Sph2Pob: Boosting Object Detection on Spherical Images with Planar Oriented Boxes Methods

Xinyuan Liu[1,2], **Hang Xu**[3*], **Bin Chen**[1,2], **Qiang Zhao**[1*], **Yike Ma**[1], **Chenggang Yan**[3] and **Feng Dai**[1†]
[1]Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China
[3]Hangzhou Dianzi University, Hangzhou, China
{liuxinyuan21s, chenbin20s, zhaoqiang, ykma, fdai}@ict.ac.cn, {hxu, cgyan}@hdu.edu.cn

Abstract

Object detection on panoramic/spherical images has been developed rapidly in the past few years, where IoU-calculator is a fundamental part of various detector components, i.e. Label Assignment, Loss and NMS. Due to the low efficiency and non-differentiability of spherical Unbiased IoU, spherical approximate IoU methods have been proposed recently. We find that the key of these approximate methods is to map spherical boxes to planar boxes. However, there exists two problems in these methods: (1) they do not eliminate the influence of panoramic image distortion; (2) they break the original pose between bounding boxes. They lead to the low accuracy of these methods. Taking the two problems into account, we propose a new sphere-plane boxes transform, called Sph2Pob. Based on the Sph2Pob, we propose (1) an differentiable IoU, Sph2Pob-IoU, for spherical boxes with low time-cost and high accuracy and (2) an agent Loss, Sph2Pob-Loss, for spherical detection with high flexibility and expansibility. Extensive experiments verify the effectiveness and generality of our approaches, and Sph2Pob-IoU and Sph2Pob-Loss together boost the performance of spherical detectors. The source code is available at https://github.com/AntXinyuan/sph2pob.

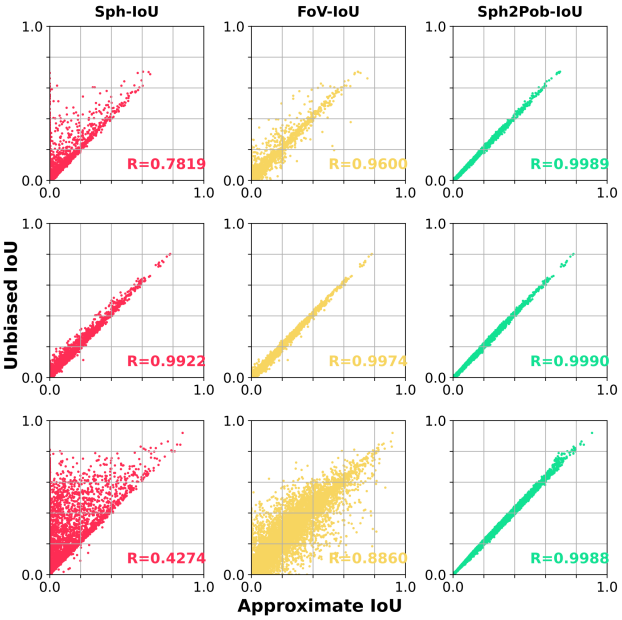

Figure 1: Consistency comparison between Unbiased-IoU [Dai *et al.*, 2022a] and other approximate IoU for spherical boxes in different cases. From left to right are respectively Sph-IoU [Zhao *et al.*, 2020], Fov-IoU [Cao *et al.*, 2022] and our proposed Sph2Pob-IoU; while from top to bottom are respectively cases corresponding all-latitudes, low-latitudes and high-latitudes. R refers to the Correlation Coefficient. Our methods achieve the best consistency with Unbiased-IoU in all cases.

1 Introduction

For the demand of comprehensive environment perception in emerging applications such as as robotics [Kang and Cho, 2019; Zhang *et al.*, 2021] and automatic driving [Yang *et al.*, 2021a; He *et al.*, 2021], object detection can be naturally extended from planar images to panoramic images, because the latter have the whole 360° view with richer information in a compact form. Unlike planar images, panoramic images usually take the sphere as the imaging surface and are stored in the ERP-format [Snyder, 1997] based on polar-coordinates in many works [Chou *et al.*, 2020; Xu *et al.*, 2022], so object detection on them is also called *spherical (object) detection*.

To tightly enclose objects on spherical images, mainstream method usually replace planar rectangular boxes with spherical rectangular boxes, which is also known as the Bounding Field of View (BFoV) [Chou *et al.*, 2020; Dai *et al.*, 2022a] directly defined on the sphere rather than plane. However, as is shown in Figure 2(a), this representation introduces some extra challenges about IoU calculation and Loss design for spherical detection. On the one hand, the calculation of spherical IoU has not been well solved, due to the complexity of intersection between two spherical boxes, which limits com-

*This work was done when Hang Xu & Qiang Zhao were at ICT.
†Corresponding author.

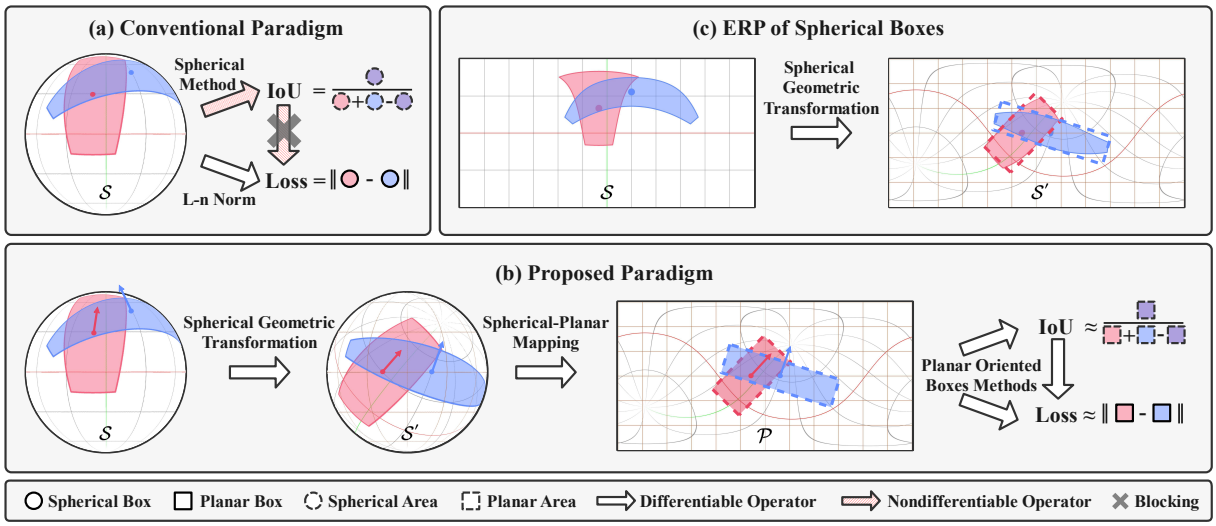

Figure 2: Overview of our proposed Sph2Pob for spherical object detection. (a) In conventional paradigm, IoU and Loss are directly calculated for spherical boxes. (b|c) In proposed paradigm, spherical boxes are transformed to planar oriented boxes, and then IoU and Loss are indirectly calculated. The latter overcomes the nondifferentiable issue of spherical IoU and enriches Loss design for spherical detection.

