CTU Journal of Innovation and Sustainable Development
ISSN 2588-1418 | e-ISSN 2815-6412

DOI:10.22144/ctujoisd.2025.054

Towards robust visual recognition for smart cities and remote sensing: A survey of regression losses in rotated object detection

Chien Thai[1,2], Mai Xuan Trang[1*], and Son Le Anh[2,3]

[1]*Phenikaa School of Computing, Phenikaa University, Viet Nam*

[2]*Phenikaa-X JSC, Phenikaa Group, Viet Nam*

[3]*Faculty of Vehicle and Energy Engineering, Phenikaa University, Viet Nam*

**Corresponding author (trang.maixuan@phenikaa-uni.edu.vn)*

Article info.

Received 30 Jun 2025
Revised 18 Aug 2025
Accepted 7 Oct 2025

Keywords

Autonomous driving, regression loss functions, rotated object detection, smart city applications

ABSTRACT

Rotated object detection (ROD), often termed oriented object detection, is essential for numerous practical tasks, including remote sensing, self-driving systems, urban surveillance, and text recognition in natural scenes. Unlike conventional object detection, ROD must estimate object orientation, making angle regression and loss function design crucial to model performance. This paper presents a comprehensive survey of regression loss functions used in ROD, categorized into coordinate-based, approximated rotated IoU-based, and Gaussian-based approaches. We analyze their theoretical foundations, practical trade-offs, and effectiveness in addressing core challenges including angle periodicity, edge ambiguity, and metric inconsistency. Representative loss functions are benchmarked on standard datasets to evaluate their suitability for various detection frameworks. By emphasizing application contexts such as smart city monitoring and environmental analysis, this survey offers practical guidance for designing robust and efficient ROD systems that support sustainable development goals.

1. INTRODUCTION

Rotated object detection (ROD) has emerged as a critical component of modern computer vision, with applications spanning remote sensing, smart city surveillance, autonomous vehicle navigation, retail environment analytics. For example, in smart cities, surveillance drones and traffic cameras frequently capture tilted or obliquely oriented vehicles and infrastructure elements. Accurate detection of these objects using ROD improves traffic analysis, accident detection, and infrastructure monitoring. Similarly, in remote sensing, precise localization of ships, buildings, or land use patterns is crucial for disaster response, maritime navigation, and environmental management. Despite recent progress in ROD, several unresolved challenges remain.

As opposed to classic object detection, which uses horizontal bounding boxes (HBB) defined by parameters (x, y, w, h) to represent the location and size of an object, rotated object detection employs oriented bounding boxes (OBB) that add an angle parameter θ (as shown in Figure 1). This representation provides a more precise localization by reducing background overlap, thereby improving detection accuracy.

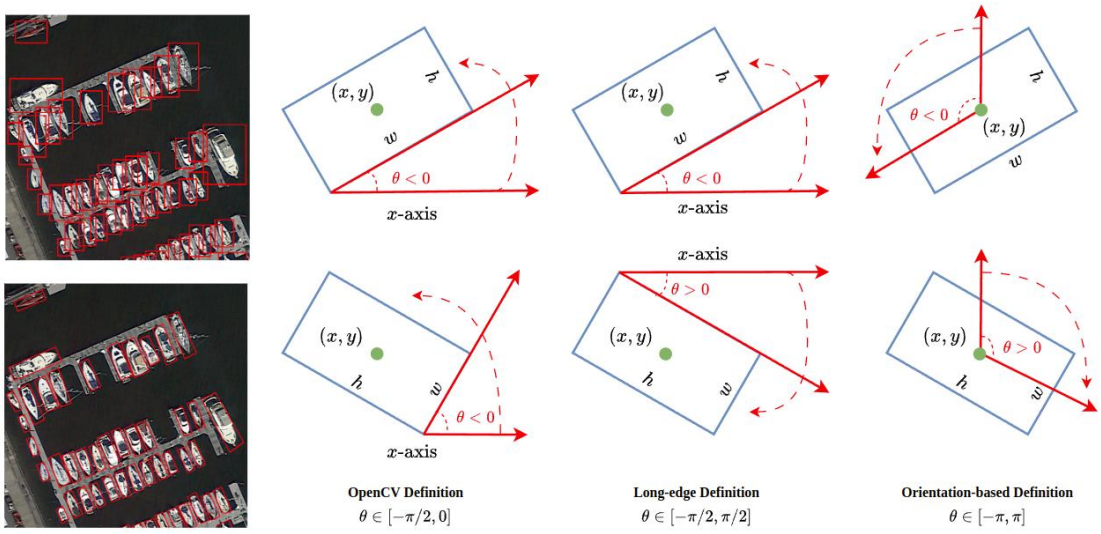

Figure 1. Illustration HBB and OBB (left), along with three types of regression-based representations of rotated bounding boxes (right)

OBBs are typically represented in two ways: (i) a point-based scheme that specifies the quadrilateral through the coordinates off its four vertices $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$, or (ii) a regression-based representation, which defines an oriented box by (x, y, w, h, θ). The regression-based method is widely adopted and supports three distinct definitions:

– OpenCV definition: The angle θ denotes the acute orientation formed by the box's width relative to the horizontal axis, constrained to $\theta \in [\pi/2\,, 0]$ radians.

– Long-edge definition: θ denotes the angle between the longer edge of the bounding box and the horizontal axis, with $\theta \in [-\pi/2\,, \pi/2]$ radians.

– Orientation-based definition: θ denotes the clockwise rotation from a reference axis to the object's orientation axis, covering entire circle with value in range $[-\pi, \pi]$ radians, thus providing a complete representation of object orientation.

