JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

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

Gaussian Label Distribution Learning for Spherical Image Object Detection

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

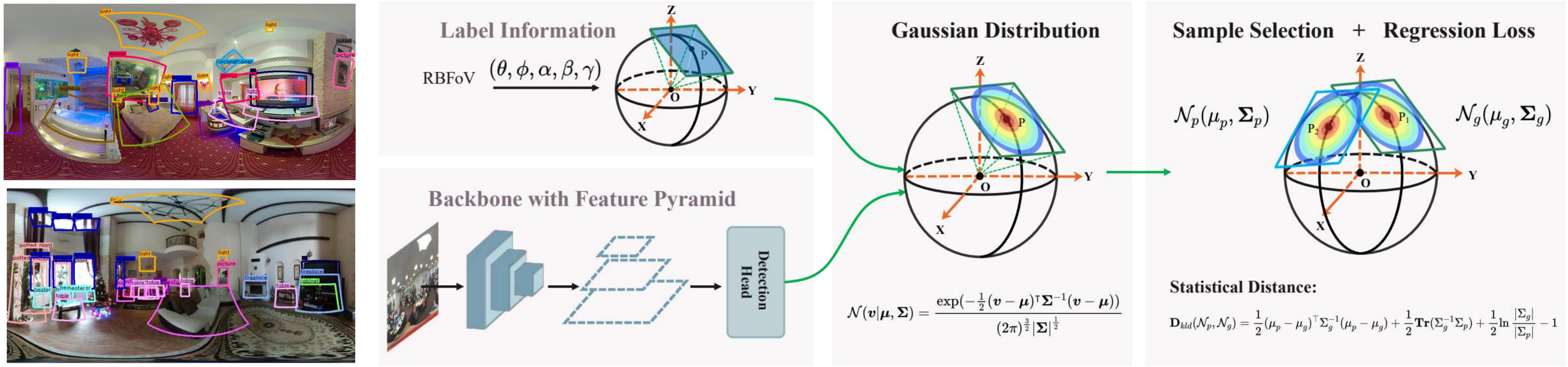
[1]Hangzhou Dianzi University, Hangzhou, China
[2]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

Vancouver, Canada. June 18-22, 2023. (TUE-AM-098)

* Corresponding authors

Overview

- Ln-norm loss has two intrinstic flaws for object objection on spherical images, i.e., independent optimization & inconsistency with IoU-metric.
- We convert spherical boxes into Gaussian Distribution on tangent plane, and design sample selection strategy (GLDL-ATSS) and joint-optimization regression loss (GLDL-Loss) based on distribution distance GLDL which is more consistent with IoU.

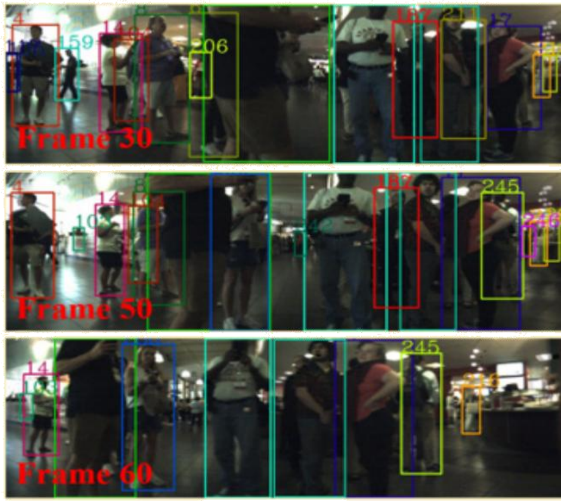


Spherical/Panoramic Image

- Spherical/Panoramic image is a natural extend of comon planar image.
- It has the whole 360° view with richer information and higher practice value.

