Hilbert Curve-Encoded Rotation-Equivariant Oriented Object Detector with Locality-Preserving Spatial Mapping

Qi Ming[1*], Liuqian Wang[2*], Juan Fang[1†], Xudong Zhao[3†], Yucheng Xu[4], Ziyi Teng[1], Yue Zhou[5], Xiaoxi Hu[6], Xiaohan Zhang[7], Yufei Guo[8]

[1] College of Computer Science, Beijing University of Technology,
[2] School of Cyber Science and Engineering, Zhengzhou University,
[3] School of Information and Electronics, Beijing Institute of Technology,
[4] Hong Kong University of Science and Technology (Guangzhou),
[5] School of Geospatial Artificial Intelligence, East China Normal University,
[6] State Key Laboratory of Intelligent Green Vehicle and Mobility, Tsinghua University,
[7] College of Information Science and Electronic Engineering, Zhejiang University,
[8] Intelligent Science & Technology Academy of CASIC
{mingqi,fangjuan}@bjut.edu.cn, jeremy.wang0126@gmail.com, zhaoxudong@bit.edu.cn

Abstract

Arbitrary-Oriented Object Detection (AOOD) has found broad applications in embodied intelligence, autonomous driving, and satellite remote sensing. However, current AOOD frameworks face challenges in ineffective feature extraction and orientation regression inaccuracy. Inspired by Hilbert curve's intrinsic locality-preserving property, we propose a flexible Hilbert curve-Encoded Rotation-Equivariant Oriented Object Detector (HERO-Det). Our innovations include: (i) a novel Hilbert curve traversal convolution paradigm with a dimensionality reduction scheme, which employs locality-preserving spatial filling curves for feature transformation, (ii) a Hilbert pyramid transformer enabling hierarchical construction of multi-scale feature sequences through space-folding operations, as well as (iii) an orientation-adaptive prediction head that decouples rotation-equivariant regression features from invariant classification cues to resolve orientation regression dilemmas in two-stage detectors. Extensive experiments show HERO-Det achieves state-of-the-art performance on AOOD benchmarks, with mAP of 79.56%, 90.64%, 90.10%, and 80.47% on DOTA, HRSC2016, SSDD, and HRSID, respectively. Performance gains in cross-task validation further demonstrate the versatility of our method to diverse vision tasks, such as medical image segmentation and 3D object detection.

Code — https://github.com/Qian-CV/HERO-Det

Introduction

The advancement of computer vision has progressively turned object detection into more specialized frameworks (Lin et al. 2017). Within this landscape, Arbitrary-Oriented Object Detection (AOOD) has emerged as a prominent research frontier, aiming to localize objects with arbitrary orientations in visual recognition tasks. AOOD has demonstrated wide applicability in various domains, including retail product recognition (Chen et al. 2020), remote sensing imagery interpretation (Yang et al. 2021c; Ming et al. 2021b), autonomous driving (Sheng et al. 2022), embodied intelligence (Arsalan et al. 2024), and scene text detection (He et al. 2021).

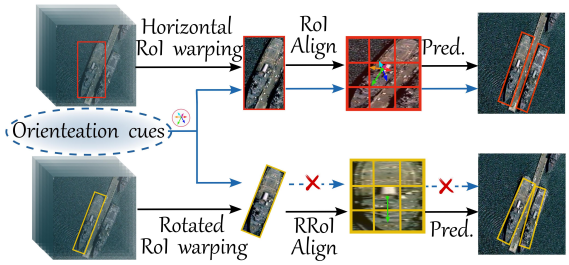


Figure 1: Illustration of orientation information flow during the RoI align process. The features extracted by HRoI Align (top) preserve object orientation cues, whereas those from the more precise RRoI Align lose orientation information.

Over the past decade, numerous detectors have been developed through carefully designed pipelines involving label assignment (Hou et al. 2022), angular representation (Yang and Yan 2022), feature extraction (Ming et al. 2021a), and loss optimization (Yang et al. 2021c), achieving state-of-the-art results on benchmarks. These approaches are generally categorized into one-stage and two-stage paradigms. While one-stage detectors exhibit computational efficiency advantages (Yang et al. 2021a), two-stage architectures have shown superior performance in complex scenarios such as occlusions, low-resolution images, or extreme precision requirements (Xie et al. 2021; Ding et al. 2019). This study adopts two-stage detectors as the primary focus.

Despite the promising potential of two-stage detectors, their performance remains constrained by basic limitations in orientation prediction and feature representation. (i) **Orientation Prediction Bottleneck**: Current two-stage AOOD frameworks typically employ rotated region proposals (RRoI)

*These authors contributed equally.
†Corresponding author.

via rotated RoI Align to extract local features (illustrated at bottom of Fig. 1), which are then simultaneously utilized for both classification and regression. However, the architecture inevitably discards critical orientation information after the RRoI cropping step, significantly amplifying the orientation error of predictions. (ii) **Feature Representation Deficiencies**: Beyond the aforementioned RRoI feature inaccuracies, two further limitations persist in current approaches: Firstly, standard convolutional kernels perform axis-aligned sliding-window operations that fail to maintain spatial coherence and cannot effectively leverage contextual dependencies. Secondly, in the second stage, directly flattening RRoI features into 1D sequences would disrupt intra-object spatial coherence, thereby degrading the regression precision.

In this study, we propose a **H**ilbert curve-**E**ncoded **R**otation-**E**quivariant **O**riented Object **Det**ector (HERO-Det) to address the above issues. The framework integrates Hilbert Curve Traversal Convolution (HCTConv) to model contextual relationships, and builds a Hilbert Pyramid Transformer (HPFormer) to aggregate multi-scale features via cross-attention. A Hilbert curve unfolding operation is then applied to reduce dimensionality while preserving the local structure of RoI features. For accurate orientation prediction, we design an Orientation-Adaptive Prediction Head (OAPH) with two mechanisms: 1) a residual oriented response network is applied to enforce rotation-equivariant feature learning for regression, and 2) a Hilbert sequence circular shifting operator that improves classification robustness with rotation-invariant features. The main contributions are as follows:

- We propose HERO-Det, a novel detection framework that exploits the locality-preserving property of the Hilbert curve, integrating a HCTConv operator, a HPformer, and a HC unfolding mechanism to achieve unified high-precision oriented object detection.
- We reveal a commonly overlooked limitation in two-stage detectors—orientation feature degradation in regression branch—and address it with a novel Orientation-Adaptive Prediction Head. This head disentangles directional representations, employing rotation-equivariant features for localization and rotation-invariant ones for classification.
- HERO-Det achieves state-of-the-art results on multiple AOOD benchmarks and shows performance gains across tasks such as image segmentation, 3D object detection.

