Gaussian Label Distribution Learning for Spherical Image Object Detection

Hang Xu[1,2], Xinyuan Liu[2], Qiang Zhao[2], Yike Ma[2], Chenggang Yan[1], Feng Dai[2]

[1]Hangzhou Dianzi University [2]Institute of Computing Technology, Chinese Academy of Sciences

杭州電子科技大學 HANGZHOU DIANZI UNIVERSITY

中国科学院计算技术研究所 INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Introduction:

- Task:
  - Design an easy-to-implement and fully differentiable Joint optimization loss for Spherical Image Object Detection.
- Challenges:
  - Spherical IoU is not differentiable, making it impossible to use as a loss function for box.
  - Currently, only independently optimized Ln loss can be used in spherical object detection.
- Contributions:
  - We explore a new regression loss function based on Gaussian Label Distribution Learning (GLDL) for spherical object detection task. It achieves a trend-level alignment with SphIoU loss and thus naturally improves the model.
  - We align the measurement between sample selection and loss regression based on the GLDL, and then construct new dynamic sample selection strategies (GLDL-ATSS) accordingly. GLDL-ATSS can alleviate the drawback of IoU threshold-based strategy (i.e., scale-sample imbalance).
  - Extensive experimental results on two datasets and popular spherical image detectors show the effectiveness of our approach.

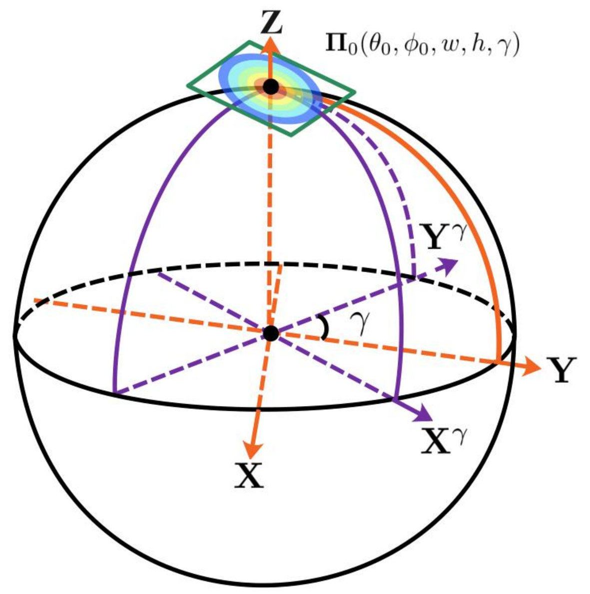
Proposed Method:

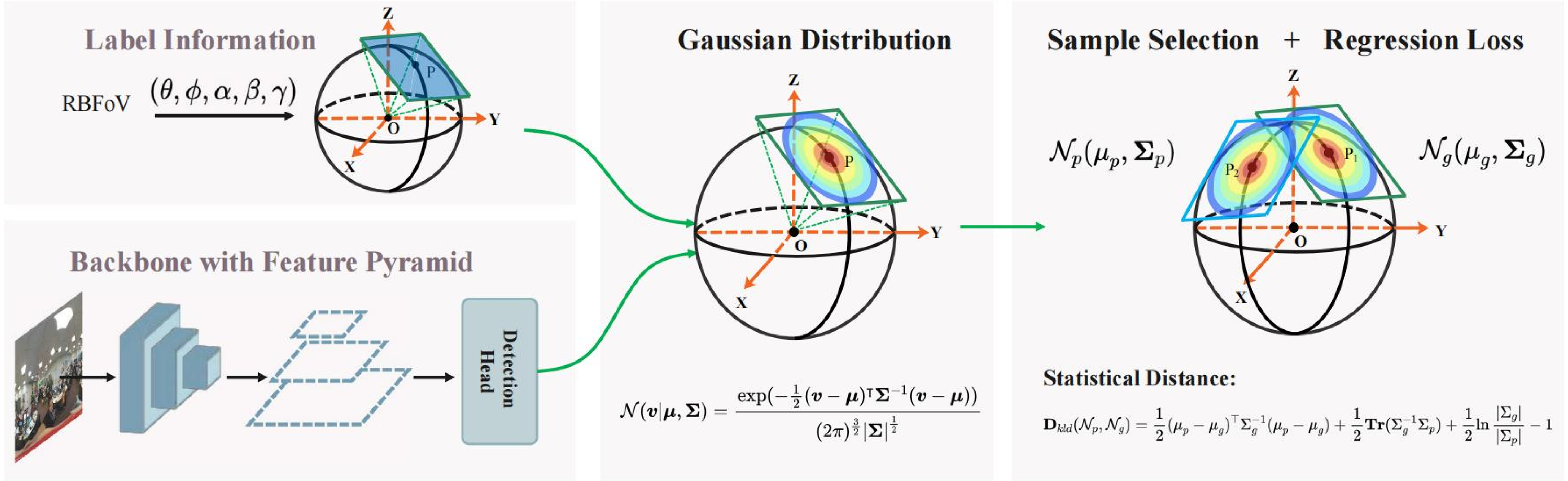
- Gaussian Label Distribution Learning
  - Firstly, transform the box in polar region to a Gaussian distribution