Environment Perception

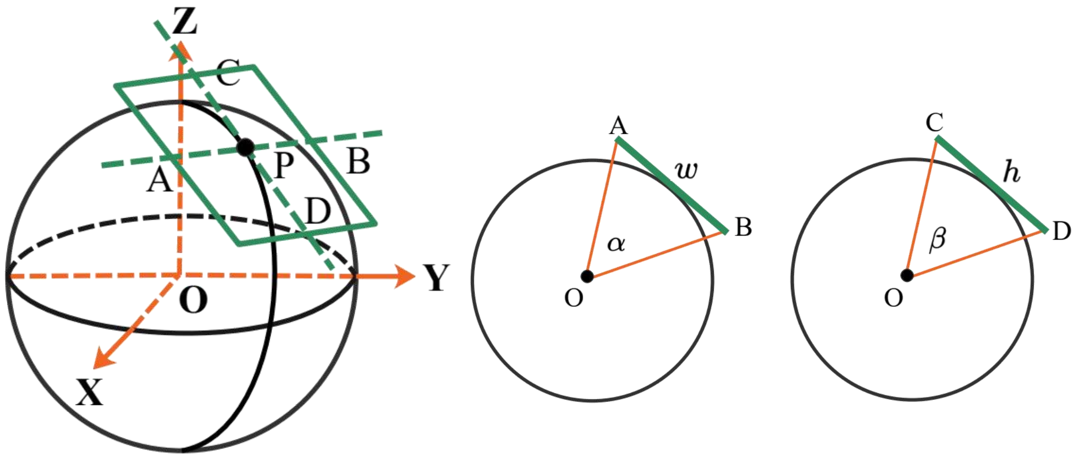
Visual Question&Answer



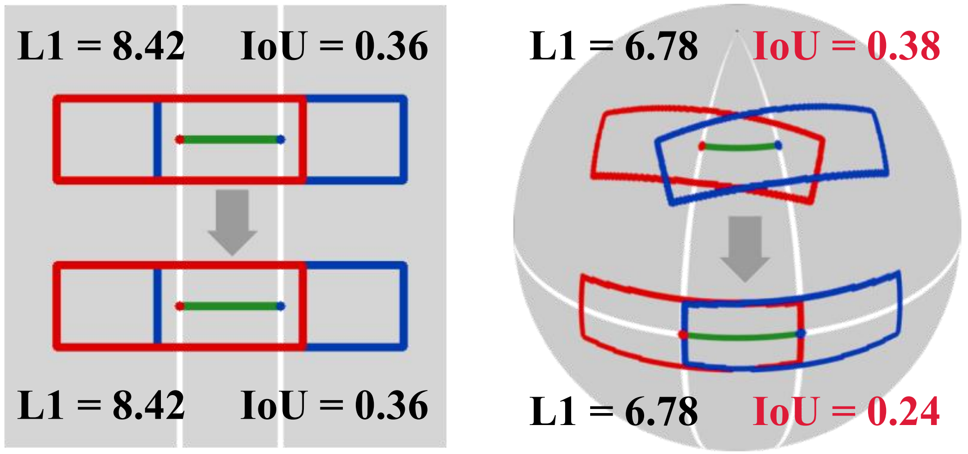
Security Tracking

Spherical Bounding Box

- Spherical bounding box is defined as $(\theta, \phi, \alpha, \beta, \gamma)$.
- P(θ, ϕ) is the tangent point of the sphere and rectangular tangent plane.
- α and β are the horizontal and vertical fields of view of the spherical bounding box.