Related Work

Arbitrary-oriented object detection. AOOD aims to localize objects with arbitrary rotations, which is essential in tasks like aerial imagery (Yang et al. 2021c), scene text detection (Sheng, Chen, and Lian 2021), and autonomous driving (Ming et al. 2023). Existing methods follow either one-stage or two-stage detection paradigms. One-stage detectors directly regress object class and OBB parameters on dense feature maps, achieving fast inference, such as RDD (Liao et al. 2018), DAL (Ming et al. 2021b), and SASM (Hou et al. 2022). Two-stage models first generate rotated proposals, then employ rotated RoI align (Ding et al. 2019) to extract regional features for subsequent classification and regression, achieving higher accuracy in complex scenes (Wang et al. 2023; Xie et al. 2021; Wang et al. 2025). For example, RoI Transformer introduces rotated RoI Align to achieve precise feature alignment with rotated proposals (Ding et al. 2019). This technique has since become a standard module in two-stage oriented object detectors. Building on this, ORCNN (Xie et al. 2021) performs eight-parameter OBB regression directly on the aligned RRoI features to enable flexible representation of oriented objects. Recent advances improve AOOD pipelines through better label assignment (Hou et al. 2022; Ming et al. 2021b), rotation-aware features (Lee et al. 2024), and tailored loss functions (Yang et al. 2021c; Ming et al. 2024) that address angle periodicity and representation ambiguity. Despite these gains, achieving both high accuracy and efficiency in AOOD remains an ongoing challenge.

Rotation sensitive networks. Accurate orientation prediction is crucial for AOOD task, as it directly affects the detection performance of rotated objects. Researchers have developed rotation-equivariant architectures that embed angular priors into the network structure (Han et al. 2021a; Zhou et al. 2017; Liao et al. 2018; Han et al. 2021b; Lee et al. 2024). S^2ANet (Han et al. 2021a) employs Active Rotating Filters (Zhou et al. 2017) to construct rotation sensitive features for regression. On this basis, RDD (Liao et al. 2018) further builds rotation-invariant features to enhance object classification. ReDet (Han et al. 2021b) and FRED (Lee et al. 2024) use group convolutions to extract rotation-equivariant features for object detection. Despite their success, the integration of such rotation-aware designs into object detection has been comparatively limited, especially in two stage detectors. Current approaches overlook the fact that RRoI features in the heads of two-stage detectors inherently lack orientation information, thereby limiting their ability to perform rotation-sensitive regression.

Space-filling curve. The space-filling curves (SFCs) are mappings from a one-dimensional domain onto a higher-dimensional space, such that the entire multi-dimensional region is densely covered. The most widely used SFCs are the Hilbert curve (Hilbert 1891) and Z-order curve (Morton 1966). The Hilbert curve is especially valued for its superior locality-preserving properties, meaning that points close together in space are likely to remain close after linearization (He and Owen 2016). SFCs have wide-ranging applications across multiple domains. Some researchers use Hilbert curve to process point cloud data to reduce GPU occupation (Chen et al. 2022). Wang *et al.* (Wang, Xu, and Song 2021) achieve high-performance tumor lesion segmentation by leveraging the spatial correspondence mapping of the Hilbert curve. Hilbert curve projection distance (Li et al. 2024) has been proposed measure the distance between distributions.

Preliminaries

We first introduce the relevant concepts of space-filling curves. The Hilbert space-filling curve provides a continuous mapping $\mathcal{H} : [0, 1] \to [0, 1]^d$ from the one-dimensional unit interval to d-dimensional unit hypercube. Following established conventions (Sagan 1994), we formally define the Hilbert curve as the limit of a sequence of mappings:

Definition: For a given dimensionality $d \geq 1$ and a positive integer k, we partition the unit interval $I = [0, 1]$ into

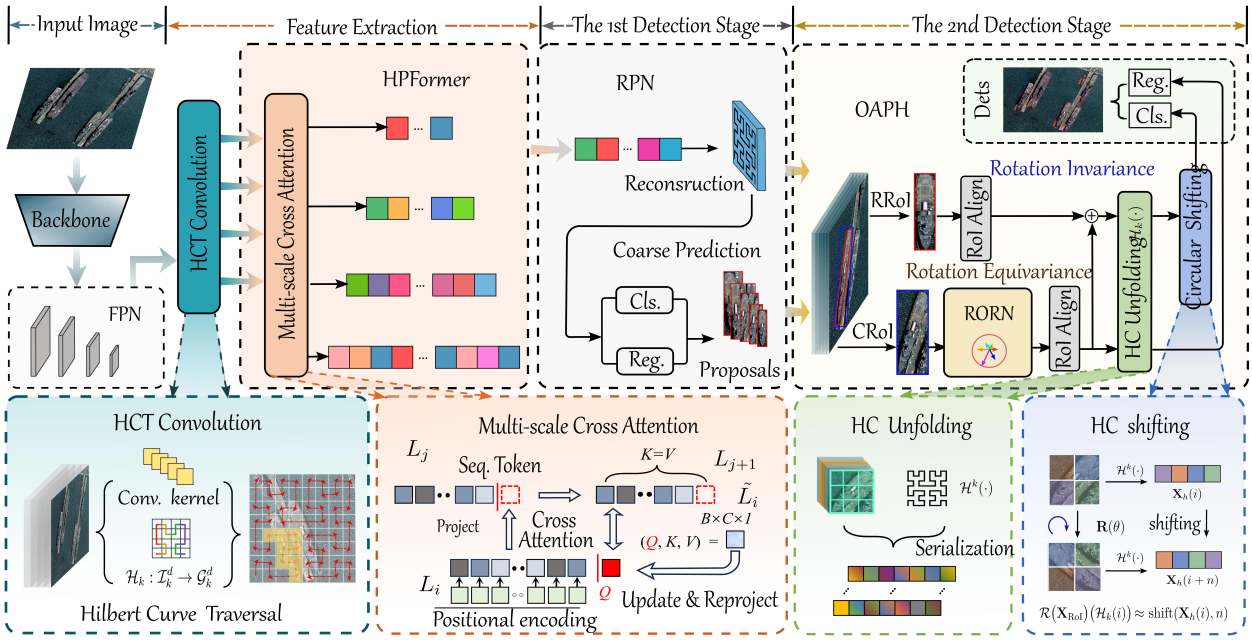


Figure 2: Overall framework of HERO-Det.