ponents relying on IoU, such as Label Assignment and NMS. From the previous literatures, some approximation of spherical IoU, such as Sph-IoU [Zhao *et al.*, 2020] and Fov-IoU [Cao *et al.*, 2022] sacrifice too much accuracy due to simplistic approximation strategies. In contrast, Unbiased-IoU [Dai *et al.*, 2022a] innovatively provides an accurate IoU, but it needs a complex DFS algorithm to handle special cases, making the whole calculation tough to either accelerate in parallel or differentiate. On the other hand, researchers almost have no choice but naive L-n Loss during training because of the complex spherical natures, and let alone popular IoU-based Loss since the Unbiased-IoU mentioned before is undifferentiable. According to the research [Yu *et al.*, 2016], this inconsistencies between the L-n Loss and the IoU metric lead to suboptimal trained models.

In this paper, we explore a novel paradigm to comprehensively boost object detection on spherical images from the perspectives of IoU and Loss with the power of methods for planar oriented boxes, as is shown in Figure 2(b|c). Reviewing Sph-IoU [Zhao *et al.*, 2020] and Fov-IoU [Cao *et al.*, 2022], we discover that their key insight is to move spherical boxes to the equator, which is the line with no distortion on ERP-images. Then, the spherical boxes are regarded as planar boxes, and the spherical IoU is approximated by planar IoU. However, in order not to change the relative position of the two spherical boxes, Sph-IoU and Fov-IoU actually move the middle of two centers to the equator of the ERP-images, as is shown in Figure 3(a|b). It means that the centers of the two spherical boxes are not at the equator in most cases, resulting in distortion in the spherical boxes. Moreover, we discover that there exists a dynamic and coupled angle (*internal angle*) between two intersecting spherical boxes, as is shown in Figure 2(b), which is another special rotation distinguished from orientations of objects (*external angle*). Taking the two problems into account, we propose a sphere-plane boxes transform algorithm, called Sph2Pob, based on spherical geometric transformation, where the boxes are exactly moved to the equator and are easier to map to planar oriented boxes, and the *internal angle* between two spherical boxes is calculated precisely. Due to the concise logic, the algorithm is easy to accelerate and differentiate. Based on Sph2Pob, spherical IoU can be approximately calculated through IoU for planar oriented boxes [Ma *et al.*, 2018; Zhou *et al.*, 2019]. As is shown in Figure 1, our Sph2Pob-IoU is superior to other methods even in the most hard cases (high-latitudes). In addition, we can also construct agent Loss based on Sph2Pob for spherical detection through Loss for planar oriented boxes [Yang *et al.*, 2021c; Yang *et al.*, 2022b]. This not only makes spherical IoU-based Loss possible, but also greatly enriches alternative of Loss for spherical detection. Overall, our contribution can be summarized as following:

- By analyzing the problems of previous approximate spherical IoU methods, we propose a sphere-plane boxes transform, called Sph2Pob, which makes spherical boxes can be treated as planar oriented boxes.
- We propose an differentiable approximate IoU for spherical boxes based on Sph2Pob with low time-cost and high accuracy, which can extend spherical IoU into multiple components, i.e. Label Assignment, Loss, NMS.
- We propose an agent Loss for spherical detection based on Sph2Pob with high flexibility and expansibility, which enriches Loss design for spherical detection.
- Extensive experimental results verify the effectiveness and generality of our approach, from which spherical detectors can get a significant boost.

2 Related Works

2.1 Spherical Object Detection

In spherical object detection, objects are distributed over spherical images covering the whole 360° view with richer

semantic information and higher practical value. However, spherical imaging surfaces tend to distort objects, which causes the difficulty of object detection from another aspect. To address these issues, researchers usually introduce additional adaptive priors to existing conventional detectors through intricate designs. Multikernel [Wang and Lai, 2019] and Reprojection R-CNN [Zhao *et al.*, 2020] are two-stage approaches whose pipeline is inherited from Faster RCNN [Ren *et al.*, 2015], while Sphere-CenterNet [Dai *et al.*, 2022a] and R-CenterNet [Xu *et al.*, 2022] are based on corresponding single-stage methods for faster detection speed.

However, there exists some more pressing problems in foundational components of spherical detectors, such as Label Assignment, Loss and NMS, due to Intersection over Union (IoU) for spherical boxes. It is easy for planar rectangular boxes to calculate IoU, but intractable for spherical rectangular boxes. No accurate IoU [Zhao *et al.*, 2020; Cao *et al.*, 2022] is proposed until SphIoU[Dai *et al.*, 2022a], but it still has obvious flaw. On the one hand, the Sph-IoU has a complex logic relied on 64-bit high-precision float computation in its offical implement. In addition, it uses a DFS algorithm that is tough to parallelize and does not support CUDA acceleration, which slows down the training and evaluation of the entire network. On the other hand, the whole calculation is non-differentiable, which prevents training benefiting from IoU-based Loss, such as IoU [Yu *et al.*, 2016], GIoU [Rezatofighi *et al.*, 2019], DIoU [Zheng *et al.*, 2020] and CIoU [Zheng *et al.*, 2021] Loss. In addition, it is hard to design other practical Loss due to spherical complexity, so the detectors only use naive L-n Loss during the training phase, which leads to poor detection performance.

2.2 Planar Oriented Object Detection

As a hot spot in research about object detection recently, planar oriented object detection has attracted wide attention in several real scenario, such as aerial images [Dai *et al.*, 2022b], scene text [Dai *et al.*, 2021], face pose [Zhang *et al.*, 2020], etc. In contrast to the popular undirected object detection, oriented object detection requires predicting a rotated rectangular box (x, y, w, h, a) with an additional angular parameter a, but introduces more challenges due to the boundary and square-like issues [Yang *et al.*, 2022a]. Fortunately, there are many solutions to address these issues, especially from the perspective of Loss function, such as IoU-Smooth L1 [Yang *et al.*, 2018], CSL [Yang and Yan, 2020], GWD [Yang *et al.*, 2021b], KLD [Yang *et al.*, 2021c], KFIoU [Yang *et al.*, 2022b], and so on. As for IoU of planar oriented rectangular boxes, although it also faces many computational challenges, there are some excellent solutions, i.e. RotatedIoU [Ma *et al.*, 2018] and differentiable SkewIoU[Zhou *et al.*, 2019].

