Chinese Journal of Electronics
vol. 35, no. 4, pp. 1380–1392, July 2026
https://doi.org/10.23919/cje.2025.00.240

Special Focus: Advanced Interpretation Models for Remote Sensing Data and Its Applications
RESEARCH ARTICLE

Space-Filling Curve-Based Fractional Fourier Feature Transformation for Oriented Object Detection in Optical and SAR Remote Sensing Imagery

Qi Ming[1], Liuqian Wang[2], Juan Fang[1], Yifan Xiao[3], Zhixin Guo[3], and Xudong Zhao[4]

1. *College of Computer Science, Beijing University of Technology, Beijing 100124, China*
2. *School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China*
3. *System Design Institute of Hubei Aerospace Technology Academy, Wuhan 430040, China*
4. *School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China*

Corresponding author: Xudong Zhao, E-mail: zhaoxudong@bit.edu.cn
Manuscript Received June 22, 2025; Accepted February 10, 2026; Published Online March 3, 2026

Abstract — In recent years, oriented object detection in remote sensing images has received increasing attention due to its wide range of application scenarios. However, in complex scenes, object features are often hard to distinguish, making accurate object detection in this case still a challenging task. In this paper, we propose a space-filling curve-based fractional fourier feature transformation detector (SF^{3}Det) to achieve high-precision oriented object detection in remote sensing images. First, we propose a fractional fourier transform based multi-domain features extraction module. By utilizing amplitude and phase information, the module can provide discriminative and enhanced features. Then, we adopt a locality-preserving feature serialization method for regional proposals in two-stage detectors. Specifically, object features are unfolded along a space-filling curve to obtain feature sequences with local correlation, which helps optimize model convergence and provides efficient local semantic cues for subsequent prediction. Finally, we apply a joint consistency-aware loss to dynamically adjust the optimization of classification and regression losses, consistently improving the performance of both tasks. Extensive experiments on optical remote sensing datasets and synthetic aperture radar image datasets demonstrate the superiority of SF^{3}Det. Our method achieves mAPs of 90.6%, 79.92%, 89.90%, and 80.40% on the HRSC2016, DOTA, SSDD, and HRSID datasets, respectively.

Keywords — Oriented object detection, Space-filling curve, Fractional Fourier transform, Remote sensing imager.

Citation — Qi Ming, Liuqian Wang, Juan Fang, *et al.*, "Space-filling curve-based fractional fourier feature transformation for oriented object detection in optical and SAR remote sensing imagery," *Chinese Journal of Electronics*, vol. 35, no. 4, pp. 1380–1392, 2026. doi: 10.23919/cje.2025.00.240.

I. Introduction

With the increasing of high-resolution remote sensing data, object detection in remote sensing images has become an essential task in the field of geospatial information analysis. Remote sensing object detection plays a critical role in various applications such as urban planning, military reconnaissance, and disaster response. In particular, the emergence of deep learning techniques, convolutional neural networks (CNNs), has made great progress in the field of computer vision. The CNN-based methods have demonstrated superior performance in extracting robust features and handling large-scale image data, leading to the impressive results in both accuracy and efficiency [1]–[10].

Object detection aims to determine the locations and categories of objects of interest in the images. The task is the basic for downstream applications. General object detection frameworks achieved remarkable success and were widely adopted in various computer

Guest Editor: Prof. Hengchao Li, Southwest Jiaotong University.

vision tasks [11]–[16]. Classical detectors such as faster R-CNN [11], YOLO [12], and FPN [13] have demonstrated strong performance in these tasks. However, the characteristics of objects in remote sensing imagery are significantly different from those in natural scenes. For example, remote sensing object detection often suffers from challenges such as complex backgrounds, large variations in object scale, and the prediction issues in arbitrary-oriented bounding boxes. To address these issues, current detectors commonly adopt oriented bounding boxes (OBB) to better represent rotated object in remote sensing imagery. A series of methods have been proposed from different steps of the detection pipeline (such as object representation [17]–[19], label assignment [20]–[22], feature extraction [23]–[25], and loss function optimization [26], [27]) to achieve accurate oriented object detection.

Despite the progress made in this area, there are still several issues remain unresolved. First, the fixed-basis signal analysis methods are unable to extract localized spatial-frequency structures, and usually overlook phase information in the detection task. Second, the Bird's-eye view of remote sensing images typically contains limited texture and semantic information, especially for small objects. In this case, it is difficult for detectors to distinguish the object via the low-quality features. Third, the object detection task consists of two different subtasks: classification and localization. However, a high classification score does not always correspond to an accurate predicted box, and therefore high-confidence predictions may not output the precise detection results. Moreover, remote sensing images contain complex spatial-spectral structures with diverse textures, scales, and orientations. The fractional Fourier transform (FrFT) provides a tunable intermediate domain between spatial and frequency representations, enabling adaptive extraction of both local textures and global spectral patterns. This balance makes FrFT particularly effective for remote sensing scenes, and this is also one of insights of our framework.

To address the above issues, we propose a space-filling curve-based fractional fourier feature transformation detector (SF^3Det) for high-precision oriented object detection in remote sensing imagery. Specifically, we first propose a FrFT based multi-domain features extraction module. This module can provide discriminative and enhanced features by utilizing amplitude and phase information. Then, we introduce a locality-preserving feature serialization (LFS) method to enhance local contextual representation of objects. LFS unfolds object features along a space-filling curve to produce feature sequences with both global receptive field and strong local correlation. This promotes effective feature learning and benefits subsequent classification and regression. Finally, to alleviate the inconsistency between the classification and regression subtasks, we employ a joint consistency-aware loss (JCA loss). The JCA loss couples the optimization of classification and localization losses during training, which helps to improve consistency between the two subtasks to ensure that high classification scores could indicate high-quality detection outputs.

In summary, we propose a novel oriented object detector, SF^3Det, with the following key contributions:

i) A FrFT based spatial-frequency joint features extraction (FrFSP) module is proposed to extract multi-domain features. By utilizing amplitude and phase information, the module can provide discriminative and enhanced features for detection tasks.

ii) A LFS method is proposed to establish spatial-contextual relationships using a space-filling curve, which extracts locality-preserving features with robust semantic representation.

iii) We apply a joint consistency-aware loss to align the optimization processes of the classification and regression subtasks, bridging the gap between confidence of classification and evaluation criteria of bounding box regression.

II. Related Work

1. Generic object detection

Object detection serves as a fundamental task in computer vision, aiming to accurately classify and locate objects of interest in the images. It has wide-ranging applications [1], [28]–[29]. Early methods relied on hand-crafted features and sliding window strategies. These approaches were limited by low accuracy and poor generalization to complex scenes. With the rising of CNNs, object detection entered a new era. Generally, deep learning-based methods are commonly categorized into two types: two-stage detectors and one-stage detectors. Two-stage methods, such as the R-CNN series [11], [30], first generate region proposals and then refine detections in the prediction head. In contrast, one-stage detectors directly predict object classes and regress bounding boxes in a dense manner, such as YOLO [12].

In these generic object detectors, they typically involve two subtasks: object classification and bounding box regression. The classification task predicts the category label of the detected object, while the regression task estimates its spatial location and scale. To improve the classification confidence, some methods have been proposed to refine the confidence scores of predictions [31]–[33]. For example, focal loss [33] and generalized focal loss [32] address the extreme class imbalance in dense object detection by down-weighting well-classified examples to focus learning on hard negatives in one stage detectors. On the other hand, to improve regression accuracy, various strategies have been proposed. For example, cascade R-CNN [34] employs a multi-stage architecture with progressively increasing intersetion over union (IoU) thresholds to refining object localization and classification step by step to improve detection quality. There are also many

IoU-based loss functions directly optimize the spatial overlap between predicted and ground truth boxes, leading to more accurate and robust bounding box regression compared to traditional ℓ_n losses [22], [26].