2^{kd} sub-intervals with length 2^{-kd} and denote the set of intervals as $\mathcal{I}_k^d$. Simultaneously, divide the d-dimensional unit hypercube $C = [0,1]^d$ into 2^{kd} congruent sub-cubes with length 2^{-k} and denote the set of suncubes as $\mathcal{G}_k^d$. Then Hilbert curve $\mathcal{H}$ is defined as the limit of the sequence of mappings $H_k : \mathcal{I}_k^d \to \mathcal{G}_k^d$ as $k \to \infty$, satisfying the following properties: (i) **Surjectivity:** For every sub-cube $C_{k,v}^d \in \mathcal{G}_k^d$, there exists a corresponding sub-interval $I_{k,u}^d \in \mathcal{I}_k^d$ such that $\mathcal{H}_k(I_{k,u}^d) = C_{k,v}^d$, where $u \in \{0,1,\ldots,2^{kd}-1\}$, and $v \in \{0,1,\ldots,2^k-1\}^d$. (ii) **Adjacency:** Sub-intervals $I_{k,u}^d$ and $I_{k,u'}^d$ that are adjacent in the one-dimensional ordering are mapped by H_k to sub-cubes $C_{k,v}^d$ and $C_{k,v'}^d$ that share a $(d-1)$-dimensional face. (iii) **Nesting:** The division of sub-intervals at the $(k-1)$-th order corresponds to subdivision of the corresponding sub-cubes at the k-th order.

Theorem 1 (Hölder Continuity)**.** *The Hilbert curve exhibits a Hölder continuity property, which guarantees that: For any two points $x, y \in [0,1]$, and their mapped points $\mathcal{H}(x), \mathcal{H}(y) \in [0,1]^d$, there exist integer $d > 1$ such that:*

$$\|\mathcal{H}(x) - \mathcal{H}(y)\| \leq 2\sqrt{d+3}|x-y|^{1/d} \tag{1}$$

where $\mathcal{H}: [0,1] \to [0,1]^d$ is the Hilbert curve mapping.

The continuity demonstrated by Theorem 1 ensures that the Hilbert curve mapping preserves local relationships: elements that are close with respect to the one-dimensional Hilbert ordering are mapped to points that are spatially proximate in the higher-dimensional space. And also, the spatial proximity works in reverse Hilbert curve mapping (Wang and Shan 2005; Bader 2012). This characteristic spatial coherence renders the Hilbert curve advantageous for applications where maintaining spatial adjacency is a key requirement (Chen and Chang 2011).

Methodology

Overall Architecture

Overall architecture of HERO-Det is shown in Fig. 2. Given an input image, we first extracted features via a backbone network, and then HCTConv is applied to obtain locality-preserving features. Next, a cross-scale attention mechanism is used to align features across layers, constructing the HPFormer for subsequent predictions. With the coarse proposals from the 1st stage, contextual RoIs are fed into OAPH to achieve rotation-aware prediction.

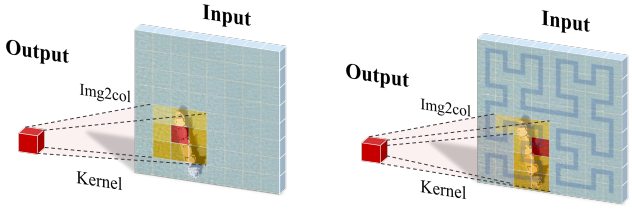
Hilbert Curve Traversal Convolution

Most object detection framework use axis-aligned convolution kernels shown in Fig. 3a to slide over feature map in a regular grid pattern (Redmon et al. 2016). However, these kernels are limited under strict rectangular receptive fields and are inherently constrained in modeling rotation variance and long-range spatial dependencies for AOOD task.

To solve the issues, we propose a Hilbert Curve Traversal Convolution (HCTConv) operator. HCTConv handles high-dimensional spatial features via 1D Hilbert curve path to preserve spatial locality. Given the input feature $\mathbf{X} \in \mathbb{R}^{H\times W\times C_{\text{in}}}$, we define a k-order Hilbert mapping $\mathcal{H}_k(\cdot)$:

$$\begin{aligned} &\mathcal{H}_k : \{0,1,\ldots,N-1\} \to \mathbb{Z}^2, \quad N = H \cdot W, \\ &\mathcal{H}_k(i) = (x_i, y_i), \quad \text{where } (x_i, y_i) \in \{0,1,\ldots,N-1\}^2. \end{aligned} \tag{2}$$

This function defines a bijective mapping from 1D index i to 2D coordinates (x_i, y_i) on the grid. Although $\mathcal{H}_k(\cdot)$ is a surjection, the discrete Hilbert curve is a bijection, making it suitable for image processing. As shown in Fig. 3b, on a



(a) Standard Conv. (b) Our HCTConv.

Figure 3: Illustration of convolution principles.

gridded image data, Hilbert curve traverses all positions in a locality-preserving manner such that spatially adjacent pixels in feature map are projected to adjacent positions in Hilbert sequence with minimal distortion in locality.

Given the mapping $\mathcal{H}_k(\cdot)$, for $t \in \{0, \ldots, N-1\}$, the HCTConv is defined as:

$$\mathbf{Y}(t, c_{\text{out}}) = \sum_{s \in \mathcal{N}(t)} \sum_{c=1}^{C_{\text{in}}} \mathbf{W}(s-t, c_{\text{in}}, c_{\text{out}}) \cdot \mathbf{X}(\mathcal{H}_k(s), c), \tag{3}$$

where $\mathcal{N}(t)$ denotes the local receptive neighborhood around t in the Hilbert curve domain (e.g., a 1D window of kernel size K), $\mathbf{W} \in \mathbb{R}^{K \times C_{\text{in}} \times C_{\text{out}}}$ is the convolutional kernel defined over the Hilbert traversal offsets. Note that the input feature dimensions do not always exactly match the size of the k-order Hilbert curve, i.e. $H, W \neq 2^{2k}$. In practice, we apply zero-padding to the spatial dimensions such that $H', W' = \min\{2^{2m} \geq H, 2^{2n} \geq W\}$. After HCTConv on the padded features, the output is cropped back to the original resolution to eliminate padding effects. An intuitive visualization is shown in Fig. 2.