3 Methodology

3.1 Overview

To solve the distortion position & internal angle problems exposed from Sph-IoU and Fov-IoU, we propose a new sphere-plane boxes transform method, called Sph2Pob. Firstly, we ensure the centers of the two spherical boxes are strictly at the equator of the ERP image to avoid the influence of distortion. Secondly, we calculate the exact *internal angle* between two spherical boxes by pose vectors of boxes, as is shown in Figure 3(c). Finally, we propose a new approximate IoU and Loss based on Sph2Pob for spherical image object detection. The overview of our method is shown in Figure 2.

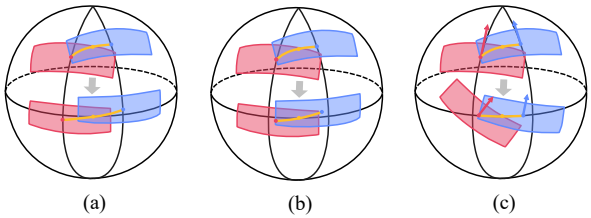

Figure 3: Different approximate calculation of spherical box IoU by moving boxes to the equator without change of (a) longitude difference [Zhao *et al.*, 2020] and (b) arc length between boxes centers [Cao *et al.*, 2022]. Apart from relative position information, our proposed method (c) also keeps relative pose information.

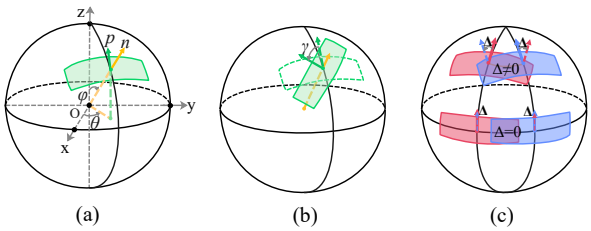

Figure 4: Spherical boxes and two types of rotation. (a) Spherical box BFoV. (b) Spherical box RBFoV with **external angle** γ. (c) The non-parallel longitude lines lead to the formation of the **internal angle** Δ of the two boxes.

3.2 Rethink Orientation of Spherical Boxes

Spherical Bounding Box

A scheme of spherical bounding box is Bounding Field of View (BFoV) $\boldsymbol{B}^{\mathcal{S}}(\theta, \phi, \alpha, \beta)$ [Chou *et al.*, 2020; Dai *et al.*, 2022a], where (θ, ϕ) is the center position, (α, β) are the width & height of the bounding box, as is shown in Figure 4(a). The recent works [Xu *et al.*, 2022; Xu *et al.*, 2023] find that the objects in spherical images often appear with arbitrary orientations, and the BFoV cannot compactly outline oriented instances. Therefore, Rotated Bounding Field of View (RBFoV) $\boldsymbol{B}^{\mathcal{S}}(\theta, \phi, \alpha, \phi, \gamma)$ is proposed as the spherical bounding box, where γ refers to the rotation angle of BFoV around center vector $\boldsymbol{n}(\theta, \phi)$, as is shown in Figure 4(b).

However, we find that in addition to the *external angle* γ about the orientation of the object, there is an *internal angle* between spherical bounding boxes whose center points are not at the same longitude, as is shown in Figure 4(c). This angle is yielded by the intrinsic nature (longitude lines are not parallel) of spherical images and boxes, and does not depend on the orientation of object. It is an ignored but important factor in current researches for approximated spherical IoU.

Internal Angle

BFoVs are strictly aligned with coordinate axes. Although the latitude lines are parallel to each other and perpendicular

Algorithm 1 Sph2Pob

Input: Two spherical boxes $\boldsymbol{B}_i^{\mathcal{S}}(\theta_i, \phi_i, \alpha_i, \beta_i, (\gamma_i))$
Output: Two planar boxes $\boldsymbol{B}_i^{\mathcal{P}}(x_i, y_i, w_i, h_i, a_i)$

1: Compute position $\boldsymbol{n}_i$ and pose $\boldsymbol{p}_i$ via Equation 1, 2.
2: Construct spherical geometric transformation $\boldsymbol{R}$.
3: Transform $\boldsymbol{n}_i, \boldsymbol{p}_i$ to $\hat{\boldsymbol{n}}_i, \hat{\boldsymbol{p}}_i$ with $\boldsymbol{R}$.
4: Compute *internal angle* Δ via Equation 8.
5: Map boxes $\hat{\boldsymbol{B}}_i^{\mathcal{S}}$ to planar boxes. $\boldsymbol{B}_i^{\mathcal{P}}$ via Equation 9
6: (optional) Modify a_i of $\boldsymbol{B}_i^{\mathcal{P}}$ with *external angle* γ_i.
7: **return** $\boldsymbol{B}_1^{\mathcal{P}}, \boldsymbol{B}_2^{\mathcal{P}}$.

to the longitude lines anywhere, the longitude lines are not parallel beyond the equator and converge at the poles, which results in the rotation of spherical boxes coupled with each other, as is shown in Figure 4(c). We call the angle of rotation as *internal angle*.

For a BFoV $\boldsymbol{B}^{\mathcal{S}} = (\theta, \phi, \alpha, \beta)$, we denote the position of the box by the center vector $\boldsymbol{n}$, where

$$\begin{aligned} \boldsymbol{n} &= \boldsymbol{n}(\theta, \phi) = [nx, ny, nz]^\top \\ &= [\sin(\phi)\cos(\theta), \sin(\phi)\sin(\theta), \cos(\phi)]^\top \end{aligned} \tag{1}$$

Then, we can represent the pose of the box by the tangent vector $\boldsymbol{p}$ of its center along the direction of longitude, where

$$\begin{aligned} \boldsymbol{p} = \boldsymbol{p}(\theta, \phi) = \frac{\partial \boldsymbol{n}(\theta, \phi)}{\partial \phi} = [px, py, pz]^\top \\ = [\cos(\phi)\cos(\theta), \cos(\phi)\sin(\theta), -\sin(\phi)]^\top \end{aligned} \tag{2}$$

If the centers of the two boxes are at the same latitude, the *internal angle* Δ can be calculated based on the pose vectors $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ of the two boxes, i.e.

$$\Delta = \arccos(\boldsymbol{p}_1 \cdot \boldsymbol{p}_2) \tag{3}$$

where $\cdot$ denotes the vector inner product.