The use of multiple definitions for rotated bounding boxes introduces inconsistencies in loss function design, leading to challenges such as boundary discontinuity, edge exchangeability, and square-like ambiguities. Different loss functions address these issues with varying degrees of effectiveness, directly impacting detection performance. This article delivers a systematic overview of regression loss functions proposed for ROD. We categorize and analyze existing approaches, examining their theoretical foundations, strengths, and limitations. Key challenges in designing effective losses are discussed, along with promising directions for future research. By offering a structured overview, this survey aims to assist researchers and practitioners in selecting suitable loss functions for real-world applications and inspire further advancements in the field. The key contributions of this study can be outlined as follows:

– Identification of key challenges in designing regression losses for rotated object detection, including boundary discontinuity, edge ambiguity, square-like cases, and loss–metric inconsistency.

– Systematic categorization and analysis of existing loss functions based on mathematical properties, with emphasis on their strengths and limitations.

– Comprehensive evaluation of representative losses on benchmark datasets, offering practical guidance for real-world deployment.

The structure of this paper is as follows: Section 2 discusses the critical issues in designing regression loss for ROD. Section 3 reviews and analyzes existing regression loss functions in detail. Section 4 reports experimental results on standard datasets and detection frameworks. Finally, Section 5 concludes the paper with key findings and outlines potential avenues for future research.

For clarity, the acronyms frequently used in this paper are summarized in Table 1.

Table 1. List of acronyms used in this paper

Acronym	Definition
ROD	Rotated Object Detection
OBB	Oriented Bounding Box
HBB	Horizontal Bounding Box
PoA	Periodicity of Angle
EoE	Exchangeability of Edges
IoU	Intersection over Union
CSL	Circular Smooth Label
DCL	Densely Coded Labels
PSC	Phase-Shifting Coder
GWD	Generalized Wasserstein Distance
KLD	Kullback-Leibler Divergence
KFIoU	Kalman Filter-based IoU
BD	Bhattacharyya Distance
DOTA	Dataset for Object deTection in Aerial images
HRSC	High resolution ship collections

2. CHALLENGES IN DESIGNING ROTATED OBJECT DETECTION LOSSES

Designing effective regression loss functions for ROD presents several unique challenges, primarily due to the complex geometry of oriented bounding boxes (OBBs). These challenges significantly affect learning stability and model accuracy in real-world applications such as aerial surveillance and traffic analysis.

2.1. Boundary discontinuity

Boundary discontinuity remains a major difficulty in ROD systems, which makes the loss value suddenly increase at the boundary caused by the parameterization of angles and edges. This issue is closely related to the of angle periodicity (PoA) and edge exchangeability (EoE).

PoA indicates the cyclical nature of angle representation, where angles like θ and $\theta + 2\pi$ denote the same orientation. In many ROD frameworks, this periodicity causes sharp discontinuities in loss values near angle wrap-around points (e.g., transitioning from 179 degrees to -180 degrees). As a result, small angular differences can lead to a significant change in loss value, causing instability during the model's training.

EoE refers to the ambiguity that arises when swapping the width and height of a rotated bounding box results in an equivalent box with a different angle. For instance, (i.e. (x, y, w, h, θ) and $(x, y, h, w, \theta + \pi/2)$) represent the same region but have different parameterizations. This can confuse the model during training, forcing it to learn unnecessarily complex transformations, such as a large rotation (clockwise) instead of a minor correction (counterclockwise), thereby reducing regression stability and accuracy.

2.2. Square-like problem

In rotated object detection, the square-like problem arises when dealing with objects that are nearly square. In such cases, different angle values can represent visually similar bounding boxes. For example, angles of 0 degrees and -89 degrees may correspond to nearly identical squares. However, standard regression losses often produce large loss values for such differences, despite minimal geometric discrepancy. This problem may hinder the model in predicting objects with aspect ratios close to 1.

2.3. Inconsistency between loss and metric

In conventional horizontal object detection, Intersection over Union (IoU) metric is commonly applied to evaluate the overlap between predicted and ground-truth bounding boxes. However, a well-known issue is that traditional regression losses (e.g. l_n-norms) do not align well with the IoU metric. When extended to rotated object detection, the direct use of IoU loss poses extra difficulties, such as increased computational complexity and non-differentiable regions in the parameter space, which hinder optimization.

The intersection of two rotated boxes forms an irregular polygon rather than a rectangle, unlike axis-aligned boxes. Computing this intersection involves multiple steps: 1) determining intersection points; 2) sorting these points in an anti-clockwise order based on their coordinates; 3) computing the polygon area using computational geometry methods (e.g. the Shoelace formula). These operations are computationally expensive and include non-differentiable components, such as conditional logic and sorting.

Moreover, the intersection operation involves conditional logic that is not differentiable with respect to the bounding box parameters, such as sort operators. Although several works attempt to use several custom operations, the computation of rotated IoU still fails to be differentiable in certain corner cases, for instance, when two boxes produce more than eight intersection points due to partial edge overlap. Therefore, it is crucial to design a simple yet fully differentiable approximation of IoU loss for rotated object detection.

3. REGRESSION LOSSES IN ROTATED OBJECT DETECTION

To overcome the aforementioned challenges, a range of regression loss functions have been proposed, each designed to minimize discrepancies between predicted and ground-truth bounding boxes while handling rotation-specific issues. These loss functions can be broadly categorized into four main types: l_n-norm, angular encoding, approximated IoU, and Gaussian distribution-based losses. Each category offers distinct advantages depending on the detection scenario and task requirements. This section presents a detailed survey of these loess functions, highlighting their mathematical properties, strengths, and limitations. An overview of the taxonomy is illustrated in Figure 2.