Compared with standard convolutions with fixed rectangular receptive field, HCTConv inherently preserves spatial correspondence and enables robust feature encoding. The locality-preserving property of the Hilbert curve mapping guarantees that nearby indices in the 1D domain correspond to spatially adjacent regions in 2D features as explained in Theorem 1. In this case, the model is able to capture fine-grained geometric patterns without disrupting local consistency, which is particularly important for AOOD since rotational and contextual sensitivity are crucial. To the best of our knowledge, this is the first work to introduce the Hilbert curve for constructing convolutional operations.

Hilbert Pyramid Transformer

With HCTConv, we extract multi-scale feature sequences and further construct the Hilbert Pyramid Transformer (HPFormer) to effectively detect multi-scale oriented objects. Given feature map P_i, HCTConv transforms it into a locally consistent feature $L_i = \text{HCTConv}(P_i)$. For each input sequence L_i, a learnable positional encoding is applied to introduce a position bias for each token. At each layer, a learnable sequence token, serving as the query (Q) in the attention computation, is appended to the feature sequence to capture global information. For every adjacent feature sequence pair (L_i , L_j), the first step is to project the sequence token from L_i to L_j to learn contextual knowledge, forming a new sequence L_{j+1}, which then serves as the key (K) and value (V) in the attention computation with the Q from the lower-level sequence. The attention-enhanced sequence token $\tilde{L}_i \in \mathbb{R}^{C \times 1}$ is computed as:

$$\tilde{L}_i = \text{softmax}\left(\frac{\left(\tilde{L}_i W_Q\right)\left(L_{j+1} W_K\right)^\top}{\sqrt{d}}\right) \cdot (L_{j+1} W_V), \tag{4}$$

where $W_Q, W_K, W_V \in \mathbb{R}^{C \times d}$ are learned projection matrices, d is the embedding dimension. After obtaining the attention-enhanced sequence token, it is reprojected to the L_i layer and used to update the sequence token. Finally, HPFormer performs feature fusion from lower to higher sequences to generate progressively refined feature representations across all scales.

The HPFormer effectively combines Hilbert-based feature encodings with cross-scale semantic fusion through cross-attention mechanism. It leverages the local continuity features obtaion via HCTConv and the adaptive weighting of attention to build cross-scale feature pyramid. The output sequence $\tilde{L}_{i+1}$ captures both detailed local structures and rich contextual semantics, enhanced by Hilbert spatial locality and attention flexibility. HPFormer ensures high-quality sequence fusion that is particularly advantageous to detect small or oriented objects under complex scenes.

Orientation-Adaptive Prediction Head

Residual Oriented Response Network. Existing two-stage AOOD frameworks often extract RRoI features through spatial transform operation (Ding et al. 2019; Xie et al. 2021; Han et al. 2021b). The RRoI feature transformation inevitably erases orientation cues by normalizing orientations, rendering features not suitable for orientation regression. We suggest that regression features should not only preserve semantic patches of the object but also encode surrounding context that implicitly supports orientation understanding for deep models. Given the oriented proposal $\mathbf{B}_{\text{RRoI}}(cx, cy, w, h, \theta)$, we define a Contextual-aware RoI (CRoI) as $\mathbf{B}_{\text{CRoI}}(cx, cy, \alpha w', \alpha h')$, where α is a scaling coefficient and $\alpha > 1$, w' and h' denote the shape of the horizontal minimum box of $\mathbf{B}_{\text{RRoI}}$. CRoI features includes background and neighboring structures which acts as spatial references for orientation prediction. Then CRoI features $\mathbf{X}_{\text{CRoI}}$ serve as the input for orientation-aware regression.

$$\begin{aligned}
&\mathbf{Y}(k) = \text{Conv}\left(\mathcal{R}_{\theta_k}(\mathbf{K}) * \tilde{\mathbf{X}}^{(k-1)}\right), \quad k = 1, \ldots, N-1, \\
&\tilde{\mathbf{X}}(k) = \tilde{\mathbf{X}}^{(k-1)} + \mathbf{Y}^{(k)}, \\
&\tilde{\mathbf{X}}(0) = \mathbf{X}_{\text{CRoI}}, \quad \tilde{\mathbf{X}}_{\text{RE}} = \tilde{\mathbf{X}}^{(N-1)}.
\end{aligned} \tag{5}$$

To accurately capture rotation-equivariant semantic cues, we apply Active Rotating Filters (Zhou et al. 2017) to build a Residual Oriented Response Network (RORN). We define a set of rotation angles $\{\theta_1, \theta_2, \ldots, \theta_{N-1}\}$, with $\theta_k = k \cdot \Delta\theta$,

where $\Delta\theta = \frac{2\pi}{N}$. For each angle θ_k, we rotate a shared convolution kernel $\mathbf{K}$ to extract orientation-specific features, and then aggregate them with the features from the previous stage as shown in Eq. (5). $\mathcal{R}_{\theta_k}(\cdot)$ applies kernel rotation, $\tilde{\mathbf{X}}(k)$ represents the fused feature at stage k, $\tilde{\mathbf{X}}_{\mathrm{RE}}$ is the rotational equivariant feature. This structure guarantees that each orientation response is modeled with full channel capacity, and all orientation cues are aggregated through dense residual paths (Huang et al. 2017).

Existing methods often use a single ARF layer with limited channel capacity per orientation (e.g., $\mathbb{R}^{H\times W\times \frac{C}{N}\times N}$) (Han et al. 2021a; Zhou et al. 2017; Liao et al. 2018). In contrast, our approach employs dense residual routes to progressively integrate orientation cues, enhancing angular continuity and enabling more precise orientation modeling. This design ensures robust rotation equivariance while preserving critical semantic structures. Moreover, the residual fusion scheme also improves gradient propagation and facilitates faster convergence during training, making the method both effective and optimization-friendly.