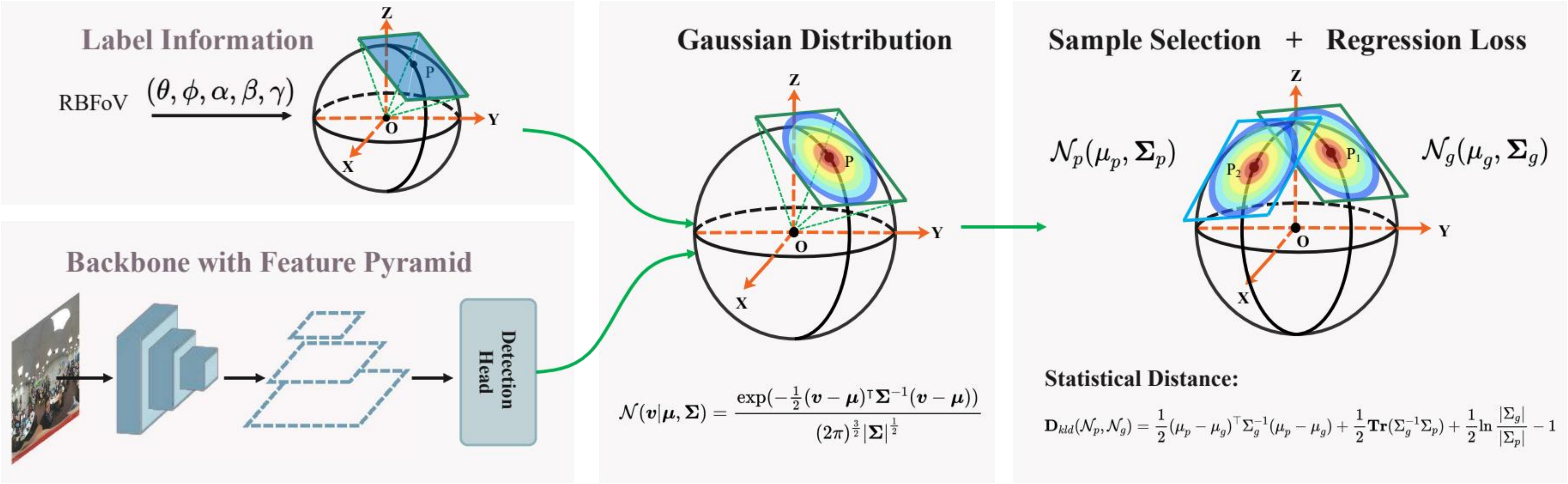


- L1 loss optimizes box parameters independently, inconsistent with metric.
- Moving spherical boxes without L1 change maybe causes sharp IoU change, different from planar boxes.



Gaussian Label Distribution Learning

- Gaussian distributions of spherical bounding boxes are constructed.
- The dynamic sample selection strategy (GLDL-ATSS) and joint-optimization regression loss (GLDL-Loss) are designed in an alignment manner on the basis of K-L divergence.



JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Mathematical Details of GLDL

1. When the object in the polar region, its tangent plane is $\mathbf{\Pi}_0$, it can be converted into a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_0, \mathbf{\Sigma}_0)$.

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

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

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

2. When the object in other region, we can establish a new coordinate system based on its tangent plane $\mathbf{\Pi}_i$.

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

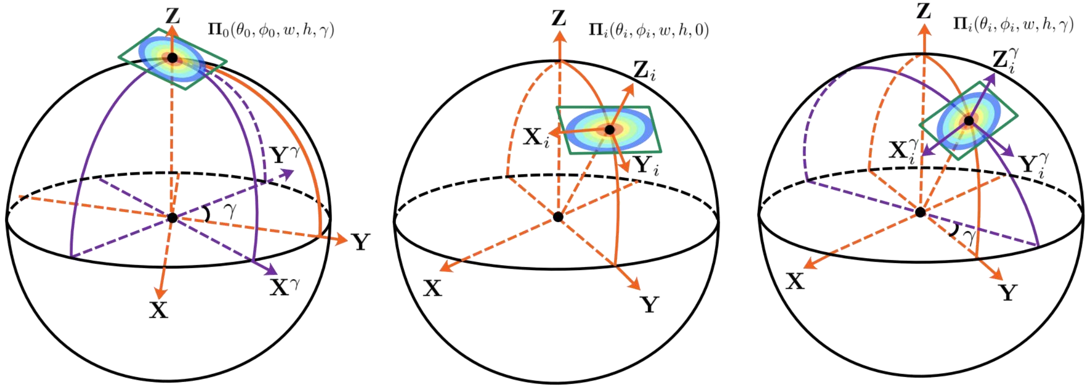
3. Gaussian distributions $\mathcal{N}(\mu_i, \mathbf{\Sigma}_i)$ can be constructed based $\mathbf{X}_i, \mathbf{Y}_i, \mathbf{Z}_i$.

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

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

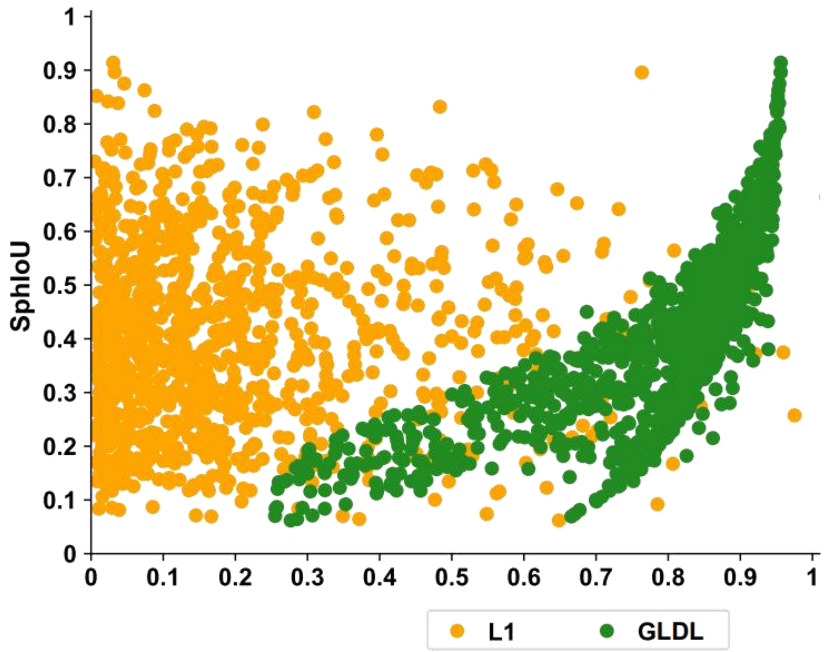
4. GLDL can calculated based on μ_i, Σ_i .