$$\boldsymbol{\mu}_0 = [\sin(\phi_0)\cos(\theta_0), \sin(\phi_0)\sin(\theta_0), \cos(\phi_0)]$$

$$\boldsymbol{\Sigma}_0 = \mathbf{R}\boldsymbol{\Lambda}\mathbf{R}^\top,$$

$$\mathbf{R} = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} \frac{w^2}{4} & 0 & 0 \\ 0 & \frac{h^2}{4} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$



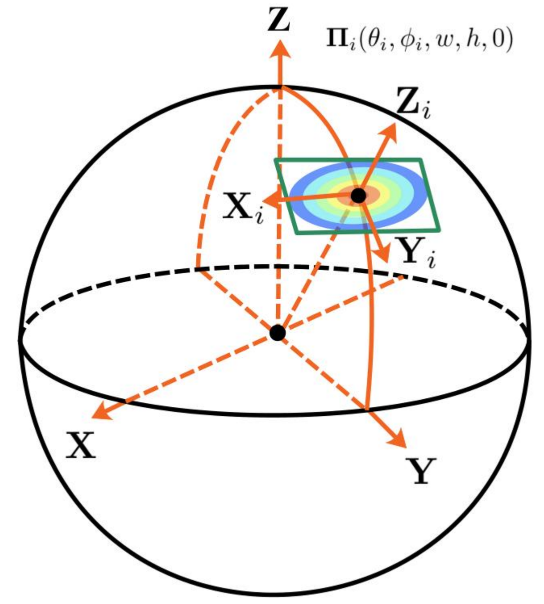


- Secondly, establish a new coordinate system

$$\begin{cases} \mathbf{X}_i = [\sin(\theta_i), -\cos(\theta_i), 0]^\top \\ \mathbf{Y}_i = [\cos(\phi_i)\cos(\theta_i), \cos(\phi_i)\sin(\theta_i), -\sin(\phi_i)]^\top \\ \mathbf{Z}_i = [\sin(\phi_i)\cos(\theta_i), \sin(\phi_i)\sin(\theta_i), \cos(\phi_i)]^\top \end{cases}$$

$$\boldsymbol{\mu}_i = (\sin(\phi_i)\cos(\theta_i), \sin(\phi_i)\sin(\theta_i), \cos(\phi_i))$$

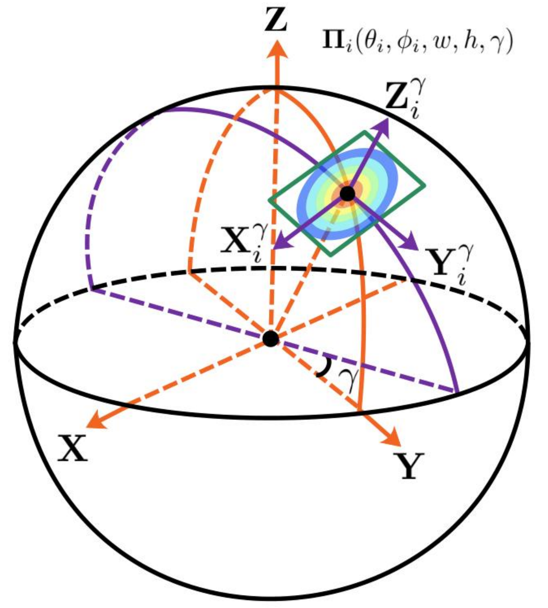
$$\boldsymbol{\Sigma}_i = \mathbf{R}(\mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^\top)\mathbf{R}^\top, \ \mathbf{T} = [\mathbf{X}_i, \mathbf{Y}_i, \mathbf{Z}_i]$$