Hilbert Curve Unfolding. With the regional grid RoI features after RoI alignment (Ding et al. 2019), we proceed to unfold each feature map into a 1D sequence for the subsequent classification and regression. We then propose the Hilbert Curve Unfolding (HC Unfolding), which employs a space-filling Hilbert curve to reorder the features while preserving their topological proximity. The unfolded sequence is $\mathbf{X}_h(t) = \tilde{\mathbf{X}}(\mathcal{H}_k(t)) \in \mathbb{R}^C$, where $\tilde{\mathbf{X}}$ denotes the input RoI features. Compared to the most commonly used raster scanning (Zhang et al. 2015) for feature flattening, HC Unfolding ensures that for adjacent indice $i, j \in \mathcal{N}(t)$ (spatial neighborhood of position t given in Eq. (3)), the distance between their mapped coordinates in the 2D domain satisfies: $\|\mathcal{H}_k(i) - \mathcal{H}_k(j)\|_2^2 < \mathrm{D}_{RS}(i,j)$, where $\mathrm{D}_{RS}(i,j)$ denotes the spatial distance under raster scanning. This local Hölder continuity in Theorem 1 implies that adjacent elements in the Hilbert sequence are also adjacent in 2D space, allowing any subsequent linear or attention-based operator to retain a strong spatial prior. As such, convolutional or attention modules applied to Hilbert sequence effectively model localized spatial interactions while enabling sequential processing.

Hilbert Circular Shifting. To further enhance the robustness of classification features, we design a latent-space feature augmentation strategy based on the structural properties of the Hilbert curve, which is called Hilbert Circular Shifting (HCS). First, we exploit the circular-shifting property of first-order Hilbert curve traversal: rotating a 2D RoI feature map by $n \cdot 90^\circ$ clockwise is equivalent to a left n-offset circular shift of its flattened sequence obtained by HC Unfolding (shown in Fig. 2). The circular shifting property is elegant, but things become more complex for higher orders. Therefore, we use the first-order Hibert curve for HCS. Given the input RoI feature map $\mathbf{X}_{\mathrm{RoI}} \in \mathbb{R}^{H\times W\times C_{\mathrm{in}}}$, its HC unfolding sequence is $\mathbf{X}_h$. Then a feature rotation by $n \cdot 90^\circ$ approximates shifting-equivalence is expressed as:

$$\mathcal{R}\big(\mathbf{X}_{\mathrm{RoI}}\big)\big(\mathcal{H}_k(i)\big) \;\approx\; \mathbf{X}_h\big(\,(i + n\tfrac{N}{4}) \bmod N\,\big), \quad n \in \mathbb{N}, \tag{6}$$

where $\mathcal{R}$ is rotation operation, N denotes the length of $\mathbf{X}_h$. The rotation augmentation could then be performed in the latent feature space presented by Hilbert curve, constructing the rotation-invariant classification features as follows:

$$\begin{aligned} \mathbf{X}_h^{(n)}(i) &= \mathbf{X}_h\big(\,(i + n\tfrac{N}{4}) \bmod N\big), \quad n = 0,1,2,3, \\ \tilde{\mathbf{X}}_{\mathrm{RI}}(i) &= \sum_{n=0}^{3} \mathbf{X}_h^{(n)}(i), \quad i = 0,\ldots,N-1. \end{aligned} \tag{7}$$

Eq. (7) yields a feature representation $\tilde{\mathbf{X}}_{\mathrm{RI}}$ that is invariant to rotations of the objects. With the rotation invariant feature sequences, the classifier would recognize identical inputs regardless of rotation, which helps to achieve higher classification accuracy and improves generalization to unseen orientations. HCS strategy offers an efficient latent-space augmentation mechanism far cheaper than direct 2D rotations. From a computational standpoint, the HCS strategy avoids per-pixel coordinate transforms or memory fetches across two dimensions like feature map rotation, and therefore yields significantly faster practical speed.

Different Variants				Metric
HCTConv	HPFormer	RORN	HCS	mAP(%)
				67.5
✓				69.1
✓	✓			70.2
		✓		68.6
		✓	✓	69.2
✓	✓	✓	✓	**70.7**

Table 1: Evaluation of components in HERO-Det.

Methods	Scanning path	Kernel	mAP(%)
Baseline	Raster Scan (Zhang et al. 2015)	3×3	67.5
Naive Unfolding	Raster Scan (Zhang et al. 2015)	1×9	67.2
	Continuous Scan (Yang et al. 2024)	1×9	67.6
SFC Unfolding	Z-order curve (Morton 1966)	1×9	68.0
	Peano curve (Peano and Peano 1990)	1×9	64.0
	Hilbert curve (Hilbert 1891)	1×7	68.8
		1×9	**69.1**

Table 2: Ablation on HCTConv settings.

ORN	RORN	HC Unfolding	mAP(%)
✓			67.1
✓		✓	68.1
	✓		67.5
	✓	✓	**68.6**

Table 3: Analysis of components in OAPH.

Experiments and Analysis

Datasets and Implementation Details

Datasets. We evaluate our method on AOOD benchmarks, including DOTA (Xia et al. 2018), HRSC2016 (Liu et al.

Flatten	Reconstruct	mAP(%)
——		67.5
Raster Scan	Raster Scan	67.2
Hilbert curve	Raster Scan	68.7
Hilbert curve	Hilbert curve	**69.1**

Table 4: Settings in HC Unfolding.

shifting	0	1	2	3
BR	43.1	43.4	43.4	**44.3**
SV	66.1	67.4	67.3	**67.5**

Table 5: Effect of shifting in HCS.

2017), SSDD (Li, Qu, and Shao 2017), HRSID (Wei et al. 2020). DOTA-v1.0 (Xia et al. 2018) is the most commonly used dataset for AOOD in remote sensing images. It includes 2806 large aerial images annotated with 188,282 instances across 15 object categories. It is divided into 1/2 for training, 1/6 for validation, and 1/3 for testing. HRSC2016 (Liu et al. 2017) includes 1061 high-resolution images for ship detection, divided into 436 training, 181 validation, and 444 testing samples. SSDD (Li, Qu, and Shao 2017) provides 1160 SAR images with over 2400 annotated ships, designed for SAR ship detection. HRSID (Wei et al. 2020) contains 560 SAR images with 16,000 rotated ship annotations for detecting small objects in complex scenes.

Implementation Details. Experiments are conducted with 4 NVIDIA RTX 4090 GPUs using PyTorch. We train the models with 16 images per minibatch using SGD (initial learning rate 0.02, momentum 0.9, weight decay 0.0001) and a 500-iteration warm-up setting. RoI scaling factor α is set to 1.2. On DOTA dataaset, we train for 12 epochs for ablations and 18 epochs for main results. Models on HRSC2016, SSDD, and HRSID are trained for 72 epochs. We conduct evaluation on the DOTA val set for ablations and on the test set for final results. Data augmentation includes random flipping, rotation, and multi-scale training and testing; only flipping is used in ablations.