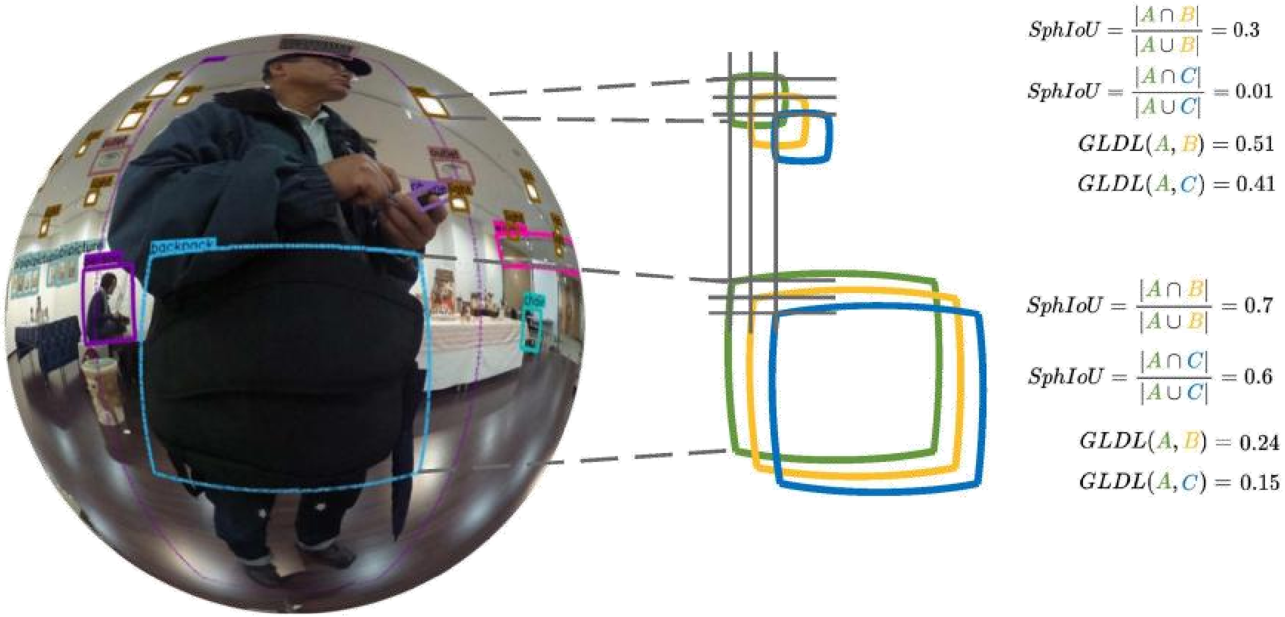
$$D_{kl}(\mathcal{N}_p, \mathcal{N}_g) = \frac{1}{2}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_g)^\top \mathbf{\Sigma}_g^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_g) + \frac{1}{2}tr(\mathbf{\Sigma}_g^{-1}\mathbf{\Sigma}_p) + \frac{1}{2}\ln\frac{|\mathbf{\Sigma}_g|}{|\mathbf{\Sigma}_p|} - 1$$



GLDL v.s. SphIoU

- GLDL has significantly greater **consistency** with SphIoU than L1-norm.
- GLDL has significantly greater **scale invariance** than SphIoU.





JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Ablation Studies

- Our method is robust to hyperparameters.

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

- Improvement dose not come from normalized function itself.