3.1. l_n-norm loss function

Regression loss functions are a core component of modern object detection frameworks. For horizontal bounding boxes, the model typically predicts four parameters for location and size (x_p, y_p, w_p, h_p) to match the ground truth (x_t, y_t, w_t, h_t), and the l_n-norm loss is calculated as:

$$\mathcal{L}_{reg} = l_n\text{-norm}(\Delta x, \Delta y, \Delta w, \Delta h)$$

where $\Delta x = x_p - x_t$, $\Delta y = y_p - y_t$, $\Delta w = ln(w_p/w_t)$, and $\Delta h = ln(h_p/h_t)$ (in anchor based detectors, Δx and Δy are normalized by size of anchor). The common l_n-norm-based loss used in object detection are L1, L2, and Smooth L1 loss. While L1 loss is the lack of sensitivity to small localization errors, L2 loss is highly sensitive to outliers. Smooth L1 (Girshick , 2015) loss combines the benefits of L1 and L2. Mathematically, the Smooth L1 loss is defined as:

$$\mathcal{L}_{\text{Smooth L1}}(x, y) = \begin{cases} 0.5(x-y)^2/\beta, & \text{if} |x-y| < \beta \\ |x-y| - 0.5 * \beta, & \text{otherwise} \end{cases}$$

In ROD, these losses are extended to include an angle parameter θ. Pan et al. (2020) directly applies the L1 loss for the regression of rotation angles $|\theta_p - \theta_t|$. To deal with the Periodicity of Angle (PoA) problem, other works transform the angle difference by trigonometric or modulo functions $\Delta\theta = f(\theta_p - \theta_t)$ or $f(\theta_p) - f(\theta_t)$. Ding et al. (2018) uses modular module $(\theta_p - \theta_t)/2\,\pi$ to adjust the angle offset target to fall in $[0,2\pi]$for the convenience of computation Han et al. (2021) simply added $k\pi$ where k is an integer, to ensure the angular difference remains within the range $[-\pi/4\,, 3\,\pi/4]$. Modulated loss (Qian et al., 2021) introduced the modulated loss, which ensures a continuous formulation by removing angle periodicity and address the ambiguity between height and width:

$$l_{mr} = \min\big(l_{n\text{-norm}}(\Delta x, \Delta y, \Delta w, \Delta h, \Delta\theta), \; l_{n\text{-norm}}(\Delta x, \Delta y, \Delta w', \Delta h', 90 - \Delta\theta)\big)$$

where $\Delta w' = w_p - h_t$ and $\Delta h' = h_p - w_t$. Other works transform the angle difference by trigonometric functions. For example, Ming et al. (2021) employs tangent transforms instead of working directly with angles to prevent abrupt changes near the angle warp-around boundary, making learning more stable for the angle difference near 0 degrees or 180 degrees. However, $tan(\theta_p - \theta_t)$ approaches infinity near $\pi/2 + k\pi$, making it unstable in some corner cases. Lang et al. (2019) computes the sine of the angle offset to ensure the loss function is smooth, differentiable, and bounded. This method is particularly useful in tasks where direction of the object is not essential (e.g. remote sensing, scene text detection, etc.) because it cannot distinguish -π and π radians. For applications such as autonomous driving, where distinguishing between forward and backward directions is crucial for accurate navigation, the sine-cosine loss computed as $l_n - \text{norm}\left(sin(\theta_p) - sin(\theta_t), cos(\theta_p) - cos(\theta_t)\right)$ is more suitable (Yin et al., 2021).

To address the Exchangeability of Edges (EoE), recent detectors adopt a long-edge definition, where h is determined as the longer side of bounding boxes, and θ is defined by long side h and the horizontal x-axis.

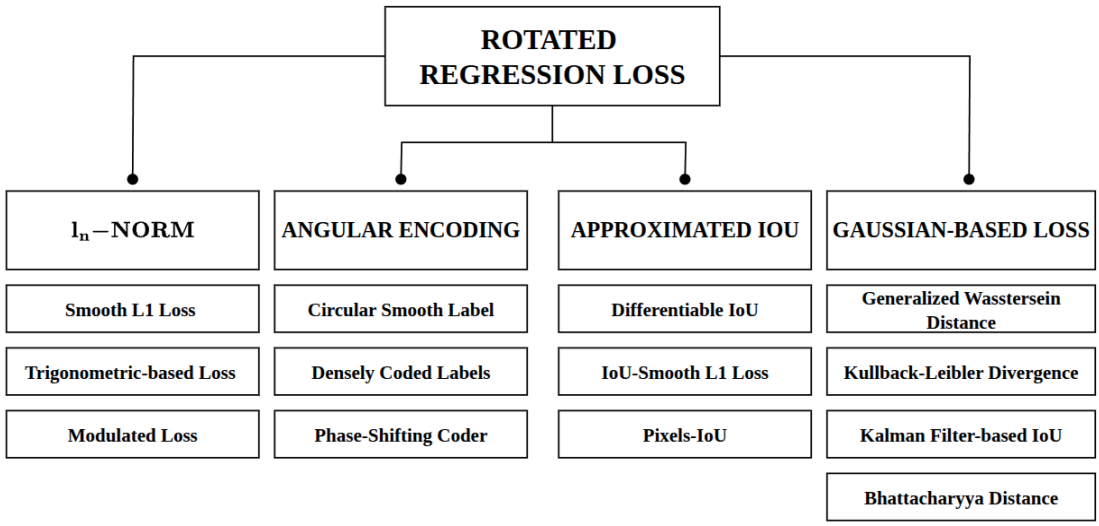

Figure 2. Taxonomy of the regression loss functions for ROD problem

3.2. Angular encoding

Angular encoding-based methods reformulate angle regression as a classification task by discretizing the angular range into a fixed number of intervals. Each interval corresponds to a class label, enabling angle prediction through classification. For example, with a 180 degrees angle range, and interval size ω degree, the number of classes is $T = 180/\omega$. However, the periodic nature of angles, directly applying standard classification loss functions may lead to poor performance, especially near boundary points. To address this problem, Yang et al. (2020) introduced a circular smooth label (CSL) method to achieve better angular representation. CSL handles the periodic nature of angles and increases error tolerance for adjacent angles by smoothing labels in a circular manner. This approach effectively mitigates boundary discontinuities and enhances detection accuracy for arbitrarily oriented objects. Given the ground truth angle θ, the circle smooth label for class x is expressed as:

$$CSL(x) = \begin{cases} g(x) & \theta - r < x < \theta + r \\ 0 & \text{otherwise} \end{cases}$$

where $g(x)$ is a window function, and r is radius. With θ is center, an ideal window function $g(x)$ is required to obey four appealing properties: periodicity $g(x) = g(x + kT), k \in N$, symmetry $(g(\theta + \epsilon) = g(\theta - \epsilon), \forall \epsilon \vee r$, maximum $(g(\theta) = 1)$, monotonically non-increasing from the center point toward both sides $(g(\theta \pm \epsilon) \leq g(\theta \pm \zeta), \forall \zeta \vee \forall \epsilon \vee r)$. Four types of window function are employed, including pulse, rectangular, triangle, and Gaussian functions. While CSL effectively handles angular discontinuity, it introduces computational overhead due to its expanded prediction layer and does not resolve the square-like problem. To tackle these issues, Yang et al. (2021) proposed a re-weighting scheme named Angle Distance and Aspect Ratio Sensitive Weighting (ADARSW) to adjust the loss function based on angular distance and aspect ratio of the object. The ADARSW is based on the sine of angle difference, which is formulated as $W_{\text{ADARSW}}(\Delta\theta) = sin\left(\alpha(\theta_g - \theta_p)\right)$ where α is set to 1 if the aspect ratio of the ground truth bounding box satisfies $h_g/w_g > r$, and $\alpha = 2$ otherwise. Here, h_g and w_g are the long and short sides of ground truth bounding boxes, respectively, and r denote the aspect ratio threshold. The final loss is the multiplication of ADARSW and classification loss. When the height and width of ground truth bounding boxes are nearly equal, the period becomes 90 degrees ($\alpha = 2$), effectively resolving the square-like problem. Moreover, Densely Coded Labels (DCL) are introduced for angle classification, which are a more compact representation compared to CSL. This encoding approach not only enhances training efficiency but also significantly improves detection accuracy. Yu et al. (2023) introduce a differentiable angle coding method named Phase-Shifting Coder (PSC), which represents angles as periodic phases to address the boundary discontinuity problem. They further introduced PSCD (Phase-Shifting Coder with Dual-frequency), which maps angular periodicity across multiple frequency domains. This

dual-frequency design effectively resolves both boundary discontinuity and square-like issues in a unified framework.

Despite their advantages, angular encoding methods have several limitations. Transforming angle regression into a classification problem increases model complexity and computational cost. Moreover, integrating these methods often requires non-trivial modifications to existing detection architectures, particularly in the loss design and training pipeline. Lastly, the inconsistency between classification-based losses and geometric evaluation metrics (e.g., IoU) remains unresolved.

3.3. Approximated IoU loss

In general object detection, IoU is a widely used metric to evaluate the performance of detection models. However, l_n-norm-based regression losses focus on minimizing differences in box parameters (e.g., center coordinates, width, height), rather than optimizing for spatial overlap. This misalignment can degrade performance in certain cases. For instance, two large boxes with minimal overlap may yield a small l_n-norm loss but a poor IoU score, while small boxes with high parameter similarity may still have low spatial overlap. To address this inconsistency, existing horizontal detectors introduced IoU-based loss, such as IoU (Yu et al. 2016), Generalized IoU (Rezatofighi et al., 2019), and Complete IoU (Zheng et al., 2020) losses. Unfortunately, these formulations are not directly applicable to oriented bounding boxes (OBBs) due to their geometric complexity.

Zhou et al. (2019) explored the computation of differentiable IoU for rotated boxes. Their approach involves: (1) determining polygon intersection points, (2) sorting them in counter-clockwise order, and (3) calculating the area using the Shoelace formula. However, this method is computationally expensive and includes non-differentiable components such as sorting. While unofficial implementations attempt to approximate differentiability using gather operators, the Rotated IoU (RIoU) remains non-differentiable in certain cases - e.g., when two OBBs have more than eight intersection points or partially coincident edges. Therefore, developing efficient and fully differentiable approximations of RIoU loss remains an open challenge.

Several works (Yang et al., 2019, 2021, 2022a); approximated Rotated IoU loss by combining l_n-norm loss and Rotated IoU:

$$\mathcal{L}(B_p, B_t) = \frac{\mathcal{L}_{reg}(B_p, B_g)}{\left|\mathcal{L}_{reg}(B_p, B_g)\right|} |f(RIoU)|$$

where $B_p(x_p, y_p, w_p, h_p, \theta_p)$ and $B_g(x_g, y_g, w_g, h_g, \theta_g)$ are predicted and ground truth bounding boxes, respectively. $f(\cdot)$ represents the loss function related to RIoU (e.g. $log(\cdot)$). $\mathcal{L}_{reg}$ is combination of Smooth L1 loss for angle and horizontal IoU loss for (x, y, w, h) parameters:

$$\begin{aligned}&\mathcal{L}_{reg}\\&= \mathcal{L}_{\text{SmoothL1}}(\theta_p, \theta_g)\\&- \text{IoU}\left(B_p(x_p, w_p, w_p, h_p), B_g(x_g, y_g, w_g, h_g)\right)\end{aligned}$$

The approximate rotated IoU can be subdivided into two parts: $\mathcal{L}_{reg}(B_p, B_g)/\left|\mathcal{L}_{reg}(B_p, B_g)\right|$ specifies the gradient propagation directions, whereas $|f(RIoU)|$ controls the magnitude of the gradient. Although the proposed loss is fully differentiable and can smooth the boundary loss jump, it still requires a complex RIoU calculation process. In addition, its gradient direction is still dominated by the angle regression; therefore , the PoA problem is not completely resolved.