More recently, there are also some detectors have emerged with new perspectives on visual modeling [35]. Carion *et al.* [35] designs the DETR that uses transformer architectures in object detection by modeling global relationships between image patches via self-attention. The DETR discards the anchor boxes and post-processing such as non-maximum suppression to build a novel framework. Subsequent works, such as deformable DETR [36], further improved convergence speed and spatial precision. Simultaneously, the vision transformer (ViT) [37] and its detection variants have demonstrated strong capability in extracting long-range dependencies and fine-grained spatial features. These transformer-based detectors represent a significant shift from traditional CNN architectures, offering new insights for structured prediction in complex vision tasks.

2. Object detection in remote sensing images

Object detection in remote sensing images is a specialized task that aims to locate and classify targets from aerial or satellite images. Compared with natural scenes, remote sensing images often cover wide areas and contain numerous densely packed, small-sized objects with arbitrary orientations. Moreover, objects may exhibit significant variations in scale, aspect ratio, and background complexity [25]. These factors pose unique challenges to detection algorithms. Most current remote sensing object detectors shares the similar frameworks with generic object detection, but there are fundamental differences. Generic detectors are typically designed for objects annotated with horizontal bounding boxes (HBBs). In contrast, remote sensing objects are usually annotated with OBBs to better capture rotated targets (e.g., ships, planes, vehicles).

Recent advances in remote sensing object detection have focused on addressing the unique challenges of detecting arbitrarily oriented, small, and densely packed remote sensing objects. These studies focus on optimizing the detection pipeline such as object representation [17], [38], label assignment strategies [20], [22], feature extraction [23], [24], and loss function design [26], [39]. First, for object representation, gliding vertex [38] improved localization precision by adjusting the positions of bounding box vertices along object contours. In terms of label assignment method, DAL [20] introduced a dynamic anchor learning mechanism to dynamically select anchors based on object geometry and feature responses with a newly proposed metric matching degree. The SLA [22] proposed a sparse label assignment strategy that selects a small set of high-quality anchors based on angle-aware cost metrics, effectively reducing label noise and improving detection of oriented objects. Feature extraction is crucial for CNN-based framework, and therefore many researches focus on building robust feature representation for oriented objects in remote sensing images. For example, S^2A-Net [23] proposed a two-step alignment mechanism to improve anchor-target alignment procedure. Specifically, a feature alignment module is applied to align the regional features according to the anchor area, and then an oriented detection module is adopted for orientation-sensitive prediction. ReDet [24] utilizes a rotation-equivariant backbone to enhance feature consistency under varying object orientations. For efficient model training, some novel loss functions have been proposed to address the angular periodicity and loss oscillation issues. The GWD [19] and KLD [27] models OBBS as two-dimensional (2D) Gaussian distributions and computes their similarity to avoid angular periodicity issue. The GCL [26] calibrated the gradient contributions of object of different sizes during training, effectively stabilizing optimization and improving the detection performance.

Recently, transformer-based frameworks have gained increasing attention in remote sensing object detection. Oriented DETR [40] adapts the DETR framework and modified the object query and prediction heads to regress rotated bounding boxes, demonstrating the feasibility of end-to-end oriented detection. OrientedFormer introduced Gaussian positional encoding for angle and size, and proposed a Wasserstein self-attention to capture geometric relations [41]. The ARS-DETR [42] enhanced angle regression for oriented objects through an angle-aware circular smooth label mechanism, a rotated deformable attention, and an aspect-ratio-weighted angle loss, thereby achieving competitive performance under stricter IoU thresholds.

III. Methodology

The overall architecture of our proposed SF^3Det is illustrated in Figure 1. In the first stage, SF^3Det employs a backbone network with a feature pyramid network (FPN) [13] for multi-scale feature extraction. Subsequently, the FrFSP module is utilized to obtain discriminative and enhanced multi-domain features. Then, region-of-interest (RoI) features are extracted via the RoI Align operation. Next, a LFS module is designed to generate robust feature representations for the subsequent prediction heads. During the training process, a JCA loss is introduced to enforce consistency optimization between the classification and regression subtasks.

1. Fractional fourier spatial-phase feature extraction

The fixed-basis analysis methods are unable to extract localized time-frequency structures. Thus, the FrFT is used for localized feature representation [1], [43], [44]. By utilizing amplitude and phase information, the multi-domain features can provide discriminative and enhanced features for the detection task. In this part, we designed a FrFSP feature extraction block as part A of Figure 2 shows, which is a transform-domain filter

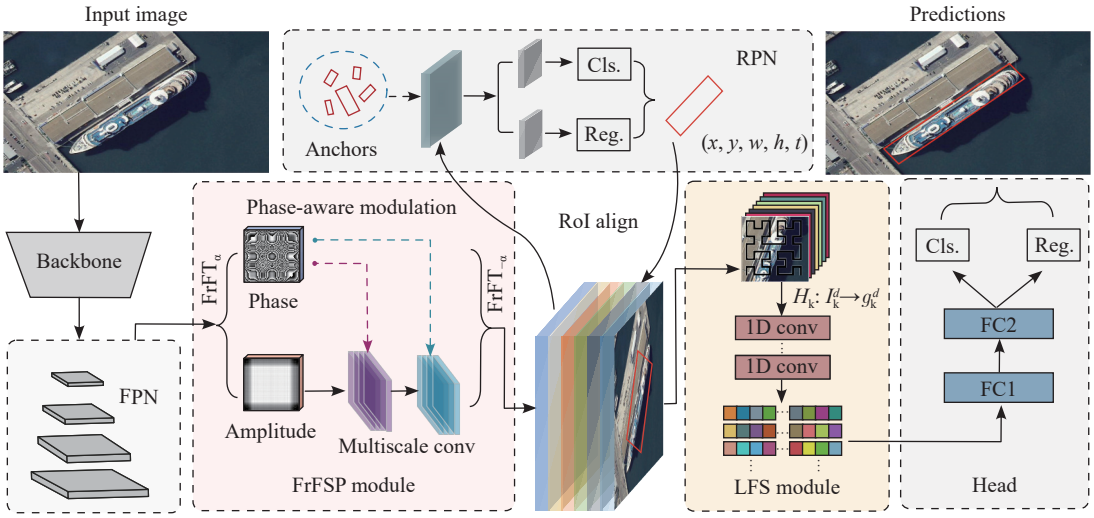


Figure 1 The architecture of the proposed SF^{3}Det.

and provide magnitude and phase features jointly.

The FrFT of a signal $f(x)$ is defined in an integral kernel form as

$$\mathcal{F}_\alpha[f](u) = \int_{-\infty}^{\infty} K_\alpha(u,x) f(x) \mathrm{d}x \tag{1}$$

where the kernel function is

$$K_\alpha(u,x) = \sqrt{\frac{1 - i\cot\alpha}{2\pi}} \exp\left(i\pi \frac{(u^2+x^2)\cos\alpha - 2ux}{\sin\alpha} \right) \tag{2}$$

Firstly, the FrFT is applied to the prepared FPN signals $\mathcal{X}^s \in \mathbb{R}^{C\times H\times W}$, and the fractional-domain features after applying FrFT are

$$X_\alpha = \mathcal{F}_\alpha(\mathcal{X}^s) = M_\alpha \odot e^{j\phi_\alpha} \tag{3}$$

where $\mathcal{F}_\alpha$ represents the 2D α-order FrFT operator, which is applied to each spatial slice $(H \times W)$ for every channel and batch. The $M_\alpha = |X_\alpha|$ is the magnitude while $\phi_\alpha = \arg(X_\alpha)$ is the phase. The $\odot$ is the Hadamard product.

