Highlights

Water Reflection Detection Using Symmetric Attention

Shuxuan Yao, Chengjia Wang, Jianyuan Sun, Junyu Dong, Xinghui Dong

- We introduce an end-to-end water reflection detection network, i.e., SAWRD-Net, which couples group-equivariant convolutions with low-rank matrix decomposition
- We develop a symmetric-attention mechanism that redistributes feature weights according to reflection cues, thereby improving the capture of symmetry-specific information and enhancing detection accuracy
- We design an MSRE block and an MD Decoder, which leverages the intrinsic symmetry of water reflection scenes while modeling long-range context via precise low-rank factorization

Water Reflection Detection Using Symmetric Attention

Shuxuan Yao[a], Chengjia Wang[b], Jianyuan Sun[c], Junyu Dong[a], Xinghui Dong[a,*]

[a]*State Key Laboratory of Physical Oceanography and the Faculty of Information Science and Engineering, Ocean University of China, 238 Songling Road, Qingdao, 266100, Shandong, China*
[b]*School of Mathematical and Computer Sciences, Heriot-Watt University, Riccarton Mains Road, EH14 4AS, Edinburgh, United Kingdom*
[c]*School of Computer Science and Informatics, De Montfort University, Leicester, LE1 9BH, United Kingdom*

Abstract

Reflections of water pose a significant challenge for computer vision systems, as standard deep learning models frequently confuse objects with their mirror images, producing spurious false positives and negatives in tasks such as object detection and semantic segmentation. As a result, detecting reflection axes in natural-water scenes is pivotal for reliable object detection and scene understanding. To mitigate this issue, we leverage the intrinsic imperfect reflective symmetry of water and introduce a Symmetry-Aware Water Reflection Detection Network, namely, SAWRD-Net, that couples dihedral group–equivariant convolutions with a matrix-decomposition decoder in an end-to-end framework. First, dihedral group convolutional layers extract geometry-consistent feature maps that explicitly encode both rotational and mirror symmetries. A Multi-scale Reflection Equivariant block then aggregates features across scales and employs a symmetric-attention mechanism to highlight reflection-relevant regions. The proposed matrix-decomposition decoder factorizes high-dimensional features into compact low-rank parameter and confidence spaces, after which the network directly regresses key-points on the reflection axis. Then a robust principal component analysis fits the final axis. Evaluated on the largest available water reflection scene data set, SAWRD-Net achieves a true-positive rate of 0.890 against human annotations, outperforming all

⋆Code and models will be made available at https://github.com/INDTLab/SAWRD-Net.

*Corresponding author

Email addresses: yaoshuxuan@stu.ouc.edu.cn (Shuxuan Yao), chengjia.wang@hw.ac.uk (Chengjia Wang), jianyuan.sun@dmu.ac.uk (Jianyuan Sun), dongjunyu@ouc.edu.cn (Junyu Dong), xinghui.dong@ouc.edu.cn (Xinghui Dong)

existing water reflection detectors.

Keywords:
Water Reflection Detection, Symmetry Detection, Line Detection, Equivariant Learning, Matrix Decomposition.

---

1. Introduction

Water reflections are ubiquitous in nature photography and constitute a prominent class of outdoor images [1, 2, 3, 4], thus accurately modeling them benefits large-scale visual understanding [5]. Reflections of water create mirror-like duplicates of scene objects [6, 7]. Ideally, the object and its reflection would exhibit perfect reflective symmetry on the water surface, but surface ripples, refraction, and light scattering introduce *imperfect (or approximate) symmetry* cause an object and its virtual image to exhibit a geometric correspondence relative to the water surface [3], forming a reflective symmetry [8]. Detecting the underlying reflection axis and separating real objects from their virtual counterparts is therefore a prerequisite for robust scene parsing, object detection [9, 10], and image segmentation in natural environments [1].

Water reflection images exhibit mirror symmetry, causing segmentation and detection networks to confuse real objects with their reflections [11]. As shown in Fig. 1, YOLOv12 [12] identifies an object and its reflection as separate instances, thus generating false positives. By localizing the reflection axis and suppressing the mirrored region, such errors can be eliminated, enabling correct recognition. In this situation, a dedicated preprocessing is necessary to identify and filter out reflective regions to achieve reliable scene understanding.

Water reflection detection originally relied on manual feature-based pipelines [2, 3, 4]. These approaches require expert tuning of gradient thresholds, color ratios, and geometric heuristics, which sharply limits their applicability in real-world scenarios. Deep learning models, in contrast, learn discriminative features directly from data and therefore offer greater robustness. However, comprehensive benchmarks and large-scale data sets for this purpose remain scarce. In the literature, the only end-to-end water reflection detection model is WRD-Net [1], which uses parallel attention for multi-scale

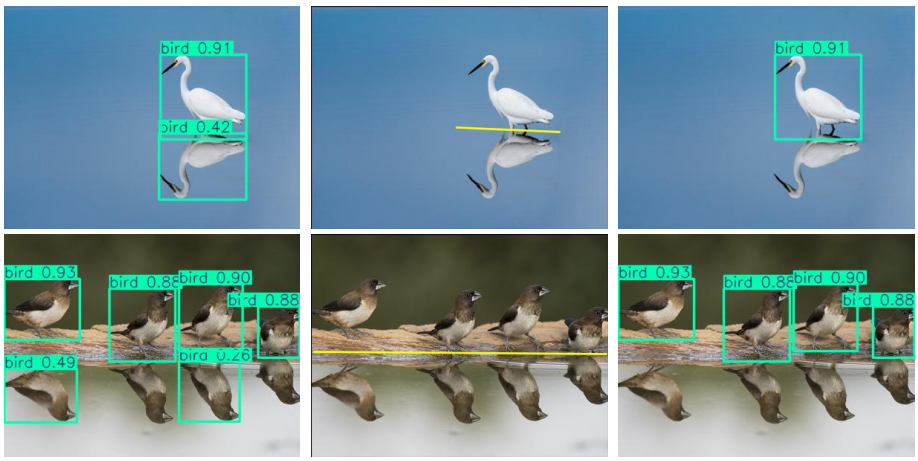

Figure 1: The case of failure in detecting water reflection images using the YOLOv12 [12] object detector (a), the reflection axis detected by the SAWRD-Net (b), and the successful detection result of the YOLOv12 [12] object detector after filtering the reflection area of the water reflection image based on the reflection axis detected by the SAWRD-Net (c).

context modeling, but implicitly learns symmetry instead of encoding mirror geometry. On the other hand, EquiSym [13] and related detectors [14, 15] enforce strict equivariance under perfect symmetry, making them vulnerable to water ripples and refraction. In contrast, SAWRD-Net balances reflection equivariance and perturbation tolerance via equivariant convolutions and low-rank decomposition.

Earlier vision pipelines achieved rotational or mirror equivariance by pairing orientation-sensitive key-point detectors with an orientation-normalization step, for example, aligning feature patches to their dominant gradient direction [16, 17]. The goal was to preserve one-to-one correspondences between features when the input image was rotated or flipped. Although effective for shallow, gradient-based descriptors, this mechanism degrades when transferred to modern deep networks. Deep feature maps undergo complex, non-linear distortions under the same transformations, breaking the assumed correspondence. Equivariant learning addresses this limitation by embedding the transformation law directly into the network. By guaranteeing predictable, structurally consistent changes throughout all layers whenever the input is rotated or reflected, it provides

a principled foundation for symmetry-aware tasks and naturally respects the geometric constraints of water reflection scenes, especially in high-resolution outdoor footage captured under uncontrolled lighting [18].

Water reflection detection focuses on modeling global yet imperfect symmetry in natural scenes. Such symmetry, caused by ripples, refraction, etc., implies object and reflection are not pixel-wise identical but globally correlated. Since matrix decomposition uses low-rank constraints to capture global symmetry while treating local perturbations as residuals, it naturally formulates approximate symmetry in water reflection scenes. Building on the established applicability of group-equivariant convolutions to symmetry detection, we construct a Multi-scale Reflection Equivariant (MSRE) block that extracts rotation- and reflection-equivariant features, complemented by a symmetric-attention mechanism that concentrates on reflection-relevant regions. Since water reflections form a global level structure and typically exhibit incomplete reflective symmetry, we further introduce a dedicated decoder, the Matrix Decomposition (MD) Decoder, which factorises the high-dimensional representation into complementary low-rank parameter and confidence sub-spaces, thereby enabling more reliable axis estimation in challenging, real-world water reflection scenes.

The main contributions of this work are threefold.