Moreover, the *internal angle* coupled between two boxes dynamically changes with latitude and longitude, where latitude determines the lower bound, while longitude difference determines the upper bound. Extremely, the angle $\Delta = 0$ (minimum) always holds on when two boxes are located at the equator ($\phi_1 = \phi_2 = \pi/2$), and it reaches maximum $|\theta_1 - \theta_2|$ at the pole ($\phi_1 = \phi_2 \in \{0, \pi\}$).

3.3 Sph2Pob: Sphere-Plane Boxes Transform

Sph2Pob is a boxes transform from sphere to plane. It consists of two parts, i.e. Spherical Geometric Transformation and Spherical-Planar Mapping, and we describe it in Algorithm 1. Given two BFoVs $\boldsymbol{B}_1^{\mathcal{S}} = (\theta_1, \phi_1, \alpha_1, \beta_1)$ and $\boldsymbol{B}_2^{\mathcal{S}} = (\theta_2, \phi_2, \alpha_2, \beta_2)$, the output of Sph2Pob is two planar oriented boxes $\boldsymbol{B}_1^{\mathcal{P}} = (x_1, y_1, w_1, h_1, a_1)$ and $\boldsymbol{B}_2^{\mathcal{P}} = (x_2, y_2, w_2, h_2, a_2)$.

Spherical Geometric Transformation (SGT)

To make the center points of the two spherical boxes strictly on the equator of the ERP image to avoid the influence of distortion, we apply a rotation transformation $\boldsymbol{R}$ to the sphere,

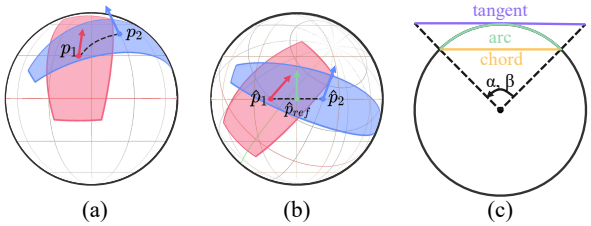

Figure 5: Technical details of Sph2Pob. A spherical rotation transformation $\boldsymbol{R}$ is used to move spherical boxes to from (a) anywhere to (b) the equator, where the arc between the two centers lie on the equator. (c) Three projected edges, i.e. tangent, arc and chord.

making the arc between the two centers lie at the equator, as is shown in Figure 5(b). For unified operation, we supply extra constraints including i) letting midpoint of the arc lie in x-axis-vector $(1, 0, 0)^\top$ and ii) letting $\boldsymbol{B_1}^{\mathcal{S}}$ on the left of $\boldsymbol{B_2}^{\mathcal{S}}$. The rotation matrix is as following:

$$\boldsymbol{R} = [\boldsymbol{v}_x, \boldsymbol{v}_y, \boldsymbol{v}_z]^\top = \left[\frac{\boldsymbol{n}_1+\boldsymbol{n}_2}{\|\boldsymbol{n}_1+\boldsymbol{n}_2\|}, \frac{\boldsymbol{n}_1-\boldsymbol{n}_2}{\|\boldsymbol{n}_1-\boldsymbol{n}_2\|}, \boldsymbol{v}_x \times \boldsymbol{v}_y\right]^\top \tag{4}$$

where $\boldsymbol{n}$ refers to the position vector of $\boldsymbol{B}^{\mathcal{S}}$, and $\times$ denotes the vector cross product. By the way, $\boldsymbol{v}_x \equiv \boldsymbol{n}(\frac{\theta_1+\theta_2}{2}, \frac{\phi_1+\phi_2}{2})$ always keeps, and even avoids the extra standardization due to the uniqueness of Euclidean coordinates.

At this time, the two spherical boxes can be moved to the equator strictly without any loss of both relative position and pose information. The position vector $\hat{\boldsymbol{n}}$ of the new spherical box can be calculated as follows:

$$\hat{\boldsymbol{n}}(\hat{\theta}, \hat{\phi}) = \boldsymbol{R}\,\boldsymbol{n}(\theta, \phi) \tag{5}$$

As can be seen from equation 3, the *internal angle* Δ is determined by the pose vector. Therefore, we first acquire the pose vectors $(\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2)$ of two new spherical boxes:

$$[\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2] = \boldsymbol{R}\,[\boldsymbol{p}_1, \boldsymbol{p}_2] \tag{6}$$

where $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are the pose vectors of the original spherical boxes, as is shown in Figure 5(a). Then, as is shown in Figure 5(b), the *internal angle* Δ of two boxes equivalent to the direct angle between the new pose vectors $\hat{\boldsymbol{p}}_1$ and $\hat{\boldsymbol{p}}_2$:

$$\Delta = \arccos(\hat{\boldsymbol{p}}_1 \cdot \hat{\boldsymbol{p}}_2) \tag{7}$$

To facilitate IoU calculation and Loss design, we decomposed the *internal angle* Δ into each box by a reference vector $\hat{\boldsymbol{p}}_{ref}$ (the pose vector of the middle of two centers), i.e.

$$\begin{aligned} \Delta &= \Delta_1 + \Delta_2 \\ &= \arccos(\hat{\boldsymbol{p}}_1 \cdot \hat{\boldsymbol{p}}_{ref}) + \arccos(\hat{\boldsymbol{p}}_2 \cdot \hat{\boldsymbol{p}}_{ref}) \end{aligned} \tag{8}$$

Spherical-Planar Mapping (SPM)

After being moved to the equator, the new spherical boxes $\hat{\boldsymbol{B}}_i^{\mathcal{S}} = (\hat{\theta}_i, \hat{\phi}_i, \hat{\alpha}_i, \hat{\beta}_i)$ in the ERP image are close to oriented planar boxes in shape, as is shown in Figure 2(c). Then we map them to the planar counterpart further. We can get the planar rotated boxes $\boldsymbol{B}_i^{\mathcal{P}} = (x_i, y_i, w_i, h_i, a_i)$ as follows:

$$\begin{cases} x_i & = \hat{\theta}_i / 2\pi \cdot W \\ y_i & = \hat{\phi}_i / \pi \cdot H \\ w_i & = f(\hat{\alpha}_i) \\ h_i & = f(\hat{\beta}_i) \\ a_i & = \Delta_i = \arccos(\hat{\boldsymbol{p}}_i \cdot \hat{\boldsymbol{p}}_{\boldsymbol{i}ref}) \end{cases} , i = 1, 2 \tag{9}$$

where W and H refer to the width and height of the ERP image, $\hat{\theta}_i$ and $\hat{\phi}_i$ refers to the latitude and longitude corresponding to the transformed position vector $\hat{\boldsymbol{n}}_i$, and $f(\cdot)$ represents the edge projection methods, such as *arc*, *chord* and *tangent*, as is shown in Figure 5(c). In the ablation study, we test these edge projection methods, details will be shown in Table 3.

