Rotation Invariant and Symmetry Aware Pixel Difference Network for Remote Sensing Object Detection

Supplementary Material

6. Technical Appendices and Supplementary Material

6.1. Mathematical Theory

6.1.1. Pixel Difference Convolution (PDC)

In remote sensing object detection, small target sizes and low contrast demand a feature extraction mechanism sensitive to subtle local variations. Standard convolution aggregates information over a receptive field without explicitly modeling local intensity changes, which may suppress critical details. To address this, we propose a Pixel Difference Convolution (PDC) rooted in Local Binary Patterns (LBP) [37], a robust descriptor for local texture and edge information.

Let $x \in \mathbb{R}^{C\times H\times W}$ be an input feature map and $w \in \mathbb{R}^{C_{\text{out}}\times C_{\text{in}}\times k\times k}$ denote the convolutional kernel. Standard convolution computes:

$$y = x * w = \sum_{i=1}^{k^2} w_i \cdot x_i, \tag{12}$$

where x_i represents the pixel values in the $k \times k$ receptive field.

PDC emphasises local contrast by considering the difference between each neighbouring pixel and a reference value q:

$$y = \sum_{\substack{i=1\\ i\neq c}}^{k^2} w_i \cdot (x_i - q), \tag{13}$$

where c is the central position. The subtraction acts as a high-pass filter that accentuates edges and intensity changes.

Connection to LBP. Given a central pixel x_c and its P neighbours $\{x_p\}_{p=0}^{P-1}$, LBP encodes local structure via:

$$\text{LBP}(x_c) = \sum_{p=0}^{P-1} s(x_p - x_c)\cdot 2^p, \quad s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

PDC relaxes the binary decision to a continuous, learnable representation through the weights w_i, bridging classical descriptors with deep feature learning. By explicitly computing $x_i - q$, PDC highlights minute intensity variations essential for detecting subtle boundaries in high-resolution remote sensing imagery.

Algorithm 1 PDC Module Pseudocode

Class PDC **inherits** nn.Module
Function INIT($in\_channels$, $out\_channels$, $kernel\_size$=3, $stride$=1, $padding$=1, $dilation$=1, $groups$=1, $bias$=False, θ=0.7)
  $conv \leftarrow$ **Conv2d**($in\_channels$, $out\_channels$, $kernel\_size, \dots$)
  $self.\theta \leftarrow \theta$
EndFunction
Function FORWARD(x)
  $out\_normal \leftarrow conv(x)$
if $|\theta| < 10^{-8}$ **then**
    return out_normal
else
    $kernel\_diff \leftarrow$ sum of $conv.weight$ over spatial dims, reshape to $(C_{out}, C_{in}, 1, 1)$
    $out\_diff \leftarrow$ **conv2d**(x, $kernel\_diff$, padding=0, groups=conv.groups)
    return $out\_normal - \theta \times out\_diff$
end if
EndFunction

6.1.2. Polar Harmonic Transform (PHT) for Symmetry-Aware Encoding

Detecting structural symmetry in aerial imagery requires operators sensitive to both angular periodicity and radial regularity. Classical approaches such as the Laplacian capture isotropic second-order variations but are blind to directional symmetry (e.g., bilateral, K-fold rotational). We adopt the Polar Harmonic Transform (PHT) [63] as the mathematical foundation for our Symmetry-aware Pixel Difference Convolution (S-PDC).

Polar Coordinate Construction. Given a square patch $\boldsymbol{x} \in \mathbb{R}^{N\times N}$, each pixel i at position (u_i, v_i) is mapped to polar coordinates centred on the patch:

$$r_i^2 = \frac{u_i^2 + v_i^2}{k^2 + \epsilon}, \qquad \theta_i = \text{atan2}(v_i,\ u_i + \epsilon), \tag{15}$$

where $k = \lfloor N/2 \rfloor$ normalises the radius and ϵ is a small constant for numerical stability.

Harmonic Basis Functions. The PHT defines a family of orthogonal basis kernels indexed by radial order $n \in \mathbb{Z}^+$ and angular order $l \in \mathbb{Z}$:

$$H_i^{(n,l)} = \cos\big(2\pi n\, r_i^2 + l\, \theta_i\big). \tag{16}$$

Each kernel encodes a specific combination of radial fre-

quency (concentric rings, controlled by n) and angular frequency (rotational periodicity, controlled by l).

