, classification for categories and regression for positions of objects. The losses $\mathcal{L}$ consist of the Focal loss $\mathcal{L}_{cls}$, ℓ_1 loss $\mathcal{L}_{L1}$ and rotate IoU loss $\mathcal{L}_{riou}$:

$$\mathcal{L} = \lambda_{cls}\mathcal{L}_{cls} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{riou}\mathcal{L}_{riou}, \tag{7}$$

where $\mathcal{L}_{cls}$, $\mathcal{L}_{L1}$, and $\mathcal{L}_{riou}$ are coefficient of corresponding losses. We adopt $\mathcal{L}_{cls} = 2.0$, $\mathcal{L}_{L1} = 2.0$, and $\mathcal{L}_{riou} = 5.0$.

4. Training configuration

The training configuration is in Tab. 1.

Config	Value
optimizer	AdamW
base learning rate	4e-5
weight decay	1e-4
optimizer momentum	β_1, β_2=0.9, 0.999
batch size	4 (2 images per GPU)
GPUs	2 NVIDA 2080ti
epochs	24
lr decay epochs	(16, 22)
warmup iter	500
warmup factor	0.333
clip gradient type	full model
clip gradient value	1.0
clip gradiant norm	2.0
data augmentation	only RandomFlip
seed	Random Seed

Table 1. DIOR-R setting.

5. Dataset

DOTA-v1.0 has 15 common categories: plane (PL), baseball diamond (BD), bridge (BR), ground track field (GTF), small vehicle (SV), large vehicle (LV), ship (SH), tennis court (TC), basketball court (BC), storage tank (ST), soccer-ball field (SBF), roundabout (RA), harbor (HA), swimming pool (SP), and helicopter (HC).

DIOR-R has 20 common categories: airplane (APL), airport (APO), baseball field (BF), basketball court (BC), bridge (BR), chimney (CH), expressway service area (ESA), expressway toll station (ETS), dam (DAM), golf field (GF), ground track field (GTF), harbor (HA), overpass (OP), ship (SH), stadium (STA), storage tank (STO), tennis court (TC), train station (TS), vehicle (VE), and windmill (WM).

6. Nomenclature

To facilitate clarity, we present a summary of symbols along with their corresponding descriptions as utilized in this study, encapsulated in Tab. 2.

7. Main results

Results on DIOR-R. The detailed results of every category on the DIOR-R are reported in Tab 3.

Notation	Description
$\boldsymbol{z}_t$	random variable
$t \in \{1, ..., T\}$	time steps
$q(\cdot\|\cdot)$	diffusion process
$p_\theta(\cdot\|\cdot)$	reverse process
$\boldsymbol{\mu}$	mean
$\boldsymbol{\Sigma}$	variance
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gaussian distribution
$\mathcal{N}(\mathbf{0}, \mathbf{I})$	standard Gaussian distribution
$\mathbf{0}$	zero matrix
$\mathbf{I}$	identity matrix
(cx, cy)	center coordinate
(w, h)	width and height
a	angle
$\mathbf{R}$	rotation matrix
$\mathbf{\Lambda}$	diagonal matrix
$SO(2)$	2D rotation group
$\mathbb{R}^2$	2D space
$\det(\cdot)$	determinant
G	a group
$\rho : X \to Y$	a function
$S_g^{(\cdot)}$	group action
m	scaling factor
$\exp(\cdot)$	exponential function
$\mathrm{PDF}(\cdot, \cdot)$	probability density function
P=$\{\mathbf{P_1}, ..., \mathbf{P}_k\}$	samples
k	number of samples
(u_1, v_1)	a sample point
$\lambda, \lambda_1, \lambda_2$	eigenvalues
i	index
α	variance schedule
β	variance schedule
σ_t	variance schedule
$\epsilon_\theta(\boldsymbol{z}_t, t)$	model
θ	model parameter
$\mathbb{E}(\cdot)$	mean
$\mathbb{V}(\cdot)$	variance

Table 2. The nomenclature with related notations.

Method	backbone	APL	APO	BF	BC	BR	CH	DAM	ETS	ESA	GF	GTF	HA	OP	SH	STA	STO	TC	TS	VE	WM	AP_{50}
One-stage																						
RetinaNet-O [26]	ResNet50	61.49	28.52	73.57	81.17	23.98	72.54	19.94	72.39	58.20	69.25	79.54	32.14	44.87	77.71	67.57	61.09	81.46	47.33	38.01	60.24	57.55
Oriented Rep [22]	ResNet50	70.03	46.11	76.12	87.19	39.14	78.76	34.57	71.80	80.42	76.16	79.41	45.48	54.90	87.82	77.03	68.07	81.60	56.83	51.57	**71.25**	66.71
DCFL [40]	ResNet50	68.60	53.10	76.70	87.10	42.10	78.60	34.50	71.50	80.80	**79.70**	79.50	47.30	**57.40**	85.20	64.60	66.40	81.50	58.90	50.90	70.90	66.80
Two-stage																						
Gliding Vertex [42]	ResNet50	65.35	28.87	74.96	81.33	33.88	74.31	19.58	70.72	64.70	72.30	78.68	37.22	49.64	80.22	69.26	61.13	81.49	44.76	47.71	65.04	60.06
RoI Transformer [9]	ResNet50	63.34	37.88	71.78	87.53	40.68	72.60	26.86	**78.71**	68.09	68.96	**82.74**	**47.71**	55.61	81.21	**78.23**	70.26	81.61	54.86	43.27	65.52	63.87
AOPG [5]	ResNet50	62.39	37.79	71.62	87.63	40.90	72.47	31.08	65.42	77.99	73.20	81.94	42.32	54.45	81.17	72.69	71.31	81.49	60.04	52.38	69.99	64.41
End-to-End																						
ARS-DETR [49]	ResNet50	68.00	**54.17**	74.43	81.65	41.13	75.66	**34.89**	73.07	81.92	76.10	78.62	36.33	55.41	84.55	70.09	72.23	81.14	**61.52**	50.57	70.28	66.12
OrientedFormer [52]	ResNet50	65.65	48.69	**78.79**	87.17	41.90	76.34	34.37	72.14	81.40	75.34	79.83	45.15	56.12	88.66	67.59	72.68	87.32	60.31	56.54	69.56	67.28
Diffusion Model																						
DiffusionDet-O [4]	ResNet50	58.84	24.25	70.10	78.93	21.20	72.25	21.93	53.19	53.82	56.26	74.26	1.71	37.82	53.01	62.66	50.48	80.36	27.92	37.44	61.80	49.91
ReDiffDet (ours)	PKINet-T	66.69	42.46	77.02	84.74	42.33	74.14	30.06	69.15	79.12	71.03	79.76	39.57	56.44	88.78	73.37	75.72	85.26	53.47	56.39	63.92	65.47
ReDiffDet (ours)	LSK-T	70.81	44.08	76.03	85.29	43.66	75.90	31.81	71.83	81.75	71.52	81.42	41.82	56.91	89.27	75.25	75.32	87.42	54.43	**58.22**	59.69	66.62
ReDiffDet (ours)	ReR50	**71.36**	49.22	71.65	**87.88**	**47.12**	**79.28**	33.35	73.37	**83.74**	70.29	80.38	43.63	57.17	**89.52**	72.39	**79.81**	**89.03**	57.34	57.32	67.23	**68.05**

Table 3. Experimental results on **DIOR-R** dataset.