GauCho: Gaussian Distributions and Cholesky Decomposition for Oriented Object Detection
Supplementary Material

José Henrique Lima Marques[2]* Jeffri Murrugarra-Llerena[1]* Claudio R. Jung[2] *Equal contribution
[1] Stony Brook University, [2] Federal University of Rio Grande do Sul
jmurrugarral@cs.stonybrook.edu, {jhlmarques,crjung}@inf.ufrgs.br

1. Possible cause for orientation discontinuity using Gaussian-based loss functions

In the paper, we hypothesize that the angular discontinuity problem with Gaussian-based loss functions recently noted in [3, 5] is caused by the OBB to Gaussian mapping. To illustrate the problem, let us consider an origin-centered ground-truth (GT) OBB with shape parameters $(w, h, \theta) = (3, 1, 89°)$ with LE encoding. Figure 1 shows the plot of the KLD loss $\mathcal{L}_{\text{KLD}}$ [4] (the same behavior happens to *any* Gaussian-based loss) as a function of θ in the range $[-90°, 90°)$. The global minimum is reached for $\theta = 89°$, but a local minimum (almost as low as the global one) is achieved for $\theta = -90°$. In fact, the corresponding OBBs are geometrically very similar, as shown on the left of Figure 1. The loss function is clearly not convex, and the network might not learn the angular information properly in this scenario.

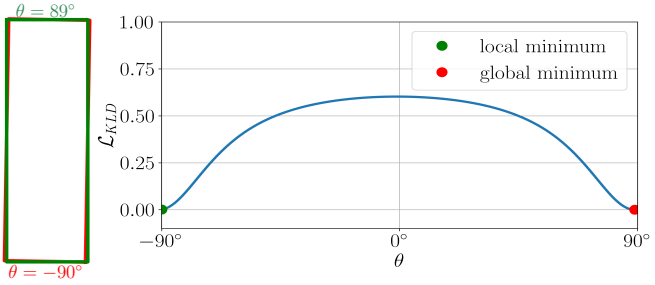

Figure 1. When regressing angular information from a Gaussian-based loss, the global angular minimum (red, $\theta = 89°$) might be close to the discontinuous counterpart (green, $\theta = -90°$).

2. Proof of Proposition 3.2

Here, we show the proof for the bound on the off-diagonal element of the Cholesky matrix. First, we revise the notation and the re-state the proposition.

Let us consider a covariance matrix expressed as a function of the eigenvalues λ_w, λ_h and the orientation θ:

$$C = \begin{bmatrix} \lambda_w \cos^2\theta + \lambda_h \sin^2\theta & \frac{1}{2}(\lambda_w - \lambda_h)\sin(2\theta) \\ \frac{1}{2}(\lambda_w - \lambda_h)\sin(2\theta) & \lambda_w \sin^2\theta + \lambda_h \cos^2\theta \end{bmatrix}. \tag{1}$$

Also, let us recall the Cholesky decomposition characterized by a lower-triangular matrix L

$$L = \begin{bmatrix} \alpha & 0 \\ \gamma & \beta \end{bmatrix} \tag{2}$$

with $\alpha, \beta > 0, \gamma \in \mathbb{R}$, such that $C = LL^T$, i.e.,

$$C = \begin{bmatrix} \alpha^2 & \alpha\gamma \\ \alpha\gamma & \beta^2 + \gamma^2 \end{bmatrix} = \begin{bmatrix} a & c \\ c & b \end{bmatrix}. \tag{3}$$

Proposition: $|\gamma| \leq \sqrt{\lambda_{max}} - \sqrt{\lambda_{min}}$

Proof. From Eqs. (1) and (3), we have that

$$\alpha^2 = a = \lambda_w \cos^2\theta + \lambda_h \sin^2\theta \tag{4}$$

Since

$$\lambda_w = \frac{\lambda_w + \lambda_h}{2} + \frac{\lambda_w - \lambda_h}{2}, \tag{5}$$

$$\lambda_h = \frac{\lambda_w + \lambda_h}{2} + \frac{\lambda_h - \lambda_w}{2}, \tag{6}$$

we can rewrite Eq. (4) as

$$\alpha^2 = \frac{\lambda_w + \lambda_h}{2}(\cos^2\theta + \sin^2\theta) \tag{7}$$

$$+ \frac{\lambda_w - \lambda_h}{2}(\cos^2\theta - \sin^2\theta) \tag{8}$$

$$= \frac{\lambda_w + \lambda_h}{2} + \frac{\lambda_w - \lambda_h}{2}\cos(2\theta) \tag{9}$$