1. We introduce SAWRD-Net, an end-to-end network for water reflection detection that couples group-equivariant convolutions with low-rank matrix decomposition. This network regresses symmetry-axis point sets and refines the axis through principal component analysis.
2. We develop a symmetric-attention mechanism that redistributes feature weights according to reflection cues, thereby improving the capture of symmetry-specific information and enhancing detection accuracy.
3. We design an MSRE block together with an MD Decoder, which leverages the intrinsic symmetry of water scenes while modeling long-range context via precise low-rank factorization, yielding robust axis estimation under imperfect symmetry.

The remainder of the paper is organized as follows: Section 2 reviews related work.

Section 3 describes the proposed SAWRD-Net and its symmetric-attention module. Section 4 details the experimental setup, and Section 5 reports the results. Section 6 concludes the paper and provides a further detailed discussion.

2. Related Work

2.1. Water Reflection Detection

Water reflection detection techniques have progressed from manual feature engineering to modern deep learning approaches. Early work by Zhang et al. [3] introduced a *flip-invariant shape descriptor* to mitigate distortions caused by surface waves. In separate work, Zhong et al. [4] proposed a motion blur invariant feature space (MBIS) with a dynamic programming axis detector; they later combined MBIS with curvelet coefficients in a dual-channel algorithm Zhong et al. [2]. These methods exploit log–polar representations and dynamic programming but remain limited by hand-crafted features. More recently, Dong et al. [1] framed water reflection detection as a symmetry axis point prediction task and proposed WRD-Net, a deep architecture that fuses a parallel-attention Vision Transformer with an *atrous* spatial pyramid. Although WRD-Net removes the need for manual features, it under-utilizes the inherent reflective symmetry of water scenes.

2.2. Reflection Symmetry Detection

Traditional methods usually generate initial symmetry assumptions based on geometric features, and then make iterative corrections by optimizing the symmetry coefficients or matching feature pairs. In contrast, Deep learning methods use neural networks to automatically learn symmetry features through end-to-end training and optimize the model parameters with the help of back-propagation to progressively improve detection accuracy.

2.2.1. Traditional Methods

Early methods for symmetry detection often relied on global image properties and geometric transforms. For instance, Marola [19] proposed an algorithm that detects the reflection axis by determining the image's centre of mass and maximising symmetry

coefficients. A Hough transform-based method for reflection symmetry axis detection was proposed in [20], which converts the problem into a line detection task in Hough space, allowing for robust detection in complex line drawings. Other approaches leveraged the frequency domain. Sun and Si [21] proposed a method based on histograms of gradient directions and Fourier transform, while the method in [22] operates by examining specific patterns in an object's Fourier transform using the angular difference function (ADF). Expanding on transform techniques, an efficient geometry-based framework was proposed in [23] that uses invariant hashing and the Hough transform to detect various symmetric configurations, including periodicity and mirror symmetry, while maintaining robustness to perspective distortions.

A second major family of traditional methods is based on local features and point matching. Loy and Eklundh [16] proposed a symmetry detection method based on this principle, where feature points are generated and matched to find symmetric pairs, detecting both mirror and rotational symmetry. This concept was extended by Lee and Liu [24] with a local feature-based algorithm for detecting curved glide-reflection symmetry in real images. An adaptive algorithm, also based on local feature point matching, was proposed in [25] to detect both reflectional and rotational symmetries. Similarly, a bilateral symmetry detection method based on the discovery of extremum curvature points along the edges was proposed in [26], which uses histograms of curvature responses to describe the characteristics. Other advanced approaches used different strategies. Elawady et al. [27] proposed a method for the detection of high-precision global symmetry axis using the Log-Gabor wavelet transform, which integrates edge, texture and color descriptors within a voting framework. Meanwhile, Cicconet et al. [28] introduced a new framework called Mirror Symmetry via Registration (MSR) to detect mirror symmetry in data. Finally, a method combining appearance and gradient information was proposed in [29], and its effectiveness was confirmed through user perception validation experiments.

2.2.2. Learning-Based Methods

Deep learning-based approaches automatically learn symmetry features through neural networks, enabling them to capture symmetry patterns at different scales and

orientations with more precision. Funk and Liu [14] proposed a deep symmetry detector, Sym-NET, which mimics human perception of non-planar symmetry by converting sparse symmetry labels into dense heat maps for the network to predict. Seo et al. [15] proposed Polar Matching Convolution (PMC), which utilizes high-dimensional matching kernels and relational descriptors to identify symmetry patterns in challenging real-world images. Most relevant to our work, Seo et al. [13] introduced an equivariant learning-based method. It employs an equivariant convolutional neural network to simultaneously detect reflective and rotational symmetries while maintaining detection accuracy under geometric transformations.

Previous methods have primarily focused on identifying symmetric patterns within the candidate directional space to better exploit the symmetrical characteristics in the images. To our knowledge, however, very few techniques employ deep learning for water reflection detection. Although WRD-Net [1] was built on top of deep learning techniques, it combined parallel attention transformers with dilated spatial pyramids, thus failing to fully capitalize on the inherent reflection symmetry characteristics of water surfaces. Some symmetry detection methods, such as EquiSym [13], leveraged the concept of equivariant learning to construct a fully equivariant network. However, these networks were designed for the detection of perfect symmetry, whereas water reflection is a quintessential example of an imperfect reflection phenomenon. Starting from symmetry detection and designed to handle both inherent symmetry in water reflection images and noise induced by external factors like water waves, we propose SAWRD-Net to boost detection performance.

3. Symmetry-Aware Water Reflection Detection Network

Feature equivariance plays a fundamental role in the detection of image symmetry. Since the symmetric properties exhibited in images remain invariant under Euclidean transformations, images subjected to such transformations should maintain symmetry detection results consistent with the original images. Using this equivariance property and combining it with the inherent reflective symmetry characteristics of water reflection images [30], the equivariance to reflection emerges as a key discriminative

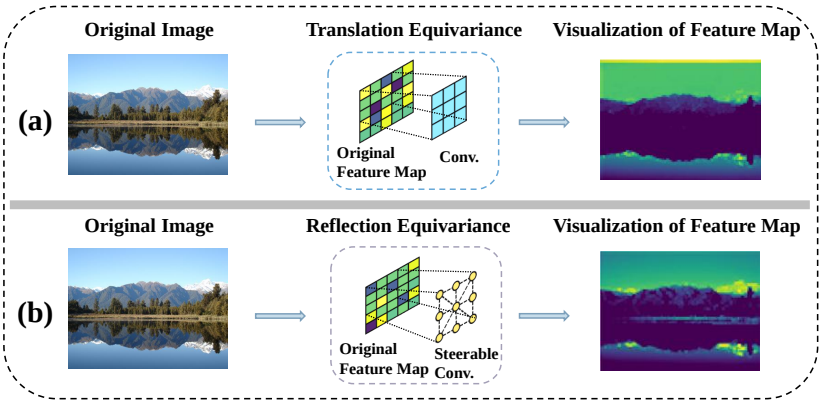

Figure 2: (a) visualizes the feature map of a water reflection image extracted using traditional convolution with translational equivariance, while (b) visualizes the feature map extracted using steerable convolution with rotational and reflective equivariance. The visualization in (b) form a clearer representation.

characteristic for water reflection detection models. Considering the decisive impact of global contextual dependencies in images on water reflection detection and the nature of water reflection as an imperfect reflective symmetry, we further employ matrix decomposition [31] to extract global information from network features. This process involves constructing a lookup dictionary and its corresponding encoding while eliminating the influence of noise, thereby effectively capturing the intrinsic correlations among features.

3.1. Preliminaries

3.1.1. Equivariant Learning

Steerable convolution is a type of convolution whose feature space consists of steerable feature fields, which obey transformation laws associated with the group representations of $E(2)$ subgroups. As illustrated in Fig. 2, the upper panel displays the feature map of the water reflection image extracted by a conventional Convolutional Neural Network (CNN), while the lower panel presents the feature map extracted by a network constructed with steerable convolutions. It can be clearly observed from the figure that the network built with steerable convolutions yields feature maps with more distinct contours and more complete symmetric features.