For the magnitude part, two Conv blocks with different window sizes $T_1 = \mathcal{C}_{5\times5}(M_\alpha)$ and $T_2 = \mathcal{C}_{7\times7}(T_1)$ are applied, where $\mathcal{C}_{k\times k}$ denotes convolutional layers with $k \times k$ kernels.

For the phase part, a phase-aware weight module

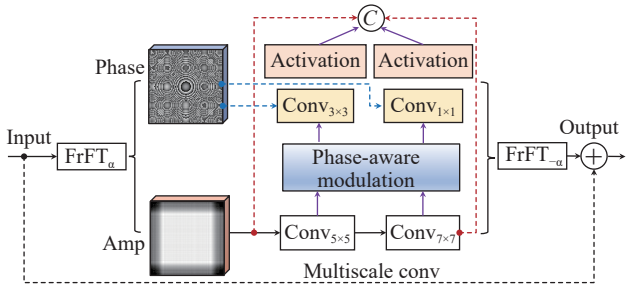


Figure 2 The pipeline of FrFSP module.

is designed to extract the multi-domain phase feature $\phi_\alpha \in \mathbb{R}^{C\times H\times W}$. Phase feature is related to the structural details of input data. The phase-aware layes are denoted as

$$\phi_{\text{out}} = \sigma(\mathcal{C}_{1\times1}(\sigma(\mathcal{C}_{3\times3}(\phi_\alpha)))) \in \mathbb{R}^{2C\times H\times W} \tag{4}$$

The output ϕ_{out} is then split along the channel dimension into two tensors, ϕ_1 and ϕ_2, each of size $\mathbb{R}^{C\times H\times W}$. The σ is the activation function.

Then, the multiple fractional phase-aware feature $\mathcal{X}^\phi \in \mathbb{R}^{C\times H\times W}$ is

$$\mathcal{X}^\phi = \phi_1 \odot T_1 + \phi_2 \odot T_2 \tag{5}$$

The motivations of designing the FrFSP module include 1) an adaptive fractional order parameters α enabling frequency feature selection, and 2) a phase-aware weighting module for optimal multi-domain feature fusion.

2. Locality-preserving feature serialization

Space-filling curves are continuous mappings from a one-dimensional (1D) interval to a multi-dimensional space that densely cover the target space [45]. The Hilbert curve [46] is a special space-filling curve which is known for its superior locality-preserving properties. Let $\mathcal{H}_k : \mathcal{I}_k \to \mathcal{G}_k^d$ denote a discrete Hilbert curve at recursion level k, which maps a 1D index to a d-dimensional grid point. A key mathematical property is that the Hilbert curve preserves spatial locality under both forward and (approximately) inverse mappings, i.e., for any two adjacent points $x, y \in \mathcal{I}_k$, their images $\mathcal{H}_k(x), \mathcal{H}_k(y)$ are also adjacent or nearby in $\mathcal{G}_k^d$. Conversely, under bounded distance conditions, the inverse mapping $\mathcal{H}_k^{-1}$ satisfies the inequality as

$$\left|\mathcal{H}_k^{-1}(Q_1) - \mathcal{H}_k^{-1}(Q_2)\right| \le \phi \cdot \|Q_1 - Q_2\|^d \tag{6}$$

where $\phi > 0$ and Q_1, Q_2 are adjacent spatial positions on the discrete grid. This locality-preserving behavior is

crucial when unfolding high-dimensional feature maps into 1D sequences, as it ensures the preservation of neighborhood relations and spatial coherence in the serialized representation. Compared to raster scan [47] or Z-order curve [48], the Hilbert curve achieves significantly better performance in terms of minimizing the maximum distance between neighbors in the original space and the serialized order, making it particularly advantageous for the convolutional operations after flattening.

As depicted in Figure 3, raster scanning and continuous scanning are simple but show poor spatial coherence between neighboring pixels. This limitation stems from their implementation as naive 1D traversal methodologies. In contrast, space-filling curves intrinsically preserve local spatial relationships during 2D image traversal, such as Hilbert curve Figure 3(e). Such geometric preservation endows the features with superior locality-preserving properties which is essential for vision tasks.

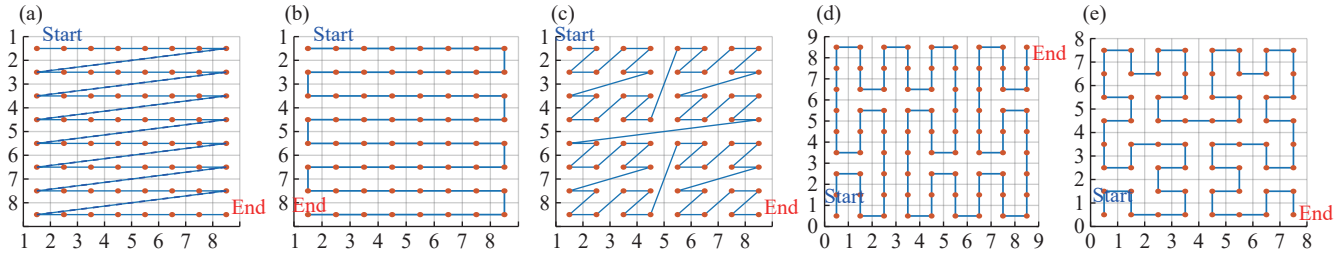


Figure 3 Comparison of different feature scanning paths. (a) Raster scanning; (b) Continuous scanning; (c) Z-order curve [48]-order curve; (d) Peano curve; (e) Hilbert curve.

Therefore, we incorporate the Hilbert curve as a special case into our proposed LFS module to enhance the spatial representation of region features. Given an RoI-aligned feature map $\boldsymbol{F} \in \mathbb{R}^{w\times h\times c}$, we first apply a discrete Hilbert curve $\mathcal{H}_k : \{0, 1, \ldots, n-1\} \to \{(x_i, y_i)\}_{i=0}^{n-1}$ to serialize the 2D spatial grid (w, h) into a 1D sequence of $n = w \cdot h$ positions, while preserving local neighborhood structure. The serialized feature $\boldsymbol{S} \in \mathbb{R}^{n\times c}$ is defined as

$$\boldsymbol{S}_i = \boldsymbol{F}\left(\mathcal{H}_k(i), :\right) = \boldsymbol{F}(x_i, y_i, :), \text{for } i = 0, 1, \ldots, n-1 \tag{7}$$

This step rearranges the spatial dimensions into a 1D sequence while maintaining spatial coherence. Next, a 1D convolutional layer $\mathcal{C}_1$ is applied to $\boldsymbol{S}$ to capture contextual dependencies in the serialized domain as

$$\boldsymbol{S}' = \mathcal{C}_1(\boldsymbol{S}), \boldsymbol{S}' \in \mathbb{R}^{n\times c'} \tag{8}$$

The output $\boldsymbol{S}'$ is subsequently aggregated using global average pooling or flattening, followed by a multi-layer perceptron (MLP) to generate the final representation $\boldsymbol{z}$ as

$$\boldsymbol{z} = \text{MLP}(\text{Flatten}(\boldsymbol{S}')) \tag{9}$$

The resulting vector $\boldsymbol{z}$ is passed into task-specific heads for classification and regression as

$$\hat{y}_{\text{cls}} = \mathcal{F}_{\text{cls}}(\boldsymbol{z}), \quad \hat{y}_{\text{reg}} = \mathcal{F}_{\text{reg}}(\boldsymbol{z}) \tag{10}$$

where $\mathcal{F}_{\text{cls}}$ and $\mathcal{F}_{\text{reg}}$ denote the classification and regression branches, respectively. This process ensures that the feature representation is both compact and locality-aware, which is particularly beneficial when the receptive field is limited and small object features need to be preserved.