- Finally, measure two distributions

$$D_{kl}(\mathcal{N}_p, \mathcal{N}_g) = \frac{1}{2}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_g)^\top \boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_g) + \frac{1}{2}tr(\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Sigma}_p) + \frac{1}{2}\ln\frac{|\boldsymbol{\Sigma}_g|}{|\boldsymbol{\Sigma}_p|} - 1$$

$$\mathcal{L}_{reg} = 1 - \frac{1}{\tau + f(D_{kl})}$$



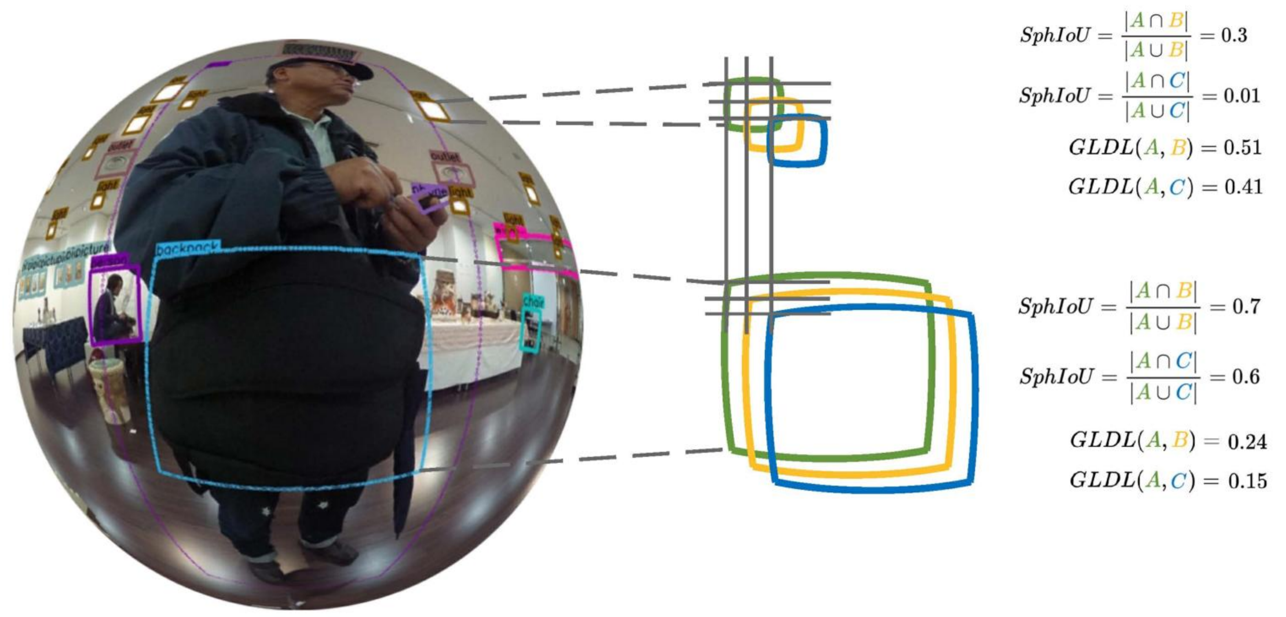
- Sample Selection Based on GLDL-ATSS

$$f(D_{kl}) = \frac{1}{c + D_{kl}(\mathcal{N}_g, \mathcal{N}_p)}$$

$$m_g^i = \frac{1}{N}\sum_{j=1}^{N} f^{i,j}(D_{kl})$$

$$v_g^i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(f^{i,j}(D_{kl}) - m_g^i)^2}$$

$$t_g^i = m_g^i + v_g^i$$



Experiments:

- Ablation study of hyper-parameters.

Dataset	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$	baseline
360-Indoor	20.3	**20.5**	20.4	20.0	19.4	17.6
PANDORA	20.1	**20.3**	20.1	19.9	19.7	17.2

Dataset	$c = 1$	$c = 2$	$c = 3$	$c = 4$	$c = 5$	baseline
360-Indoor	21.5	**21.8**	21.7	21.4	21.2	20.1
PANDORA	20.9	**21.3**	21.2	21.1	22.8	19.6

- Ablation study of each component.