The Euclidean group, also known as the rigid motion group or isometry group, constitutes the mathematical structure describing all distance-preserving transformations in the Euclidean space. These transformations include translations, rotations, reflections, and their arbitrary combinations. The Euclidean group is usually denoted as $E(n)$, where n is the dimension of the Euclidean space. It consists of all transformations $f : \mathbb{R}^n \to \mathbb{R}^n$ that satisfy

$$||f(x) - f(y)|| = ||x - y||, \forall x, y \in \mathbb{R}^n, \tag{1}$$

where $||.||$ is the Euclidean norm. Here we are mainly concerned with the group $E(n)$ of isometric transformations on the plane $\mathbb{R}^2$. The Euclidean group, $E(n)$, is composed of the translation group $\left(\mathbb{R}^2, +\right)$ and the orthogonal group $O(2) = \left\{O \in \mathbb{R}^{2\times2} | O^T O = id_{2\times2}\right\}$ via the semi-direct product operation, i.e.

$$E(2) \cong (\mathbb{R}^2, +) \rtimes O(2). \tag{2}$$

We mainly consider subgroups of the form $(\mathbb{R}^2, +) \rtimes G$, where $G \leq O(2)$ [32].

The feature space of a traditional CNN is a collection of multi-channel feature maps that represent the image features at particular scales or frequencies. The weight-sharing property of CNN convolutional kernels ensures translation equivariance in the feature space, i.e., a translation of input features results in corresponding translated output features.

In contrast, the feature space of a steerable CNN is the steerable feature field $f : \mathbb{R}^2 \to \mathbb{R}^c$, which associates a c-dimensional feature vector, $f(x) \in \mathbb{R}^c$, with each point x on the plane $\mathbb{R}^2$, and is related to the transformation law of $E(n)$. For example, for vector eigenfields $v : \mathbb{R}^2 \to \mathbb{R}^2$ (as in a gradient image), the action of $E(n)$ is to move each pixel to a new position and also to change its direction by the action of $g \in G$. While the transformation law of a general feature field $f : \mathbb{R}^2 \to \mathbb{R}^c$ is completely determined by its type ρ, the $\rho : G \mapsto GL(\mathbb{R}^c)$ of a steerable feature field is a group representation that specifies how the c channels of each feature vector $f(x)$ are mixed in the feature space. Like the feature space of a conventional CNN that contains multiple channels, the feature space of a steerable CNN is composed of multiple feature fields $f_i : \mathbb{R}^2 \to \mathbb{R}^{c_i}$, each of which has its own type $\rho_i : G \to GL(\mathbb{R}^{c_i})$.

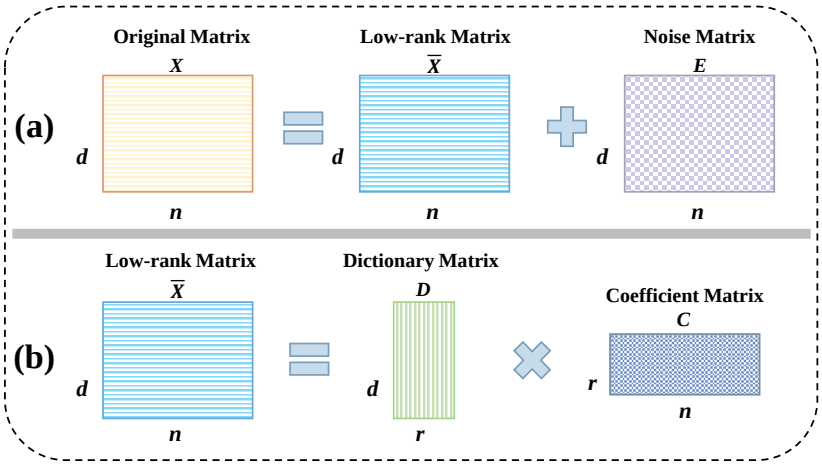

Figure 3: Formulation of matrix decomposition. Specifically, (a) indicates that the original matrix is composed of the sum of a low-rank matrix and a noise matrix, while (b) illustrates that the low-rank matrix is constructed through the multiplication of a dictionary matrix and a coefficient matrix.

Steerable convolution [32] not only exhibits translation equivariance but also demonstrates equivariance to reflection and rotation. Therefore, the most general equivariant linear mapping between steerable feature spaces transforming under ρ_{in} and ρ_{out} is given by a G-steerable kernel $k : \mathbb{R}^2 \rightarrow \mathbb{R}^{c_{out} \times c_{in}}$ that satisfies the kernel constraint [32]:

$$k(gx) = \rho_{out}(g)k(x)\rho_{in}(g^{-1}) \forall g \in G, x \in \mathbb{R}^2. \tag{3}$$

This constraint ensures the correct transformation behavior of input and output feature fields.

3.1.2. Matrix Decomposition

Matrix decomposition is a mathematical process that decomposes a given matrix into the product or sum of several matrices with specific mathematical properties. In the context of global context modeling, compared to self-attention mechanisms, matrix decomposition demonstrates relative advantages in computational efficiency and performance. From a mathematical perspective, matrix decomposition algorithms represent input images as linear combinations of basis images, achieving low-rank embedding reconstruction through submatrix decomposition of the input representation.

If the image data are arranged into a large matrix

$$X = [x_1, \ldots, x_n] \in \mathbb{R}^{d \times n}, \tag{4}$$

where each column x_i corresponds to a pixel vector of an image, d is the total number of pixels in the image, and n is the number of images. It is often assumed that these image data imply a low-dimensional subspace structure, or consist of multiple subspaces. Specifically, there exist a dictionary matrix D and a coefficient matrix C, i.e.,

$$\begin{aligned} D &= [d_1, \ldots, d_r] \in \mathbb{R}^{d \times r}, \\ C &= [c_1, \ldots, c_n] \in \mathbb{R}^{r \times n}, \\ X &\approx DC, \end{aligned} \tag{5}$$

where the column vectors of the dictionary matrix D can be treated as the base features of the images, and the coefficient matrix C denotes the coefficients of the individual images that are linearly combined from these features [31]. As illustrated in Fig. 3, this matrix decomposition process is depicted as:

$$\overline{X} = DC, \quad \text{where} \quad X = \overline{X} + E. \tag{6}$$

$\overline{X} \in \mathbb{R}^{d \times n}$ is the output low-rank reconstruction matrix, representing the main structural features of the image data, and $E \in \mathbb{R}^{d \times n}$ is the noise matrix to be discarded, containing the high-frequency noises or non-primary features in the image. It is assumed here that the recovered matrix $\overline{X}$ has the low-rank property, i.e., its rank satisfies [31]:

$$rank(\overline{X}) \leq min(rank(D), rank(C)) \leq r \ll min(d, n). \tag{7}$$

In summary, the dictionary matrix, D, contains the basic features of the image; the coefficient matrix, C, reconstructs the image by combining these base features with additive noise, E. By making different assumptions about the structure, various matrix decomposition models can be derived for tasks such as feature extraction, dimensionality reduction, denoising, and image representation learning.

3.2. Water Reflection Detection Network

This section details the proposed Symmetry-Aware Water Reflection Detection Network. The overall architecture, illustrated in Fig. 4, is composed of three key

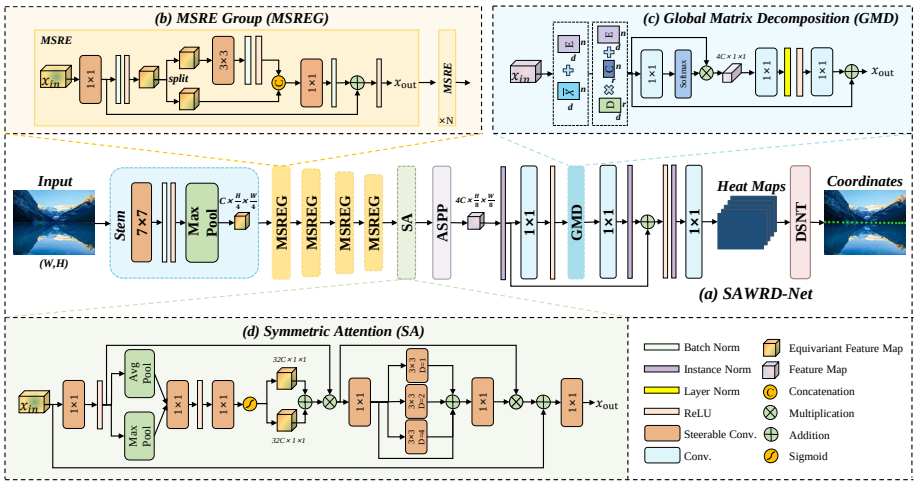

Figure 4: Illustration of the proposed water reflection detection network, SAWRD-Net. (a) depicts the overall architecture of the model. (b) shows the detailed structure of the proposed MSRE block. (c) illustrates the detailed structure of the Global Matrix Decomposition module within the MD decoder. (d) presents the detailed structure of the proposed symmetric attention mechanism.

components: the Multi-scale Reflection Equivariant block, the Symmetric Attention mechanism, and the Matrix Decomposition Decoder.

