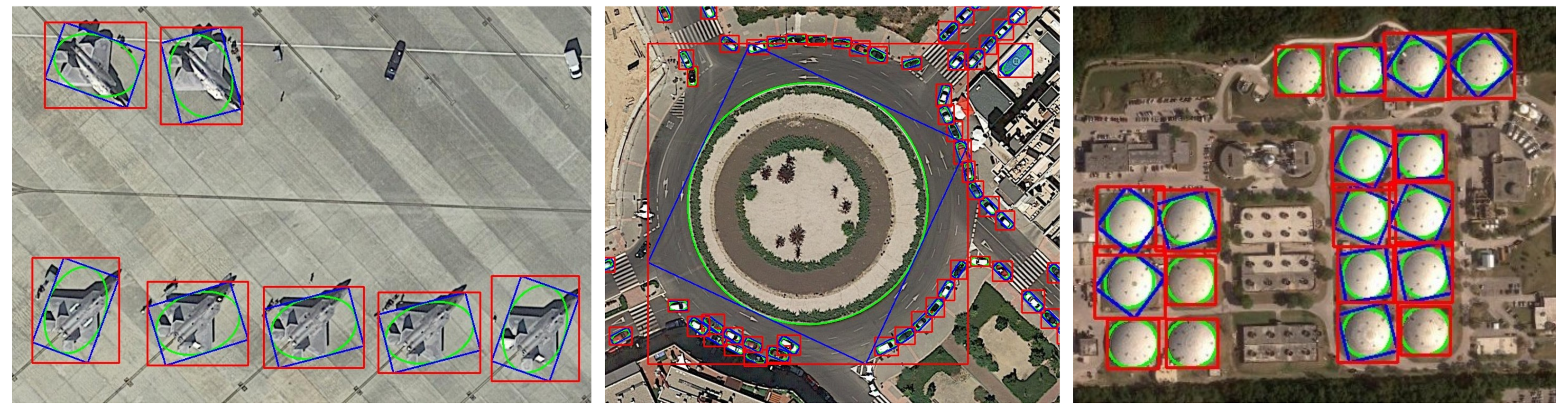
Probabilistic Intersection-over-Union for Training and Evaluation of Oriented Object Detectors

Jeffri Murrugarra-Llerena, Lucas N. Kirsten , Luis Felipe Zeni, Claudio R. Jung
Federal University of Rio Grande do Sul[1]
{jeffri.mllerena, lucasnkir, luis.zeni, crjung}@inf.ufrgs.br

CVPR SEATTLE, WA JUNE 17-21, 2024

Motivation

- Object detection is a **crucial problem in image processing and computer vision**, with significant advances in recent years using deep learning.
- However, most object detectors explore **Horizontal Bounding Boxes (HBBs)** to encode the object shape/location, which might contain substantial portions of the background.
- As an alternative, object detectors based on **Oriented Bounding Boxes (OBBs)** are emerging.



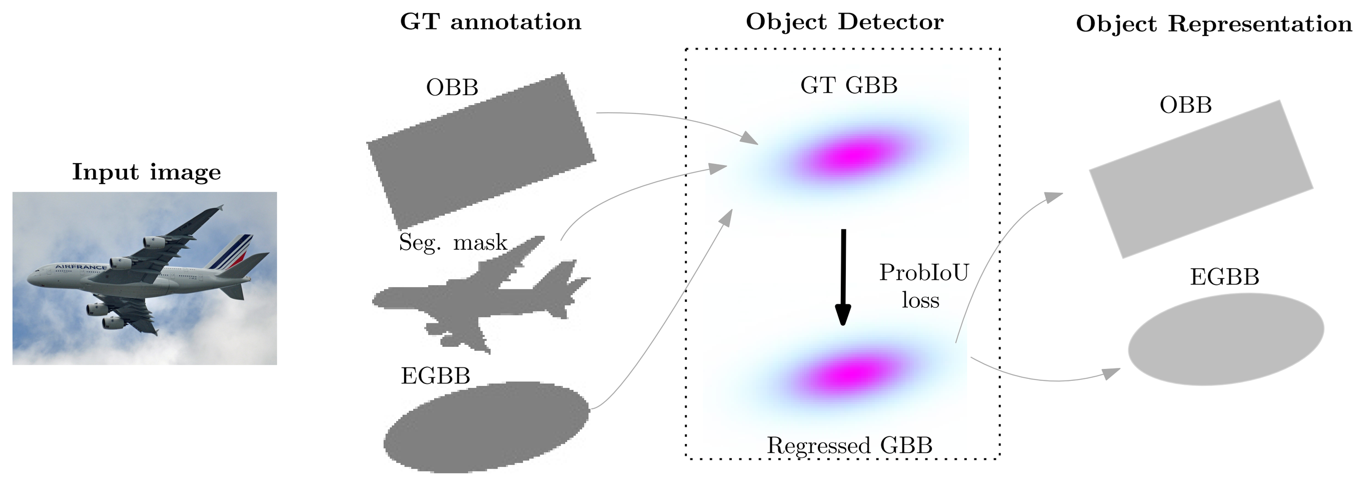
Key idea

- We propose **Probabilistic IoU (ProbIoU)** considers fuzzy object representations as probability density functions.