Chen et al. (2020) introduced the Pixels Intersection over Union loss (PIoU) to improve OBB regression by approximating IoU in a pixel-wise manner. PIoU loss simply counts the number of overlap pixels by employing a differentiable kernel function that computes the accumulated contributions of interior overlapping pixels. However, PIoU loss involves more complex calculations than traditional loss functions (e.g. l_n-norm loss). Its performance is highly affected by the grid size; thus, it may not always handle occlusions or overlapping objects optimally.

3.4. Gaussian distribution-based loss

Recent studies propose a unified and advanced approach to overcome boundary discontinuity and the square-shaped ambiguities by emplying Gaussian distribution. In this framework, the conventional oriented bounding box representation $B(x, y, w, h, \theta)$ is reformulated as a bivariate Gaussian distribution $N(\mu, \Sigma)$ where the mean $\mu = (x, y)$ represents the object center, and the covariance matrix $\Sigma^{1/2} = RSR^T$, with R denoting the rotation matrix and S the diagonal eigenvalue matrix. The specific derivations of R and S are defined as follows:

$$R = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix},\ S = \begin{bmatrix} \frac{h}{2} & 0 \\ 0 & \frac{w}{2} \end{bmatrix}$$

An important advantage of Gaussian formulation is that the orientation is captured via trigonometric encoding, thereby eliminating issues caused by angular periodicity. Furthermore, the parameters of OBB are co-optimized, enabling interdependence throughout the training process. For comparing two multivariate Gaussian distributions, metrics such as Generalize Wastersein Distance (GWD) (Yang et al., 2021), Kullback-Leiber Divergence (KLD) (Yang et al., 2021), Kalman Filter-based IoU (KFIoU) (Yang et al., 2022b), and Bhattacharyya Distance (BD) (Thai et al., 2025) are commonly adopted. The bivariate Gaussian formulation of bounding boxes introduces following characteristics that effectively tackle certain challenges in computing regression losses for ROD:

– Property 1: Two bounding boxes defined as $B(x, y, w, h, \theta)$ and $B(x, y, h, w, \theta - \pi/2)$ correspond to the same Gaussian form, i.e., $\Sigma(w, h, \theta) = \Sigma(h, w, \theta - \pi/2)$, which eliminates the problem of edge exchangeability (EoE).

– Property 2: The bounding boxes defined as $B(x, y, w, h, \theta)$ and $B(x, y, w, h, \theta - k\pi)$, with $k \in Z$, produces equivalent Gaussian forms. This property resolves angle periodicity (PoA) problem.

– Property 3: if width and height of the bounding box are similar ($w \approx h$), $\Sigma(w, h, \theta) = \Sigma(w, h, \theta - k\,\pi/2)$ where k is an integer. This property prevents square-like problem.

Generalize Wasserstein Distance (GWD) (Yang et al., 2021) is a valuable metric for measuring the distance between two probability distributions, particularly useful in comparing Gaussian distributions due to its consideration of both the mean and covariance. For two multivariate Gaussian distributions $X_p \sim N(\mu_p, \Sigma_p)$ and $X_t \sim N(\mu_t, \Sigma_t)$, the GWD can be formally expressed as:

$$W_2^2(X_p, X_t) = |\mu_p - \mu_t|^2 + Tr\left(\Sigma_p + \Sigma_t - 2\left(\Sigma_p^{1/2}\Sigma_t\Sigma_p^{1/2}\right)^{1/2}\right)$$

Because GWD is sensitive to large error, Yang et al. (2021) applies a non-linear transformation $g(d) = 1 - 1/(\tau + f(d))$ to normalize the distance with range [0,1], where $\tau \geq 1$ is hyper-parameter, and $f(\cdot)$ is a non-linear function (e.g. sqrt, ln, etc.). GWD provides a differentiable approximation of RIoU loss, effectively mitigating the boundary discontinuity and square-like problems.

4. EXPERIMENTAL

4.1. Dataset

Experiments were performed on two popular datasets for oriented object detection: DOTA (Xia et al., 2018) and HRSC2016 (Liu et al., 2017). DOTA dataset contains 2,806 large-scale aerial images collected from multiple platforms and sensors, encompassing 15 object categories. The dataset is partitioned into 1,411 training images, 458 validation images, and the remainder images reserved for testing. Since the ground-truth labels for the test set are unavailable, results must be submitted to an official evaluation server. HRSC2016, on the other hand, is a prominent dataset for high-resolution remote sensing with a focus ship detection. It includes of over 1,000 high-resolution images covering a wide variety of vessel types under diverse and challenging conditions. Each image is densely annotated with bounding boxes, orientations, and precise ship locations.

4.2. Evaluation protocol

The experiment setup was implemented using the MMRotate framework (Zhou et al., 2022). All evaluations were carried out with RetinaNet (Lin et al., 2017) and R3Det (Yang et al., 2021) both utilizing ResNet50 (He et al., 2016) architecture as backbone network. For the DOTA-v1.0 dataset, input images were resized to 1024 × 1024 pixels, whereas HRSC2016 samples were adjusted to 800 × 800. The preprocessing pipeline included normalization along with extensive data augmentation, such as random cropping and flipping operations (horizontal, vertical and diagonal) applied with a probability of 0.25.

Model training was carried out for 20 epochs on DOTA-v1.0 dataset and 50 epochs on HRSC2016. AdamW (Ilya Loshchilov & Frank Hutter, 2019) was adopted as optimize, initializing with learning rate of 1e-4, gradually reduced to1e-8 using the cosine annealing strategy, to ensure stable convergence. The experiments were conducted with a batch size of 2.