Ablation Studies

Component-wise Ablations. To evaluate the HERO-Det framework, we conduct ablation studies on each component of the detector on the DOTA dataset. The experimental results are summarized in Tab. 1. First, we apply the HCTConv operator to multi-scale features yields locality-preserving representations, leading to a 1.6% improvement over the baseline. Incorporating cross-scale attention to build HPFormer further enhances local feature representation, boosting performance by an additional 1.1%. We then evaluate units of the OAPH. Specifically, the RORN effectively captures orientation-aware features to achieve rotation-equivariant regression, improving mAP by 1.1%. On this basis, the HCS strategy introduces rotation-invariant features for classification, yielding a further gain of 0.6%. Integrating all components into HERO-Det achieves an overall mAP of 70.7%, demonstrating their compatibility and effectiveness.

Analysis of HCTConv Operator. We conduct ablations about mechanism of HCTConv and present detailed experimental results in Tab. 2. The baseline model adopts a standard 3×3 convolution and applies the raster scanning (Zhang et al. 2015) over input features. Similarly, we flatten the feature into a sequence and apply convolution guided by a raster scanning, but observe negligible performance improvement. The result suggests that the spatial scanning path plays a crucial role, as it determines how local features are aggregated. Therefore, we explore various scanning strategies in Tab. 2. For a fair comparison, most convolution designs share the same receptive field. First, we adopt the Continuous Scanning (Yang et al. 2024), which connects features across rows to enhance contextual association. Continuous Scanning improves mAP by 0.4%, which supports our hypothesis. Then, we try the space-filling curves for convolution. Compared to Z-order curve (Morton 1966) and Peano curve (Peano and Peano 1990), the Hilbert curve (Hilbert 1891) exhibits superior locality-preserving properties, effectively capturing local feature dependencies during the convolution process. As a result, it achieves the best mAP of 69.1%.

Evaluation of OAPH Module. As a core component of HERO-Det, we conduct ablation studies on the functional Unit of OAPH, with results shown in Tab. 3. Compared to the niave ORN framework build on ARF (Zhou et al. 2017), RORN introduces dense residual connections and continuous angular transformations, yielding a performance gain of 0.4 points. Furthermore, instead of directly flattening the rotation-equivariant features, HC Unfolding preserves local feature correlations. It provides more effective semantic cues for subsequent fully connected layers, leading to a 1.1% improvement based on RORN. Since HPFormer also involves Hilbert curve-based unfolding and reconstruction, we analyze their effect on performance. As shown in Tab. 4, using HCTConv with Hilbert curve reconstruction achieves the best mAP of 69.1%, which is higher than Raster Scanning by 1.9%. Even with mismatched input-output mappings (Hilbert in, Raster out), performance still improves by 1.5%, indicating that Hilbert curve's local continuity helps preserve semantic structure through transformation.

Effect of HCS Strategy. The HCS strategy introduces a hyperparameter n to control the number of rotation angles using feature shifting. We analyze its impact in Tab. 5. In general, objects with large aspect ratios are more sensitive to orientation changes. Therefore, we report performance on narrow objects such as bridges (BR) and small vehicles (SV). Clearly, as n increases, rotation-invariant features can better capture similar patterns across varying orientations, leading to more robust classification. When $n = 3$, mAP improves by 1.2% and 1.4% on bridges and small vehicles, respectively.

Comparision with Other State-of-the-art Models

The DOTA dataset (Xia et al. 2018) is currently the most widely used large-scale benchmark for AOOD in remote sensing imagery. We report the performance of our method on DOTA dataset and compare it with existing advanced approaches, as shown in Tab. 6. HERO-Det achieves state-of-the-art mAP of 79.56%, outperforming many existing advanced methods and demonstrating its superiority. Visual-

	Methods	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP(%)
One Stage Detectors	GCL (Ming et al. 2024)	89.65	85.13	43.25	77.41	**81.25**	77.93	86.69	90.90	86.93	84.49	64.13	65.77	68.14	78.51	61.31	76.10
	R^3Det (Yang et al. 2021b)	89.80	83.77	48.11	66.77	78.76	83.27	87.84	90.82	85.38	85.51	65.67	62.68	67.53	78.56	72.62	76.47
	DAL (Ming et al. 2021b)	89.69	83.11	55.03	71.00	78.30	81.90	88.46	90.89	84.97	87.46	64.41	65.65	76.86	72.09	64.35	76.95
	DCL (Yang et al. 2021a)	89.26	83.60	53.54	72.76	79.04	82.56	87.31	90.67	86.59	86.98	67.49	66.88	73.29	70.56	69.99	77.37
	GWD (Yang et al. 2021c)	89.06	84.32	55.33	77.53	76.95	70.28	83.95	89.75	84.51	86.06	**73.47**	67.77	72.60	75.76	74.17	77.43
	O-Rep. (Li et al. 2022)	89.11	82.32	56.71	74.95	80.70	83.73	87.67	90.81	87.11	85.85	63.60	68.60	75.95	73.54	63.76	77.63
	KLD (Yang et al. 2021d)	88.91	85.23	53.64	81.23	78.20	76.99	84.58	89.50	86.84	86.38	71.69	68.06	75.95	72.23	**75.42**	78.32
	S^2A-Net (Han et al. 2021a)	89.28	84.11	56.95	**79.21**	80.18	82.93	89.21	90.86	84.66	87.61	71.66	68.23	78.58	78.20	65.55	79.15
	SASM (Hou et al. 2022)	89.54	**85.94**	57.73	78.41	79.78	84.19	**89.25**	90.87	85.80	87.27	63.82	67.81	**78.67**	79.35	69.37	79.17
Two Stage Detectors	RoI-Trans. (Ding et al. 2019)	88.64	78.52	43.44	75.92	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
	SCRDet (Yang et al. 2019)	89.98	80.65	52.09	68.36	68.36	60.32	72.41	90.85	**87.94**	86.86	65.02	66.68	66.25	68.24	65.21	72.61
	FRED (Lee et al. 2024)	89.37	82.12	50.84	73.89	77.58	77.38	87.51	90.82	86.30	84.25	62.54	65.10	72.65	69.55	63.41	75.56
	CSL (Yang and Yan 2020)	**90.25**	85.53	54.64	75.31	70.44	73.51	77.62	90.84	86.15	86.69	69.60	68.04	73.83	71.10	68.93	76.17
	RSDet (Qian et al. 2021)	89.93	84.45	53.77	74.35	71.52	78.31	78.12	**91.14**	87.35	86.93	65.64	65.17	75.35	**79.74**	63.31	76.34
	COBB (Xiao et al. 2024)	89.52	84.98	54.99	72.16	77.71	82.81	88.10	90.81	85.45	85.62	63.89	66.15	76.64	70.13	59.05	76.53
	SCRDet++ (Yang et al. 2022)	90.05	84.39	55.44	73.99	77.54	71.11	86.05	90.67	87.32	**87.08**	69.62	68.90	73.74	71.29	65.08	76.81
	AProNet (Zheng et al. 2021)	88.77	84.95	55.27	78.40	76.65	78.54	88.45	90.83	86.56	87.01	65.62	70.29	75.43	78.17	67.28	78.16
	HERO-Det	85.71	82.41	**60.05**	75.46	80.88	**85.51**	88.62	90.56	84.02	85.26	69.30	**74.85**	78.53	79.33	73.00	**79.56**

Table 6: Comparision with state-of-the-arts on the DOTA-v1.0 dataset.

