视觉有向目标的高精度检测

VALSE 2023 无锡

杨学

yangxue-2019-sjtu@sjtu.edu.cn

https://yangxue0827.github.io/
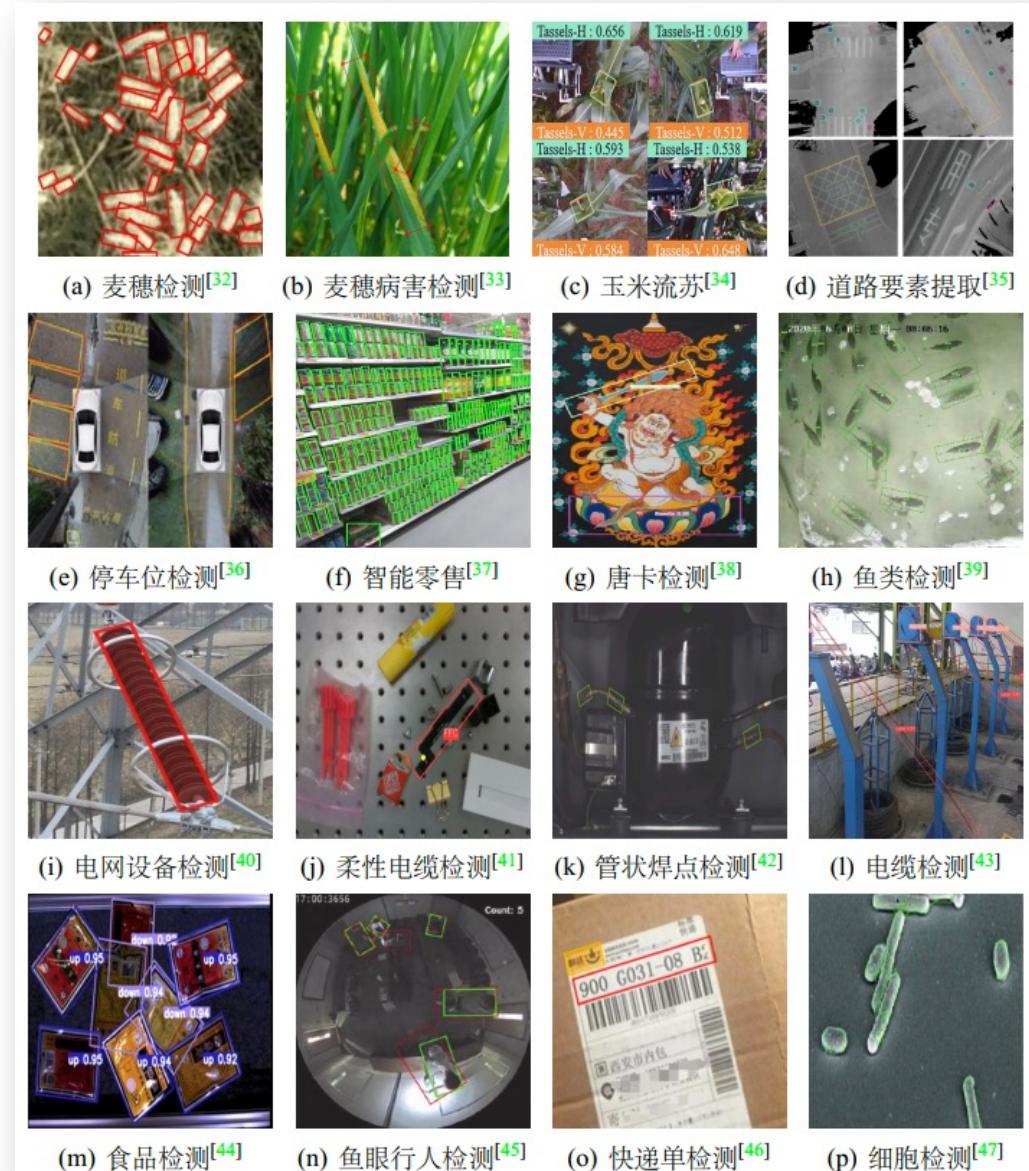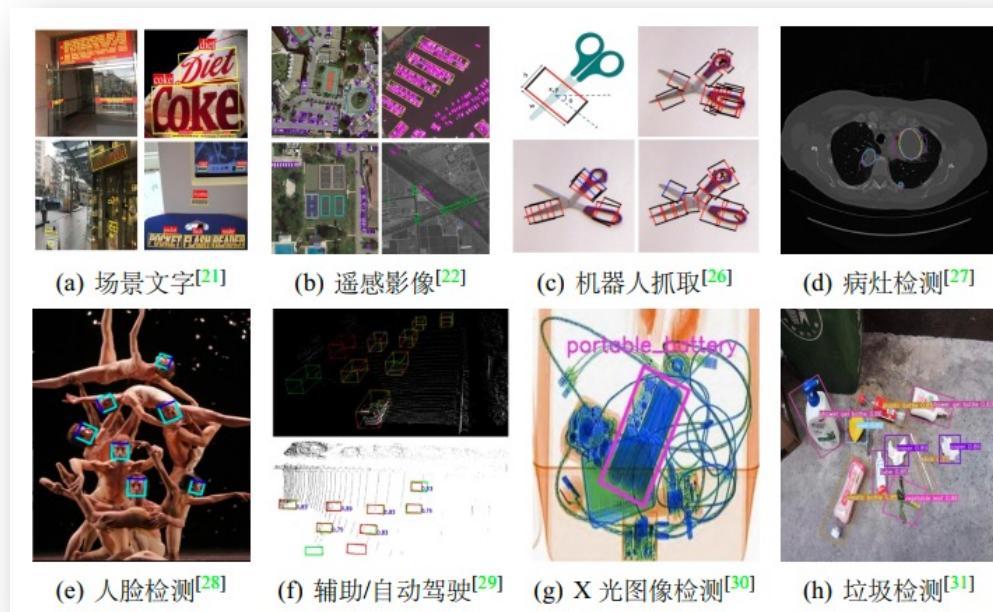
目标检测

有向目标检测

- 遥感目标检测
- 场景文字检测
- 人脸检测
- 智能零售
- 3D目标检测
- 机器人抓取
- 农业检测
- ……

(a) 场景文字[21]　(b) 遥感影像[22]　(c) 机器人抓取[26]　(d) 病灶检测[27]

(e) 人脸检测[28]　(f) 辅助/自动驾驶[29]　(g) X光图像检测[30]　(h) 垃圾检测[31]

(a) 麦穗检测[32]　(b) 麦穗病害检测[33]　(c) 玉米流苏[34]　(d) 道路要素提取[35]

(e) 停车位检测[36]　(f) 智能零售[37]　(g) 唐卡检测[38]　(h) 鱼类检测[39]

(i) 电网设备检测[40]　(j) 柔性电缆检测[41]　(k) 管状焊点检测[42]　(l) 电缆检测[43]

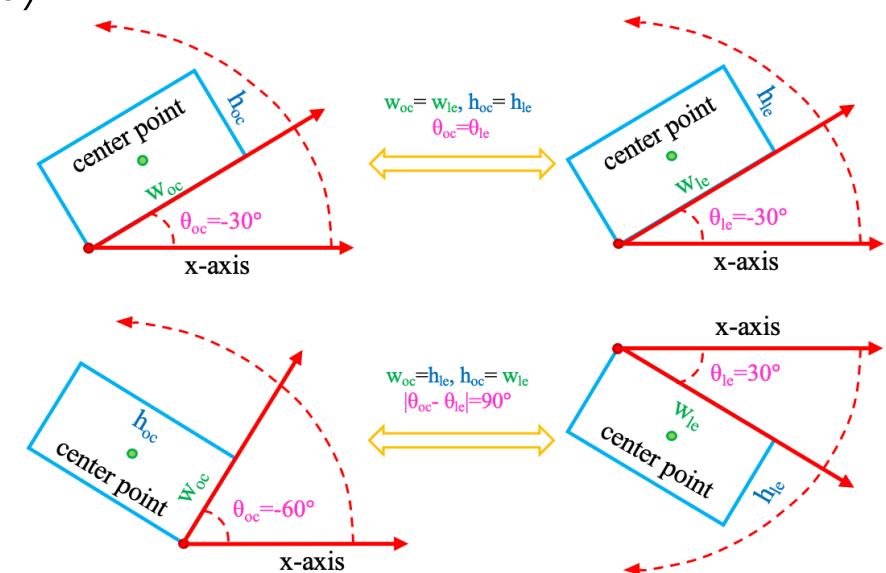(m) 食品检测[44]　(n) 鱼眼行人检测[45]　(o) 快递单检测[46]　(p) 细胞检测[47]

旋转框的定义

- **旋转框的定义方式**
  - **OpenCV 定义法**：$(x, y, w_{oc}, h_{oc}, \theta_{oc})$, $\theta_{oc} \in [-90, 0)$
  - **长边定义法**：$(x, y, w_{le}, h_{le}, \theta_{le})$, $\theta_{le} \in [-90, 90)$