Dataset	Backbone	$\mathcal{S}_{ss}$	$\mathcal{L}_{reg}$	AP_{50}
360-Indoor	R-101	$\mathcal{S}_{IoU}$ (Fixed)	$\mathcal{L}_{L1}$	17.6
	R-101	$\mathcal{S}_{IoU}$ (Fixed)	$\mathcal{L}_{GLDL}$	20.7 (+3.1)
	R-101	$\mathcal{S}_{GLDL}$ (Fixed)	$\mathcal{L}_{GLDL}$	22.8 (+5.2)
	R-101	$\mathcal{S}_{IoU}$ (ATSS)	$\mathcal{L}_{L1}$	20.1
	R-101	$\mathcal{S}_{IoU}$ (ATSS)	$\mathcal{L}_{GLDL}$	22.3 (+2.2)
	R-101	$\mathcal{S}_{GLDL}$ (ATSS)	$\mathcal{L}_{GLDL}$	25.0 (+4.9)
PANDORA	R-101	$\mathcal{S}_{IoU}$ (Fixed)	$\mathcal{L}_{L1}$	17.2
	R-101	$\mathcal{S}_{IoU}$ (Fixed)	$\mathcal{L}_{GLDL}$	21.4 (+4.2)
	R-101	$\mathcal{S}_{GLDL}$ (Fixed)	$\mathcal{L}_{GLDL}$	22.7 (+5.5)
	R-101	$\mathcal{S}_{IoU}$ (ATSS)	$\mathcal{L}_{L1}$	19.6
	R-101	$\mathcal{S}_{IoU}$ (ATSS)	$\mathcal{L}_{GLDL}$	23.4 (+3.8)
	R-101	$\mathcal{S}_{GLDL}$ (ATSS)	$\mathcal{L}_{GLDL}$	25.2 (+5.6)

- Comparison of the performance of different methods.

Method	Backbone	$\mathcal{S}_{ss}$		$\mathcal{L}_{reg}$		360-Indoor			PANDORA		
		$\mathcal{S}_{IoU}$	$\mathcal{S}_{GLDL}$	$\mathcal{L}_{L1}$	$\mathcal{L}_{GLDL}$	AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}
Multi-Kernel [26]	R-101	✓		✓		4.7	11.1	2.8	4.2	10.8	2.2
	R-101	✓			✓	7.2(+2.5)	14.2(+4.1)	5.4(+2.4)	7.8(+3.6)	15.6(+4.8)	4.3(+2.1)
	R-101		✓	✓		6.8(+2.1)	13.9(+2.8)	4.7(+1.9)	6.2(+2.0)	14.5(+3.7)	3.9(+1.7)
	R-101		✓		✓	9.3(+4.6)	17.2(+6.1)	6.6(+3.8)	10.2(+6.0)	17.6(+6.8)	6.9(+4.4)
Sphere-SSD [4]	R-101	✓		✓		2.9	7.8	1.4	2.3	7.7	1.5
	R-101	✓			✓	5.6(+2.7)	10.8(+3.0)	4.2(+2.8)	5.9(+3.6)	12.3(+4.6)	4.9(+3.4)
	R-101		✓	✓		4.9(+2.0)	10.2(+2.4)	3.7(+2.3)	4.1(+1.8)	9.8(+2.1)	3.2(+1.7)
	R-101		✓		✓	7.8(+4.9)	12.6(+4.8)	5.4(+4.0)	8.0(+5.7)	13.8(+6.1)	6.8(+5.3)
Reprojection R-CNN [36]	R-101	✓		✓		5.0	15.3	1.9	4.2	14.7	1.8
	R-101	✓			✓	7.5(+2.5)	18.2(+2.9)	3.8(+1.9)	7.9(+3.7)	18.7(+4.0)	4.5(+2.7)
	R-101		✓	✓		7.1(+2.1)	17.8(+2.5)	3.2(+1.3)	6.8(+2.6)	17.4(+2.7)	3.0(+1.2)
	R-101		✓		✓	10.8(+5.8)	22.5(+7.2)	5.3(+3.4)	11.1(+6.9)	22.8(+8.1)	5.8(+4.0)
Sphere-CenterNet [5]	R-101			✓		10.0	24.8	6.0	-	-	-
	R-101				✓	11.2(+1.1)	26.1(+1.3)	7.4(+1.4)	-	-	-
R-CenterNet [27]	R-101			✓		-	-	-	7.3	22.7	2.6
	R-101				✓	-	-	-	8.7(+1.4)	24.3(+1.6)	4.5(+1.9)