From Eqs. (1) and (3), we have that

$$\gamma^2 = \frac{c^2}{\alpha^2} = \frac{1}{2}\frac{(\lambda_w - \lambda_h)^2 \sin^2(2\theta)}{(\lambda_w + \lambda_h) + (\lambda_w - \lambda_h)\cos(2\theta)}. \tag{10}$$

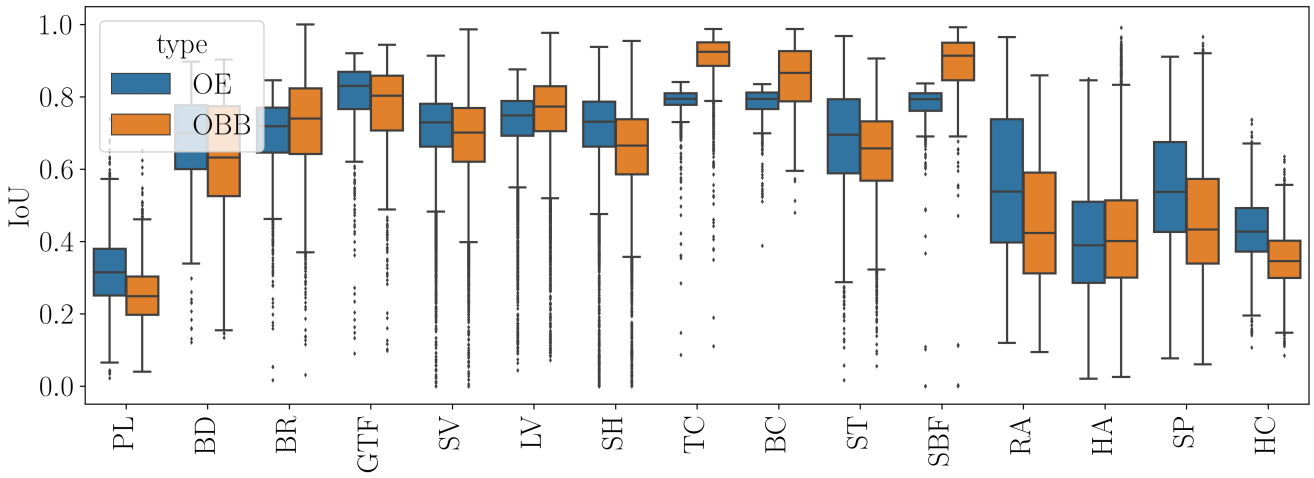

Figure 2. IoU between OBB and OE representations with the corresponding segmentation masks in DOTA.

Defining $x = \cos(2\theta)$, we have that $x \in [-1, 1]$. We can express γ^2 as a function of x, given by

$$\gamma^2 = f(x) = \frac{1}{2} \frac{(\lambda_w - \lambda_h)^2 (1 - x^2)}{(\lambda_w + \lambda_h) + (\lambda_w - \lambda_h) x}, \tag{11}$$

so that

$$f'(x) = \frac{(\lambda_h - \lambda_w)^2 \left(\lambda_h x^2 - 2\lambda_h x + \lambda_h - \lambda_w x^2 - 2\lambda_w x - \lambda_w\right)}{2\left(\lambda_h x - \lambda_h - \lambda_w x - \lambda_w\right)^2}$$

The only solution of $f'(x) = 0$ in the interval $[-1, 1]$ is given by

$$x^* = \frac{\lambda_h + \lambda_w - 2\sqrt{\lambda_h}\sqrt{\lambda_w}}{\lambda_h - \lambda_w}. \tag{12}$$

Since $f(-1) = f(1) = 0$ and γ^2 is non-negative, the global maximum occurs at $x = x^*$. The maximum value of γ^2 is given by

$$f(x^*) = \lambda_h + \lambda_w - 2\sqrt{\lambda_h}\sqrt{\lambda_w} = (\sqrt{\lambda_w} - \sqrt{\lambda_h})^2. \tag{13}$$