Sph2Pob-Transform for RBFoV

Given two RBFoVs $\boldsymbol{B_1}^{\mathcal{S}} = (\theta_1, \phi_1, \alpha_1, \beta_1, \gamma_1)$ and $\boldsymbol{B_2}^{\mathcal{S}} = (\theta_2, \phi_2, \alpha_2, \beta_2, \gamma_2)$, the output of Sph2Pob is still two planar boxes $\boldsymbol{B_1}^{\mathcal{P}} = (x_1, y_1, w_1, h_1, a_1)$ and $\boldsymbol{B_2}^{\mathcal{P}} = (x_2, y_2, w_2, h_2, a_2)$, where (x_i, y_i, w_i, h_i) are the same as in equation 9. Due to *external angle* is equivalent to planar rotation angle in ERP-images, the angle a_i can be obtained by simple addition:

$$a_i = \Delta_i + \gamma_i, i = 1, 2 \tag{10}$$

It can be found that our method is compatible with the two spherical bounding boxes BFoV and RBFoV.

3.4 A New Paradigm for Spherical Detection

Training of Spherical Detectors with Planar Methods

Common object detectors can be divided into two types, i.e. anchor-based and anchor-free ones, where the former adopts boxes (also called anchors) as priors, while the latter adopts points as priors. In general, anchor-based detectors performs better than anchor-free detectors, but they need calculate IoUs between anchors and ground-truth-boxes to assign pos / neg labels to anchors, and only positive samples are used in Loss calculation for boxes-prediction. Thus, the accuracy of IoU directly affect the sample quality and and subsequent training effect. By the way, time-cost of IoU should be considered due to the huge number of anchors (typical 200k). In addition, Loss design also significantly affect the training effect, regardless of the type of detectors.

So the transform from spherical boxes to planar oriented boxes is actually a huge leap forward, and we can utilize the power of planar oriented methods to comprehensively boost spherical detection from the perspectives of IoU and Loss.

IoU Approximation of Spherical Boxes

Sph2Pob-Transform first brings a new IoU approximation method. IoU for planar oriented boxes [Ma *et al.*, 2018; Zhou *et al.*, 2019] has been studied extensively, and some mainstream deep learning frameworks have also implemented efficient differentiable CUDA operator for it. Using these operators, we can calculate the IoU of the spherical boxes indirectly, i.e

$$IoU^{\mathcal{S}}(\boldsymbol{B}_1^{\mathcal{S}}, \boldsymbol{B}_2^{\mathcal{S}}) \approx IoU^{\mathcal{P}}(Sph2Pob(\boldsymbol{B}_1^{\mathcal{S}}, \boldsymbol{B}_2^{\mathcal{S}})) \tag{11}$$

Later, we will verify through experiments that this IoU has tiny approximation error and high computational efficiency, so it can directly replace the existing slow SphIoU in various components, including Label Assignment, Loss and NMS.

Agent Spherical Loss based on Planar Boxes

Due to proposed differentiable Sph2Pob-Transform, those well-done IoU-based Loss [Yu *et al.*, 2016; Rezatofighi *et al.*, 2019; Zheng *et al.*, 2020; Zheng *et al.*, 2021] commonly used on the plane can be introduced into spherical detection. Furthermore, the existing [Yang *et al.*, 2021b; Yang *et al.*, 2021c; Yang *et al.*, 2022b] and future Loss designed for the planar oriented detection can also be easily introduced here, i.e.

$$Loss^{\mathcal{S}}(\boldsymbol{B}_1^{\mathcal{S}}, \boldsymbol{B}_2^{\mathcal{S}}) \approx Loss^{\mathcal{P}}(Sph2Pob(\boldsymbol{B}_1^{\mathcal{S}}, \boldsymbol{B}_2^{\mathcal{S}})) \tag{12}$$

The agent Loss is extremely flexible and extensible, greatly enriching methods for spherical detection. From the macroscopic perspective, spherical detection can be regarded as a special case of planar oriented detection in a sense.

4 Experiments

4.1 Datasets

360-Indoor [Chou *et al.*, 2020] is a classic real-world dataset for spherical object detection. It consists of 3,335 indoor spherical images and 89,148 spherical box annotations (BFoV of 4 parameters) among 37 categories. Before 360-Indoor, evaluations were made with synthetic data, which did not reflect the complex scenes of the real world.

PANDORA [Xu *et al.*, 2022] is a recently released real-world dataset with oriented spherical box annotations. It consists of 3,000 spherical images and 94,353 oriented boxes annotations (RBFov of 5 parameters) among 47 categories.

By default, the proportion of the train/valid/test set is 1/2, 1/6, and 1/3 in the two datasets, and the images are all resized to 512×1024 from 960×1920.

4.2 Evaluation Metrics

Methods are evaluated using the standard COCO style Average Precision (AP) [Lin *et al.*, 2014], which is also the convention throughout the field of object detection. Following current mainstream works in spherical detection [Chou *et al.*, 2020; Xu *et al.*, 2022], we adopt AP_{50} as main metric, while AP & AP_{75} are just auxiliary metric. Naturally, these APs here are calculated based on spherical IoU to adapt to the spherical boxes. Even though our proposed Sph2Pob-IoU has high consistency with Unbiased-IoU [Dai *et al.*, 2022a], we still use Unbiased-IoU to compute the IoU for evaluation to make a fair comparison.

4.3 Implementation Details

All approaches are implemented in PyTorch [Paszke *et al.*, 2017] with `mmcv/mmdet/mmrotate` [Chen *et al.*, 2019; Zhou *et al.*, 2022] repository, and we make adaptation to the models for spherical boxes with 4/5-parameters. We use 4 NVIDIA 2080Ti GPUs with a batch size of 16 and a input resolution of 512×1024. SGD is adopted to optimize the models with momentum set to 0.9 and weight decay set to 0.0005. All evaluated models are trained for 120 epochs with an initial learning rate of 0.001 which is then divided by 10 at 80 epochs and again at 110 epochs. All other settings follow the default settings recommended in `mmdet`.

4.4 Comparisons with State-of-the-art

We make a comprehensive comparison between our proposed Sph2Pob algorithm and other State-of-the-art methods from three aspects, i.e. IoU consistency, IoU time-cost, and performance of detectors trained with IoU-calculators.