- **转换关系**

$$D_{le}(w_{le}, h_{le}, \theta_{le}) = \begin{cases} D_{oc}(w_{oc}, h_{oc}, \theta_{oc}), & w_{oc} \geq h_{oc} \\ D_{oc}(h_{oc}, w_{oc}, \theta_{oc} + 90°), & otherwise \end{cases}$$
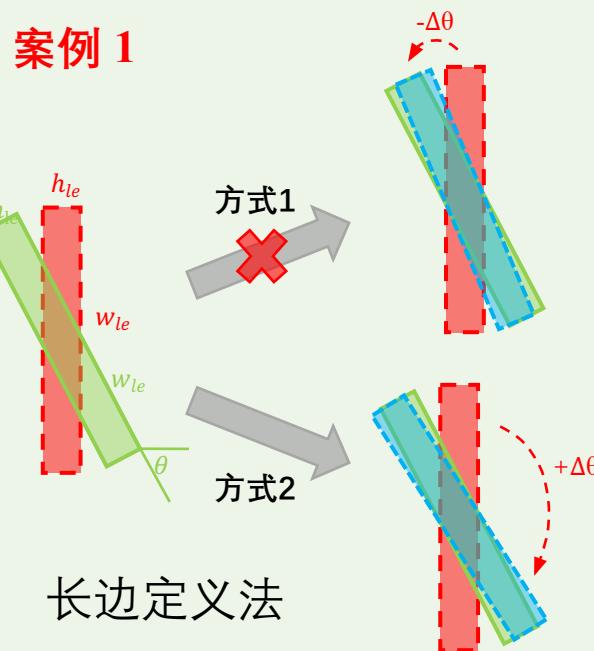
$$D_{oc}(w_{oc}, h_{oc}, \theta_{oc}) = \begin{cases} D_{le}(w_{le}, h_{le}, \theta_{le}), & \theta_{le} \in [-90°, 0°) \\ D_{le}(h_{le}, w_{le}, \theta_{le} - 90°), & otherwise \end{cases}$$

问题与挑战

● 边界不连续问题

案例 1

方式1 ❌

Anchor/Proposal: $(0,0,70,10,-90°)$
Ground-Truth: $(0,0,70,10,65°)$
Predict box: $(0,0,70,10,-115°)$

$w = w, h = h, |\theta - \theta| = 180°$
IoU< G, P >≈ 1
Smooth-L1 Loss< G, P >PoA≫ 0

方式2
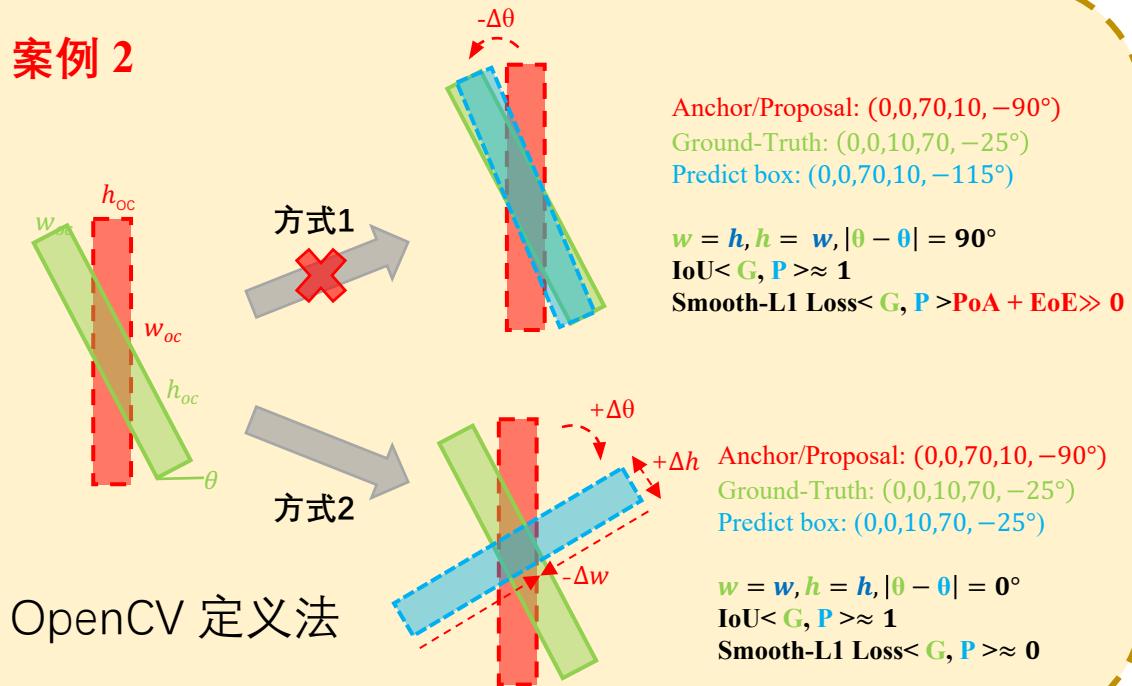
Anchor/Proposal: $(0,0,70,10,-90°)$
Ground-Truth: $(0,0,70,10,65°)$
Predict box: $(0,0,70,10,65°)$

$w = w, h = h, |\theta - \theta| = 0°$
IoU< G, P >≈ 1
Smooth-L1 Loss< G, P > ≈ 0

长边定义法

案例 2

方式1 ❌

Anchor/Proposal: $(0,0,70,10,-90°)$
Ground-Truth: $(0,0,10,70,-25°)$
Predict box: $(0,0,70,10,-115°)$
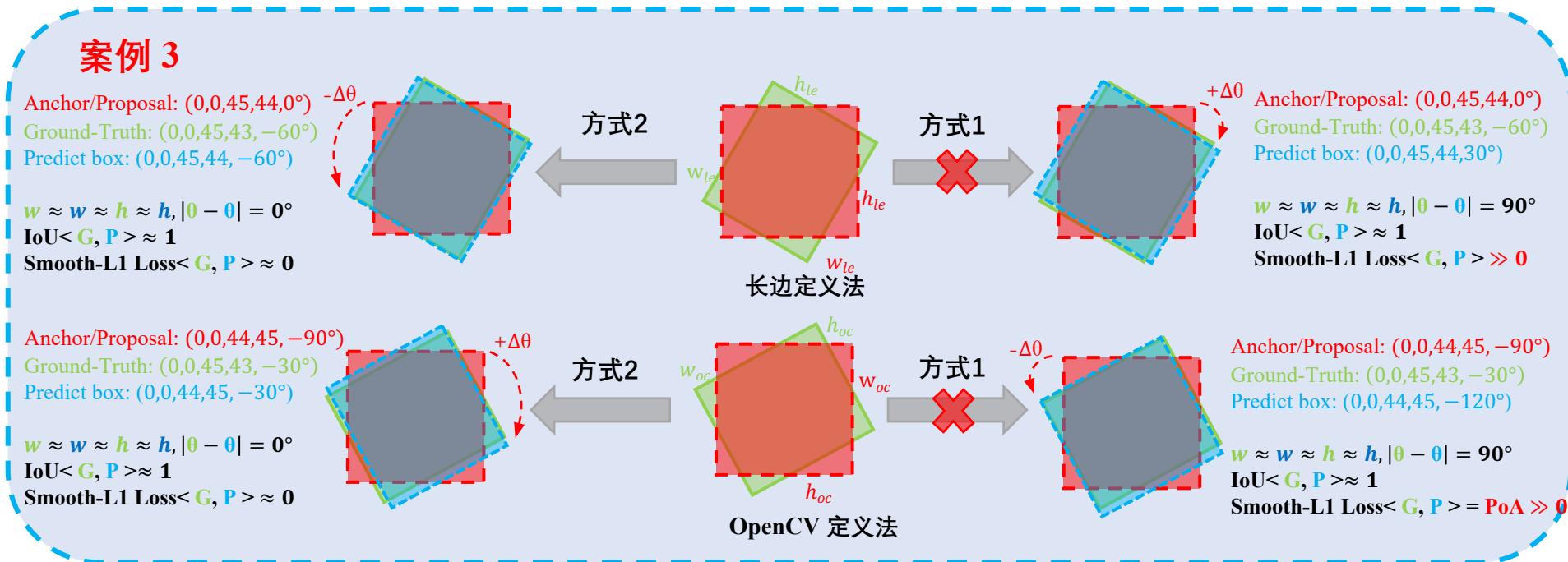
$w = h, h = w, |\theta - \theta| = 90°$
IoU< G, P >≈ 1
Smooth-L1 Loss< G, P >PoA + EoE≫ 0

方式2

Anchor/Proposal: $(0,0,70,10,-90°)$
Ground-Truth: $(0,0,10,70,-25°)$
Predict box: $(0,0,10,70,-25°)$

$w = w, h = h, |\theta - \theta| = 0°$
IoU< G, P >≈ 1
Smooth-L1 Loss< G, P > ≈ 0

OpenCV 定义法

问题与挑战

- ## 类正方形检测问题

案例 3

Anchor/Proposal: $(0,0,45,44,0°)$
Ground-Truth: $(0,0,45,43,-60°)$
Predict box: $(0,0,45,44,-60°)$

$w \approx w \approx h \approx h, |\theta - \theta| = 0°$
IoU< G, P > $\approx$ 1
Smooth-L1 Loss< G, P > $\approx$ 0

方式2

$-\Delta\theta$

h_{le}

w_{le}

h_{le}

w_{le}

长边定义法

方式1

$+\Delta\theta$

Anchor/Proposal: $(0,0,45,44,0°)$
Ground-Truth: $(0,0,45,43,-60°)$
Predict box: $(0,0,45,44,30°)$