In the evaluation process, we adopted dataset-specific metrics to assess detection performance. For the DOTA dataset, we used AP_{50} (Average Precision at 50% IoU threshold) as the primary evaluation metric. Meanwhile, for the HRSC2016

dataset, we employed $AP_{50}:AP_{95}$, which considers Average Precision across multiple IoU thresholds ranging from 50% to 90%. This metrics provide a robust assessment of the models' ability to localize rotated objects under various overlap conditions.

Due to computational constraints, we evaluated a selected set of representative loss functions from each category, with an emphasis on Gaussian-based losses, which have shown strong performance in addressing ROD-specific challenges. This focused approach allows us to draw meaningful comparisons while maintaining practical feasibility.

4.3. Experimental results

Table 2 provides a detailed evaluation of ROD losses on the DOTA-1.0 dataset. For RetinaNet, a variety of loss types are used, including the l_n-norm with Smooth L1 loss (Girshick, 2015), two Angular Encoding methods: CSL (Yang et al., 2020) and DCL (Yang et al., 2021), an Approximate IoU method with IoU-Smooth L1 loss (Yang et al., 2019), and several Gaussian-based losses such as GWD (Yang et al., 2021), KLD (Yang et al., 2021), KFIoU (Yang et al., 2022b), and BD Loss (Thai et al., 2025).

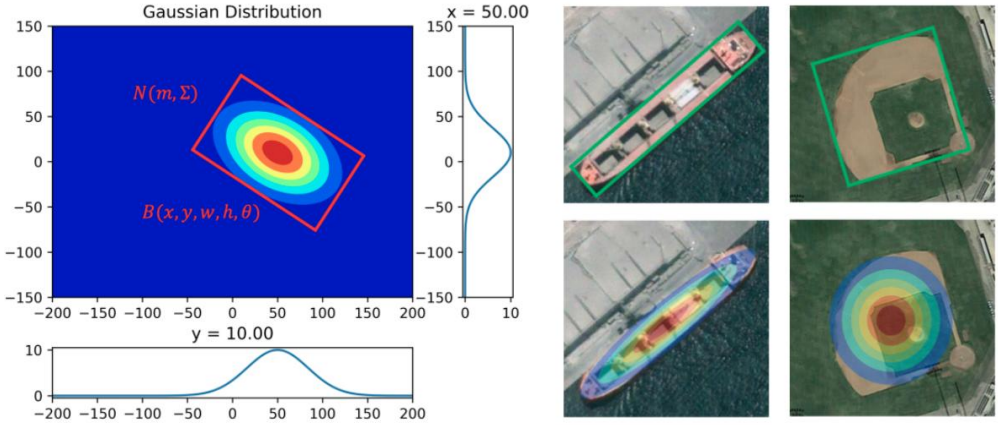

Figure 3. A conceptual illustration of representing a rotated bounding box using a bivariate Gaussian distribution (Yang et al., 2021)

Kullback-Leibler Divergence (KLD): (Yang et al., 2021) shows that GWD loss does not obey the scale-invariant property of IoU, and introduces a suitable regression loss for ROD named KLD loss. When applied to two-dimensional Gaussian distributions, the KLD provides a measure of how one distribution diverges from another, capturing both shifts in mean and differences in covariance structure (conceptulized in Figure 3). Given two Gaussian distributions, $X_p\ N(\mu_p, \Sigma_p)$ and $X_t\ N(\mu_t, \Sigma_t)$, the KLD from X_p to X_t is computed as:

$$\mathrm{KLD}(X_t||X_p) = \frac{1}{2}\left(\log \frac{|\Sigma_p|}{|\Sigma_t|} - d + \mathrm{Tr}(\Sigma_p^{-1}\Sigma_t) + (\mu_p - \mu_t)^T \Sigma_p^{-1}(\mu_p - \mu_t)\right)$$

Because each term of $D_{KL}(X_p \,||\, X_t)$ is partial parameter coupling, all parameters of bounding boxes are jointly optimized. This ensures that the loss function does not become overly sensitive to under-fitting in any of the parameters, thereby maintaining a balanced optimization process. Similar to GWD, the final regression loss is obtained by applying the non-linear transformation $g(\cdot)$, $L_{KLD}(X_p \,||\, X_t) = g\left(D_{KL}(X_p \,||\, X_t)\right)$.

Kalman Filter-based IoU (KFIoU) (Yang et al., 2022b): KFIoU estimates the overlap between two Gaussian distributions using Kalman filter to achieve better trend-level alignment with Rotated IoU. The overlap area is defined as $\alpha N(\mu, \Sigma) = N(\mu_1, \Sigma_1)N(\mu_2, \Sigma_2)$ where $\mu = \mu_1 + K(\mu_2 - \mu_1)$, $\Sigma = \Sigma_1 - K\Sigma_1$, and $K = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}$ is the Kalman gain. The overlap area $\alpha N(\mu, \Sigma)$ is not a Gaussian-like distribution when $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$ are far away. Therefore, authors use an additional center point loss L_c to allow the entire loss to continue optimizing in non-overlapping cases. The overlap is defined as:

$$\mathrm{KFIoU} = \frac{\mathcal{V}(\Sigma)}{\mathcal{V}(\Sigma_1) + \mathcal{V}(\Sigma_2) - \mathcal{V}(\Sigma)} = \frac{|\Sigma|^{1/2}}{|\Sigma_1|^{1/2}| + |\Sigma_2|^{1/2} - |\Sigma|^{1/2}}$$

where $\mathcal{V}(\Sigma) = 2^n|\Sigma|^{1/2}$ (n is the number of dimensions) denotes the volume of the rotated

bounding box. The overall regression is total of $\mathcal{L}_c = \ln((\mu_2 - \mu_1)^T \Sigma_1^{-1}(\mu_2 - \mu_1) + 1)$ and $\mathcal{L}_{kf} = e^{1-KFIoU} - 1$. By using Kalman filter, this loss achieves better trend-level alignment with Rotated IoU than GWD and KLD loss, as verified by measuring the error variance.