The advantages of using the LFS module are threefold. First, intuitively, by adopting a locality-preserving Hilbert curve, the spatial coherence among neighboring pixels is retained even after serialization. This is critical to maintain meaningful spatial structures subsequent prediction. Second, the 1D convolution applied over the serialized sequence helps to effective build long-range dependencies with significantly reduced computational overhead compared to full 2D convolutions. Third, the LFS module provides a flexible plug-in component that can be seamlessly integrated into existing two-stage detection frameworks. It enhances the robustness of feature representations, especially for small or densely packed objects where fine-grained spatial cues are required. Furthermore, LFS requires no additional learnable parameters for position modeling, thereby avoiding potential overfitting risk. In conclusion, the LFS module offers a principled and efficient approach to improving RoI-level feature representation while maintaining strong spatial locality.

3. Joint consistency-aware loss

Current object detectors typically treat classification and regression as independent tasks, optimizing them separately via cross-entropy loss and regression loss (e.g., Smooth L_1 loss [30], IoU loss [26]). However, the independent learning of the two sutasks may lead to inconsistent predictions. For instance, a predicted bounding box might achieve a high classification score while having poor localization accuracy, or vice versa. These inconsistency would degrade the final detection performance after non-maximum suppression (NMS).

To illustrate this, we first review the conventional

two-stage detection loss as

$$\mathcal{L} = \frac{1}{N}\Bigg(\sum_{j\in N} \mathrm{CE}(p_j, y_j) + \sum_{i\in N_{\mathrm{pos}}} \Big(\mathcal{L}_{\mathrm{reg}}(b_i, \hat{b}_i) + \mathcal{L}_{\mathrm{rpn}}(r_i, \hat{b}_i) \Big) \Bigg) \tag{11}$$

where p_i and y_i denote the predicted classification score and ground truth label, respectively. The b_i and $\hat{b}_i$ are the regression output and its ground truth from the detection head, while r_i denote proposals from the RPN. The N is the all candidate samples, and N_{pos} denotes the positives. The $\mathrm{CE}(\cdot)$ is the cross-entropy loss and $\mathcal{L}_{\mathrm{reg}}(\cdot)$ is the Smooth L_1 loss used for oriented bounding box regression.

To mitigate the inconsistency between classification and regression tasks, inspird by harmonic loss [49] for one-stage detectors, we design a joint consistency-aware (JCA) loss for two stage detectors, which incorporates mutual guidance between classification and regression tasks using adaptive modulation weights. Specifically, for each output box predicted from a positive sample as

$$\mathcal{L}^i_{\mathrm{JCA}} = (1+\lambda_{\mathrm{r}}) \cdot \mathrm{CE}(p_i, y_i) + (1+\lambda_{\mathrm{c}}) \cdot \Big(\mathcal{L}_{\mathrm{rpn}}(r_i, \hat{b}_i) + \mathcal{L}_{\mathrm{reg}}(b_i, \hat{b}_i) \Big) \tag{12}$$

where the modulation coefficients λ_{r} and λ_{c} are dynamically adjusted according to the current state of the regression and classification branches as

$$\begin{aligned} \lambda_{\mathrm{r}} &= e^{-\left(\mathcal{L}_{\mathrm{rpn}}(r_i,\hat{b}_i)+\mathcal{L}_{\mathrm{reg}}(b_i,\hat{b}_i)\right)}, \\ \lambda_{\mathrm{c}} &= e^{-\mathrm{CE}(p_i,y_i)} \end{aligned} \tag{13}$$

The JCA loss introduces a symmetric supervision mechanism. When the regression quality is poor, indicated by large values of $\mathcal{L}_{\mathrm{reg}}$ and $\mathcal{L}_{\mathrm{rpn}}$, the classification loss is down-weighted through a smaller λ_{r}. This adjustment encourages the network to focus on improving the regression branch first. Conversely, if the classification confidence is low, the regression branch is down-weighted by λ_{c}, promoting the network to enhance class representation learning initially. In cases where both branches perform well, the values of λ_{r} and λ_{c} are both high, providing strong and positive supervision signals. However, for hard positive samples where both losses are large, both weights become small, which helps prevent the model from overfitting to noisy.

To better understand how the JCA loss affects the training process, we first analyze the gradient of $\mathcal{L}^i_{\mathrm{JCA}}$ with respect to p_i as

$$\frac{\partial \mathcal{L}^i_{\mathrm{JCA}}}{\partial p_i} = (1+\lambda_{\mathrm{r}}) \cdot \frac{\partial \mathrm{CE}(p_i, y_i)}{\partial p_i} + \left(\mathcal{L}_{\mathrm{rpn}} + \mathcal{L}_{\mathrm{reg}}\right) \cdot \frac{\partial \lambda_{\mathrm{c}}}{\partial p_i} \tag{14}$$

Given that $\mathrm{CE}(p_i, y_i) = -\log(p_i)$, it follows that $\lambda_{\mathrm{c}} = e^{\log(p_i)} = p_i$. Thus,

$$\frac{\partial \lambda_{\mathrm{c}}}{\partial p_i} = 1, \quad \frac{\partial \mathrm{CE}(p_i, y_i)}{\partial p_i} = -\frac{1}{p_i}$$

Substituting these back gives as

$$\frac{\partial \mathcal{L}^i_{\mathrm{JCA}}}{\partial p_i} = -\frac{1+\lambda_{\mathrm{r}}}{p_i} + \left(\mathcal{L}_{\mathrm{rpn}} + \mathcal{L}_{\mathrm{reg}}\right) \tag{15}$$

This expression indicates that the classification gradient depends not only on the classification error but is also modulated by the regression loss. Specifically, when the regression loss $(\mathcal{L}_{\mathrm{rpn}} + \mathcal{L}_{\mathrm{reg}})$ is large, the absolute value of the gradient on p_i decreases, which suppresses the confidence scores for poorly localized predictions.

Then, we analyze the gradient of the JCA loss with respect to the regression prediction. Take the final prediction $\hat{b}_i$ for example, the regression loss is $(1+\lambda_{\mathrm{c}}) \cdot \mathcal{L}_{\mathrm{reg}}(b_i, \hat{b}_i)$, where $\lambda_{\mathrm{c}} = e^{-\mathrm{CE}(p_i, y_i)}$ depends on the classification error. The gradient of the loss with respect to $\hat{b}_i$ is

$$\frac{\partial \mathcal{L}^i_{\mathrm{JCA}}}{\partial \hat{b}_i} = (1+\lambda_{\mathrm{c}}) \cdot \frac{\partial \mathcal{L}_{\mathrm{reg}}(b_i, \hat{b}_i)}{\partial \hat{b}_i} + \mathcal{L}_{\mathrm{reg}}(b_i, \hat{b}_i) \cdot \frac{\partial \lambda_{\mathrm{c}}}{\partial \hat{b}_i} \tag{16}$$

Since λ_{c} only depends on the classification loss and does not depend on $\hat{b}_i$, we have

$$\frac{\partial \lambda_{\mathrm{c}}}{\partial \hat{b}_i} = 0 \tag{17}$$

Therefore, the gradient simplifies to

$$\frac{\partial \mathcal{L}^i_{\mathrm{JCA}}}{\partial \hat{b}_i} = (1+\lambda_{\mathrm{c}}) \cdot \frac{\partial \mathcal{L}_{\mathrm{reg}}(b_i, \hat{b}_i)}{\partial \hat{b}_i} \tag{18}$$

It indicates that the regression loss gradient is modulated by the factor $(1+\lambda_{\mathrm{c}})$. When the classification error is high, λ_{c} is small, thus reducing the regression gradient to prevent overfitting to poorly classified samples. Conversely, when classification error is low, the regression gradient is amplified, encouraging the model to focus on bounding box regression.