3.2.1. Multi-scale Reflection Equivariant Convolutional Block

Water reflection normally manifests an approximate mirror symmetry. However, the reflection patterns often exhibit discrete rotational or flipping deformations due to water surface fluctuations and variations in camera viewing angles and illumination directions during the imaging processes. To rigorously preserve the equivariance to these spatial transformations during feature extraction, while simultaneously balancing representation ability and computational efficiency, we employ $E(2)$-equivariant convolutions [32] based on the dihedral group, D_8, to construct a MSRE block, where the D_8 group comprises eight discrete rotations (with angles as integer multiples of $\frac{2\pi}{8}$) and reflection transformations. We first perform a linear transformation along the channel dimension of input features through 1×1 equivariant convolutions.

Subsequently, the transformed feature tensor is partitioned into two mutually exclusive subspaces along the channel dimension. In the first subspace, we employ 3×3

equivariant dilated convolution kernels that expand the receptive field while preserving group representation constraints, effectively capturing multi-scale contextual information. The second subspace retains original features to preserve low-level detail information. The processed outputs from both subspaces are then reintegrated through channel-wise concatenation, followed by a 1×1 equivariant convolution to achieve unified fusion of multi-level features.

This design paradigm simultaneously enables feature reuse for enhanced discriminative power and hierarchical extraction of multi-scale features through diversified receptive fields. Such architecture ensures effective multi-scale feature learning while strictly maintaining equivariance properties.

3.2.2. Symmetric Attention Mechanism

In general, the scale of water reflection varies while the reflective area often occupies only a small portion of the image. The intensity and morphology of the reflective area are influenced by external factors, such as illumination conditions, water quality, and surface fluctuations. Therefore, a detection network should be able to selectively enhance reflection-related characteristics while suppressing irrelevant background regions. To this end, we propose a composite Symmetric Attention (SA) mechanism inspired by the Equivariant Steerable Convolutional Neural Network (ESCNN) introduced by Weiler and Cesa [32]. ESCNN uses channel attention [33] and multi-scale spatial features to represent complex reflection patterns.

Our channel attention module utilizes a symmetric dual-path structure to extract global feature statistics. It applies Pointwise Adaptive Average Pooling and Pointwise Adaptive Max Pooling operations in parallel to capture both the average activation levels and the salient feature responses of the feature maps, thereby characterizing the channel-wise information distribution.

The feature transformation is then performed by a two-layer shared parameter bottleneck structure. This structure first uses an equivariant convolution 1×1 to compress the channel dimension to $\frac{1}{4}$ of its original size, enhancing the interaction of features. After a non-linear activation with an Equivariant ReLU, another 1×1 equivariant convolution restores the original channel dimension. Finally, the channel attention weights

are generated via a pointwise non-linear function, formally expressed as:

$$H_{n+1} = H_n \otimes \Big(\sigma\Big(\mathcal{W}_2(\delta(\mathcal{W}_1(\text{AvgPool}(H_n))))) + \mathcal{W}_2(\delta(\mathcal{W}_1(\text{MaxPool}(H_n))))\Big)\Big), \tag{8}$$

where $\otimes$ represents element-wise multiplication, H represents the feature map, H_n represents the feature map before input, and H_{n+1} represents the feature map after input. $\mathcal{W}_1$ and $\mathcal{W}_2$ represent equivariant convolutions 1×1 for the reduction and expansion of the dimensionality, respectively, δ denotes the activation of ReLU and σ represents the sigmoid function.

To capture multi-scale reflection features, a spatial feature extraction module with heterogeneous receptive fields is introduced. This module adopts a multi-branch architecture. The basic branch employs a standard 3×3 equivariant group convolution (dilation rate = 1) to capture fine-grained local details. Two multi-scale branches utilize 3×3 equivariant dilated convolutions with dilation rates of 2 and 4, respectively, to progressively expand the receptive field and model spatial context at different scales, and padding is used to ensure that the feature maps output by each branch have the same size. The outputs of all branches are aggregated via residual connections and fused using a 1×1 equivariant convolution. The resulting features are then multiplied element-wise with the original input features.

This multi-scale design enables the model to capture both local details and global distribution characteristics of water reflections, significantly improving its adaptability. As visualized in Fig. 5, this process effectively enhances the regions of interest. The operation can be formally expressed as

$$H_{n+1} = H_n \otimes \left(\mathcal{W}_c\left(\sum_{d\in\{1,2,4\}} \mathcal{W}_d(H_n)\right)\right), \tag{9}$$

where $\otimes$ represents element-wise multiplication, H represents the feature map, H_n represents the feature map before input, and H_{n+1} represents the feature map after input. $\mathcal{W}_d$ denotes the convolution operation with the dilation rate d, and $\mathcal{W}_c$ represents the feature fusion layer. This architecture explicitly establishes cross-scale representations, bridging pixel-level details with region-level context.

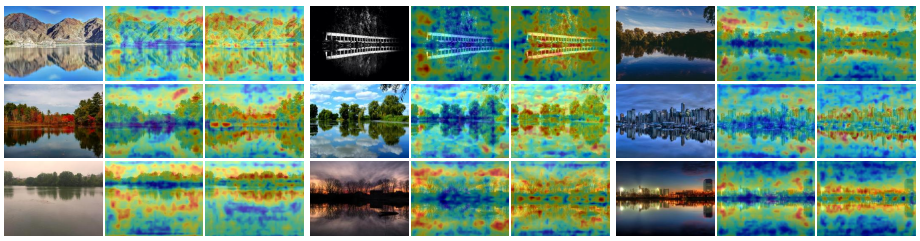

Figure 5: Visualization of attention maps. Here, column "Image" represents the input images fed into the model, column "Before" indicates the visualization of feature maps prior to the application of symmetric attention, and column "After" shows the visualization of feature maps following the utilization of symmetric attention. The visualization results clearly demonstrate that our symmetric attention mechanism can effectively enhance the weights of the regions of interest.

3.2.3. Matrix Decomposition Decoder

Global contextual dependency is decisive in water reflection detection, as most reflections exhibit long-range characteristics and are rendered as an imperfect reflective symmetry due to external factors such as water waves and illumination. Convolutional operations, while effective for local pattern extraction, inherently lack the ability to explicitly model these global spatial relationships. In contrast, matrix decomposition offers significant advantages for this task. Its intrinsic low-rank property effectively models the global contextual dependencies in water reflection images, and it is capable of eliminating the influence of noise through its decomposition process.

Although the self-attention mechanism can also model global dependencies, it has two limitations in water reflection detection. First, the space complexity of it is $O(n^2d)$ (where $n = H \times W$ is the number of spatial locations and d is the number of channels). Second, it is susceptible to noise interference. In contrast, matrix decomposition has a space complexity of only $O(ndr)$ (where the low-rank rank $r \ll n$), and can explicitly separate noise from the global structure.

Therefore, our decoder is designed to model long-range dependencies and filter noise by building on the principles of matrix decomposition [31]. The process begins by preparing the features: we first apply Instance Normalization [34] to standardize the statistics of each channel, followed by a 1×1 convolution with a ReLU activation to linearly map the features into a suitable embedding space. The core of the decoder is

a Global Matrix Decomposition (GMD) block that performs low-rank signal recovery. This module takes the embedded feature maps and decomposes them into the product of a dictionary matrix and a coefficient matrix.

Specifically, the features are modeled as the product of a dictionary and coefficients plus noise. A low-rank reconstruction is obtained by alternately updating the coefficients and the dictionary for six iterations using multiplicative update rules. The objective function converges linearly and stabilizes after six iterations. The computational complexity is $O(ndr)$. Each forward propagation takes around 15 milliseconds, and the memory usage is one-fifth that of the self-attention mechanism. This process is key to effectively extracting the essential global correlations from the scene while simultaneously suppressing the high-frequency noise caused by water surface distortions.

In the GMD block, after obtaining the low-rank signal subspace, this subspace is subsequently utilized to aggregate global contextual information. First, a 1×1 convolution and a softmax function are applied to the subspace representation to generate attention-like weights. These weights are then used in a pooling mechanism to produce a globally aggregated feature vector through weighted averaging. This vector represents a summary of the entire scene's global context. Next, this global context vector is refined and fused with the original local features. A bottleneck structure is employed to perform non-linear transformations on the aggregated vector, which captures complex inter-channel dependencies. These global contextual features are then fused with the original position-wise features through element-wise addition. This critical step enriches the local features with an understanding of the global scene, allowing the model to make more informed predictions.