$w \approx w \approx h \approx h, |\theta - \theta| = 90°$
IoU< G, P > $\approx$ 1
Smooth-L1 Loss< G, P > $\gg$ 0

Anchor/Proposal: $(0,0,44,45,-90°)$
Ground-Truth: $(0,0,45,43,-30°)$
Predict box: $(0,0,44,45,-30°)$

$w \approx w \approx h \approx h, |\theta - \theta| = 0°$
IoU< G, P > $\approx$ 1
Smooth-L1 Loss< G, P > $\approx$ 0

方式2

$+\Delta\theta$

h_{oc}

w_{oc}

w_{oc}

h_{oc}

OpenCV 定义法

方式1

$-\Delta\theta$

Anchor/Proposal: $(0,0,44,45,-90°)$
Ground-Truth: $(0,0,45,43,-30°)$
Predict box: $(0,0,44,45,-120°)$

$w \approx w \approx h \approx h, |\theta - \theta| = 90°$
IoU< G, P > $\approx$ 1
Smooth-L1 Loss< G, P > = PoA $\gg$ 0

长宽比越小，IoU对角度越不敏感

问题与挑战

● 评估与损失不一致问题
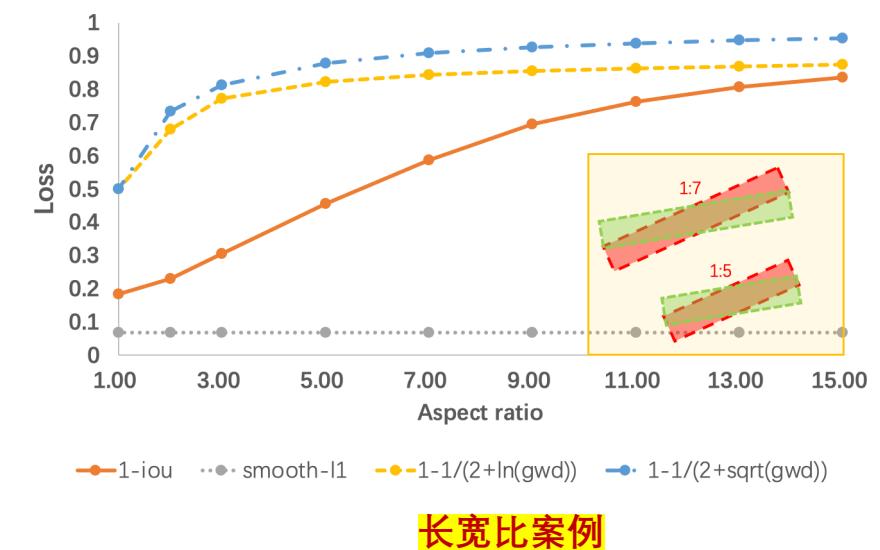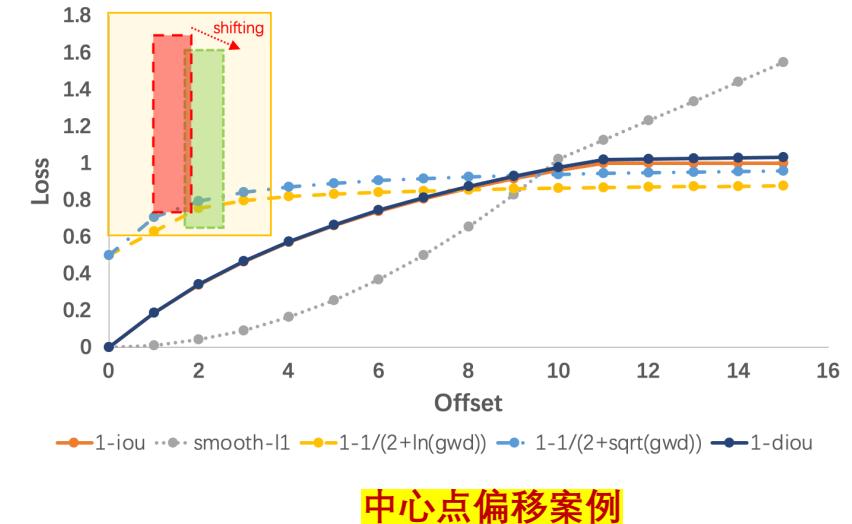

角度差案例

SkewIoU Loss

SkewGIoU Loss

中心点偏移案例

长宽比案例
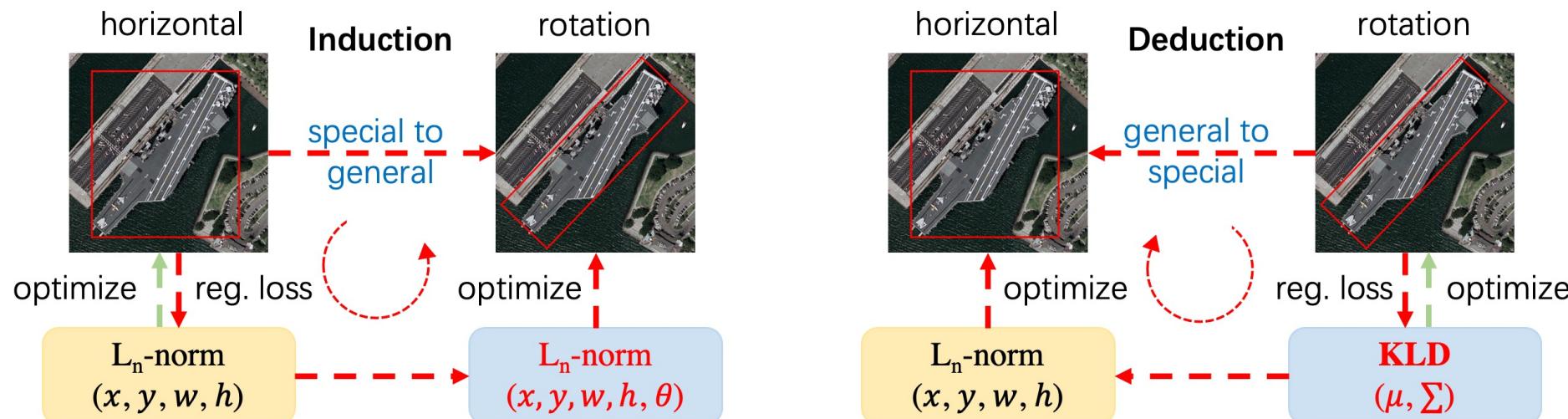
模型设计范式

- 有向目标检测器的两种设计范式

  - 归纳范式 (Induction)

  - 演绎范式 (Deduction)

X. Yang, et al. "Rethinking Rotated Object Detection with Gaussian Wasserstein Distance Loss." In ICML 2021.
X. Yang, et al. "Learning High-Precision Bounding Box for Rotated Object Detection via Kullback-Leibler Divergence." In NeurIPS 2021.

模型设计范式

- 归纳范式 (Induction)

  - 对于常见的通用检测模型（水平框检测），模型通常是通过回归四个偏移量的形式来进行框位置和大小的预测：

$$t_x^p = \frac{x_p - x_a}{w_a}, t_y^p = \frac{y_p - y_a}{h_a}, t_w^p = \ln\left(\frac{w_p}{w_a}\right), t_h^p = \ln\left(\frac{h_p}{h_a}\right)$$

  来匹配实际的偏移量：

$$t_x^t = \frac{x_t - x_a}{w_a}, t_y^t = \frac{y_t - y_a}{h_a}, t_w^t = \ln\left(\frac{w_t}{w_a}\right), t_h^t = \ln\left(\frac{h_t}{h_a}\right)$$

  - 借鉴于此，目前绝大多数的旋转目标检测在上面的基础上加上了角度参数的回归：

$$t_\theta^p = f(\theta_p - \theta_a), t_\theta^t = f(\theta_t - \theta_a)$$

模型设计范式

● 归纳范式 (Induction)

  ● 回归损失采用L_n-norm：

$$L_{reg} = l_n\text{-norm}\left(\Delta t_x, \Delta t_y, \Delta t_w, \Delta t_h, \Delta t_\theta\right)$$

  where $\Delta t_x = t_x^p - t_x^t = \frac{\Delta x}{w_a}$, $\Delta t_y = t_y^p - t_y^t = \frac{\Delta y}{h_a}$, $\Delta t_w = t_w^p - t_w^t = \ln(w_p/w_t)$, $\Delta t_h = t_h^p - t_h^t = \ln(h_p/h_t)$, and $\Delta t_\theta = t_\theta^p - t_\theta^t = \Delta\theta$.