Loss	**Normalized Function**	**360-Indoor**	**PANDORA**
		AP_{50}	**AP_{50}**
Smooth L1	w/	13.7	12.9
	w/o	**17.6**	**17.2**

- GLDL-ATSS and GLDL-Loss can cooperate with each other to improve the detection performance.

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)**

Main Results

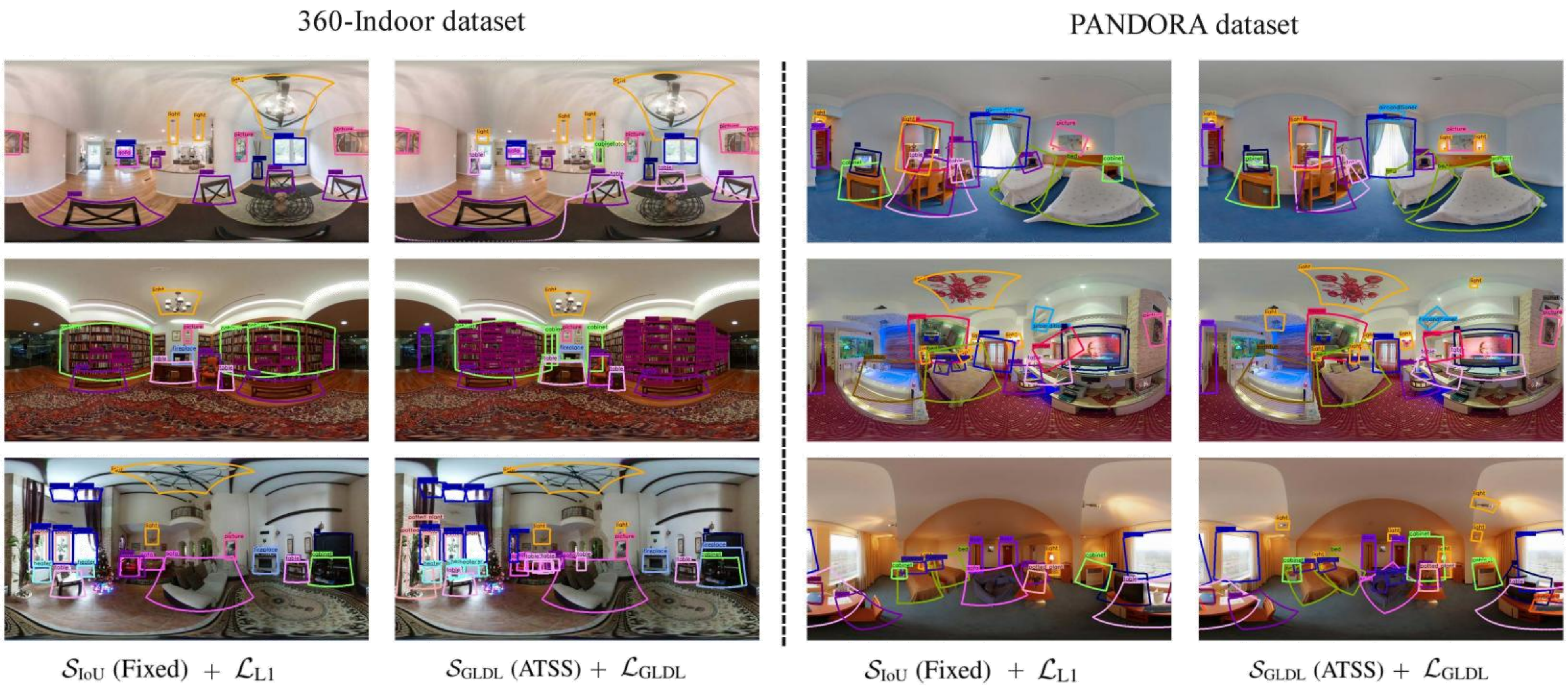
- GLDL-ATSS and GLDL-loss improve various detectors on two mainstream datasets.

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**)

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Visualized Comparsions

- GLDL-ATSS and GLDL-loss help detector to get more precise predicted results.



JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Thank You !

- Paper: Gaussian Label Distribution Learning for Spherical Image Object Detection

- Concat:
  - Hang Xu: hxu@hdu.edu.cn
  - Qiang Zhao: zhaoqiang@ict.ac.cn
  - Feng Dai: fdai@ict.ac.cn
- Home of our Lab: http://iipl.net.cn/index/index.aspx