Why Cosine Harmonics Encode Symmetry. A pattern with K-fold rotational symmetry satisfies $f(r,\theta) = f(r,\theta{+}2\pi/K)$. The harmonic kernel $H^{(n,l)}$ shares this periodicity when l is a multiple of K, producing a maximal inner product with such patterns while non-symmetric textures interfere destructively. This selective amplification is the inductive bias needed for aerial objects such as aircraft (bilateral, $K{=}2$) and roundabouts (radial, large K).

Multi-Order Learnable Combination. To capture diverse symmetry types, we combine multiple harmonic orders with trainable coefficients $\alpha_{n,l}$:

$$H_i^{\text{comb}} = \sum_{(n,l)\in\mathcal{O}} \alpha_{n,l}\, \cos\!\big(2\pi n\, r_i^2 + l\,\theta_i\big), \qquad (17)$$

where $\mathcal{O} = \{(n_1,l_1),\ldots,(n_M,l_M)\}$ is a predefined order set and $\alpha_{n,l}$ are initialised uniformly and updated via backpropagation.

S-PDC Integration. In S-PDC, each non-central pixel's contribution is modulated by the combined harmonic kernel:

$$y = \sum_{(n,l)\in\mathcal{O}} \alpha_{n,l} \sum_{i\neq c}^{N^2} w_i\, \cos\!\big(2\pi n\, r_i^2 + l\,\theta_i\big)\,(x_i - q). \quad (18)$$

This construction ensures rotational alignment, suppresses isotropic noise, and emphasises angular-radial regularity — vital for detecting symmetric or oriented targets.

Theoretical Properties. PHT-based kernels in S-PDC inherit several strong mathematical properties:

- **Directional Selectivity:** Angular frequency l models arbitrary rotation symmetries.
- **Radial Locality:** Radial frequency n encodes near vs. far structure relative to the patch centre.
- **Orthogonality:** Kernels with different (n,l) are linearly independent, ensuring basis separability.
- **Efficiency:** Closed-form cosine basis enables fast, fully differentiable implementation.

6.1.3. SO(2) Rotation Group and Rotation-Invariant Convolution

SO(2) is the Lie group of 2D rotations, consisting of all $2{\times}2$ orthogonal matrices with determinant 1. For any angle θ, the rotation matrix is:

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \qquad (19)$$

Group Properties.

- **Closure:** $R_{\theta_1} R_{\theta_2} = R_{\theta_1+\theta_2}$.
- **Identity:** $R_0 = I$.
- **Inverse:** $R_\theta^{-1} = R_{-\theta}$.

From S-PDC to RIS-PDC. While S-PDC detects symmetry features effectively, applying SO(2) rotations makes the feature detection invariant to arbitrary orientations — crucial for remote sensing where objects may appear at any angle. For each of n sampled rotation angles θ_j, we rotate the convolution kernel via R_{θ_j} and apply S-PDC:

$$y_{\theta_j} = \sum_{(n,l)\in\mathcal{O}} \alpha_{n,l} \sum_{\substack{i=1\\ i\neq c}}^{N^2} w_i{\cdot}H_i^{(n,l)}{\cdot}(R_{\theta_j}\, x_i{-}q) \quad (\text{RIS-PDC}_{\theta_j}) \qquad (20)$$

The final rotation-invariant feature map averages all directions:

$$y_{\text{final}} = \frac{1}{n}\sum_{j=1}^{n} y_{\theta_j} \quad (\text{RIS-PDC}) \qquad (21)$$

In practice, $n{=}8$ provides sufficient angular sampling to approximate continuous SO(2) symmetry while minimising interpolation artifacts on discrete grids.

6.2. Extended Ablation Studies

6.2.1. Ablation of RIS-PDC Variants

We explore the role of each submodule within RIS-PDC (rotation-invariance via SO(2), symmetry modelling via PHT, and pixel difference convolution) and demonstrate that each component is beneficial independently. Introducing rotation invariance alone reduces angular bias and improves recall for arbitrarily oriented objects. PHT-based harmonic modulation alone sharpens salient edges and captures structural symmetry, enhancing precision for small or complex objects. PDC contributes gradient-level information that enriches boundary delineation. Their joint use in RIS-PDC yields the largest gains through complementary interplay: rotation invariance secures orientation robustness, PHT modulation amplifies symmetry-consistent edges, and PDC encapsulates fine-grained local structures. RIS-PDC gains 1.18% mAP over Vanilla Conv with no parameter increase. PHT modulation and SO(2) invariance jointly achieve 77.75% mAP on DOTA-v1.0.