  ● 五个参数的优化和目标本身形状等关联不大，使得我们需要根据不同的数据集特点进行权重的调整。比如大长宽比目标可能需要着重关注角度参数，小目标则需要关注中心点参数，因此这些参数的轻微偏移都会造成这些目标预测精准度（IoU）的急剧下降。
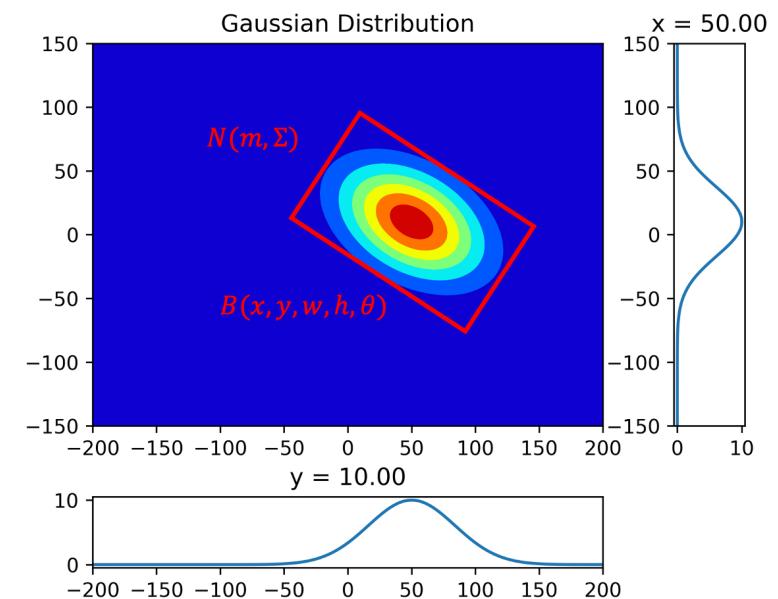
模型设计范式

- ## 演绎范式 (Deduction)

  - ### 特点1：易实现、可微分

  - ### 特点2：在角度边界情况下平滑

  - ### 特点3：与 IoU 度量高度一致

$$\Sigma^{1/2} = \mathbf{RSR}^\top$$

$$= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \frac{w}{2} & 0 \\ 0 & \frac{h}{2} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

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

$$\mathbf{m} = (x, y)^\top$$

Property 1: $\Sigma^{1/2}(w, h, \theta) = \Sigma^{1/2}(h, w, \theta - \frac{\pi}{2})$; 解决OpenCV定义法边界问题

Property 2: $\Sigma^{1/2}(w, h, \theta) = \Sigma^{1/2}(w, h, \theta - \pi)$; 解决长边定义法边界问题

Property 3: $\Sigma^{1/2}(w, h, \theta) \approx \Sigma^{1/2}(w, h, \theta - \frac{\pi}{2})$, if $w \approx h$. 解决类正方形检测问题

Gaussian Distribution

$N(m, \Sigma)$

$B(x, y, w, h, \theta)$

模型设计范式

- **演绎范式：Wasserstein Distance**

  - 通用公式：

$$\mathbf{D}_w(\mathcal{N}_p, \mathcal{N}_t)^2 = \underbrace{\|\boldsymbol{\mu}_p - \boldsymbol{\mu}_t\|_2^2}_{\text{center distance}} + \underbrace{\mathbf{Tr}(\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_t - 2(\boldsymbol{\Sigma}_p^{1/2} \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_p^{1/2})^{1/2})}_{\text{coupling terms about } h_p, w_p \text{ and } \theta_p}$$

  - 水平特殊情况：

$$\begin{aligned}
\mathbf{D}_w^h(\mathcal{N}_p, \mathcal{N}_t)^2 &= \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_t\|_2^2 + \|\boldsymbol{\Sigma}_p^{1/2} - \boldsymbol{\Sigma}_t^{1/2}\|_F^2 \\
&= (x_p - x_t)^2 + (y_p - y_t)^2 + \left((w_p - w_t)^2 + (h_p - h_t)^2\right)/4 \\
&= l_2\text{-norm}(\Delta x, \Delta y, \Delta w/2, \Delta h/2)
\end{aligned}$$

模型设计范式

● 演绎范式：Wasserstein Distance

模型设计范式

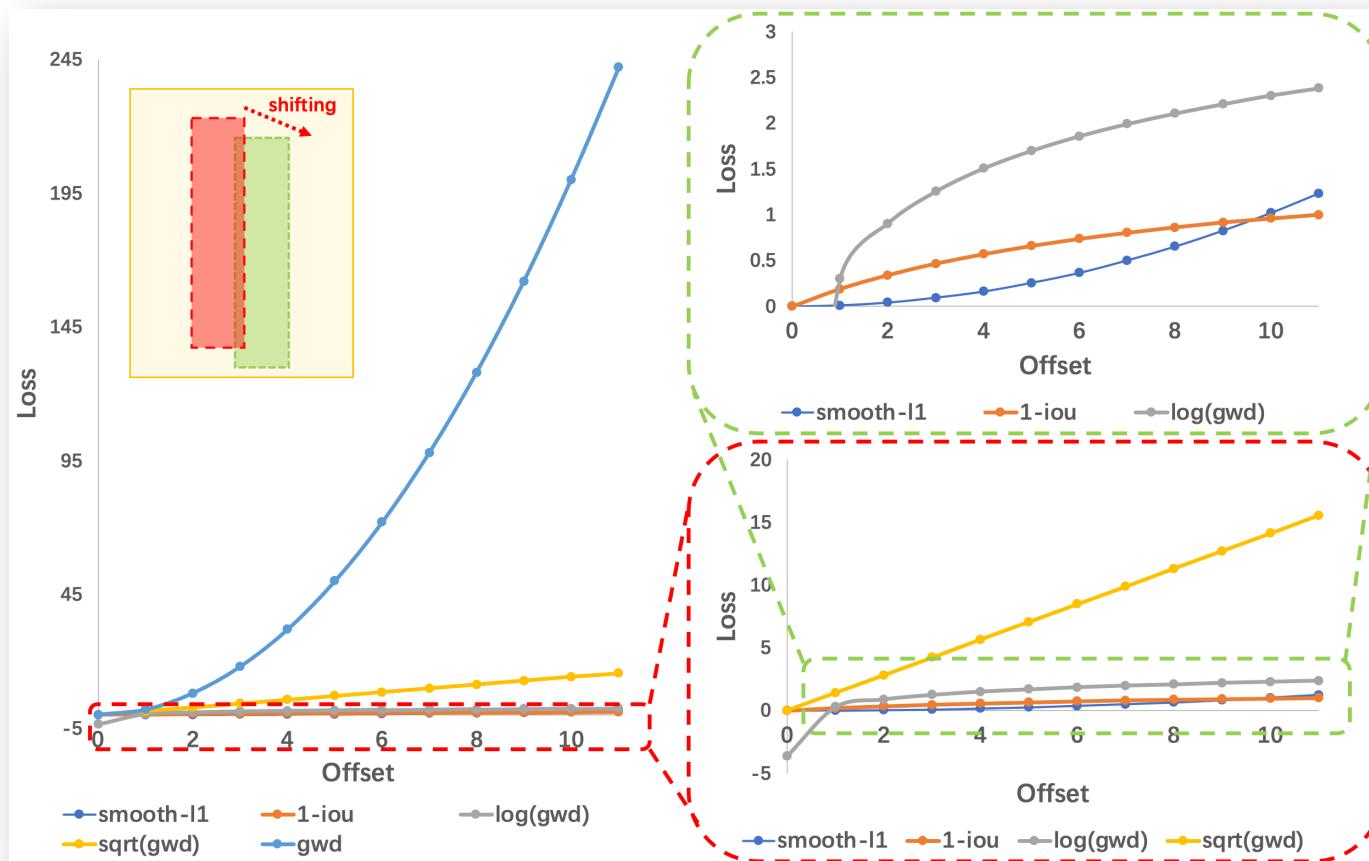- **演绎范式：Wasserstein Distance**

  - 通用公式：

$$\mathbf{D}_w(\mathcal{N}_p, \mathcal{N}_t)^2 = \underbrace{\|\boldsymbol{\mu}_p - \boldsymbol{\mu}_t\|_2^2}_{\text{center distance}} + \underbrace{\mathbf{Tr}(\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_t - 2(\boldsymbol{\Sigma}_p^{1/2}\boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_p^{1/2})^{1/2})}_{\text{coupling terms about } h_p,\, w_p \text{ and } \theta_p}$$

  - 水平特殊情况：

$$
\begin{aligned}
\mathbf{D}_w^h(\mathcal{N}_p, \mathcal{N}_t)^2 &= \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_t\|_2^2 + \|\boldsymbol{\Sigma}_p^{1/2} - \boldsymbol{\Sigma}_t^{1/2}\|_F^2 \\
&= (x_p - x_t)^2 + (y_p - y_t)^2 + \left((w_p - w_t)^2 + (h_p - h_t)^2\right)/4 \\
&= l_2\text{-norm}(\Delta x, \Delta y, \Delta w/2, \Delta h/2)
\end{aligned}
$$

  - 损失函数：

$$L_{gwd} = 1 - \frac{1}{\tau + f(\mathbf{D}_w^2)}, \quad \tau \geq 1$$

模型设计范式

- 演绎范式：**Kullback-Leibler Divergence**

  - 通用公式：

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

  或

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

  - 水平特殊情况：

  $$\begin{aligned}
  \mathbf{D}_{kl}^h(\mathcal{N}_p||\mathcal{N}_t) &= \frac{1}{2}\left(\frac{w_p^2}{w_t^2} + \frac{h_p^2}{h_t^2} + \frac{4\Delta^2 x}{w_t^2} + \frac{4\Delta^2 y}{h_t^2} + \ln\frac{w_t^2}{w_p^2} + \ln\frac{h_t^2}{h_p^2} - 2\right) \\
  &= 2l_2\text{-norm}(\Delta t_x, \Delta t_y) + l_1\text{-norm}(\Delta t_w, \Delta t_h) + \frac{1}{2}l_2\text{-norm}(\frac{1}{\Delta t_w}, \frac{1}{\Delta t_h}) - 1
  \end{aligned}$$