$$\mu = \frac{1}{N}\int_{x\in\Omega} x\mathrm{d}x, \Sigma = \frac{1}{N}\int_{x\in\Omega}(x-\mu)(x-\mu)^T\mathrm{d}x, \tag{1}$$

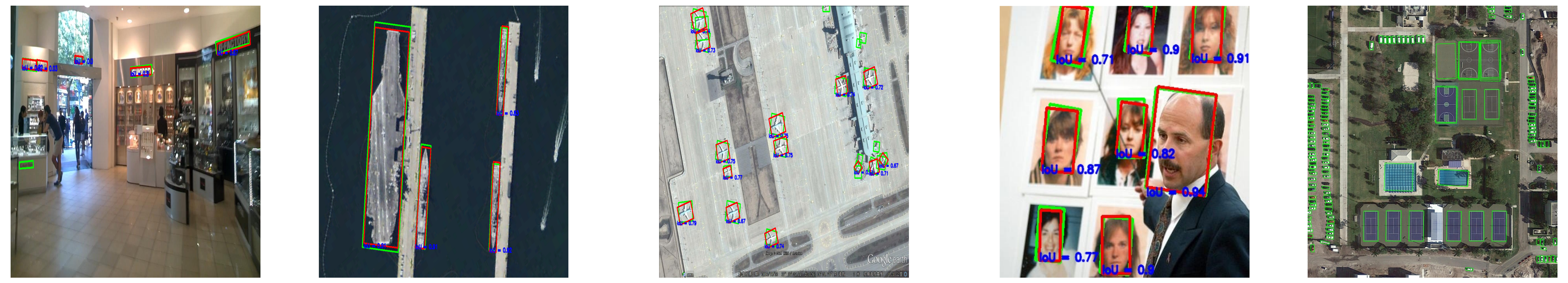
$$\Sigma = \begin{bmatrix} a & c \\ c & b \end{bmatrix} = R_\theta \begin{bmatrix} a' & 0 \\ 0 & b' \end{bmatrix} R_\theta^T = \begin{bmatrix} a'\cos^2\theta + b'\sin^2\theta & \frac{1}{2}(a'-b')\sin 2\theta \\ \frac{1}{2}(a'-b')\sin 2\theta & a'\sin^2\theta + b'\cos^2\theta \end{bmatrix}, \tag{2}$$

- ProbIoU reduces to a **differentiable closed-form expression** (with Bhattacharyya Coefficient coefficient with Hellinger distance) that can be used as a localization loss term.



Dataset Collection

- We consider several OBB datasets: **ICDAR2015** (text in the wild), **HRSC2016** (satellite images of ships), **UCAS-AOD** (satellite images of car and airplanes), **FDDB** (faces in the wild), and **DOTAv1** (satellite images of 15 classes).



(a) ICDAR (b) HRSC (c) UCAS-AOD (d) FDDB (e) DOTAv1

Approach

- If we use a fuzzy object representation based on GBBs, we can calculate B_D (distance measure) present closed-form expressions in terms of the GBB parameters.
- Considering that $p \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $q \sim \mathcal{N}(\mu_2, \Sigma_2)$ are Gaussian distributions with

$$\mu_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \ \Sigma_1 = \begin{bmatrix} a_1 & c_1 \\ c_1 & b_1 \end{bmatrix}, \ \mu_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \ \Sigma_2 = \begin{bmatrix} a_2 & c_2 \\ c_2 & b_2 \end{bmatrix}, \tag{3}$$

obtain:

$$B_D = \frac{1}{8}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2) + \frac{1}{2}\ln\left(\frac{\det\Sigma}{\sqrt{\det\Sigma_1\det\Sigma_2}}\right), \Sigma = \frac{1}{2}(\Sigma_1+\Sigma_2). \tag{4}$$

since:

$$B_C = e^{-B_D}, H_D(p,q) = \sqrt{1-B_C(p,q)}, \text{ProbIoU}(p,q) = 1 - H_D(p,q). \tag{5}$$

- ProbIoU ranges from 0 to 1, with the following properties:

Loss	Implementation	Scale Invariance	Metric Properties	Hyper parameters
r-IoU	Hard	✓	✓	–
Smooth ℓ_1	Easy	×	×	–
GWD	Easy	×	×	τ, $f(\cdot)$
KLD	Easy	✓	×	τ, $f(\cdot)$
ProbIoU	Easy	✓	✓	–

Experiments

- We use the MAP metric to compare our method with other SOTA methods.