Bhattacharyya Distance (BD) (Thai et al., 2025) introduces a promising alternative regression loss function based on Bhattacharyya distance, which is designed to measure the partial overlap between two probability distributions. The Bhattacharyya Distance between two multivariate Gaussian distributions $N(\mu_p, \Sigma_p)$ and $N(\mu_t, \Sigma_t)$ is calculated as follows:

$$D_B(X_p, X_t) = \alpha \frac{1}{8}(\mu_p - \mu_t)^T \Sigma^{-1}(\mu_p - \mu_t) + \frac{1}{2}\ln\left(\frac{|\Sigma|}{\sqrt{|\Sigma_p| \cdot |\Sigma_t|}}\right)$$

where $\Sigma = (\Sigma_p + \Sigma_t)/2$ is the average of two covariance matrices, and α is a hyper-parameter to adjust the impact of mean different term. The final regression loss is $\mathcal{L}_{\mathcal{BD}} = 1 - 1/(1 + \sqrt{D_B})$. Unlike KLD and GWD, BD satisfies all desirable IoU loss properties: non-negativity, symmetry, triangle inequality, and scale-invariance.

Table 2. Result on the DOTA-1.0 test set (measured by mean AP_{50})

Model	Loss Type	Loss	AP_{50}
RetinaNet (Lin et al., 2017)	l_n-norm	Smooth L1 (Girshick, 2015)	68.43
	Angular Encoding	CSL (Yang et al., 2020)	69.51
	Angular Encoding	DCL (Yang et al., 2021)	69.79
	Approximate IoU	IoU-Smooth L1 (Yang et al., 2019)	69.49
	Gaussian-based	GWD Loss (Yang et al., 2021)	70.07
	Gaussian-based	KLD Loss (Yang et al., 2021)	70.31
	Gaussian-based	KFIoU Loss (Yang et al., 2022b)	69.96
	Gaussian-based	BD Loss (Thai et al., 2025)	**71.86**
R3Det (Yang et al., 2021)	l_n-norm	Smooth L1 (Girshick, 2015)	69.80
	Gaussian-based	GWD Loss (Yang et al., 2021)	72.82
	Gaussian-based	KLD Loss (Yang et al., 2021)	72.12
	Gaussian-based	KFIoU Loss (Yang et al., 2022b)	72.60
	Gaussian-based	BD Loss (Thai et al., 2025)	**73.41**

The performance of RetinaNet varies with each loss type, with the Bhattacharyya Distance Loss achieving the highest AP_{50} of 71.86, indicating its effectiveness in improving model accuracy. On the other hand, R3Det is evaluated using similar loss settings, except Angular Encoding and Approximate IoU-based losses. Gaussian-based loss functions demonstrate superior performance, with the BD Loss achieving the highest AP_{50} score of 73.41 for this model. This suggests that the Gaussian-based loss functions significantly enhance detection capabilities. Overall, the table highlights the importance of selecting appropriate loss functions and demonstrates the potential of Gaussian-based methods in boosting object detection performance.

Table 3 shows the effectiveness of Gaussian-based loss functions, particularly BD Loss, in enhancing object detection performance on the HRSC2016 dataset. For RetinaNet, Gaussian-based losses outperform the l_n–norm with Smooth L1 loss, with BD Loss achieving the highest AP_{50}: AP_{95} score of 56.25. This indicates the superior ability of BD Loss in capturing complex spatial relationships and providing robust performance, better than GWD (52.33), KLD (54.78), and KFIoU (49.15). In the R3Det model, Gaussian-based losses are consistently more effective, with BD Loss again delivering the top performance at 57.86. This further confirms BD Loss's enhanced accuracy and robustness, surpassing GWD (56.42), KLD (57.79), and KFIoU (55.41). BD Loss demonstrates outstanding performance across both models, emphasizing its effectiveness in enhancing object detection on HRSC2016. The consistent superiority of BD Loss among Gaussian loss types emphasizes its capacity to better handle the intricacies of object localization and classification.

Table 3. Evaluation on HRSC2016 dataset. The evaluation metric is mean AP_{50}:AP_{95}

Model	Loss Type	Loss	AP_{50}:AP_{95}
RetinaNet (Lin et al., 2017)	l_n-norm	Smooth L1 (Girshick, 2015)	44.82
	Gaussian-based	GWD Loss (Yang et al., 2021)	52.33
	Gaussian-based	KLD Loss (Yang et al., 2021)	54.78
	Gaussian-based	KFIoU Loss (Yang et al., 2022b)	49.15
	Gaussian-based	BD Loss (Thai et al., 2025)	**56.25**
R3Det (Yang et al., 2021)	l_n-norm	Smooth L1 (Girshick, 2015)	53.68
	Gaussian-based	GWD Loss (Yang et al., 2021)	56.42
	Gaussian-based	KLD Loss (Yang et al., 2021)	57.79
	Gaussian-based	KFIoU Loss (Yang et al., 2022b)	55.41
	Gaussian-based	BD Loss (Thai et al., 2025)	**57.86**

5. CONCLUSION

In this study, a thorough review of regression loss functions for rotated object detection was presented, focusing on how they address key challenges such as boundary discontinuity, the squared-like ambiguities, and inconsistencies between optimization objectives and evaluations metric. Among the surveyed loss functions, Gaussian-based losses were emphasized due to their notable effectiveness in improving detection accuracy.

Comprehensive experiments on the DOTA and HRSC2016 benchmarked were conducted to evaluate representative loss functions across different categories. The evaluation demonstrates that Gaussian-based methods, with Bhattacharyya Distance in particular, achieve higher precision and resilience in object localization.