Method	Consistency			Time-cost		Detection		
	$R_{all}\uparrow$	$R_{low}\uparrow$	$R_{high}\uparrow$	$T_{cpu}\downarrow$	$T_{cuda}\downarrow$	AP↑	$AP_{50}\uparrow$	$AP_{75}\uparrow$
Sph	0.7819	0.9922	0.4274	0.0364	0.0033	10.7	24.3	7.8
Fov	0.9600	0.9974	0.8860	0.0372	0.0034	10.9	25.0	7.9
Sph2Pob	0.9989	0.9990	0.9988	2.2275	0.0096	11.5	25.7	8.2
Unbiased	1.0000	1.0000	1.0000	46.4417	-	-	-	-

Table 1: Comprehensive comparison between our proposed Sph2Pob algorithm and other State-of-the-art methods.

Angle	Error↓(mean±std)	R↑
original	0.0025±0.0086	0.9946
equator	0.0016±0.0042	0.9989
project	0.0017±0.0043	0.9987

Table 2: Approximate error and correlation coefficient of IoU with different angle calculation.

Edge	Error↓(mean±std)	R↑
arc	0.0016±0.0042	0.9989
chord	0.0023±0.0063	0.9974
tangent	0.0086±0.0192	0.9681

Table 3: Approximate error and correlation coefficient of IoU with different edge projection.

Consistency between Unbiased-IoU and different approximate-IoU. To compare the consistency between unbiased-IoU and different approximate methods, we design a experiment, as is shown in Figure 1 and Table 1. We randomly generate $10,000$ box-pairs in three cases, and calculate Unbiased-IoU [Dai *et al.*, 2022a] and three approximate-IoUs. In addition, we calculate *Pearson Product-Moment Correlation Coefficient* R to describe the consistency quantitatively. We can find that our methods are significant superior to previous method with higher consistency, and our Sph2Pob-IoU is very stable even in the most difficult *high-latitude* case, in which other methods show varying degradation. In detail, FoV-IoU is more accurate than Sph-IoU due to better retention of relative position information, but it is inferior than our Sph2Pob-IoU due to the lack of precise relative position and pose information, which is vital especially in *high-latitude* case.

Time cost of different IoU-calculators. Apart from consistency, computational speed should be taken into consideration for the practical application of IoU in training. We performed time-cost tests for the various IoU calculations in two devices, i.e. CPU and CUDA, as is shown in Table 1. Note that Unbiased-IoU does not support CUDA acceleration due to its inherent defect. Here, we randomly generate $1,000,000$ box-pairs in *all-latitude* case as we did in the previous *consistency* experiment, and record the total computational time(unit: second) of IoU for these box-pairs. According to Table 1, Sph-IoU and Fov-IoU is slightly fast but with a huge sacrifice of accuracy; while Unbiased-IoU is accurate, but consumes too much time. Limited by the current mainstream implementation of IoU for planar oriented boxes, our Sph2Pob-IoU computes slowly on the CPU, but achieves speed comparable to Sph-IoU and Fov-IoU on CUDA. Obviously, our Sph2Pob-IoU keeps the best balance between speed and accuracy, which gives it a practical advantage.

Performance of detectors with different IoU-calculators. Detection performance is the focus of attention in practical applications, so we also compare the performance of detectors trained with different IoU-calculators in various components, as is shown in Table 1. Note that both Sph-IoU and Fov-IoU are designed for BFoV rather than RBFoV, so we just compare these methods on 360-Indoor dataset. As for Unbiased-IoU, nondifferentiable operators can block IoU-based Loss, and slow computation makes Label-Assignment unacceptably time consuming, so we have to provide a empty result. It shows that our method can more effectively improve the detection performance than others.

Label Assignment	Loss	NMS	360-Indoor			PANDORA		
			AP↑	$AP_{50}\uparrow$	$AP_{75}\uparrow$	AP↑	$AP_{50}\uparrow$	$AP_{75}\uparrow$
			9.8	22.2	7.0	10.4	23.8	6.9
✓			10.2	23.0	7.8	10.3	24.3	6.6
	✓		11.0	25.4	7.8	10.6	24.5	6.9
		✓	9.8	22.1	6.8	10.4	23.9	6.9
✓	✓		11.5	25.7	8.2	10.5	25.3	7.0
✓	✓	✓	11.6	26.1	8.4	10.6	25.7	7.1

Table 4: Ablation study of different components with our proposed Sph2Pob-IoU. Unchecked components use planar IoU in ERP.

4.5 Ablation Studies for Sph2Pob

Different internal angle calculation. We conduct an experiment to explore the effect of the *internal angle* calculation method on the Sph2Pob-IoU, as is shown in Table 2. We calculate the IoU approximation error and correlation coefficient with $1,000,000$ random *all-latitude* box-pairs based on the three different angle calculation methods, i.e. *original*, *equator* and *project*. The *original* refers to calculate *internal angle* at the original boxes' positions (Figure 4(a)), whether the boxes are at the same latitudes or not. The *equator* refers to calculate at the equator after moving boxes (Figure 4(b)), and the *project* refers to append extra projection for *pose vector* onto the yOz plane before the angle calculation at the equator. The results show that *original* is clearly worse than other two ways, but *equator* is almost on a par with *project*. It implies it necessary to move to the equator even if only for calculation of the *internal angle*. In fact, both *equator* and *project* satisfy the hidden requirement that internal angle should be zero when the boxes are at the same longitude line, but only *original* dose not.

Different edge projection. We perform another experiment to explore the effect of the edge projection method on the Sph2Pob-IoU, as is shown in Table 3. We adopt the same metrics as internal angle calculation experiment based on the three edge projection methods, i.e. *arc*, *chord* and *tangent*, as is shown in Figure 4(c). The results show that simple *arc* is sightly superior to *chord*, and surpasses *tangent* with a large margin. This may be because the tangent is longer a lot than the corresponding arc/chord when field-of-view α, β is large.

Baselines with/without Sph2Pob. We use one-stage anchor-based detector Retinanet [Lin *et al.*, 2017] with adaption for spherical detection as base detector in following experiments. Here, we provide two baselines trained with naive L1 Loss for spherical boxes and Sph2Pob-L1 Loss for transformed planar boxes. As is shown in Table 5, the latter is just slightly worse than the former, which verifies the transform does not hurt boxes information a lot, and stronger Loss could compensate this decline.