Finally, an output module projects the fused features into the final heat maps. This begins with a linear transformation, using a 1×1 convolution and a ReLU activation, to reduce the parameter count. Subsequently, Instance Normalization [34] is reapplied to stabilize the channel-wise statistics of the output, followed by a final 1×1 convolution for the channel-wise linear transformation that generates the heat maps. This overall decoder design excels at modeling the long-range dependencies inherent in water reflection. By explicitly extracting the underlying low-rank data structure, it

demonstrates superior performance in water reflection detection tasks.

4. Experimental Setup

To rigorously validate the performance of our proposed SAWRD-Net and to dissect the contributions of its core architectural components, we designed a multi-faceted experimental protocol. Our evaluation is centered on the large-scale Water Reflection Scene Data set (WRSD) benchmark [1], where SAWRD-Net is compared against a comprehensive suite of state-of-the-art methods. We also conduct a series of detailed ablation studies to systematically isolate and quantify the impact of each architectural innovation. All other design choices and the training/evaluation protocol follow the setup defined earlier, enabling unambiguous attribution of each contribution.

4.1. Data Set

As the largest currently available data set for this task, WRSD [1] contains 2,117 high-quality images, each with accurately manually annotated symmetry axis points. The data set is characterized by its diversity. Specifically, the symmetry axes are presented at various tilt angles, and the scenes encompass a rich variety of content with a wide range of scale variations, providing a comprehensive set of test samples.

4.2. Evaluation Metric

The primary evaluation metric is the True Positive (TP) Rate [35], which considers both the angular deviation ($\Delta\theta$) and the midpoint distance (Δd) between the predicted and ground-truth axes. A prediction is classified as TP if its angular deviation is below a threshold θ and its distance deviation is below a threshold d; otherwise, it is marked as a False Positive (FP). The TP Rate is then calculated as:

$$\text{TP Rate} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{10}$$

Following the settings in [1], we set the distance threshold d to 2.5% of the image height (i.e., $d = 7.5$ for a 300-pixel tall image) and the angle threshold θ to $\frac{\pi}{60}$ radians (i.e., $\theta = 3°$). To assess the stability and reliability of each method, we also report the mean and standard deviation of the angle and distance differences for all test images.

4.3. Baselines

To comprehensively and fairly evaluate SAWRD-Net, we compare it against five families of baseline methods: (i) a task-specific water reflection detector, WRD-Net [1]; (ii) four detectors for reflection or rotational-symmetry [14, 15, 13]; (iii) ten strong dense-prediction or segmentation backbones [36, 37, 38, 39, 40, 41, 42, 43, 44, 45]; (iv) five classical symmetry methods [16, 46, 28, 27, 29]; and (v) two typical object detectors [12, 47]. We adapt each method to the water reflection *axis-point* task using the authors' released code where available, and compare predicted axes with the ground-truth data.

4.4. Implementation Details

The images are divided into two categories, in which the reflection axis is approximately horizontal or vertical. All images are fed into the network at the original resolution of 400×300 pixels. Data augmentation is not used during the training process. Model weights are saved at the end of each training epoch and early stopping is not employed. We use the Adam optimizer. The learning rate, weight decay and mini-batch size are set to $5{\times}10^{-4}$, 10^{-4} and 2, respectively. The learning rate remains constant throughout the training process without a decay strategy. In total, each model is trained for 100 epochs. A three-fold *image-level* cross-validation is conducted, which is identical to that used by WRD-Net [1]. Performance metrics are computed over all test images across the three folds using the same TP Rate definition and thresholds ($\theta = \pi/60$, $d = 2.5\%$ of image height).

SAWRD-Net is implemented in Python 3.8.20 with PyTorch 1.11.0 and CUDA 11.5. All experiments are performed on a single NVIDIA RTX 3090 GPU. All baselines are implemented in PyTorch, with an identical DSNT[48] head appended to regress reflection-axis keypoints. Although PMCNet [15] and EquiSym [13] provide pre-trained weights on symmetric data sets, the training on the WRSD [1] can improve detection accuracy. To ensure fair comparison, we do not use pre-trained weights unless otherwise specified. That is, all baselines are trained *from scratch* under a unified protocol.

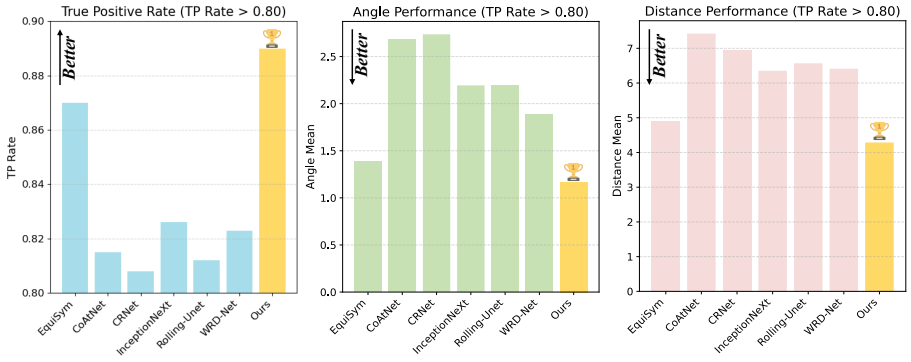

Figure 6: Visualization of comparison with state-of-the-art methods. We visualized models with a TP Rate greater than 0.80, demonstrating the superiority of our method across various evaluation metrics such as TP Rate, angle, and distance.

5. Experimental Results

Following the settings detailed in Section 4, experiments were performed. This section elucidates the subsequent results.

5.1. Comparison with State-of-the-Art Methods

Table 1 demonstrates the performance of the proposed SAWRD-Net against 22 state-of-the-art methods. Results marked with † are quoted from [1]; all other entries were obtained under the same experimental settings (Sec. 4-4). Under the standard thresholds ($\theta = \pi/60$, $d = 2.5\%$ of image height), SAWRD-Net achieves a TP Rate [35] of 0.890 and obtains the best angle and distance statistics in terms of mean and standard deviation. None of the baseline methods outperform SAWRD-Net. Fig. 6 visualizes the TP Rate, angle mean and distance mean values derived using the models which produce a TP Rate greater than 0.80. The results show that our method performs best across the three evaluation metrics.

Deep learning methods generally surpass traditional symmetry detectors on WRSD. Traditional approaches often assume near-perfect symmetry and degrade under the imperfect symmetry of natural water reflections, which are affected by ripples, refraction, and scattering. For dense-prediction baselines, we evaluate strong image-segmentation models because our task involves regressing 20 points along the reflection axis and thus

Table 1: Comparison between the baselines and SAWRD-Net in terms of different performance metrics.

Method	TP Rate [35]	Angle			Distance		
		Mean	Var.	Std.	Mean	Var.	Std.
Loy and Eklundh's† [16] (ECCV 2006)	0.682	4.478	274.684	16.574	41.270	48727.560	220.743
Cicconet et al.'s† [46] (CVPR 2014)	0.512	32.000	1736.433	41.671	28.298	1435.399	37.887
Cicconet et al.'s† [28] ICCV 2017)	0.173	24.754	1194.133	34.556	38.372	1274.303	35.697
Elawady et al.'s† [27] (ICCV 2017)	0.184	43.768	1936.014	44.000	26.120	697.579	26.412
Gnutti et al.'s† [29] (TIP 2021)	0.529	5.257	377.350	19.426	22.407	1169.404	34.197
Sym-VGG † [14] (ICCV 2017)	0.330	11.138	807.850	28.423	19.866	437.702	20.921
Sym-ResNet† [14] (ICCV 2017)	0.594	4.866	175.251	13.238	13.077	496.902	22.314
PMCNet† [15] (ICCV 2021)	0.670	5.533	228.725	15.124	12.664	589.244	24.274
UNet† [36] (MICCAI 2015)	0.461	5.860	224.266	14.976	18.830	576.884	24.018
XNet† [37] (SPIE 2019)	0.234	9.160	301.653	17.368	30.607	727.304	26.969
EquiSym [13] (CVPR 2022)	0.870	1.389	61.655	7.852	4.898	175.530	13.248
CoAtNet [38] (NeurIPS 2021)	0.815	2.681	158.055	12.572	7.423	288.507	16.985
CRNet [39] (JKSUCI 2024)	0.808	2.736	133.182	11.540	6.948	292.675	17.107
FocalNet [40] (NeurIPS 2022)	0.772	3.317	166.455	12.901	7.852	273.159	16.527
HRNet [41] (CVPR 2019)	0.795	2.706	120.416	10.973	7.763	328.204	18.116
InceptionNeXt [42] (CVPR 2024)	0.826	2.194	108.900	10.435	6.353	215.829	14.691
Rolling-Unet [43] (AAAI 2024)	0.812	2.195	101.183	10.058	6.568	238.807	15.453
STViT [44] (CVPR 2023)	0.378	8.667	570.263	23.880	18.088	442.536	21.036
YOLOv12 [12] (NeurIPS 2025)	0.785	2.802	125.893	11.220	7.763	266.749	16.332
YOLO26 [47] (arXiv 2025)	0.794	3.108	154.023	12.410	7.358	255.281	15.977
SegFormer [45] (NeurIPS 2021)	0.798	2.008	124.696	11.700	6.373	251.449	15.857
WRD-Net †[1] (PR 2024)	0.823	1.882	81.395	9.022	6.413	250.276	15.820
SAWRD-Net (Ours)	**0.890**	**1.168**	**52.916**	**7.272**	**4.287**	**147.818**	**12.158**

The results of the baselines marked with † were directly obtained from [1].

relates to pixel-level prediction. Although these backbones perform well at pixel classification, they lack explicit symmetry priors and therefore yield lower TP Rates [35].