Finally, we have that

$$\max |\gamma| = \sqrt{\max \gamma^2} = \sqrt{\lambda_{max}} - \sqrt{\lambda_{min}}, \tag{14}$$

□

3. Comparison Between OBBs and OEs in DOTA

To show that Oriented Ellipses (OEs) can be used as an alternative to Oriented Bounding Boxes (OEs) for representing typical objects in oriented object detection, we performed a study based on the DOTA 1.0 dataset [2], which provides OBB annotations. From each OBB, we generated an OE representation with the same orientation of the OBB and semi-axis composed of half of the OBB dimensions, as explained in the paper. For each annotation, represented as both OBB and OE, we computed the IoU with the segmentation masks provided in [1].

Figure 2 shows the IoU values for OE and OBBs considering all 15 categories of the DOTA dataset. The median IoU value computed with OEs is higher than the IoU using OBBs in nine of the sixteen categories: PL, BD, GTF, SV, SH, ST, RA, SP, and HC. In particular, we highlight the relatively low IoU values for RA (`roundabout`) using OEs. The main cause is the discrepancy between the OBB and segmentation mask annotations, as illustrated in Figure 3. In Figures 3a and 3b, the mask comprises only the roundabout, but the OBB also includes the surrounding street. In Figure 3c, both OBB and mask comprise the roundabout and street, while in Figure 3d, they comprise only the roundabout.

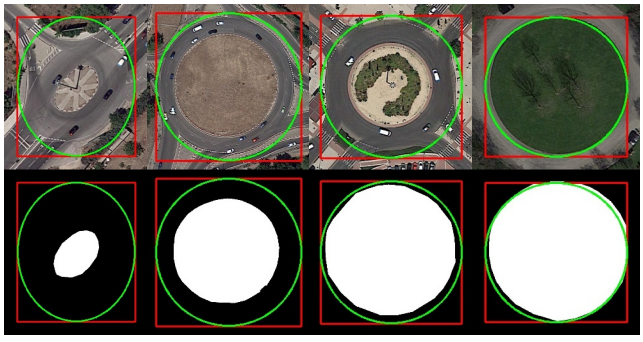

Figure 3. Examples of inconsistencies between OBB and segmentation masks for `roundabouts` in DOTA and iSAID.

4. Flowchart and Visual Results

Figure 4 provides an overview of our method and how we incprporate Cholesky parameters to free-anchor and anchor-based detectors. For free anchor detectors, the Cholesky parameters were regressed directly from the networks and can be mapping to Gaussian parameters to calculate a loss functions as well decoded to box or ellipse. In anchor based detectors, we establish a relationship between HBB anchors and the Cholesky parameters to perform the training and test.

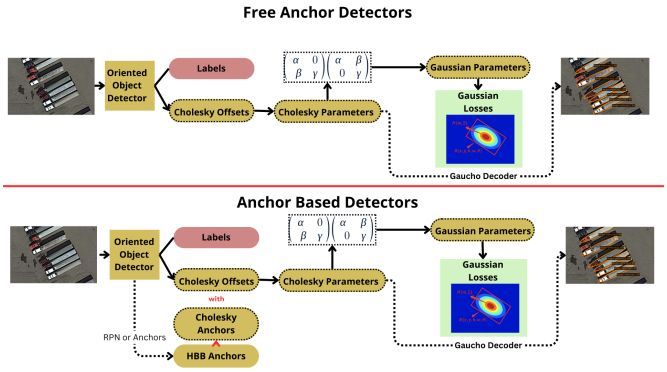

Figure 4. Example of GauCho for oriented object detection for free anchors and anchor-based detectors. For anchor-based approach, a relationship between HBB anchors and Cholesky was proposed. While free anchors detectors use the Cholesky decomposition directly.

Also, we show some results of oriented object detection using GauCho. We selected some representative images of the tested datasets (DOTA, HRSC and UCAS-AOD) and showed the results as both OBBs and OEs.

Figure 5 shows a visual comparison of FCOS-GauCho and FCOS-Baseline over rotated images of the HRSC dataset aiming to evaluate the rotation equivariance (RE) assumption. Although the results of both detectors are mostly coherent, using the GauCho head yields better orientation consistency (see the maximum orientation error for each detector).