Figure 4: Visualization of detections on DOTA dataset.

Task	3D Object detection	Image Segmentation
Model	GCIoU (Ming et al. 2023)	U-Net (Zhang, Liu, and Wang 2018)
Dataset	KITTI (Geiger et al. 2013)	Pancreatic Tumor (Ming and Xiao 2024)
Metric	Moderate AP (%)	Dice (%)
Baseline	78.5	67.6
+ HCTConv	$78.9^{(\uparrow 0.4)}$	$68.7^{(\uparrow 1.1)}$

Table 7: Experiments on other vision tasks.

ization results on DOTA are shown in Fig .4. Meanwhile, HERO-Det achieves the mAP of 90.64% on HRSC2016 dataset (Liu et al. 2017), proving its advantage in detecting objects with large aspect ratios. Additionally, it obtains mAP of 90.10% and 80.47% on the SAR image datasets SSDD (Li, Qu, and Shao 2017) and HRSID (Wei et al. 2020), respectively, achieving state-of-the-art performance and showing its robustness across diverse AOOD scenarios.

Extensions to Other Vision Tasks

The locality-preserving property of the Hilbert curve could extends well to other vision tasks, as shown in Tab. 7. By replacing standard convolutions with HCTConv and reconstructing features via Hilbert mapping, we achieve minimal model changes to improve the performance. On 3D object detection benchmark KITTI dataset (Geiger et al. 2013), HCTConv improves the baseline moderate AP by 0.4%, and on medical image segmentation task, it yields a 1.1% Dice gain on the pancreatic tumor dataset (Ming and Xiao 2024). These results demonstrate the broad applicability of the proposed method across vision tasks.

Conclusion

In this paper, we present HERO-Det, a novel two-stage oriented object detector for AOOD task. By leveraging the locality-preserving properties of the Hilbert curve, we design a HCTConv operator and then build a HPformer to improve feature representation with minimal spatial disruption. Furthermore, the OAPH decouples rotation-equivariant and invariant cues, effectively addressing the orientation regression bottleneck in two-stage frameworks. Extensive experiments on multiple AOOD benchmarks and vision tasks demonstrate the superior accuracy and generalizability of our mehtod.

Broader impacts. This study systematically explores Hilbert curve-based locality preservation for AOOD and demonstrates its broad applicability. The approach could generalize well to vision tasks like image segmentation, classification, and 3D vision. It can also be be flexibly extended to architectures involve image serialization such as ViT (Dosovitskiy et al. 2020), vision Mamba (Liu et al. 2024), and Swin Transformer (Liu et al. 2021) for further optimization.

Limitations. Several properties of the Hilbert curve remain underexplored. HCS strategy only considers feature shifting under fixed angles, while modeling equivariant shifts under arbitrary rotations would be more complex in this case.

References

Arsalan, M.; Imran, A.; Shah, S. S.; Malaika, S.; and Iqbal, Z. A. 2024. Deep Learning Based Autonomous Robotic Grasping and Sorting System for Industry 4.0. In *2024 International Conference on Robotics and Automation in Industry (ICRAI)*, 1–6. IEEE.

Bader, M. 2012. *Space-filling curves: an introduction with applications in scientific computing*, volume 9. Springer Science & Business Media.

Chen, H.-L.; and Chang, Y.-I. 2011. All-nearest-neighbors finding based on the Hilbert curve. *Expert Systems with Applications*, 38(6): 7462–7475.

Chen, W.; Zhu, X.; Chen, G.; and Yu, B. 2022. Efficient point cloud analysis using hilbert curve. In *European Conference on Computer Vision*, 730–747. Springer.

Chen, Z.; Chen, K.; Lin, W.; See, J.; Yu, H.; Ke, Y.; and Yang, C. 2020. Piou loss: Towards accurate oriented object detection in complex environments. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, 195–211. Springer.

Ding, J.; Xue, N.; Long, Y.; Xia, G.-S.; and Lu, Q. 2019. Learning RoI transformer for oriented object detection in aerial images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2849–2858.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Geiger, A.; Lenz, P.; Stiller, C.; and Urtasun, R. 2013. Vision meets robotics: The kitti dataset. *The international journal of robotics research*, 32(11): 1231–1237.

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

Han, J.; Ding, J.; Xue, N.; and Xia, G.-S. 2021b. Redet: A rotation-equivariant detector for aerial object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2786–2795.

He, M.; Liao, M.; Yang, Z.; Zhong, H.; Tang, J.; Cheng, W.; Yao, C.; Wang, Y.; and Bai, X. 2021. MOST: A multi-oriented scene text detector with localization refinement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8813–8822.

He, Z.; and Owen, A. B. 2016. Extensible grids: uniform sampling on a space filling curve. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(4): 917–931.

Hilbert, D. 1891. Über die stetige abbildung einer linie aufein flächenstück. *Mathematische Annalen*, 38: 459–460.

Hou, L.; Lu, K.; Xue, J.; and Li, Y. 2022. Shape-adaptive selection and measurement for oriented object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 923–932.

Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Lee, C.; Son, J.; Shon, H.; Jeon, Y.; and Kim, J. 2024. FRED: Towards a full rotation-equivariance in aerial image object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 2883–2891.

Li, J.; Qu, C.; and Shao, J. 2017. Ship detection in SAR images based on an improved faster R-CNN. In *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)*, 1–6. IEEE.

Li, T.; Meng, C.; Xu, H.; and Yu, J. 2024. Hilbert curve projection distance for distribution comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7): 4993–5007.

Li, W.; Chen, Y.; Hu, K.; and Zhu, J. 2022. Oriented reppoints for aerial object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1829–1838.