By incorporating regression-aware classification supervision and classification-aware regression supervision within a unified joint loss, the JCA loss effectively mitigates the training imbalance between classification and regression tasks.

IV. Expriments

1. Implementation details

All experiments were conducted on high-performance computing nodes with two NVIDIA GeForce RTX 4090 GPUs (24 GB GDDR6X VRAM for each). The proposed method were implemented with PyTorch 2.1.0 compiled against CUDA 12.1. We used stochastic gradient descent as optimizer to train the model with

momentum set to 0.9. Initial learning rate was set to 0.02, decaying by an order of magnitude at 50% and 75% of the total training duration. The batch size is set to 8. We used the backbone pretrained on ImageNet [50] in all experiments. Data augmentation was applied during training for better performance. Our adopted augmentation strategies include: random flipping, random rotation, HSV color space perturbation, and stochastic multi-scale resizing. For DOTA dataset, we employed test-time multi-scale augmentation with three-scales $\{0.5, 1.0, 1.5\}$.

In our experiments, we use average precision $(\mathrm{AP})_{50}$ and AP_{75} for performance evaluation. The AP_{50} and AP_{75} denote the AP at IoU thresholds of 0.50 and 0.75, respectively. The AP_{50} reflects the model's general detection ability, while AP_{75} emphasizes precise localization results.

We used four datasets to verify the effectiveness of the proposed method, including two optical remote sensing image datasets HRSC2016 [51] and DOTA [52], and two synthetic-aperture radar (SAR) image datasets SSDD [53] and HRSID [54].

The HRSC2016 is the benchmark for ship recognition in high-resolution satellite imagery. It contains 1061 sub-meter resolution images acquired by Google Earth. The dataset adopts hierarchical labeling with three class tiers: background, generic ships, and specialized warship subtypes. Official partitions include 436 training images, 181 validation images, and 444 test images. We train the model for 72 epochs on HRSC2016 dataset.

The DOTA is the most commonly used public aerial image dataset for oriented object detection. It contains 2806 images ($\sim 4000 \times 4000$ pixels) collected from Google Earth and JL-1 satellites. There are 188,282 rotated bounding box annotations across 15 categories: plane (PL), baseball diamond (BD), bridge (BR), ground track field (GTF), small vehicle (SV), large vehicle (LV), ship (SH), tennis court (TC), basketball court (BC), storage tank (ST), soccer ball field (SBF), roundabout (RA), harbor (HA), swimming pool (SP), and helicopter (HC). We train the model for 12 epochs on DOTA dataset.

The SSDD is a SAR dataset specialized for ship detection research. It contains 1160 SAR images from Sentinel-1, TerraSAR-X, and RadarSat-2 platforms. The dataset contains more than 2000 ship targets with bounding box annotations. We train the model for 72 epochs on SSDD dataset.

The HRSID is a popular benchmark for high-resolution SAR ship detection. The dataset contains 5604 images and 16951 ships. We train the model for 72 epochs on HRSID dataset.

2. Ablation study

i) Component-wise ablations

To validate the effectiveness of the proposed modules, we conducted ablation studies on the individual components of SF^{3}Det. As shown in Table 1, we report AP_{50} and AP_{75} to respectively reflect general detection performance and high-precision object localization capability. Specifically, the proposed FrFSP module achieves 83.5% by extracting both amplitude and phase feature in fractional frequency domains, leading to 0.8% AP_{50} gains over the baseline. On this basis, we further incorporate the LFS module to introduce spatially semantic local feature correlation, which boosts AP_{50} by an additional 5.2%. Finally, the application of the JCA loss enhances the consistency between the classification and regression branches, ensuring that high classification confidence corresponds to accurate localization. In this way, JCA loss yields a 1.2% improvement in AP_{75}. These results demonstrate that proposed modules contribute to consistent performance improvements.

Table 1 Experiments on different modules of the proposed SF^{3}Det

Models			Metrics	
FrFSP module	LFS module	JCA loss	AP_{50}(%)	AP_{75}(%)
			82.7	69.5
✓			83.5	69.6
	✓		87.1	69.5
✓	✓		88.7	69.9
✓	✓	✓	**89.1**	**71.1**

Note: The bold fonts indicate the best performances.

ii) Effect of fractional fourier spatial-phase feature extraction module

As shown in Table 2, we evaluate the FrFSP module's performance under varying fractional orders. The table illustrates that the module achieves its best AP_{50} of 83.5% at 0.4 fractional order and its best AP_{75} of 70.3% at 0.6 fractional order. The spatial domain achieves AP_{50} values of 80.7% and the Fourier domain with order 1.0 achieves 80.8%. Compared to these traditional transforms, the FrFSP performs better by providing multi-domain features that leverage both amplitude and phase information. The adaptive ability of the FrFSP module allows it for selecting optimal fractional domain feature, which contribute to the improvement of detection performance.

Table 2 Effects of different settings in FrFSP module

Order	AP_{50}(%)	AP_{75}(%)
0.0 (Spatial domain)	80.7	69.1
0.2	80.8	69.6
0.4	**83.5**	69.6
0.5	83.3	69.5
0.6	82.8	**70.3**
0.8	80.9	69.7
1.0 (Frequency domain)	80.8	69.6

Note: The bold fonts indicate the best performances.

iii) Effect of locality-preserving feature serialization

To further discuss the performance of the LFS module, we conducted experiments to evaluate the impact of different serialization strategies. The results are presented in Table 3. The default kernel size is with $d = 1$. First, compared to traditional scanning methods such as raster scan and continuous scan —which tend to disrupt the spatial correlations of local features—space-filling curves enable spatially coherent feature sequence unfolding, thus demonstrating superior performance. For instance, both the Z-order curve and the Hilbert curve outperform the raster scan in terms of AP_{50} by 1.6%. Furthermore, we explored convolution kernels with varying receptive fields and observed that excessively large receptive fields may actually lead to performance degradation by 0.4%. In general, a larger receptive field is beneficial to capture long-range dependencies in standard convolutions. However, the Hilbert curve provides a feature traversal that inherently preserves local continuity, and a large stride would instead disrupt this local property, significantly weakening the spatial dependencies between adjacent pixels. Therefore, for SFCs that scan along complex paths, a large stride can actually degrade model performance.

Table 3 Effect of different serialization methods in LFS module

Scanning path	Kerner size	AP_{50}(%)	AP_{75}(%)
Raster scan [47]	1*9	85.5	69.3
Continuous scan	1*9	85.8	69.2
Z-order curve [48]	1*9	87.0	67.6
Hilbert curve [46]	1*9	**87.1**	**69.5**
	1*11	85.4	69.5
	1*3 (d=2)	85.6	67.7
	1*9 (d=2)	86.7	69.3

Note: The bold fonts indicate the best performances.