模型设计范式

- ## 演绎范式：Kullback-Leibler Divergence

  - ### KLD主要三项的具体表达式：

$$(\boldsymbol{\mu}_p - \boldsymbol{\mu}_t)^\top \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_t) = \frac{4(\Delta x \cos\theta_t + \Delta y \sin\theta_t)^2}{w_t^2} + \frac{4(\Delta y \cos\theta_t - \Delta x \sin\theta_t)^2}{h_t^2}$$

$$\mathbf{Tr}(\boldsymbol{\Sigma}_t^{-1} \boldsymbol{\Sigma}_p) = \frac{h_p^2}{w_t^2}\sin^2\Delta\theta + \frac{w_p^2}{h_t^2}\sin^2\Delta\theta + \frac{h_p^2}{h_t^2}\cos^2\Delta\theta + \frac{w_p^2}{w_t^2}\cos^2\Delta\theta$$

$$\ln\frac{|\boldsymbol{\Sigma}_t|}{|\boldsymbol{\Sigma}_p|} = \ln\frac{h_t^2}{h_p^2} + \ln\frac{w_t^2}{w_p^2}$$

其中 $\Delta x = x_p - x_t, \Delta y = y_p - y_t, \Delta\theta = \theta_p - \theta_t$ 。

模型设计范式

- ## 高精度分析 （KLD > GWD > L_n）：

  - 不失一般性，我们令 $\theta_t = 0$，对 KLD 的 μ_p 求导数：

    $$\frac{\partial \mathbf{D}_{kl}(\mu_p)}{\partial \mu_p} = \left(\frac{4}{w_t^2} \Delta x, \frac{4}{h_t^2} \Delta y \right)^\top$$

  - 当 $\theta_t \neq 0$ 时，目标的偏移量（ Δx 和 Δy）的梯度会根据角度进行动态调整以提供更好的优化。相比之下，GWD 和 L_2 关于偏移量的梯度分别是：

    $$\frac{\partial \mathbf{D}_w(\mu_p)}{\partial \mu_p} = (2\Delta x, 2\Delta y)^\top$$

    GWD

    $$\frac{\partial L_2(\mu_p)}{\partial \mu_p} = \left(\frac{2}{w_a^2} \Delta x, \frac{2}{h_a^2} \Delta y \right)^\top$$

    L_2

- 高精度分析 （KLD > GWD > L_n）：

  - 对 KLD 的 h_p 和 w_p 求导数：

$$\frac{\partial \mathbf{D}_{kl}(\mathbf{\Sigma}_p)}{\partial \ln h_p} = \frac{h_p^2}{h_t^2}\cos^2 \Delta\theta + \frac{h_p^2}{w_t^2}\sin^2 \Delta\theta - 1, \quad \frac{\partial \mathbf{D}_{kl}(\mathbf{\Sigma}_p)}{\partial \ln w_p} = \frac{w_p^2}{w_t^2}\cos^2 \Delta\theta + \frac{w_p^2}{h_t^2}\sin^2 \Delta\theta - 1$$

  - 我们可以看到，两边 h_p 和 w_p 梯度和角度差 $\Delta\theta$ 有关。当 $\Delta\theta$=0 时：

$$\frac{\partial \mathbf{D}_{kl}(\mathbf{\Sigma}_p)}{\partial \ln h_p} = \frac{h_p^2}{h_t^2} - 1, \frac{\partial \mathbf{D}_{kl}(\mathbf{\Sigma}_p)}{\partial \ln w_p} = \frac{w_p^2}{w_t^2} - 1$$

  - 这意味着较小的目标尺度会匹配到更大的损失。 这是符合认知的，因为较短的边需要更高的匹配精度。

模型设计范式

- **高精度分析 （KLD > GWD > L_n）：**

  - 对 θ 求导数：

  $$\frac{\partial \mathbf{D}_{kl}(\mathbf{\Sigma}_p)}{\partial \ln h_p} = \frac{h_p^2}{h_t^2}\cos^2\Delta\theta + \frac{h_p^2}{w_t^2}\sin^2\Delta\theta - 1, \quad \frac{\partial \mathbf{D}_{kl}(\mathbf{\Sigma}_p)}{\partial \ln w_p} = \frac{w_p^2}{w_t^2}\cos^2\Delta\theta + \frac{w_p^2}{h_t^2}\sin^2\Delta\theta - 1$$
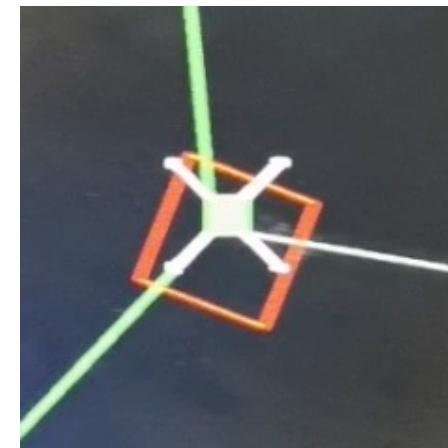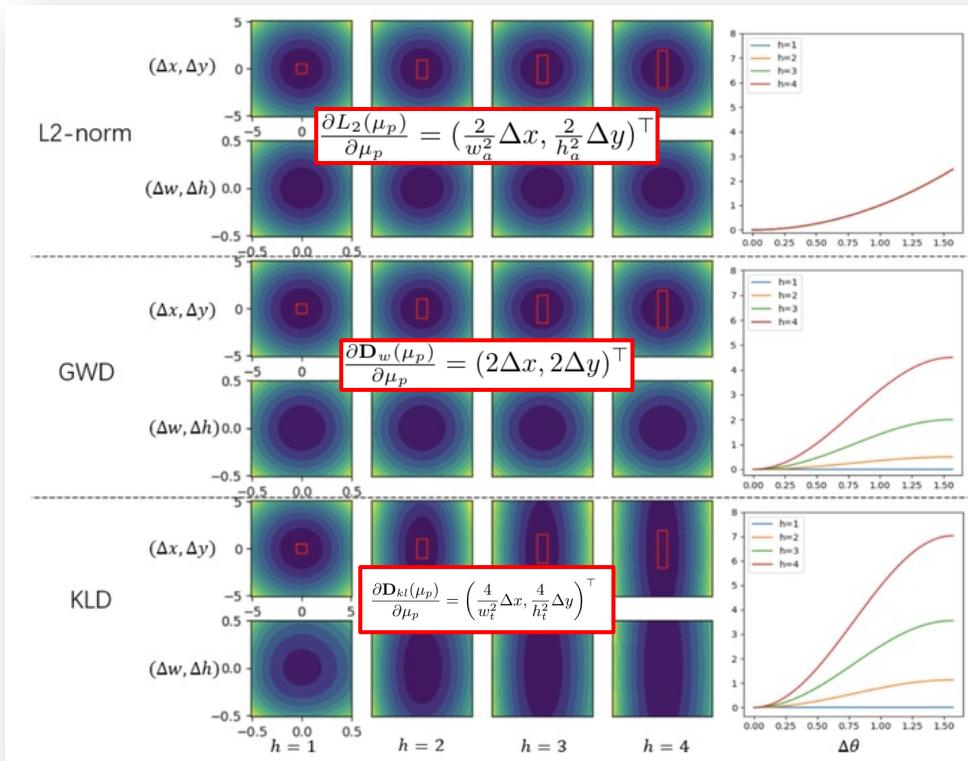
  - 角度差 $\Delta\theta$ 的优化又和两边 h_p 和 w_p 有关。当 $h_p{=}h_t$，$w_p{=}w_t$ 时：

  $$\frac{\partial \mathbf{D}_{kl}(\mathbf{\Sigma}_p)}{\partial \theta_p} = \left(\frac{h_t^2}{w_t^2} + \frac{w_t^2}{h_t^2} - 2\right)\sin 2\Delta\theta \geq \sin 2\Delta\theta$$

  - 当目标长宽比慢慢变大的时候，整个式子的值就会变大，也就是意味着对角度优化更加看重。这个优化机制是非常好的，我们知道对于长宽比越大的目标来说，它受角度差的影响就越大，IoU 会产生急剧下降。