Loss	AP_{50}	AP_{60}	AP_{75}	AP_{85}	AP	AP_{50}	AP_{60}	AP_{75}	AP_{85}	AP
	ICDAR2015 (Retinanet)					ICDAR2015 (R^3Det)				
Smooth ℓ_1	72.91	65.06	35.10	7.62	37.84	76.66	68.34	33.56	5.82	38.62
GWD	73.49	65.83	32.94	7.71	37.55	76.60	68.42	35.44	7.47	39.34
KLD	75.60	67.62	35.87	7.61	39.21	77.40	70.09	35.44	6.93	39.83
ProbIoU	76.64	69.02	36.39	8.91	40.14	77.54	70.67	35.89	6.49	40.03
	HRSC2016 (RetinaNet)					HRSC2016 (R^3Det)				
Smooth ℓ_1	81.80	77.71	47.30	9.73	45.77	84.09	71.37	28.77	2.27	37.89
GWD	80.48	76.67	48.55	14.73	47.11	89.06	86.69	61.01	8.65	52.56
KLD	86.21	84.19	70.83	20.79	55.89	90.25	89.86	77.16	23.27	61.12
r-IoU	86.20	83.40	69.80	37.20	58.31	88.10	79.00	63.50	23.10	53.33
ProbIoU	87.03	85.22	71.49	24.67	56.79	90.23	90.02	78.46	25.42	61.65
	UCAS-AOD (RetinaNet)					FDDB (RetinaNet)				
Smooth ℓ_1	95.97	90.27	56.30	15.66	54.62	98.09	95.39	83.67	63.68	72.73
GWD	95.24	85.60	24.60	1.83	43.58	97.72	96.20	89.45	73.24	77.68
KLD	96.46	92.51	43.37	6.45	51.09	97.56	96.25	90.37	76.04	79.01
r-IoU	93.50	87.70	56.50	17.10	54.09	97.80	97.10	91.30	77.10	79.90
ProbIoU	96.61	92.61	49.47	7.26	52.59	98.17	96.54	89.78	76.08	79.01

- For DOTAv1 we use its evaluation server (only AP50).

Model	Loss	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	AP_{50}
R-50 Retinanet	Smooth ℓ_1	88.78	71.40	43.85	62.32	67.00	55.70	71.49	90.08	78.94	79.81	54.81	62.62	54.54	66.74	48.23	66.43
	GWD	89.06	72.56	43.56	63.85	64.66	57.55	73.87	89.59	76.73	77.56	48.51	59.35	54.90	66.90	52.06	66.05
	KLD	89.58	74.99	44.95	64.62	74.38	73.41	84.33	89.94	78.14	78.36	53.08	61.95	63.46	66.44	52.17	69.99
	r-IoU	88.07	73.30	38.13	61.81	78.51	70.38	85.39	90.75	81.43	83.12	47.66	61.51	58.79	69.66	55.21	69.58
	ProbIoU	89.06	74.09	45.50	64.96	72.61	72.28	82.62	89.69	76.99	76.35	54.43	61.77	63.73	67.27	53.61	69.66
R-50 R^3Det	Smooth ℓ_1	89.16	75.61	45.11	61.39	76.40	72.05	83.81	89.75	81.16	82.19	58.75	61.37	55.95	67.01	44.13	69.59
	GWD	89.57	76.67	45.80	58.69	76.44	74.92	86.03	89.41	80.03	82.28	51.03	60.07	54.26	66.53	54.45	69.74
	KLD	88.87	75.92	43.20	64.96	75.20	72.53	84.53	90.09	75.78	80.10	48.83	60.71	62.94	63.58	56.32	69.57
	r-IoU	87.63	76.35	48.15	58.99	78.91	79.22	87.38	90.81	79.90	83.27	48.86	63.00	64.51	69.42	58.76	71.68
	ProbIoU	89.06	77.81	46.19	64.26	76.27	79.05	86.35	89.96	79.18	82.72	55.23	63.52	59.16	66.52	49.64	70.86

Contributions

- Our localization loss function ProbIoU **presents better results** than the popular smooth ℓ_1 loss and competitive or better than GWD and KLD
- Note that distances such as GWD and KLD **perform poorly as localization terms, and require empirical non-linear mappings with an additional hyper-parameter**
- As future work, **we explore a 3D version of the GBB for volumetric object detection**.