We can also analyze results from the view of spatial continuity. Different space-filling curves exhibit distinct abilities to preserve spatial locality, which critically influences serialized feature coherence. Let $\mathcal{C}$ denote a curve mapping $\mathcal{C} : \mathbb{Z}^2 \to \mathbb{R}$. The locality preservation can be quantified as $\mathcal{L}(\mathcal{C}) = \frac{1}{N}\sum_{i=1}^{N-1} \|p_{i+1} - p_i\|_2$, where p_i and p_{i+1} are adjacent positions in the traversal order. Empirically, Z-order and raster scans yield large $\mathcal{L}$ due to frequent discontinuities, while continuous and Hilbert curves progressively reduce $\mathcal{L}$ through smoother traversal. The Hilbert curve achieves the minimum $\mathcal{L}$, providing superior locality preservation and consequently stronger feature consistency, which explains its best performance among all curves.

In addition, we explored the performance of using the LFS module to serialize features into sequences and then reconstruct them into 2D local representations for prediction. Typically, the serialization and reconstruction strategies are expected to be consistent for spatially alignment. However, we found that adopting different strategies for serialization and reconstruction can yield the unexpected performance gains. As shown in Table 4, reconstructing features serialized by raster scanning using a Hilbert curve led to increases of 1.6% in AP_{50} and 3.1% in AP_{75}. Furthermore, when features were unfolded using the Z-order curve and reconstructed via the Hilbert curve, AP_{50} was significantly improved by an additional 3.9%. We suggest that for feature blocks in RoI representations, mismatched serialization and reconstruction paths push the hidden layers to learn dependencies between local feature blocks. Such dependencies effectively capture long-range relationships on top of local representations, thereby enhancing the robustness of features and ultimately improving detection performance.

Table 4 Performance evaluation of serialization-reconstruction combinations

Serialization	Reconstruction	AP_{50}(%)	AP_{75}(%)
Raster scan [47]	Continuous scan	82.8	66.3
Raster scan [47]	Z-order curve [48]	84.5	68.6
Raster scan [47]	Hilbert curve [46]	84.4	**69.4**
Z-order curve [48]	Hilbert curve [46]	**88.3**	67.4

Note: The bold fonts indicate the best performances.

3. Main results

i) Results on HRSC2016 dataset

The results on HRSC2016 dataset are shown in Table 5. The HRSC2016 contains numerous ships with large aspect ratios, and therefore the accurate detection of these ship are challenging. As a result, SF^3Det achieves the mAP of 90.60%, surpassing many recent state-of-the-art methods. Visualization of detection results are shown in Figure 4.

Table 5 Comparison with other methods on the HRSC2016 dataset

Methods	Reference	mAP (%)
DAL [20]	AAAI 2021	89.77
GWD [19]	ICML 2021	89.85
OSKDet [55]	CVPR 2022	89.98
TIOE-Det [39]	ISPRS&RS 2023	90.16
S^2A-Net [23]	TGRS 2021	90.17
GCL [26]	TGRS 2024	90.19
SASM [56]	AAAI 2022	90.27
DDDet [14]	TGRS 2024	90.27
O-RepPoints [57]	CVPR 2022	90.38
SGKLD [58]	TGRS 2023	90.40
ReDet [24]	CVPR 2021	90.46
O-RCNN [59]	ICCV 2021	90.50
TSConv [60]	TNNLS 2024	90.59
SF^3Det (Ours)	–	**90.60**

Note: The bold fonts indicate the best performances.

Figure 4 Detections on HRSC2016 dataset.

ii) Results on HRSID dataset and SSDD dataset

The SSDD and HRSID datasets are designed for high-precision oriented object detection in SAR imagery. Challenges such as low image resolution, significant noise, and cluttered backgrounds make precise detection in SAR scenes particularly difficult. As shown in Table 6 SF^{3}Det achieves mAP of 89.90% on SSDD and 80.40% on HRSID, establishing new state-of-the-art benchmarks. Detection examples are illustrated in Figure 5.

iii) Results on DOTA dataset

The DOTA dataset [52] serves as the most prominent large-scale benchmark for aerial oriented object detection in remote sensing images. We evaluate the performance on this benchmark and compare it with many current advanced methods, as presented in Table 7. SF^{3}Det achieves a leading mAP of 79.92%, surpassing numerous state-of-the-art models and highlighting its strong detection capability. Visual examples on the DOTA dataset are shown in Figure 6.

Table 6 Comparison with current methods on SSDD and HRSID datasets

Methods	Reference	mAP(%)	
		SSDD	HRSID
R-Retinanet [33]	ICCV 2017	78.32	68.07
S^{2}ANet [23]	TGRS 2021	89.57	79.44
ReDet [24]	CVPR 2021	90.04	79.84
R^{3}Det [61]	AAAI 2021	81.04	70.92
O-RCNN [59]	ICCV 2021	89.45	71.38
RR-GWD [19]	ICML 2021	78.55	74.43
RR-KLD [27]	NeurIPS 2021	84.18	75.46
SASM [56]	AAAI 2022	88.98	76.09
O-RP [57]	CVPR 2022	89.94	79.70
RR-PSC [62]	CVPR 2023	79.14	67.82
CF-ORNet [63]	TGRS 2023	89.85	74.40
SF^{3}Det (Ours)	–	**89.90**	**80.40**

Note: The bold fonts indicate the best performances.

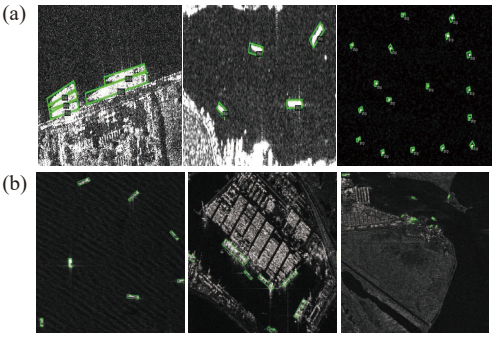


Figure 5 Detection results on the SAR datasets. (a) SSDD; (b) HRSID.

Table 7 Comparision with state-of-the-arts on the DOTA dataset

	Methods	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP(%)
One stage detectors	GCL [26]	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^{3}Det [61]	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 [20]	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 [64]	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 [19]	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. [57]	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 [27]	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^{2}A-Net [23]	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 [56]	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. [65]	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 [66]	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 [67]	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 [18]	**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 [68]	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 [69]	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++ [70]	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 [71]	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
	SF^{3}Det	88.44	83.82	**60.00**	78.95	80.83	**84.69**	88.73	90.53	85.87	86.56	69.16	67.71	**78.98**	**82.51**	72.09	**79.92**

Note: The bold fonts indicate the best performances.

Figure 6 Visualization of detections on DOTA dataset.

V. Conclusion

In this paper, We propose SF^{3}Det, a novel framework that integrates space-filling curve and fractional Fourier feature transformation to addresses the challenges of oriented object detection in remote sensing images. First, a FrFT based multi-domain features extraction module is proposed. By utilizing amplitude and phase information, the module can provide discriminative and enhanced features. Next, a novel locality-preserving feature serialization method is designed to use space-filling curves to build geometrically coherent features with spatial contiguity. Then, a joint consistency-aware loss is applied to address task-misalignment in multi-task learning by establishing dynamic cross-task supervision during training. Comprehensive experiments on remote sensing datasets prove the state-of-the-art performance of the proposed method.

Furthermore, we suggest that the deterministic traversal pattern of SFCs may limit adaptability to irregular or non-convex spatial structures. Their inherent ability to preserve locality and hierarchical continuity makes them particularly effective for structured feature encoding. Therefore, we will explore integrating SFCs with adaptive or data-driven scanning strategies could further enhance their flexibility, offering new perspectives for efficient spatial reasoning in future vision systems.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) (Grant No. 62401049).