In terms of the two state-of-the-art YOLO-based detectors, i.e., YOLOv12 [12] and YOLO26 [47], we discarded the detection head that generates prediction bounding boxes but using the DSNT [48] head to predict points of the symmetric axis. Since YOLO-based methods were originally designed for sparse bounding box detection, they cannot produce high-resolution dense features or perform explicit global geometric modeling.

Among reflection-symmetry detection models, EquiSym [13] performs best with a TP Rate of 0.870, which is inferior to SAWRD-Net. This gap reflects the difference between idealized mirror symmetry targeted by generic symmetry detectors and the

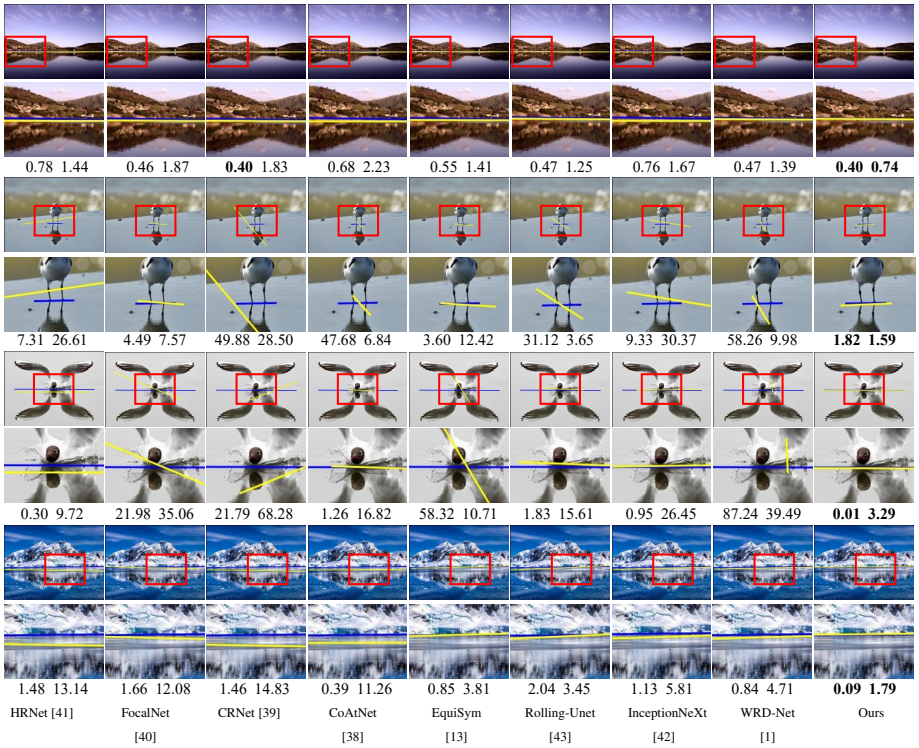

Figure 7: Qualitative results of water reflection detection on the WRSD. Within each group of the figure, the upper row shows the detection results of eight baselines and SAWRD-Net, while the lower row shows the zoomed-in view of a selected region in the corresponding image. The yellow lines represent the axes detected by the models, while the blue lines denote the ground-truth. The two values displayed below an image indicate the angle and distance calculated between the detected axis and the ground-truth axis.

imperfect symmetry present in water scenes. By encoding symmetry in the encoder and suppressing noise via matrix decomposition in the decoder to recover low-rank structure, SAWRD-Net achieves the top TP Rate value. Compared to WRD-Net [1], SAWRD-Net introduces a symmetric attention mechanism and equivariant convolutions to explicitly model reflection symmetry, while replacing computationally intensive ViT branches with the efficient MSRE groups and matrix decomposition decoder, thus reducing computational cost.

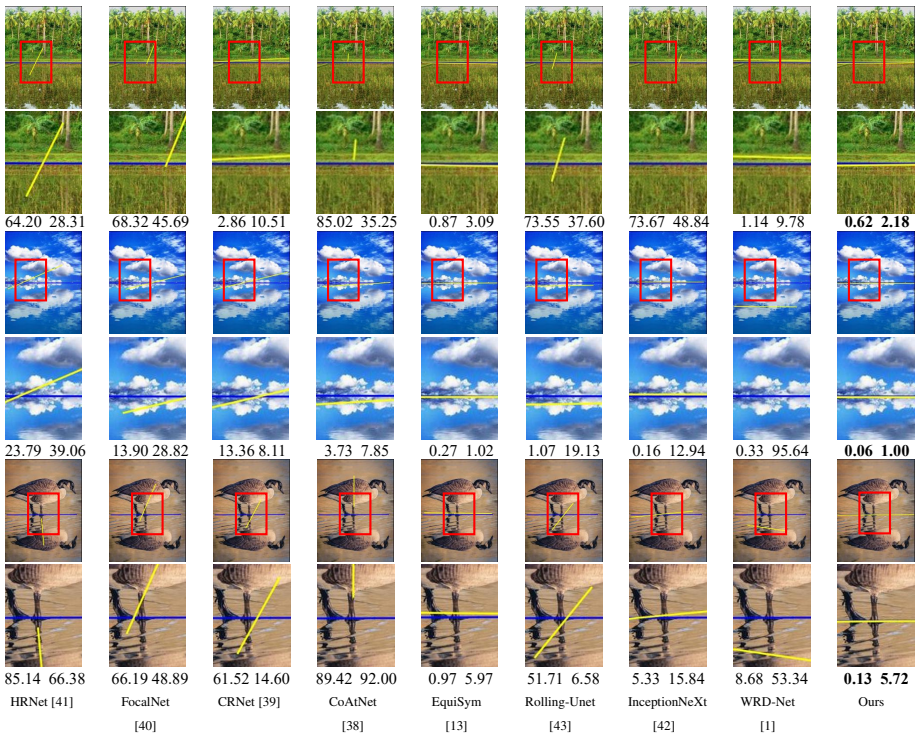

Figure 8: Qualitative results of water reflection detection on the WRSD. Note that the images displayed here are those used in the experiment after being rotated clockwise by 90°. Within each group of the figure, the upper row shows the detection results of eight baselines and SAWRD-Net, while the lower row shows the zoomed-in view of a selected region in the corresponding image. The yellow lines represent the axes detected by the models, while the blue lines denote the ground-truth. The two values displayed below an image indicate the angle and distance calculated between the detected axis and the ground-truth axis.

5.2. Qualitative Results

The qualitative comparison between SAWRD-Net and eight baseline methods (HR-Net [41], FocalNet [40], CRNet [39], CoAtNet [38], EquiSym [13], Rolling-Unet [43], InceptionNeXt [42], and WRD-Net [1]) is shown in Figs. 7 and 8. These baselines were selected based on their top TP Rate rankings in Table 1. Figure 7 displays the original experimental images, whereas Fig. 8 presents the same images rotated 90° clockwise to aid visualization.

Across the examples, SAWRD-Net exhibits closer alignment with the ground-truth annotations and consistently smaller angular and spatial deviations than the compared

Figure 9: Examples of the water reflectance images suffered from three typical interferences, including ripples, light scattering, and partial reflections. Within each group, the degree of the interference increases from the left to the right.

methods. In challenging scenarios involving water waves or blurred reflective surfaces, SAWRD-Net maintains robust performance relative to the baselines. Taken together, these observations indicate that SAWRD-Net achieves higher detection accuracy and robustness than the existing baseline methods.