6.2.2. Ablation of Different Rotation Directions

Traditional strategies (methods "A"–"C") rely on a small set of orientations (e.g., $0°/90°/180°/270°$) that align with image axes to minimise interpolation artifacts. Increasing granularity with these methods risks introducing padding/interpolation errors. Our SO(2)-based approach (method "D") provides a principled approximation of continuous rotational symmetry. Eight orientations capture sufficient diversity without over-fragmenting the rotation space: fewer under-sample oblique orientations, while more (e.g., 16) dilute discriminative power with excessive overhead. 8 directions provide the optimal accuracy-efficiency balance. Increasing to 16 yields minimal gains (+0.04% mAP) while increasing latency.

Algorithm 2 Pseudocode for the RIS-PDC

Class RIS-PDC:
Procedure COMPUTEHARMONICKERNELS(N, $\mathcal{O}$, α)
  $k \leftarrow \lfloor N/2 \rfloor$
  for each position i in $N \times N$ grid **do**
    $(u_i, v_i) \leftarrow$ offset from patch center
    $r_i^2 \leftarrow (u_i^2 + v_i^2)/(k^2 + \epsilon)$
    $\theta_i \leftarrow \text{atan2}(v_i,\ u_i + \epsilon)$
  end for
  $H \leftarrow \mathbf{0}$ {combined harmonic kernel}
  for each $(n, l) \in \mathcal{O}$ **do**
    $H \mathrel{+}= \alpha_{n,l} \cdot \cos(2\pi n\, r^2 + l\, \theta)$
  end for
  Return H
EndProcedure
Procedure S-PDC(x, $weights$, H, q)
  $diff \leftarrow x_i - q, \quad \forall\, i \neq c$ {pixel differences}
  $y \leftarrow \text{Conv2d}(diff,\ weights \odot H)$ {harmonic-modulated conv}
  Return y
EndProcedure
Procedure APPLYLIEGROUPROTATION(x, $weights$, H, q)
  $angles \leftarrow$ 8 evenly spaced values from 0 to 2π
  Initialize list $outputs \leftarrow [\,]$
  for each $angle \in angles$ **do**
    $R \leftarrow$ ROTATIONMATRIX2D($angle$)
    $W_\theta \leftarrow$ ROTATEKERNEL($weights$, R)
    $out \leftarrow$ S-PDC(x, W_θ, H, q)
    Append out to $outputs$
  end for
  Return $\frac{1}{8} \sum outputs$
EndProcedure
Procedure ROTATIONMATRIX2D($angle$)
  Return $\begin{bmatrix} \cos(angle) & -\sin(angle) \\ \sin(angle) & \cos(angle) \end{bmatrix}$
EndProcedure
Procedure ROTATEKERNEL($weights$, R)
  Initialize $rotatedKernel$ (same shape, zeros)
  for each kernel position (h, w) **do**
    $h' \leftarrow h \cdot R[0,0] + w \cdot R[0,1]$
    $w' \leftarrow h \cdot R[1,0] + w \cdot R[1,1]$
    if (h', w') is within valid range **then**
      $rotatedKernel(:,:,h',w') \leftarrow weights(:,:,h,w)$
    end if
  end for
  Return $rotatedKernel$
EndProcedure

6.3. Additional Visualisation

6.3.1. Feature Activation Comparison

Figure 10 presents side-by-side detection outputs and heat maps for the Baseline (LSKNet-T) and RIS-PiDiNet-T at four rotation angles (0°, 90°, 180°, 270°). As rotation angle increases, the Baseline shows degraded spatial coherence with weaker boundary responses, while RIS-PiDiNet-T maintains sharper edges and more salient activations across all orientations. Figure 10 shows that RIS-PiDiNet-T concentrates on structural edges, effectively suppressing background noise compared to the baseline.

6.3.2. Rotation Equivariance Analysis

Figure 11 compares detection outputs and response heat maps between LSKNet-T and RIS-PiDiNet-T on inputs rotated by 0°, 90°, 180°, and 270°. While the Baseline exhibits fluctuating confidence and spatially inconsistent heat maps (position drifts, missed targets), RIS-PiDiNet-T maintains consistently accurate predictions regardless of orientation. This invariance stems from the explicit integration of polar harmonic modulation and structured difference encoding in RIS-PDC, which jointly model angular periodicity and radial locality. Figure 11 demonstrates robust rotation equivariance. Response peaks remain centered across orientations, reducing localization variance.