References

[1] X. D. Zhao, M. M. Zhang, R. Tao, *et al.*, "Fractional Fourier image transformer for multimodal remote sensing data classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 2, pp. 2314–2326, 2024.

[2] X. D. Zhao, R. Tao, W. Li, *et al.*, "Recent developments in fractional information extraction theory and methods of hyperspectral image," *Acta Electronica Sinica*, vol. 50, no. 12, pp. 2874–2883, 2022.

[3] J. Su, F. Wang, and W. Zhuang, "An improved YOLOv7 tiny algorithm for vehicle and pedestrian detection with occlusion in autonomous driving," *Chinese Journal of Electronics*, vol. 34, no. 1, pp. 282–294, 2025.

[4] R. Zhang, C. Xie, and L. W. Deng, "A fine-grained object detection model for aerial images based on YOLOv5 deep neural network," *Chinese Journal of Electronics*, vol. 32, no. 1, pp. 51–63, 2023.

[5] C. Tang, Y. Q. Xia, and L. H. Dou, "Mobility prediction based tracking of moving objects in wireless sensor networks," *Chinese Journal of Electronics*, vol. 32, no. 4, pp. 793–805, 2023.

[6] P. Sun, Y. B. Zheng, W. Y. Xu, *et al.*, "Completing missing entities: Exploring consistency reasoning for remote sensing object detection," *IEEE Transactions on Image Processing*, vol. 35, pp. 569–584, 2026.

[7] P. H. Duan, S. S. Hu, X. D. Kang, *et al.*, "Shadow removal of hyperspectral remote sensing images with multiexposure fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, article no. 5537211, 2022.

[8] X. D. Zhao, Q. Ming, Y. X. Yang, *et al.*, "Chirplet Fourier analysis network for cross-scene classification of multi-source remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, article no. 553411 4, 2025.

[9] N. Q. Liu, X. Xu, Y. Y. Su, *et al.*, "PointSAM: Pointly-supervised segment anything model for remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, article no. 5608515, 2025.

[10] N. Chen, B. Y. Li, Y. Q. Wang, *et al.*, "Motion and appearance decoupling representation for event cameras," *IEEE Transactions on Image Processing*, vol. 34, pp. 5964–597 7, 2025.

[11] S. Q. Ren, K. M. He, R. Girshick, *et al.*, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proceedings of the 29th International Conference on Neural Information Processing Systems*, Montreal, Canada, pp. 91–99, 2015.

[12] J. Redmon, S. Divvala, R. Girshick, *et al.*, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 779–788, 2016.

[13] T. Y. Lin, P. Dollár, R. Girshick, *et al.*, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 936–944, 2017.

[14] Q. Ming, L. J. Miao, Z. Q. Zhou, *et al.*, "Not all boxes are equal: Learning to optimize bounding boxes with discriminative distributions in optical remote sensing images,"

IEEE Transactions on Geoscience and Remote Sensing, vol. 62, article no. 5622514, 2024.

[15] P. H. Duan, T. C. Shan, X. D. Kang, *et al.*, "Spectral super-resolution in frequency domain," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 7, pp. 12338–12348, 2025.

[16] X. D. Kang, P. H. Duan, J. E. Li, *et al.*, "Efficient swin transformer for remote sensing image super-resolution," *IEEE Transactions on Image Processing*, vol. 33, pp. 6367–6379, 2024.

[17] Q. Ming, L. J. Miao, Z. Q. Zhou, *et al.*, "Optimization for arbitrary-oriented object detection via representation invariance loss," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, article no. 8021505, 2022.

[18] X. Yang and J. C. Yan, "Arbitrary-oriented object detection with circular smooth label," in *Proceedings of the 16th European Conference on Computer Vision*, Glasgow, UK, pp. 677–694, 2020.

[19] X. Yang, J. C. Yan, Q. Ming, *et al.*, "Rethinking rotated object detection with Gaussian Wasserstein distance loss," in *Proceedings of the 38th International Conference on Machine Learning*, Online, pp. 11830–11841, 2021.

[20] Q. Ming, Z. Q. Zhou, L. J. Miao, *et al.*, "Dynamic anchor learning for arbitrary-oriented object detection," in *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, Online, pp. 2355–2363, 2021.

[21] J. J. Song, L. J. Miao, Q. Ming, *et al.*, "Fine-grained object detection in remote sensing images via adaptive label assignment and refined-balanced feature pyramid network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 71–82, 2023.

[22] Q. Ming, L. J. Miao, Z. Q. Zhou, *et al.*, "Sparse label assignment for oriented object detection in aerial images," *Remote Sensing*, vol. 13, no. 14, article no. 2664, 2021.

[23] J. M. Han, J. Ding, J. Li, *et al.*, "Align deep features for oriented object detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, article no. 560251 1, 2022.

[24] J. M. Han, J. Ding, N. Xue, *et al.*, "ReDet: A rotation-equivariant detector for aerial object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, pp. 2785–2794, 2021.

[25] Q. Ming, L. J. Miao, Z. Q. Zhou, *et al.*, "CFC-Net: A critical feature capturing network for arbitrary-oriented object detection in remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, article no. 5605814, 2022.

[26] Q. Ming, L. J. Miao, Z. Q. Zhou, *et al.*, "Gradient calibration loss for fast and accurate oriented bounding box regression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, article no. 5611015, 2024.

[27] X. Yang, X. J. Yang, J. R. Yang, *et al.*, "Learning high-precision bounding box for rotated object detection via kullback-leibler divergence," in *Proceedings of the 35th International Conference on Neural Information Processing Systems*, Online, article no. 1405, 2021.

[28] N. Q. Liu, X. Xu, T. Celik, *et al.*, "Transformation-invariant network for few-shot object detection in remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, article no. 5625314, 2023.

[29] N. Q. Liu, X. Xu, Y. J. Gao, *et al.*, "Semi-supervised object detection with uncurated unlabeled data for remote sensing images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 129, article no. 103814, 2024.

[30] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 1440–1448, 2015.

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, *et al.*, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 2818–2826, 2016.

[32] X. Li, W. H. Wang, L. J. Wu, *et al.*, "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Vancouver, BC, Canada, article no. 1763, 2020.

[33] T. Y. Lin, P. Goyal, R. Girshick, *et al.*, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, pp. 2999–3007, 2017.

[34] Z. W. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 6154–6162, 2018.

[35] N. Carion, F. Massa, G. Synnaeve, *et al.*, "End-to-end object detection with transformers," in *Proceedings of the 16th European Conference on Computer Vision*, Glasgow, UK, pp. 213–229, 2020.

[36] X. Z. Zhu, W. J. Su, L. W. Lu, *et al.*, "Deformable DETR: Deformable transformers for end-to-end object detection," in *Proceedings of the 9th International Conference on Learning Representations*, Online, 2021.

[37] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proceedings of the 9th International Conference on Learning Representations*, Online, 2021.

[38] Y. C. Xu, M. T. Fu, Q. M. Wang, *et al.*, "Gliding vertex on the horizontal bounding box for multi-oriented object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 4, pp. 1452–1459, 2021.

[39] Q. Ming, L. J. Miao, Z. Q. Zhou, *et al.*, "Task interleaving and orientation estimation for high-precision oriented object detection in aerial images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 241–25 5, 2023.

[40] T. L. Ma, M. Y. Mao, H. H. Zheng, *et al.*, "Oriented object detection with transformer," *arXiv preprint*, arXiv: 2106.03146, 2021.