Different components with Sph2Pob-IoU. Benefiting from low computational cost and high consistency, our

Loss	360-Indoor			PANDORA		
	AP↑	AP_{50} ↑	AP_{75} ↑	AP↑	AP_{50} ↑	AP_{75} ↑
L1	10.2	23.0	7.8	10.3	24.3	6.6
L1†	9.9	21.9	7.7	10.1	23.7	6.8
GWD† [Yang *et al.*, 2021b]	6.8	14.5	5.6	5.9	12.3	5.0
KLD† [Yang *et al.*, 2021c]	9.5	21.5	6.8	10.3	23.5	7.1
KFIoU† [Yang *et al.*, 2022b]	8.5	19.7	6.2	9.6	23.2	5.6
IoU† [Yu *et al.*, 2016]	9.8	22.1	6.8	10.4	24.8	6.9
GIoU† [Rezatofighi *et al.*, 2019]	10.5	23.9	7.8	10.3	24.7	6.8
DIoU† [Zheng *et al.*, 2020]	11.0	24.6	8.2	10.4	24.8	7.0
CIoU† [Zheng *et al.*, 2021]	11.5	25.7	8.2	10.5	25.3	7.0

Table 5: Performance of detectors with different Losses. The Losses with mark† utilize our proposed Sph2pob-Transform, which come from planar oriented detection.

Detector	Loss	360-Indoor			PANDORA		
		AP↑	AP_{50} ↑	AP_{75} ↑	AP↑	AP_{50} ↑	AP_{75} ↑
Faster R-CNN	L1	12.5	28.1	9.1	11.0	27.8	6.2
	CIoU†	12.9	29.1	9.4	11.3	28.6	7.1
SSD	L1	10.8	27.6	6.3	9.5	25.8	4.6
	CIoU†	12.0	28.7	8.0	10.5	26.9	6.0
FCOS	L1	8.8	20.2	6.7	7.7	19.7	4.4
	CIoU†	9.2	21.0	7.0	8.8	21.2	5.6

Table 6: Performance of different detectors with our Sph2Pob-Loss. The methods with mark† utilize our proposed Sph2pob-Transform

Sph2Pob-IoU can be applied to multiple components of spherical detection, i.e. Label Assignment, Loss and NMS, not limited to evaluation as previous Unbiased-IoU [Dai *et al.*, 2022a]. Before this, methods usually use IoU of planar enclosed rectangular boxes in ERP-images to serve these components. Table 4 shows that our Sph2Pob-IoU can comprehensively improve spherical detection from two aspects (Label Assignment and Loss), but fails to significantly improve from NMS. This may be because NMS just filter out redundancy boxes based on IoU, and a coarse IoU is enough to identify them.

Different Losses with Sph2Pob transform. Limited by complex spherical nature, only L1 Loss is available in practice of spherical detection. Luckly, flexible Sph2Pob transform allows us to easily use Losses designed for planar oriented detection. To compare with more SOTA Losses, we introduce some non-IoU-based Losses (GWD, KLD, KFIoU) for training detectors, as is shown in Table 5. They are representative works for planar oriented detection recently, but proposed for remote sensing images rather than natural scene images. Maybe this gap causes poor performance in our setting. By contrast, IoU-based Losses (especially CIoU-Loss) outperform other Losses in different degree as our expectation, which verifies that our method is simple but effective.

Different detectors with Sph2Pob-Loss. To verify the generality of our methods, we compare the performance of different detectors with our Sph2Pob-Loss. We choose three typical detectors (with adaption for spherical detection), including two-stage anchor-based detector Faster-RCNN [Ren *et al.*, 2015], one-stage anchor-based detector SSD [Liu *et al.*, 2016] and one-stage anchor-free detector FCOS [Tian *et al.*, 2019]. As is shown in Table 6, simply replacing the Loss with Sph2Pob-Loss can improve all detectors, and it shows that our approach is simple but very universal.

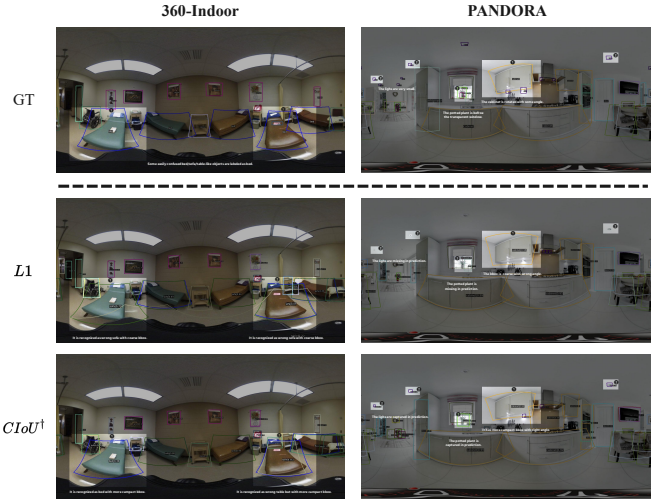

Figure 6: Qualitative comparison on the 360-Indoor and PANDORA dataset. The detector is the RetinaNet with different Losses. The methods with mark† utilizes our proposed Sph2pob-Transform. It is recommended to zoom in 5×(500%) at least for visual details of qualitative samples.

Qualitative Results. We provide some visualized results for qualitative comparison between detectors trained with naive L1 Loss and Sph2Pob-CIoU Loss to verify the effectiveness of our approach. As is shown in Figure 6, the detector trained with Sph2Pob-Loss usually predicts more compact boxes enclose objects. It has more possibility to capture small objects even in high-latitude regions, which is hard to realize when training with L1 Loss. In addition, our method can predict more precise *external angle* for oriented objects, benifit from the combination with planar oriented methods.

5 Conclusion

By analyzing the problems of previous approximate spherical IoU methods, this paper proposes a sphere-plane boxes transform, called Sph2Pob, making spherical boxes can be treated as planar oriented boxes. Based on the Sph2Pob, our proposed differentiable IoU for spherical boxes, Sph2Pob-IoU, has low time-cost and high accuracy. Besides, an agent Loss for spherical detection, Sph2Pob-Loss, is easily constructed based on the Sph2Pob, which makes the choices of Loss for spherical detection get richer. Extensive experiments also verify the effectiveness and universality of our approaches, and we hope that spherical detection will advance along with planar oriented detection from now on.

Acknowledgments

This work is supported by National Key R&D Program of China (2022YFD2001601), National Natural Science Foundation of China (62072438, U1936110), Strategic Priority Research Program of Chinese Academy of Sciences (XDA28040000, XDA28120000).

References

[Cao *et al.*, 2022] Miao Cao, Satoshi Ikehata, and Kiyoharu Aizawa. Field-of-view iou for object detection in 360° images. *arXiv e-prints*, pages arXiv–2202, 2022.

[Chen *et al.*, 2019] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.

[Chou *et al.*, 2020] Shih-Han Chou, Cheng Sun, Wen-Yen Chang, Wan-Ting Hsu, Min Sun, and Jianlong Fu. 360-indoor: Towards learning real-world objects in 360deg indoor equirectangular images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 845–853, 2020.