Figure 6 shows detection results using FCOS-GauCho with ProbIoU loss for UCAS-AOD using both OBB and OE representations. This particular image shows the known *decoding ambiguity* problem for square-like objects when Gaussian-based loss functions are used: the orientation of the planes cannot be retrieved using OBB representations, leading to larger discrepancies between predictions and GT annotations (Figure 6a). On the other hand, the proposed OEs are fully compatible with GauCho and Gaussian-based loss functions, since the orientation has a small impact on ellipses with small aspect ratios, as shown in Figure 6b. Figure 7 shows a similar result for the DOTA dataset: Figures 7a-d illustrate detection results as OBBs, while Figures 7e-h depict the same detections as OEs. Unfortunately, we do not have access to GT annotations of the test set in DOTA. However, we note that FCOS-GauCho produces coherent results.

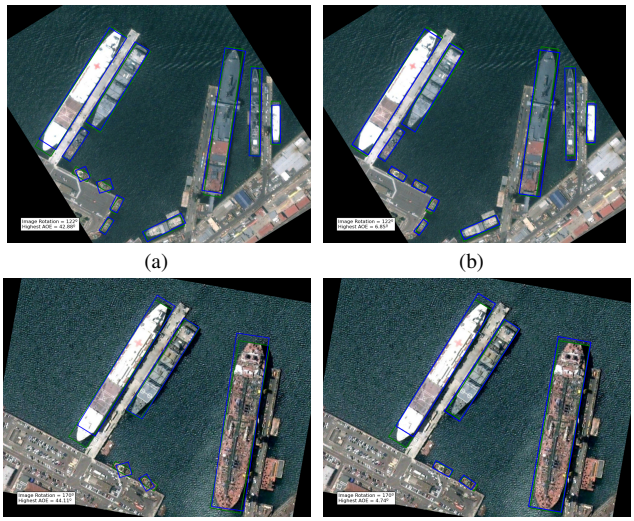

Figure 5. Visual comparison of FCOS-Baseline (a, c) and FCOS-GauCho (b, d) under rotated images of the HRSC dataset. Blue boxes represent detections that matched a ground truth box with IoU $\geq$ 0.5 and green boxes represent the matched ground truth boxes. Highest AOE refers to the highest achieved absolute orientation error of a detection in relation to its matched ground truth.

5. Performance Comparison

In this section, we analyze and provide insight into the training and testing time associated with Gaucho compared to traditional models. In the training stage, model size is the same for an OBB head and GauCho since both need to regress three shape-related parameters plus the centroid. Regarding the computational cost, training GauCho with a Gaussian loss is slightly lighter than an OBB head since Eq. (5) is simpler than Eqs. (1)-(2) from main paper. For inference, GauCho is slightly heavier since we must compute the eigendecomposition of a 2×2 matrix.

However, these differences are small compared to the cost of the backbone. Table 1 presents a test time comparison for all our trained methods in HRSC2016 dataset on two different machines: Machine 1 uses a NVIDIA GeForce RTX 3090 GPU, while Machine 2 uses a NVIDIA Titan Xp GPU.

References

[1] Syed Waqas Zamir, Aditya Arora, Akshita Gupta, Salman Khan, Guolei Sun, Fahad Shahbaz Khan, Fan Zhu, Ling Shao, Gui-Song Xia, and Xiang Bai. iSAID: A large-scale dataset

(a)

(b)

Figure 6. Detection outputs (blue) for FCOS-GauCho on the UCAS-AOD dataset and GT annotations (green), using (a) OBB and (b) OE representations.

Table 1. Comparison of different detectors using the original OBB head and the proposed GauCho with single-scale training/testing. AP values are computed with OBB or OE representations, and the best result for each detector-loss is shown in bold.