[41] J. Q. Zhao, Z. Y. Ding, Y. Zhou, *et al.*, "OrientedFormer: An end-to-end transformer-based oriented object detector in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, article no. 564081 6, 2024.

[42] Y. Zeng, Y. S. Chen, X. Yang, *et al.*, "ARS-DETR: Aspect ratio-sensitive detection transformer for aerial oriented object detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, article no. 5610315, 2024.

[43] S. C. Pei, M. H. Yeh, and C. C. Tseng, "Discrete fractional Fourier transform based on orthogonal projections," *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1335–1348, 1999.

[44] H. Yu, J. Huang, L. Z. Li, *et al.*, "Deep fractional Fourier transform," in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, New Orleans, LA, USA, article no. 3181, 2023.

[45] H. Sagan, *Space-Filling Curves*. Springer, New York, NY, USA, 2012.

[46] D. Hilbert, "Ueber die stetige Abbildung einer Line auf ein Flächenstück," *Mathematische Annalen*, vol. 38, no. 3, pp. 459–460, 1891.

[47] J. M. Zhang, S. Sclaroff, Z. Lin, *et al.*, "Minimum barrier salient object detection at 80 FPS," in *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 1404–1412, 2015.

[48] G. M. Morton, *A Computer Oriented Geodetic Data Base and A New Technique in File Sequencing*. IBM Ltd. , Ottawa, Canada, 1966.

[49] K. Y. Wang and L. Zhang, "Reconcile prediction consistency for balanced object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, QC, Canada, pp. 3611–3620, 2021.

[50] O. Russakovsky, J. Deng, H. Su, *et al.*, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[51] Z. K. Liu, L. Yuan, L. B. Weng, *et al.*, "A high resolution optical satellite image dataset for ship recognition and some new baselines," in *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*, Porto, Portugal, pp. 324–331, 2017.

[52] G. S. Xia, X. Bai, J. Ding, *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 3974–3983, 2018.

[53] J. W. Li, C. W. Qu, and J. Q. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proceedings of 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)*, Beijing, China, pp. 1–6, 2017.

[54] S. J. Wei, X. F. Zeng, Q. Z. Qu, *et al.*, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.

[55] D. C. Lu, D. M. Li, Y. L. Li, *et al.*, "OSKDet: Orientation-sensitive keypoint localization for rotated object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, pp. 1172–1182, 2022.

[56] L. P. Hou, K. Lu, J. Xue, *et al.*, "Shape-adaptive selection and measurement for oriented object detection," in *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, Online, pp. 923–932, 2022.

[57] W. T. Li, Y. J. Chen, K. X. Hu, *et al.*, "Oriented RepPoints for aerial object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, pp. 1819–1828, 2022.

[58] Z. H. Li, B. Hou, Z. T. Wu, *et al.*, "Complete rotated localization loss based on super-Gaussian distribution for remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, article no. 5618614, 2023.

[59] X. X. Xie, G. Cheng, J. B. Wang, *et al.*, "Oriented R-CNN for object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, QC, Canada, pp. 3500–3509, 2021.

[60] Z. C. Huang, W. Li, X. G. Xia, *et al.*, "Task-wise sampling convolutions for arbitrary-oriented object detection in aerial images," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 3, pp. 5204–5218, 2025.

[61] X. Yang, J. C. Yan, Z. M. Feng, *et al.*, "R3Det: Refined single-stage detector with feature refinement for rotating object," in *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, Online, pp. 3163–3171, 2021.

[62] Y. Yu and F. P. Da, "Phase-shifting coder: Predicting accurate orientation in oriented object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, BC, Canada, pp. 13354–13363, 2023.

[63] L. Q. Wang, J. Zhang, J. M. Tian, *et al.*, "Efficient fine-grained object recognition in high-resolution remote sensing images from knowledge distillation to filter grafting," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, article no. 4701016, 2023.

[64] X. Yang, L. P. Hou, Y. Zhou, *et al.*, "Dense label encoding for boundary discontinuity free rotation detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, pp. 15814–15824, 2021.

[65] J. Ding, N. Xue, Y. Long, *et al.*, "Learning RoI transformer for oriented object detection in aerial images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, pp. 2844–2853, 2019.

[66] X. Yang, J. R. Yang, J. C. Yan, *et al.*, "SCRDet: Towards more robust detection for small, cluttered and rotated objects," in *Proceedings of the IEEE International Conference on Computer Vision*, Seoul, Korea (South), pp. 8231–8240, 2019.

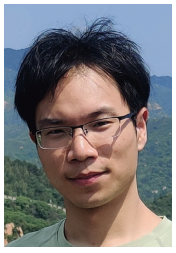
[67] C. Lee, J. Son, H. Shon, *et al.*, "FRED: Towards a full rotation-equivariance in aerial image object detection," in *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, Vancouver, Canada, pp. 2883–2891, 2024.

[68] W. Qian, X. Yang, S. L. Peng, *et al.*, "Learning modulated loss for rotated object detection," in *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, Online, pp. 2458–2466, 2021.

[69] Z. K. Xiao, G. Y. Yang, X. Yang, *et al.*, "Theoretically achieving continuous representation of oriented bounding boxes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 16912–16922, 2024.

[70] X. Yang, J. C. Yan, W. L. Liao, *et al.*, "SCRDet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 2384–2399, 2023.

[71] X. W. Zheng, W. L. Zhang, L. X. Huan, *et al.*, "AProNet: Detecting objects with precise orientation from aerial images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 181, pp. 99–112, 2021.



Qi Ming is an Associate Professor in the College of Computer Science, Beijing University of Technology (BJUT), Beijing, China. In 2018, he received the B.S. degree from the School of Automation, Beijing Institute of Technology (BIT), Beijing, China. In 2024, he received the Ph.D. degrees from BIT. His research interests include edge intelligence, computer vision, and the remote sensing.
(Email: chaser.ming@gmail.com)

Liuqian Wang received the Ph.D. degree in electronic science and technology. He is currently an Associate Professor with the School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou, China. His research focuses on remote sensing image interpretation, intelligent transportation system, and multimodal visual modeling study. (Email: jeremy.wang0126@gmail.com)

Juan Fang received the M.S. degree from Jilin University of Technology in 1997 and the Ph.D. degree from Beijing University of Technology in 2005. In 1997, she joined the College of Computer Science, Beijing University of Technology. From 2015, she is the professor of Beijing University of Technology. She currently works with the College of Computer Science, Beijing University of Technology. (Email: fangjuan@bjut.edu.cn)

Yifan Xiao is a Researcher in System Design Institute of Hubei Aerospace Technology Academy. She received the B.E. and M.S. degrees from Shandong University, Jinan, China, in 2015 and 2018, respectively. She received the Ph.D. degree from Ghent University, Belgium, in 2023. Her research interests include computer vision and image processing. (Email: yifanxiao.gent@gmail.com)

Zhixin Guo is a researcher in System Design Institute of Hubei Aerospace Technology Academy. He received the B.E. and M.S. degrees from Shandong University, Jinan, China, in 2013 and 2016, respectively. He received the Ph.D. degree from Ghent University, Belgium, in 2021. His research interests include computer vision and pattern recognition.
(Email: sduzxguo@gmail.com)

Xudong Zhao received the B.S. degree from Beijing Institute of Technology, Beijing, China, in 2016, where he received the Ph.D. degree in 2022. He also received the Ph.D. degree in computer science engineering with Ghent University, Ghent, Belgium, in 2023. His research interests include fractional signal processing and remote sensing.
(Email: zhaoxudong@bit.edu.cn)