[Dai *et al.*, 2021] Pengwen Dai, Sanyi Zhang, Hua Zhang, and Xiaochun Cao. Progressive contour regression for arbitrary-shape scene text detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7393–7402, 2021.

[Dai *et al.*, 2022a] Feng Dai, Bin Chen, Hang Xu, Yike Ma, Xiaodong Li, Bailan Feng, Peng Yuan, Chenggang Yan, and Qiang Zhao. Unbiased iou for spherical image object detection. In *Proceedings of AAAI Conference on Artificial Intelligence*, pages 508–515, 2022.

[Dai *et al.*, 2022b] Longgang Dai, Hongming Chen, Yufeng Li, Caihua Kong, Zhentao Fan, Jiyang Lu, and Xiang Chen. Tardet: Two-stage anchor-free rotating object detector in aerial images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4267–4275, 2022.

[He *et al.*, 2021] Yuhang He, Wentao Yu, Jie Han, Xing Wei, Xiaopeng Hong, and Yihong Gong. Know your surroundings: Panoramic multi-object tracking by multimodality collaboration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2969–2980, 2021.

[Kang and Cho, 2019] Kyoungkook Kang and Sunghyun Cho. Interactive and automatic navigation for 360 video playback. *ACM Transactions on Graphics (TOG)*, 38(4):1–11, 2019.

[Lin *et al.*, 2014] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[Lin *et al.*, 2017] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[Liu *et al.*, 2016] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[Ma *et al.*, 2018] Jianqi Ma, Weiyuan Shao, Hao Ye, Li Wang, Hong Wang, Yingbin Zheng, and Xiangyang Xue. Arbitrary-oriented scene text detection via rotation proposals. *IEEE Transactions on Multimedia*, 20(11):3111–3122, 2018.

[Paszke *et al.*, 2017] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[Ren *et al.*, 2015] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[Rezatofighi *et al.*, 2019] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 658–666, 2019.

[Snyder, 1997] John P Snyder. *Flattening the earth: two thousand years of map projections*. University of Chicago Press, 1997.

[Tian *et al.*, 2019] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9627–9636, 2019.

[Wang and Lai, 2019] Kuan-Hsun Wang and Shang-Hong Lai. Object detection in curved space for 360-degree camera. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3642–3646. IEEE, 2019.

[Xu *et al.*, 2022] Hang Xu, Qiang Zhao, Yike Ma, Xiaodong Li, Peng Yuan, Bailan Feng, Chenggang Yan, and Feng Dai. Pandora: A panoramic detection dataset for object with orientation. In *ECCV*, 2022.

[Xu *et al.*, 2023] Hang Xu, Xinyuan Liu, Qiang Zhao, Yike Ma, Chenggang Yan, and Feng Dai. Gaussian label distribution learning for spherical image object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[Yang and Yan, 2020] Xue Yang and Junchi Yan. Arbitrary-oriented object detection with circular smooth label. In *European Conference on Computer Vision*, pages 677–694. Springer, 2020.

[Yang *et al.*, 2018] Xue Yang, Hao Sun, Kun Fu, Jirui Yang, Xian Sun, Menglong Yan, and Zhi Guo. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. *Remote Sensing*, 10(1):132, 2018.

[Yang *et al.*, 2021a] Kailun Yang, Jiaming Zhang, Simon Reiß, Xinxin Hu, and Rainer Stiefelhagen. Capturing omni-range context for omnidirectional segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1376–1386, 2021.

[Yang *et al.*, 2021b] Xue Yang, Junchi Yan, Qi Ming, Wentao Wang, Xiaopeng Zhang, and Qi Tian. Rethinking rotated object detection with gaussian wasserstein distance loss. In *International Conference on Machine Learning*, pages 11830–11841. PMLR, 2021.

[Yang *et al.*, 2021c] 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. *Advances in Neural Information Processing Systems*, 34:18381–18394, 2021.

[Yang *et al.*, 2022a] Xue Yang, Gefan Zhang, Xiaojiang Yang, Yue Zhou, Wentao Wang, Jin Tang, Tao He, and Junchi Yan. Detecting rotated objects as gaussian distributions and its 3-d generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[Yang *et al.*, 2022b] Xue Yang, Yue Zhou, Gefan Zhang, Jitui Yang, Wentao Wang, Junchi Yan, Xiaopeng Zhang, and Qi Tian. The kfiou loss for rotated object detection. *arXiv preprint arXiv:2201.12558*, 2022.

[Yu *et al.*, 2016] Jiahui Yu, Yuning Jiang, Zhangyang Wang, Zhimin Cao, and Thomas Huang. Unitbox: An advanced object detection network. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 516–520, 2016.

[Zhang *et al.*, 2020] Shifeng Zhang, Cheng Chi, Zhen Lei, and Stan Z Li. Refineface: Refinement neural network for high performance face detection. *IEEE transactions on pattern analysis and machine intelligence*, 43(11):4008–4020, 2020.

[Zhang *et al.*, 2021] Xue Zhang, Gene Cheung, Yao Zhao, Patrick Le Callet, Chunyu Lin, and Jack ZG Tan. Graph learning based head movement prediction for interactive 360 video streaming. *IEEE Transactions on Image Processing*, 30:4622–4636, 2021.

[Zhao *et al.*, 2020] Pengyu Zhao, Ansheng You, Yuanxing Zhang, Jiaying Liu, Kaigui Bian, and Yunhai Tong. Spherical criteria for fast and accurate 360 object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12959–12966, 2020.

[Zheng *et al.*, 2020] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, and Dongwei Ren. Distance-iou loss: Faster and better learning for bounding box regression. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 12993–13000, 2020.

[Zheng *et al.*, 2021] Zhaohui Zheng, Ping Wang, Dongwei Ren, Wei Liu, Rongguang Ye, Qinghua Hu, and Wangmeng Zuo. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Transactions on Cybernetics*, 2021.

[Zhou *et al.*, 2019] Dingfu Zhou, Jin Fang, Xibin Song, Chenye Guan, Junbo Yin, Yuchao Dai, and Ruigang Yang. Iou loss for 2d/3d object detection. In *2019 International Conference on 3D Vision (3DV)*, pages 85–94. IEEE, 2019.

[Zhou *et al.*, 2022] Yue Zhou, Xue Yang, Gefan Zhang, Jiabao Wang, Yanyi Liu, Liping Hou, Xue Jiang, Xingzhao Liu, Junchi Yan, Chengqi Lyu, Wenwei Zhang, and Kai Chen. Mmrotate: A rotated object detection benchmark using pytorch. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2022.