To verify that RIS-PDC is architecture-agnostic, we integrate it into two Transformer-based detectors: LocalMamba-T [16] and RF-DETR [42]. As shown in Table 8, RIS-PDC yields consistent improvements across both architectures, confirming its universality as a plug-and-play operator. Results on LocalMamba and RF-DETR confirm that RIS-PDC is architecture-agnostic, providing local geometric priors that complement global attention.

Table 8. Universality of RIS-PDC on Transformer-based detectors (DOTA-v1.0 single-scale).

Method	mAP (%)	Δ
LSKNet-T	74.83	-
RIS-PiDiNet-T	**76.92**	**+2.09**
LocalMamba-T	78.02	-
+ RIS-PDC (Ours)	**79.76**	**+1.74**
RF-DETR	76.85	-
+ RIS-PDC (Ours)	**78.69**	**+1.84**

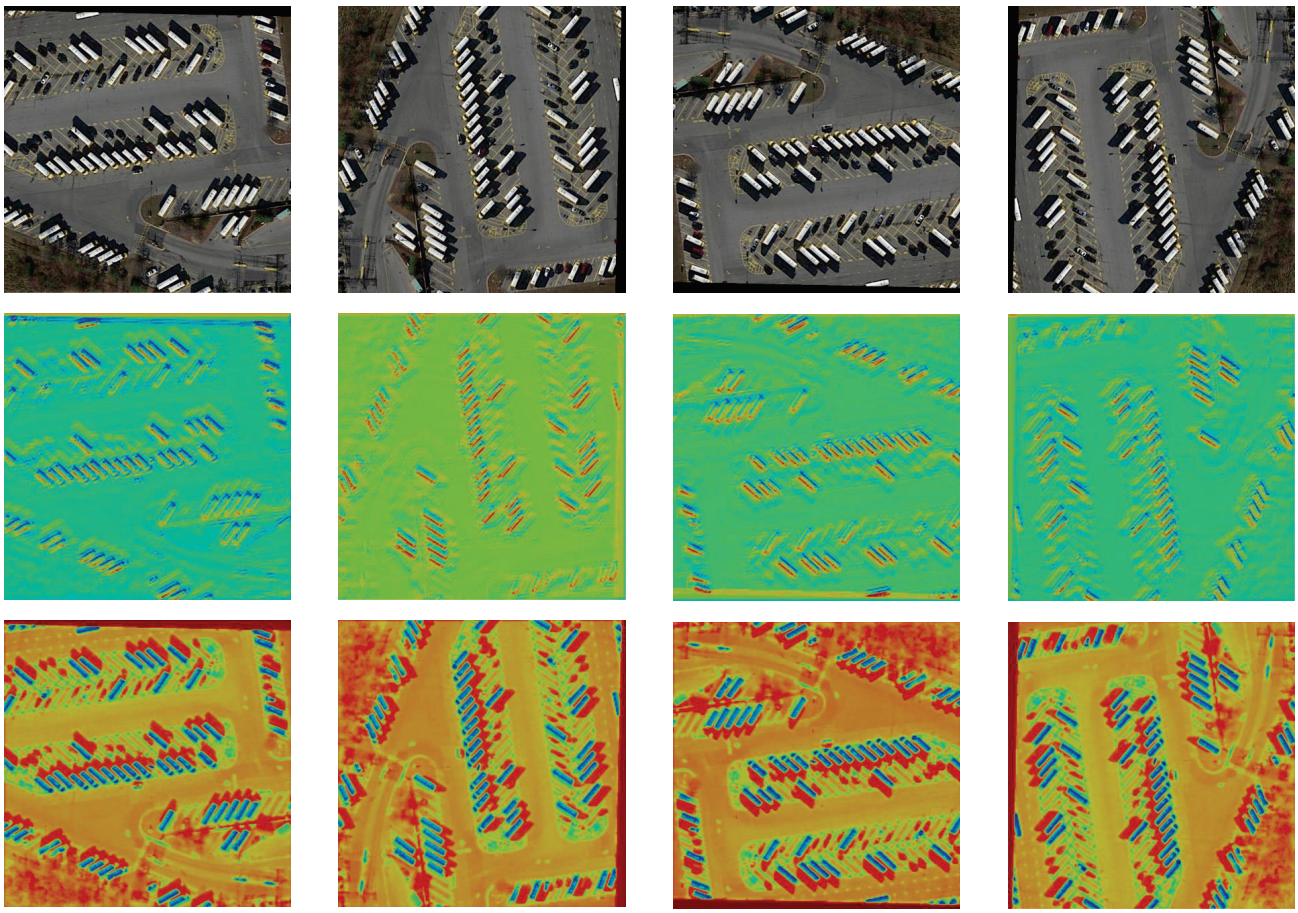

Figure 10. The first row shows the original aerial images. The second row presents the Baseline (LSKNet-T) visualisation, and the third row shows RIS-PiDiNet-T. Compared to the Baseline, RIS-PiDiNet produces clearer feature responses along object boundaries.