模型设计范式

● 高精度分析 （KLD > GWD > L_n）：

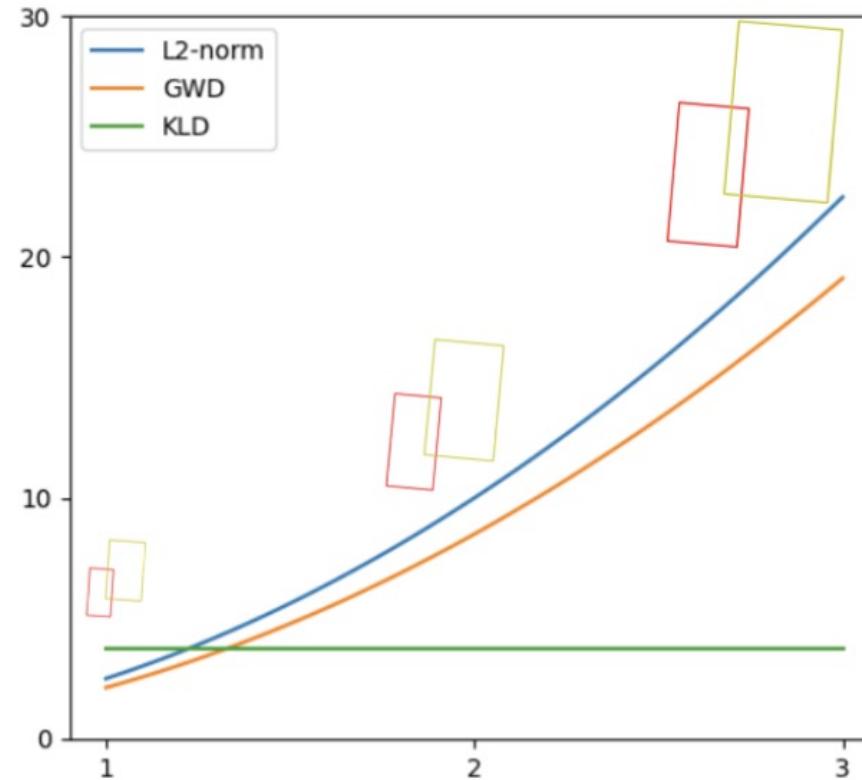**KLD Loss (square-like)**

KLD Loss

- 尺度不变性证明

  - 很明显 GWD 和 L$_2$ **不具有尺度不变性**。

  - 对于两个已知的高斯分布 $\mathbf{X}_p \sim \mathcal{N}_p(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$ 和 $\mathbf{X}_t \sim \mathcal{N}_t(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ 假设有一个满秩的矩阵 $\mathbf{M}$ (|$\mathbf{M}$|≠0)，有 $\mathbf{X}_{p'} = \mathbf{M}\mathbf{X}_p \sim \mathcal{N}_p(\mathbf{M}\boldsymbol{\mu}_p, \mathbf{M}\boldsymbol{\Sigma}_p\mathbf{M}^\top), \mathbf{X}_{t'} = \mathbf{M}\mathbf{X}_t \sim \mathcal{N}_t(\mathbf{M}\boldsymbol{\mu}_t, \mathbf{M}\boldsymbol{\Sigma}_t\mathbf{M}^\top)$

  - 我们将其分别标记为 $\mathcal{N}_{p'}$ 和 $\mathcal{N}_{t'}$，那么它们的 KLD 计算如下：

$$
\begin{aligned}
\mathbf{D}_{kl}(\mathcal{N}_{p'}||\mathcal{N}_{t'}) &= \frac{1}{2}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_t)^\top \mathbf{M}^\top (\mathbf{M}^\top)^{-1} \boldsymbol{\Sigma}_t^{-1} \mathbf{M}^{-1} \mathbf{M}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_t) \\
&+ \frac{1}{2}\mathbf{Tr}\left((\mathbf{M}^\top)^{-1}\boldsymbol{\Sigma}_t^{-1}\mathbf{M}^{-1}\mathbf{M}\boldsymbol{\Sigma}_p\mathbf{M}^\top\right) + \frac{1}{2}\ln\frac{|\mathbf{M}||\boldsymbol{\Sigma}_t||\mathbf{M}^\top|}{|\mathbf{M}||\boldsymbol{\Sigma}_p||\mathbf{M}^\top|} - 1 \\
&= \frac{1}{2}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_t)^\top \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_t) + \frac{1}{2}\mathbf{Tr}\left(\mathbf{M}^\top(\mathbf{M}^\top)^{-1}\boldsymbol{\Sigma}_t^{-1}\mathbf{M}^{-1}\mathbf{M}\boldsymbol{\Sigma}_p\right) + \frac{1}{2}\ln\frac{|\boldsymbol{\Sigma}_t|}{|\boldsymbol{\Sigma}_p|} - 1 \\
&= \mathbf{D}_{kl}(\mathcal{N}_p||\mathcal{N}_t)
\end{aligned}
$$

模型设计范式

- 尺度不变性证明

    - 因此 **KLD 具有仿射不变性的**。当 **M=kI** 时， KLD 的尺度不变性就被证明了。

实验分析

● 在 4 种数据集和 2 种检测器上进行了高精度检测实验，KLD 具有绝对优势。

TABLE 5
High-precision detection experiment under different regression loss. 'R', 'F' and 'G' indicate random rotation, flipping, and graying, respectively. The resolution of HRSC2016, MSRA-TD500, ICDAR2015 and FDDB are 500×500, $800 \times 1,000$, $800 \times 1,000$ and 800×800, respectively.

Base Detector	Dataset	Data Aug.	Reg. Loss	$\text{Hmean}_{50}/\text{AP}_{50}$	$\text{Hmean}_{60}/\text{AP}_{60}$	$\text{Hmean}_{75}/\text{AP}_{75}$	$\text{Hmean}_{85}/\text{AP}_{85}$	$\text{Hmean}_{50:95}/\text{AP}_{50:95}$
RetinaNet [15]	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)
			BCD	86.38 (+2.10)	85.32 (+10.58)	68.50 (+20.08)	15.67 (+3.11)	55.09 (+7.33)
			KLD	**87.45 (+3.17)**	**86.72 (+11.98)**	**72.39 (+23.97)**	**27.68 (+15.12)**	**57.80 (+10.04)**
R³Det [18]			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)
			BCD	**90.06 (+1.54)**	**89.75 (+10.74)**	**76.24 (+32.82)**	23.42 (+18.84)	60.26 (+14.08)
			KLD	89.97 (+1.45)	89.73 (+10.72)	**77.38 (+33.96)**	**25.12 (+20.54)**	**61.40 (+15.22)**
RetinaNet [15]	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)
			BCD	75.24 (+4.26)	69.50 (+7.08)	**48.13 (+11.40)**	**20.33 (+7.77)**	**45.26 (+7.37)**
			KLD	**76.96 (+5.98)**	**70.08 (+7.66)**	46.95 (+10.22)	19.59 (+7.03)	45.24 (+7.35)
	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)
			BCD	**76.63 (+6.85)**	**71.07 (+6.92)**	43.10 (+6.13)	10.24 (+1.53)	**42.78 (+5.05)**
			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)
			BCD	**78.03 (+3.20)**	72.50 (+3.04)	45.44 (+3.42)	10.53 (-1.06)	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)**
R³Det [18]		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)
			BCD	**79.02 (+4.74)**	**72.82 (+4.70)**	**45.68 (+9.95)**	10.42 (+2.41)	**44.22 (+5.12)**
			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)
			BCD	**80.49 (+4.96)**	**74.73 (+5.04)**	**45.42 (+7.73)**	**10.89 (+1.86)**	**44.55 (+3.99)**
			KLD	79.63 (+4.63)	73.30 (+3.61)	43.51 (+5.82)	10.61 (+1.58)	43.61 (+3.05)
RetinaNet [15]	FDDB	F	Smooth L1	95.92	87.50	55.81	12.67	52.77
			GWD	97.44 (+1.52)	94.68 (+7.18)	80.84 (+25.03)	36.38 (+23.71)	65.77 (+13.00)
			BCD	96.67 (+0.75)	94.60 (+7.10)	83.09 (+27.28)	40.72 (+28.05)	67.03 (+14.26)
			KLD	**97.51 (+1.59)**	**95.40 (+7.90)**	**85.33 (+29.52)**	**42.20 (+29.53)**	**68.01 (+15.24)**

实验分析

- 在一些更具有挑战性的数据集上进行了验证实验，包括 DOTA-v1.5 和 DOTA-v2.0（包含很多像素值小于 10 的目标），KLD 依旧表现出色。

Table 6: Accuracy comparison between different rotation detectors on DOTA dataset. † and ‡ represent the large aspect ratio object and the square-like object, respectively. The bold **red** and **blue** fonts indicate the top two performances respectively. D_{oc} and D_{le} represent OpenCV Definition ($\theta \in [-90°, 0°)$) and Long Edge Definition ($\theta \in [-90°, 90°)$) of RBox.

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 [46]	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 [46]	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 [35]	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) [47]	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**
R³Det	-	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) [47]	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**

实验分析

- 在水平检测任务上（COCO 数据集），KLD 也是和 GIoU 等常见损失函数保持差不多的水平。

Table 6: Performance evaluation of KLD on classic horizontal detection.

Detector	Reg. Loss	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l	Detector	Reg. Loss	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
RetinaNet	Smooth L1	37.2	56.6	39.7	21.4	41.1	48.0	Faster RCNN	Smooth L1	37.9	58.8	41.0	22.4	41.4	49.1
	GIoU	37.4	56.7	39.7	22.2	41.7	48.1		GIoU	**38.3**	58.7	41.5	22.5	41.7	49.7
	KLD	**38.0**	56.4	40.6	23.3	43.2	49.3		KLD	38.2	58.7	41.7	22.6	41.8	49.3

- 对 KLD 不同变体在两个数据集上进行了实验，发现最后的效果是差不多的，排除了不对称性对结果的干扰。

Table 2: Ablation of different KLD-based regression loss form. The based detector is RetinaNet.