5.3. Robustness to Different Interferences

Despite achieving the highest TP Rate value, SAWRD-Net may encounter performance degradation when different interferences occur, such as ripples, light scattering, and partial reflections. To evaluate the robustness of SAWRD-Net to these interferences, we first randomly selected 100 images per interference type from the WRSD [1]. In total, three subsets were obtained. Given a subset, six example images are shown in Fig. 9. SAWRD-Net was then applied to each subset. In terms of the ripple, light scattering, and partial reflection interferences, the TP Rate values obtained were 0.85, 0.91 and 0.87, respectively. As displayed in Fig. 9, SAWRD-Net can still detect the axis of reflectance in many challenging scenarios interfered by ripples, light scattering, and partial reflections. These results demonstrate that SAWRD-Net is robust to these interferences.

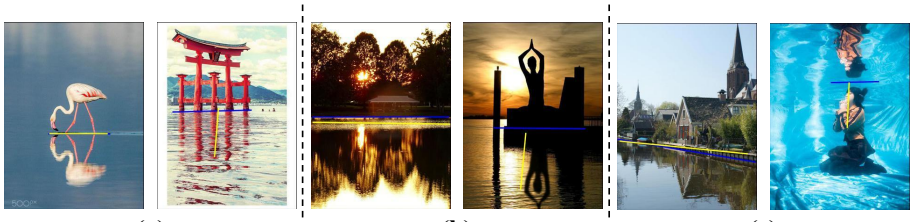

Figure 10: With regard to each of the three interferences, two failure cases are shown, in which (a) shows the detection results for different degrees of ripples; (b) presents the detection results under different degrees of light scattering; and (c) displays the detection results for different degrees of partial reflections.

5.4. Failure Cases

In terms of the three interferences, three sets of failure cases are shown in Fig. 10, respectively. Within each set, two images are shown with progressively increasing interference degrees. For ripple interference, the Canny edge detector [49] and the Hough transform [50] are used to extract ripple edges, and the degree of ripples is characterized by the range of the midpoint positions of the edges. Weak ripples (142 pixels) cause only minor deviations, while strong ripples (232 pixels) lead to larger detection errors. Regarding light scattering interference, the impact of illumination is evaluated using the ratio of saturated pixels and Sobel-based texture energy. Both overexposure and underexposure affect detection accuracy. However, weak illumination (only 5% saturated pixels) reduces texture and contrast, resulting in poor detection performance. For partial reflection interference, the threshold-based binarization and connected component analysis are adopted to analyze the degree of partial reflections. Low partial reflections (largest connected component ratio of 92%, 5 fragments) still yield coarse detection results, while high partial reflections (ratio of 30.5%, 18 fragments) may lead to a complete failure.

To address the above-mentioned failures, future improvements could be implemented as follows. Deformable convolutions can be introduced for strong ripple deformations. Graph attention mechanisms can be applied to overcome fragmented reflections. A self-calibrated illumination module or frequency-domain features can be used to alleviate extreme lighting conditions. These solutions have potential to improve the

Table 2: Comparison between the baselines and our SAWRD-Net in terms of computational complexity (FLOPs), number of parameters and average inference time (S).

Method	FLOPs (G)	Params (M)	Average Inference Time (S)
Loy and Eklundh's† [16] (ECCV 2006)	N/A	N/A	0.2551
Cicconet et al.'s† [46] (CVPR 2014)	N/A	N/A	121.5111
Cicconet et al.'s† [28] (ICCV 2017)	N/A	N/A	0.4110
Elawady et al.'s† [27] (ICCV 2017)	N/A	N/A	4.7615
Gnutti et al.'s† [29] (TIP 2021)	N/A	N/A	1079.8299
Sym-VGG † [14] (ICCV 2017)	90.98	37.87	0.0126
Sym-ResNet† [14] (ICCV 2017)	156.20	19.95	0.0456
PMCNet† [15] (ICCV 2021)	227.62	16.76	0.0322
UNet† [36] (MICCAI 2015)	56.83	13.40	0.0107
XNet† [37] (SPIE 2019)	95.41	26.64	0.0155
EquiSym [13] (CVPR 2022)	42.15	16.44	0.0239
CoAtNet [38] (NeurIPS 2021)	12.85	17.87	0.0095
CRNet [39] (JKSUCI 2024)	18.99	36.51	0.0267
FocalNet [40] (NeurIPS 2022)	109.45	59.84	0.0213
HRNet [41] (CVPR 2019)	43.26	65.40	0.0372
InceptionNeXt [42] (CVPR 2024)	46.58	86.69	0.0139
Rolling-Unet [43] (AAAI 2024)	4.80	1.78	0.0214
STViT [44] (CVPR 2023)	14.47	25.52	0.0213
YOLOv12 [12] (NeurIPS 2025)	0.92	2.14	0.0112
YOLO26 [47] (arXiv 2025)	0.88	2.27	0.0089
SegFormer [45] (NeurIPS 2021)	3.45	3.72	0.0078
WRD-Net †[1] (PR 2024)	74.06	19.03	0.0789
SAWRD-Net (Ours)	51.11	23.97	0.0434

The results of the baselines marked with † were directly obtained from [1].

detection stability of our SAWRD-Net within an end-to-end framework.

5.5. Performance Analysis

As shown in Table 2, SAWRD-Net achieves a good balance among computational complexity, number of parameters, and inference speed. Compared to WRD-Net [1], it reduces FLOPs by around 31% and also improves inference speed, with a moderate number of parameters. In particular, the GPU memory usage is around 587 MB during the inference stage while the peak memory usage (with $batchsize = 2$) is around 2.9 GB in the training process. In general, SAWRD-Net demonstrates low computational complexity and moderate memory requirements, making it suitable for practical deployment.

Table 3: TP Rate with and without equivariant convolution for training.

Design	TP Rate [35]
w/o Equiv.	0.797
w/ Equiv.	**0.890**

5.6. Ablation Studies

Unless otherwise noted, each ablation experiment varies a single factor under the unified protocol in Section 4.4. We report the TP Rate as defined in [35].

5.6.1. $E(2)$-Equivariant Convolutions

The evaluation demonstrates the efficacy of equivariant learning for water reflection detection. We conducted a controlled comparison between models implemented with ESCNN [32] and conventional CNN architectures, keeping all configurations identical except for the convolution operation. As shown in Table 3, the conventional CNN model achieved a TP Rate [35] of 0.797, whereas the equivariant-convolution variant attained 0.890. These results substantiate that adopting $E(2)$-equivariant convolutions positively impacts performance, with the complete architecture yielding optimal detection capability.

5.6.2. Dihedral Group

To evaluate the effectiveness of the ESCNN [32] with different dihedral groups used by our SAWRD-Net, we conducted a comparative experiment, which involves two models implemented based on the D_4 and D_8 groups, respectively. The other configurations were kept unchanged. As shown in Table 4, the D_4-based model achieved a TP Rate value of 0.888, which is slightly lower than that derived using the model with D_8. This finding demonstrates that the choice of D_8 is reasonable.

5.6.3. Effect of Symmetric Attention

We conduct an ablation of the SA mechanism inserted between the backbone and the ASPP module [51] under identical experimental settings. As shown in Table 5, the

Table 4: TP Rate with different dihedral groups during the training stage.

Design	TP Rate [35]
w/ D_4	0.888
w/ D_8	**0.890**

Table 5: TP Rate with and without symmetric attention for training.

Design	TP Rate [35]
w/o SA	0.878
w/ SA	**0.890**

model without SA achieves a TP Rate [35] of 0.878, whereas enabling SA improves the TP Rate to 0.890, yielding an absolute gain of +0.012. These results indicate that SA contributes positively to performance, supporting its inclusion in the complete architecture.

5.6.4. Matrix-Decomposition Decoder

To evaluate the performance of the proposed Matrix Decomposition Decoder, we conducted targeted ablation experiments. We first compared the MD decoder with the Hamburger decoder [31]. Although Hamburger has demonstrated strong performance as a general-purpose decoder in segmentation tasks, we observed limitations on the specific task of water reflection detection. Building on the general architectural pattern of Hamburger, we introduced an MD decoder tailored to the characteristics of water reflection detection. As shown in Table 6, using the standard Hamburger yielded a TP Rate [35] of 0.878, whereas employing the MD decoder improves the TP Rate to 0.890 (absolute gain +0.012), validating the effectiveness of the proposed design.

Subsequently, to further examine the decoding advantages of the matrix-decomposition mechanism, we compared the MD decoder with the EquiSym decoder under a fixed encoder. Notably, EquiSym is a fully symmetric architecture constructed entirely from equivariant convolutions; since our encoder also employs equivariant convolutions, this

Table 6: TP Rate with hamburger or our decoder for training.