Machine	Model	Mean FPS (↑Avg)	Test time (↓Avg)
Machine 1	Roi Trans	**38.00**	25,78
	GauCho Roi Trans	**38.00**	**26.00**
	RetinaNet	**51.85**	**19.00**
	GauCho RetinaNet	49.30	20.00
	R3DET	**36.00**	**27.00**
	GauCho R3DET	32.81	30.00
	FCOS	**54.62**	**18.00**
	GauCho FCOS	52.62	18.33
Machine 2	Roi Trans	**37.67**	**26.53**
	GauCho Roi Trans	37.33	26.79
	RetinaNet	**32.59**	**30.67**
	GauCho RetinaNet	30.66	32.59
	R3DET	**23.86**	**41.97**
	GauCho R3DET	21.51	46.51
	FCOS	**16.38**	**61.08**
	GauCho FCOS	16.24	61.58

for instance segmentation in aerial images. In *CVPRW*, pages 28–37, 2019. 2

[2] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. DOTA: A large-scale dataset for object detection in aerial images. In *CVPR*, 2018. 2

[3] Hang Xu, Xinyuan Liu, Haonan Xu, Yike Ma, Zunjie Zhu, Chenggang Yan, and Feng Dai. Rethinking boundary discontinuity problem for oriented object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17406–17415, 2024. 1

[4] Xue Yang, Xiaojiang Yang, Jirui Yang, Qi Ming, Wentao Wang, Qi Tian, and Junchi Yan. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence. In *NeurIPS*, pages 18381–18394, 2021. 1

[5] Yi Yu and Feipeng Da. On boundary discontinuity in angle regression based arbitrary oriented object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1

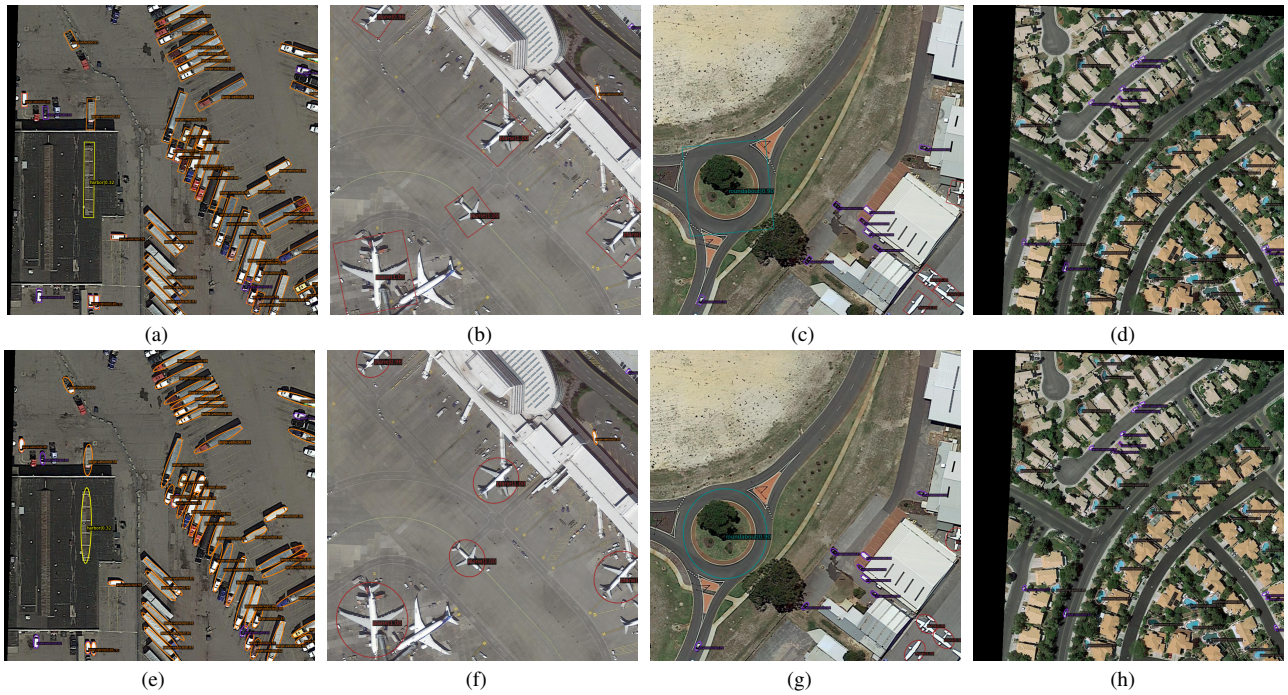

Figure 7. Qualitative comparison of FCOS-Gaucho using OBBs (top) and OEs (bottom) from some images from the DOTA dataset. Better seen zoomed.