Liao, M.; Zhu, Z.; Shi, B.; Xia, G.-s.; and Bai, X. 2018. Rotation-sensitive regression for oriented scene text detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5909–5918.

Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125.

Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Jiao, J.; and Liu, Y. 2024. Vmamba: Visual state space model. *Advances in neural information processing systems*, 37: 103031–103063.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.

Liu, Z.; Yuan, L.; Weng, L.; and Yang, Y. 2017. A high resolution optical satellite image dataset for ship recognition and some new baselines. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, volume 2, 324–331.

Ming, Q.; Miao, L.; Ma, Z.; Zhao, L.; Zhou, Z.; Huang, X.; Chen, Y.; and Guo, Y. 2023. Deep dive into gradients: Better optimization for 3d object detection with gradient-corrected iou supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5136–5145.

Ming, Q.; Miao, L.; Zhou, Z.; and Dong, Y. 2021a. CFC-Net: A critical feature capturing network for arbitrary-oriented object detection in remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–14.

Ming, Q.; Miao, L.; Zhou, Z.; Song, J.; and Pizurica, A. 2024. Gradient Calibration Loss for Fast and Accurate Oriented Bounding Box Regression. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–15.

Ming, Q.; and Xiao, X. 2024. Towards Accurate Medical Image Segmentation With Gradient-Optimized Dice Loss. *IEEE Signal Processing Letters*, 31: 191–195.

Ming, Q.; Zhou, Z.; Miao, L.; Zhang, H.; and Li, L. 2021b. Dynamic anchor learning for arbitrary-oriented object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2355–2363.

Morton, G. M. 1966. A computer oriented geodetic data base and a new technique in file sequencing.

Peano, G.; and Peano, G. 1990. *Sur une courbe, qui remplit toute une aire plane*. Springer.

Qian, W.; Yang, X.; Peng, S.; Yan, J.; and Guo, Y. 2021. Learning modulated loss for rotated object detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 2458–2466.

Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.

Sagan, H. 1994. Hilbert's space-filling curve. In *Space-filling curves*, 9–30. Springer.

Sheng, H.; Cai, S.; Zhao, N.; Deng, B.; Huang, J.; Hua, X.-S.; Zhao, M.-J.; and Lee, G. H. 2022. Rethinking IoU-based optimization for single-stage 3D object detection. In *European Conference on Computer Vision*, 544–561. Springer.

Sheng, T.; Chen, J.; and Lian, Z. 2021. Centripetaltext: An efficient text instance representation for scene text detection. *Advances in Neural Information Processing Systems*, 34: 335–346.

Wang, J.; and Shan, J. 2005. Space filling curve based point clouds index. In *Proceedings of the 8th International Conference on GeoComputation*, 551–562.

Wang, L.; Li, J.; Zhang, J.; Zhuo, L.; and Tian, Q. 2025. Position Guided Dynamic Receptive Field Network: A Small Object Detection Friendly to Optical and SAR Images. *IEEE Transactions on Circuits and Systems for Video Technology*.

Wang, L.; Xu, N.; and Song, J. 2021. Decoding intra-tumoral spatial heterogeneity on radiological images using the Hilbert curve. *Insights into Imaging*, 12: 1–10.

Wang, L.; Zhang, J.; Tian, J.; Li, J.; Zhuo, L.; and Tian, Q. 2023. Efficient fine-grained object recognition in high-resolution remote sensing images from knowledge distillation to filter grafting. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–16.

Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; and Shi, J. 2020. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation. *Ieee Access*, 8: 120234–120254.

Xia, G.-S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; and 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*, 3974–3983.

Xiao, Z.; Yang, G.; Yang, X.; Mu, T.; Yan, J.; and Hu, S. 2024. Theoretically achieving continuous representation of oriented bounding boxes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16912–16922.

Xie, X.; Cheng, G.; Wang, J.; Yao, X.; and Han, J. 2021. Oriented R-CNN for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, 3520–3529.

Yang, C.; Chen, Z.; Espinosa, M.; Ericsson, L.; Wang, Z.; Liu, J.; and Crowley, E. J. 2024. PlainMamba: Improving Non-Hierarchical Mamba in Visual Recognition. In *35th British Machine Vision Conference 2024, BMVC 2024, Glasgow, UK, November 25-28, 2024*. BMVA.

Yang, X.; Hou, L.; Zhou, Y.; Wang, W.; and Yan, J. 2021a. Dense label encoding for boundary discontinuity free rotation detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 15819–15829.

Yang, X.; and Yan, J. 2020. Arbitrary-oriented object detection with circular smooth label. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, 677–694. Springer.

Yang, X.; and Yan, J. 2022. On the arbitrary-oriented object detection: Classification based approaches revisited. *International Journal of Computer Vision*, 130(5): 1340–1365.

Yang, X.; Yan, J.; Feng, Z.; and He, T. 2021b. R3det: Refined single-stage detector with feature refinement for rotating object. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 3163–3171.

Yang, X.; Yan, J.; Liao, W.; Yang, X.; Tang, J.; and 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.; Yan, J.; Ming, Q.; Wang, W.; Zhang, X.; and Tian, Q. 2021c. Rethinking rotated object detection with gaussian wasserstein distance loss. In *International Conference on Machine Learning*, 11830–11841. PMLR.

Yang, X.; Yang, J.; Yan, J.; Zhang, Y.; Zhang, T.; Guo, Z.; Sun, X.; and Fu, K. 2019. Scrdet: Towards more robust detection for small, cluttered and rotated objects. In *Proceedings of the IEEE International Conference on Computer Vision*, 8232–8241.

Yang, X.; Yang, X.; Yang, J.; Ming, Q.; Wang, W.; Tian, Q.; and Yan, J. 2021d. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence. *Advances in Neural Information Processing Systems*, 34: 18381–18394.

Zhang, J.; Sclaroff, S.; Lin, Z.; Shen, X.; Price, B.; and Mech, R. 2015. Minimum barrier salient object detection at 80 fps. In *Proceedings of the IEEE international conference on computer vision*, 1404–1412.

Zhang, Z.; Liu, Q.; and Wang, Y. 2018. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5): 749–753.

Zheng, X.; Zhang, W.; Huan, L.; Gong, J.; and Zhang, H. 2021. AProNet: Detecting objects with precise orientation from aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181: 99–112.

Zhou, Y.; Ye, Q.; Qiu, Q.; and Jiao, J. 2017. Oriented response networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 519–528.