Dataset	$\mathbf{D}_{kl}(\mathcal{N}_p \| \mathcal{N}_t)$	$\mathbf{D}_{kl}(\mathcal{N}_t \| \mathcal{N}_p)$	$\mathbf{D}_{kl\_min}(\mathcal{N}_p \| \mathcal{N}_t)$	$\mathbf{D}_{kl\_max}(\mathcal{N}_p \| \mathcal{N}_t)$	$\mathbf{D}_{js}(\mathcal{N}_p \| \mathcal{N}_t)$	$\mathbf{D}_{jeffreys}(\mathcal{N}_p \| \mathcal{N}_t)$
DOTA-v1.0	70.17	70.64	**70.71**	70.55	69.67	70.56
HRSC2016	82.83	83.82	83.60	82.70	**84.06**	83.66

实验分析

● 最后在 DOTA-v1.0 的 SOTA 实验中，取得了当时所发表论文里的最高性能。

Table 8: AP on different objects on DOTA-v1.0. Here R-101 denotes ResNet-101 (likewise for R-50, R-152), and RX-101 and H-104 represent ResNeXt101 [50] and Hourglass-104 [51], respectively. MS indicates that multi-scale training/testing is used. Red and blue indicate the top two performances.

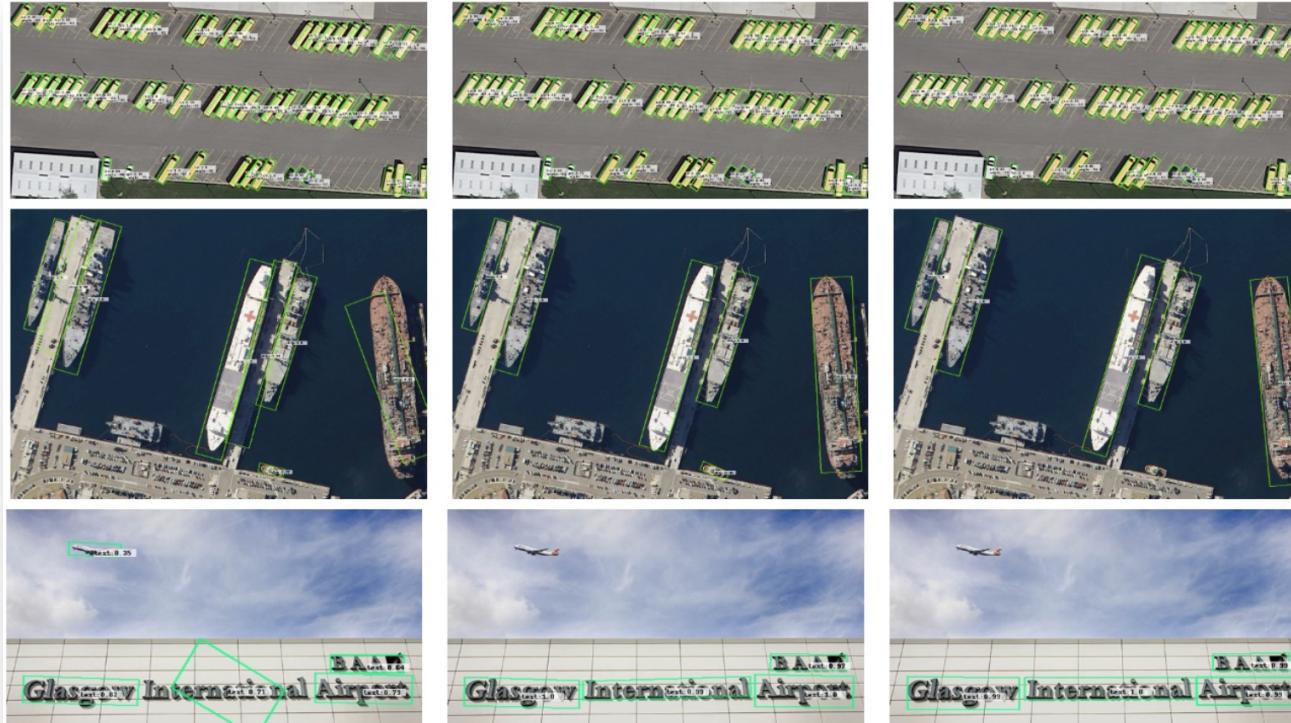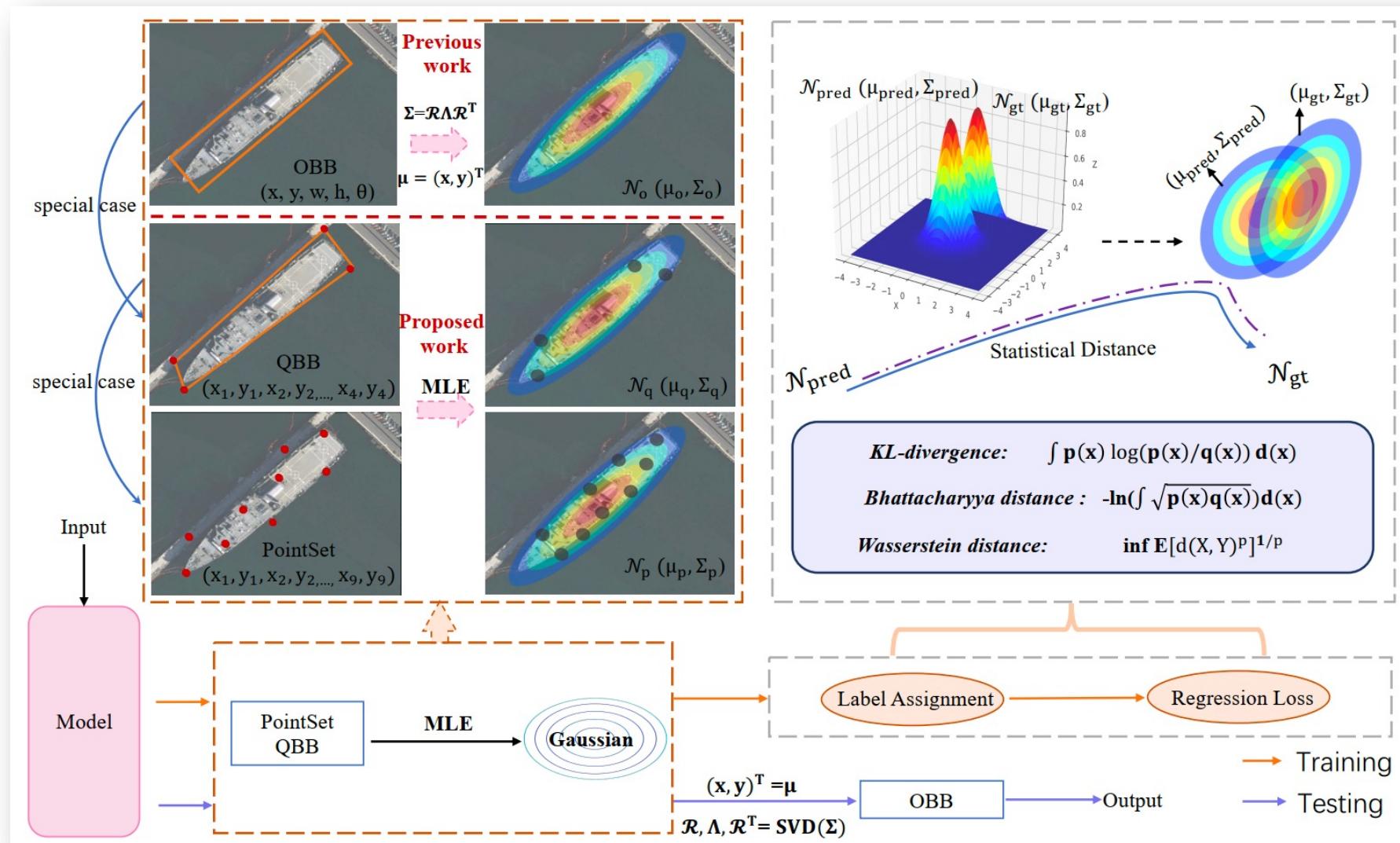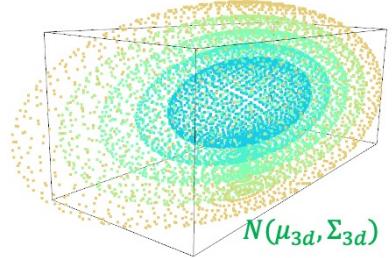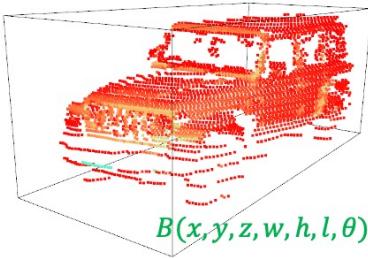|	Method	Backbone	MS	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	AP_{50}
Two-stage	ICN [32]	R-101	✓	81.40	74.30	47.70	70.30	64.90	67.80	70.00	90.80	79.10	78.20	53.60	62.90	67.00	64.20	50.20	68.20
	RoI-Trans. [12]	R-101	✓	88.64	78.52	43.44	75.92	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
	SCRDet [3]	R-101	✓	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
	Gliding Vertex [52]	R-101		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 [53]	RX-101	✓	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 OBB [54]	R-101	✓	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
	FPN-CSL [4]	R-152	✓	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
	RSDet-II [46]	R-152	✓	89.93	84.45	53.77	74.35	71.52	78.31	78.12	91.14	87.35	86.93	65.64	65.17	75.35	79.74	63.31	76.34
	SCRDet++ [55]	R-101	✓	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
	ReDet [56]	ReR-50	✓	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
Single-stage	PIoU [33]	DLA-34 [57]		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
	O^2-DNet [58]	H-104	✓	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 [15]	R-101	✓	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 [59]	R-101	✓	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 [60]	R-101	✓	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 [14]	H-104	✓	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
	PolarDet [61]	R-101	✓	89.65	87.07	48.14	70.97	78.53	80.34	87.45	90.76	85.63	86.87	61.64	70.32	71.92	73.09	67.15	76.64
	RDD [62]	R-101	✓	89.15	83.92	52.51	73.06	77.81	79.00	87.08	90.62	86.72	87.15	63.96	70.29	76.98	75.79	72.15	77.75
	GWD [5]	R-152	✓	89.06	84.32	55.33	77.53	76.95	70.28	83.95	89.75	84.51	86.06	73.47	67.77	72.60	75.76	74.17	77.43
	KLD	R-50		88.91	83.71	50.10	68.75	78.20	76.05	84.58	89.41	86.15	85.28	63.15	60.90	75.06	71.51	67.45	75.28
		R-50	✓	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
Refine-stage	CFC-Net [34]	R-101	✓	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
	R^3Det [29]	R-152	✓	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
	DAL [15]	R-50	✓	89.69	83.11	55.03	71.00	78.30	81.90	88.46	90.89	84.97	87.46	64.41	65.65	76.86	72.09	64.35	76.95
	DCL [47]	R-152	✓	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 [35]	R-50	✓	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
	S^2A-Net [13]	R-101	✓	89.28	84.11	56.95	79.21	80.18	82.93	89.21	90.86	84.66	87.61	71.66	68.23	78.58	78.20	65.55	79.15
	R^3Det-GWD [5]	R-152	✓	89.66	84.99	59.26	82.19	78.97	84.83	87.70	90.21	86.54	86.85	73.04	67.56	76.92	79.22	74.92	80.19
	R^3Det-KLD	R-50		88.90	84.17	55.80	69.35	78.72	84.08	87.00	89.75	84.32	85.73	64.74	61.80	76.62	78.49	70.89	77.36
		R-50	✓	89.90	84.91	59.21	78.74	78.82	83.95	87.41	89.89	86.63	86.69	70.47	70.87	76.96	79.40	78.62	80.17
		R-152	✓	89.92	85.13	59.19	81.33	78.82	84.38	87.50	89.80	87.33	87.00	72.57	71.35	77.12	79.34	78.68	80.63