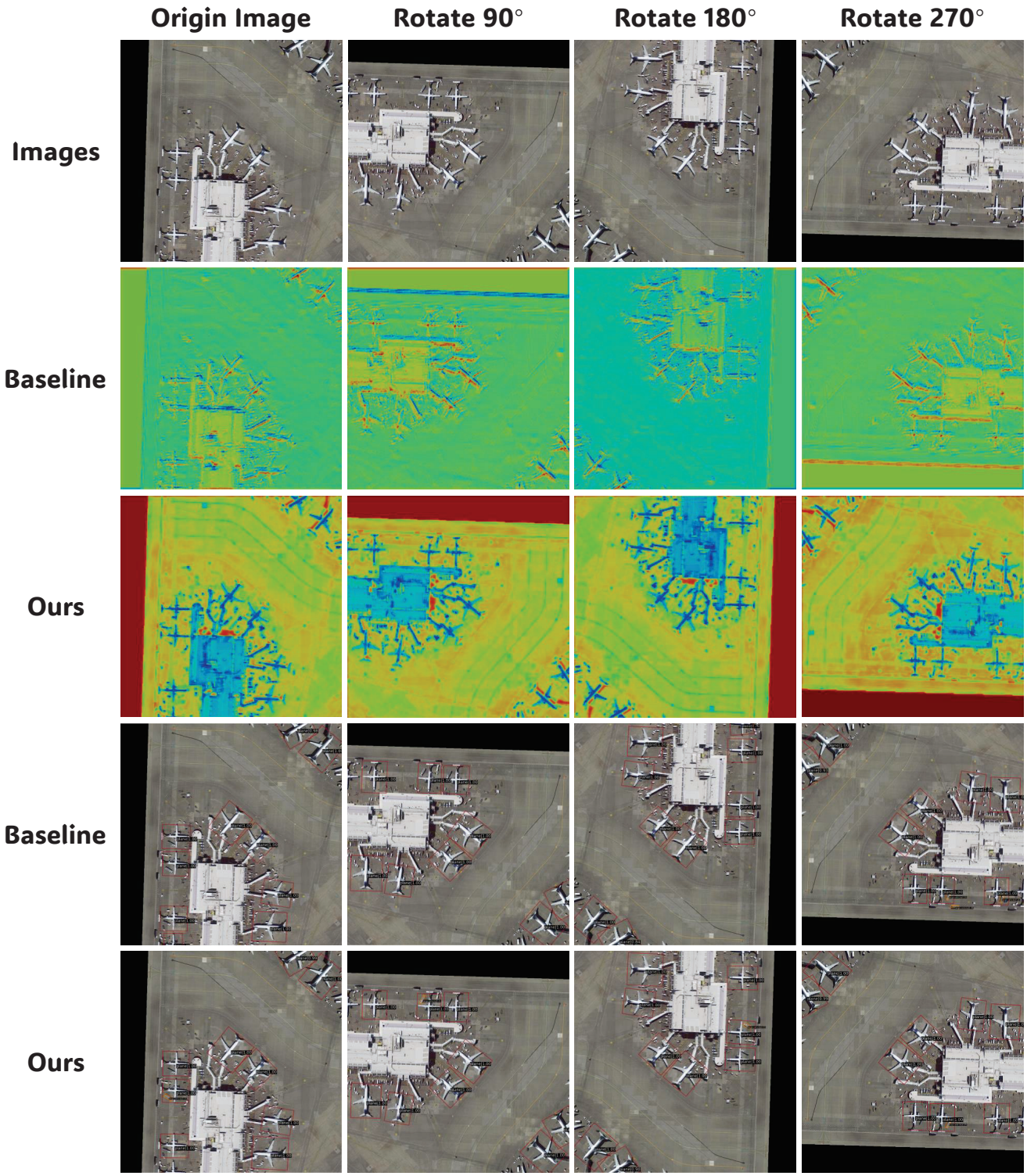

Figure 11. The detection outputs and heat maps for Baseline (LSKNet-T) and Ours (RIS-PiDiNet-T) across rotations of 0°, 90°, 180°, and 270°. RIS-PiDiNet consistently maintains high detection accuracy with uniformly distributed heat map responses.