For future research, a promising direction is the combination of multiple loss functions to leverage their complementary advantages. Hybrid approaches that integrate angular-based losses, Gaussian-based losses and approximated IoU losses could further enhance robustness and accuracy in rotated object detection. By strategically combining these losses, hybrid regression losses can simultaneously address their individual weaknesses. For instance, a weighted loss could balance the stability of angular regression, the probabilistic robustness of Gaussian, and the metric alignment of approximation IoU-based losses. Alternatively, an adaptive mechanism could dynamically adjust the contribution of each component depending on object characteristics. Such hybridization has the potential to deliver improved convergence stability, robustness against boundary discontinuities, and superior accuracy across diverse datasets.

Furthermore, we intend to expand the experimental scope by incorporating additional datasets, exploring varied detection architectures, and testing a wider range of regression loss functions. These efforts will allow for a more complete assessment of robustness and generalization, while offering deeper insights into the interaction between loss formulation, data properties, and model design.

Additionally, achieving better computational efficiency without sacrificing precision continues to be an important objective for practical deployment. It is anticipated that this study will provide a useful basic for future research on regression loss design in rotated object detection.

REFERENCES

Chen, Z., Chen, K., Lin, W., See, J., Yu, H., Ke, Y., & Yang, C. (2020, August). Piou loss: Towards accurate oriented object detection in complex environments. In *the European Conference on Computer Vision* (pp. 195-211). Cham: Springer International Publishing.

Ding, J., Xue, N., Long, Y., Xia, G. S., & Lu, Q. (2018). Learning RoI transformer for detecting oriented objects in aerial images. *arXiv preprint arXiv:1812.00155*.

Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440-1448).

Han, J., Ding, J., Li, J., & Xia, G. S. (2021). Align deep features for oriented object detection. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-11.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition* (pp. 770-778).

Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., & Beijbom, O. (2019). Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of*

the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12697-12705).

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2980-2988).

Liu, Z., Yuan, L., Weng, L., & Yang, Y. (2017, February). A high resolution optical satellite image dataset for ship recognition and some new baselines. In *International Conference on Pattern Recognition Applications and Methods* (Vol. 2, pp. 324-331). SciTePress.

Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Ming, Q., Zhou, Z., Miao, L., Zhang, H., & Li, L. (2021, May). Dynamic anchor learning for arbitrary-oriented object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 3, pp. 2355-2363).

Pan, X., Ren, Y., Sheng, K., Dong, W., Yuan, H., Guo, X., … & Xu, C. (2020). Dynamic refinement network for oriented and densely packed object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11207-11216).

Qian, W., Yang, X., Peng, S., Yan, J., & Guo, Y. (2021, May). Learning modulated loss for rotated object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 3, pp. 2458-2466).

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). 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* (pp. 658-666).

Thai, C., Trang, M. X., Ninh, H., Ly, H. H., & Le, A. S. (2025). Enhancing rotated object detection via anisotropic Gaussian bounding box and Bhattacharyya distance. *Neurocomputing*, *623*, 129432.

Xia, G. S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., … & Zhang, L. (2018). DOTA: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3974-3983).

Yang, X., Yang, J., Yan, J., Zhang, Y., Zhang, T., Guo, Z., … & Fu, K. (2019). Scrdet: Towards more robust detection for small, cluttered and rotated objects. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 8232-8241).

Yang, X., & Yan, J. (2020, August). Arbitrary-oriented object detection with circular smooth label. In *European Conference on Computer Vision* (pp. 677-694). Cham: Springer International Publishing.

Yang, X., Hou, L., Zhou, Y., Wang, W., & Yan, J. (2021). Dense label encoding for boundary discontinuity free rotation detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15819-15829).

Yang, X., Yan, J., Feng, Z., & He, T. (2021, May). R3det: Refined single-stage detector with feature refinement for rotating object. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 4, pp. 3163-3171).

Yang, X., Yan, J., Ming, Q., Wang, W., Zhang, X., & Tian, Q. (2021, July). Rethinking rotated object detection with Gaussian Wasserstein distance loss. In *International Conference on Machine Learning* (pp. 11830-11841). PMLR.

Yang, X., Yang, X., Yang, J., Ming, Q., Wang, W., Tian, Q., & Yan, J. (2021). Learning high-precision bounding box for rotated object detection via Kullback-Leibler divergence. *Advances in Neural Information Processing Systems, 34*, 18381-18394.

Yang, X., Yan, J., Liao, W., Yang, X., Tang, J., & He, T. (2022). Scrdet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(2), 2384-2399.

Yang, X., Zhou, Y., Zhang, G., Yang, J., Wang, W., Yan, J., … & Tian, Q. (2022). The KFIoU loss for rotated object detection. *arXiv preprint arXiv:2201.12558*.

Yin, T., Zhou, X., & Krahenbuhl, P. (2021). Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11784-11793).

Yu, J., Jiang, Y., Wang, Z., Cao, Z., & Huang, T. (2016, October). Unitbox: An advanced object detection network. In *Proceedings of the 24th ACM International Conference on Multimedia* (pp. 516-520).

Yu, Y., & Da, F. (2023). Phase-shifting coder: Predicting accurate orientation in oriented object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13354-13363).

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020, April). Distance-IoU loss: Faster and better learning for bounding box regression. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 07, pp. 12993-13000).

Zhou, D., Fang, J., Song, X., Guan, C., Yin, J., Dai, Y., & Yang, R. (2019, September). IoU loss for 2d/3d object detection. In *the 2019 International Conference on 3D Vision (3DV)* (pp. 85-94). IEEE.

Zhou, Y., Yang, X., Zhang, G., Wang, J., Liu, Y., Hou, L., … & Chen, K. (2022, October). Mmrotate: A rotated object detection benchmark using PyTorch. In *Proceedings of the 30th ACM International Conference on Multimedia* (pp. 7331-7334).