实验分析

● 高精度分析 （KLD > GWD > L_n）：
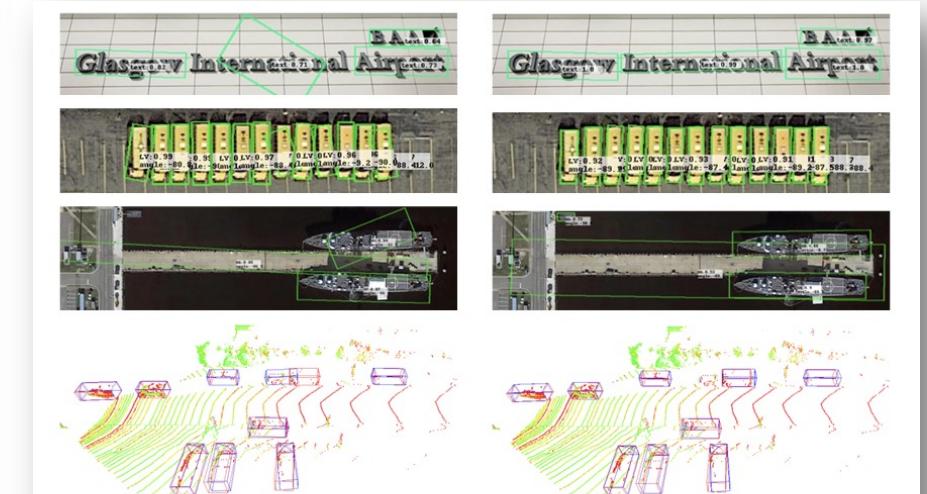
Fig. 1. Detection results comparison (top: 2-D, bottom: 3-D) at the boundary condition (i.e. horizontal or vertical rotation) between Smooth L1 loss-based (left) and the Gaussian-based (right) detectors. See illustration in Fig. 2 for the Gaussian-based bounding box detection.
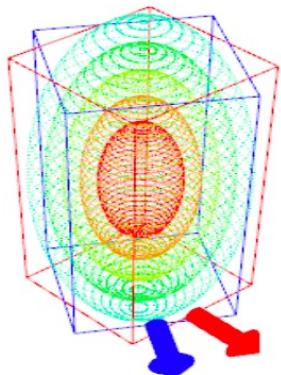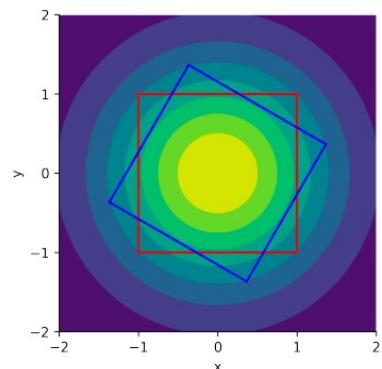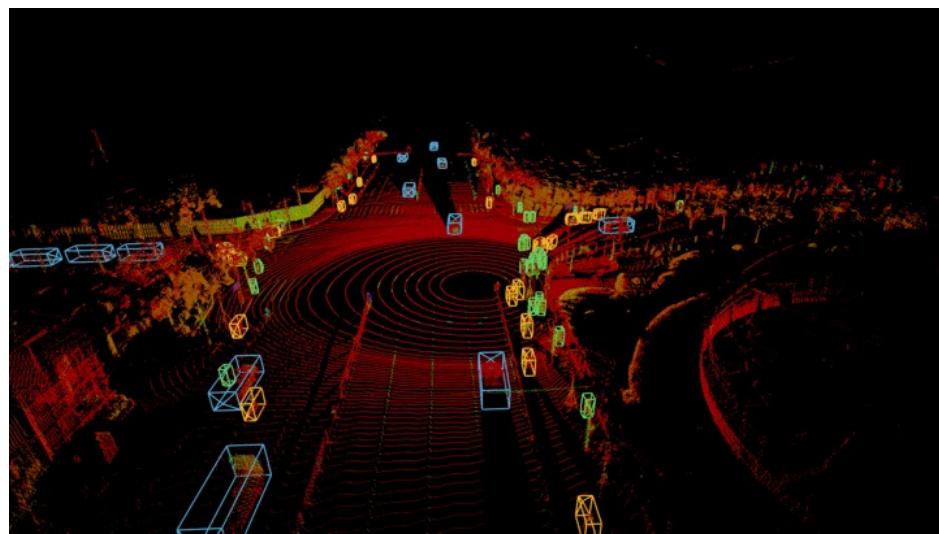
Hou L, Lu K, Yang X, et al. G-Rep: Gaussian Representation for Arbitrary-Oriented Object Detection. arXiv preprint arXiv:2205.11796, 2022.

$B(x,y,z,w,h,l,\theta)$

$N(\mu_{3d}, \Sigma_{3d})$
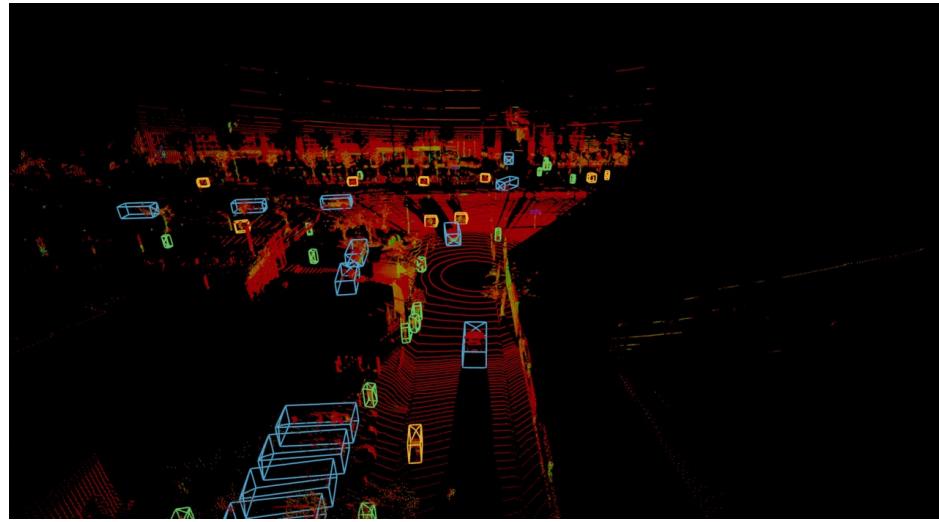
3D高斯表示

(b) 3-D BBox with square shape in top-view e.g. pedestrian.

(c) Top view of the 3-D BBox and the heading is arbitrary given the isotropic 2-D Gaussian.

类正方形退化问题

X. Yang, et al. "Detecting Rotated Objects as Gaussian Distributions and Its 3-D Generalization ." In TPAMI 2023.

谢 谢！

VALSE 2023 无锡

杨学

yangxue-2019-sjtu@sjtu.edu.cn

https://yangxue0827.github.io/