Module	TP Rate [35]
w/ Hamburger	0.878
w/ Our Decoder	**0.890**

Table 7: Impact of matrix decomposition on water reflection detection.

Design	TP Rate [35]
w/o MD	0.879
w/ MD	**0.890**

comparison isolates the impact of the decoder architecture while keeping the encoder unchanged. As reported in Table 7, the EquiSym decoder attains a TP Rate of 0.879, whereas the MD decoder achieves 0.890 (absolute gain +0.011). These results confirm the advantage of matrix decomposition for handling water reflection features and also suggest synergistic effects between the network's symmetry-aware encoder and the proposed decoder.

6. Conclusion

We presented SAWRD-Net, a symmetry-aware network for water reflection detection that integrates $E(2)$-equivariant learning with matrix decomposition. Building on symmetry-detection principles, SAWRD-Net leverages the imperfect reflective symmetry of natural water scenes by using equivariant convolutions and a symmetric attention mechanism to extract geometry-consistent, discriminative features and emphasize reflection-relevant regions. Treating axis localization as a structured modeling problem, we adopted a matrix decomposition decoder to capture global context and suppress noise, yielding a complementary synergy with the encoder. On the WRSD benchmark, SAWRD-Net attained state-of-the-art performance (TP Rate 0.890 at $\theta = \pi/60$, $d = 2.5\%$ of image height), and ablation experiments verified the contributions of

equivariant features, SA placement, and the MD design. However, SAWRD-Net struggles with some challenging scenarios, such as severe ripples, extreme illumination, and fragmented reflections. In our future work, we will manage to address these challenges.

CRediT authorship contribution statement

Shuxuan Yao: Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing - original draft; **Chengjia Wang:** Formal analysis, Methodology, Writing - Review & Editing; **Jianyuan Sun:** Writing – review & editing; **Junyu Dong:** Resources, Funding acquisition; **Xinghui Dong:** Conceptualization, Funding acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Writing - Review & Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was supported in part by the National Natural Science Foundation of China (NSFC) (No. 42576200) and in part by the Key Research and Development Program of Shandong Province, China (No. 2024ZLGX06)

Data availability

Data will be made available on request.

References

[1] H. Dong, H. Qi, H. Zhou, J. Dong, X. Dong, WRD-Net: Water Reflection Detection using a parallel attention transformer, Pattern Recognition 152 (2024) 110467.

[2] S.-H. Zhong, Y. Liu, Y. Liu, C.-S. Li, Water reflection recognition based on motion blur invariant moments in curvelet space, IEEE Transactions on Image Processing 22 (2013) 4301–4313.

[3] H. Zhang, X. Guo, X. Cao, Water reflection detection using a flip invariant shape detector, in: 2010 20th International Conference on Pattern Recognition, IEEE, 2010, pp. 633–636.

[4] S.-h. Zhong, Y. Liu, L. Shao, F.-l. Chung, Water reflection recognition via minimizing reflection cost based on motion blur invariant moments, in: Proceedings of the 1st ACM International Conference on Multimedia Retrieval, 2011, pp. 1–8.

[5] T. P. Nguyen, H. P. Truong, T. T. Nguyen, Y.-G. Kim, Reflection symmetry detection of shapes based on shape signatures, Pattern Recognition 128 (2022) 108667.

[6] M. Zha, Y. Pei, G. Wang, T. Li, Y. Yang, W. Qian, H. T. Shen, Weakly-supervised mirror detection via scribble annotations, in: Proceedings of the AAAI conference on artificial intelligence, volume 38, 2024, pp. 6953–6961.

[7] T. Huang, B. Dong, J. Lin, X. Liu, R. W. Lau, W. Zuo, Symmetry-aware transformer-based mirror detection, in: Proceedings of the aaai conference on artificial intelligence, volume 37, 2023, pp. 935–943.

[8] D. Podgorelec, I. Kolingerová, L. Lovenjak, B. Žalik, AHiLS—An Algorithm for Establishing Hierarchy among Detected Weak Local Reflection Symmetries in Raster Images, Symmetry 16 (2024) 442.

[9] H. Tang, C. Yuan, Z. Li, J. Tang, Learning attention-guided pyramidal features for few-shot fine-grained recognition, Pattern Recognition 130 (2022) 108792.

[10] H. Tang, Z. Li, D. Zhang, S. He, J. Tang, Divide-and-conquer: Confluent triple-flow network for RGB-T salient object detection, IEEE transactions on pattern analysis and machine intelligence 47 (2024) 1958–1974.

[11] H. Huang, X. Zhou, J. Cao, R. He, T. Tan, Vision transformer with super token sampling, arXiv preprint arXiv:2211.11167 (2022).

[12] Y. Tian, Q. Ye, D. Doermann, YOLOv12: Attention-Centric Real-Time Object Detectors, arXiv preprint arXiv:2502.12524 (2025).

[13] A. Seo, B. Kim, S. Kwak, M. Cho, Reflection and rotation symmetry detection via equivariant learning, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 9539–9548.

[14] C. Funk, Y. Liu, Beyond planar symmetry: Modeling human perception of reflection and rotation symmetries in the wild, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 793–803.

[15] A. Seo, W. Shim, M. Cho, Learning to discover reflection symmetry via polar matching convolution, in: Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 1285–1294.

[16] G. Loy, J.-O. Eklundh, Detecting symmetry and symmetric constellations of features, in: Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part II 9, Springer, 2006, pp. 508–521.

[17] V. S. N. Prasad, L. S. Davis, Detecting rotational symmetries, in: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, volume 2, IEEE, 2005, pp. 954–961.

[18] T. Cohen, M. Welling, Group equivariant convolutional networks, in: International conference on machine learning, PMLR, 2016, pp. 2990–2999.

[19] G. Marola, On the detection of the axes of symmetry of symmetric and almost symmetric planar images, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1989) 104–108.

[20] H. Ogawa, Symmetry analysis of line drawings using the Hough transform, Pattern Recognition Letters 12 (1991) 9–12.

[21] C. Sun, D. Si, Fast reflectional symmetry detection using orientation histograms, Real-Time Imaging 5 (1999) 63–74.

[22] Y. Keller, Y. Shkolnisky, An Algebraic Approach to Symmetry Detection., in: ICPR (3), 2004, pp. 186–189.

[23] T. Tuytelaars, A. Turina, L. Van Gool, Noncombinatorial detection of regular repetitions under perspective skew, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2003) 418–432.

[24] S. Lee, Y. Liu, Curved glide-reflection symmetry detection, IEEE transactions on pattern analysis and machine intelligence 34 (2011) 266–278.

[25] D. Cai, P. Li, F. Su, Z. Zhao, An adaptive symmetry detection algorithm based on local features, in: 2014 IEEE Visual Communications and Image Processing Conference, IEEE, 2014, pp. 478–481.

[26] I. Atadjanov, S. Lee, Bilateral symmetry detection based on scale invariant structure feature, in: 2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2015, pp. 3447–3451.

[27] M. Elawady, C. Ducottet, O. Alata, C. Barat, P. Colantoni, Wavelet-Based Reflection Symmetry Detection via Textural and Color Histograms., in: ICCV Workshops, 2017, pp. 1725–1733.

[28] M. Cicconet, D. G. Hildebrand, H. Elliott, Finding mirror symmetry via registration and optimal symmetric pairwise assignment of curves: Algorithm and results, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 1759–1763.

[29] A. Gnutti, F. Guerrini, R. Leonardi, Combining appearance and gradient information for image symmetry detection, IEEE Transactions on Image Processing 30 (2021) 5708–5723.

[30] H. Weyl, Symmetry, Princeton University Press, 2015.

[31] Z. Geng, M.-H. Guo, H. Chen, X. Li, K. Wei, Z. Lin, Is Attention Better Than Matrix Decomposition?, in: International Conference on Learning Representations, 2021.

[32] M. Weiler, G. Cesa, General e (2)-equivariant steerable cnns, Advances in neural information processing systems 32 (2019).

[33] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3–19.

[34] D. Ulyanov, A. Vedaldi, V. Lempitsky, Instance normalization: The missing ingredient for fast stylization, arXiv preprint arXiv:1607.08022 (2016).

[35] R. Nagar, S. Raman, Reflection symmetry detection by embedding symmetry in a graph, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 2147–2151.

[36] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer, 2015, pp. 234–241.

[37] J. Bullock, C. Cuesta-Lázaro, A. Quera-Bofarull, XNet: a convolutional neural network (CNN) implementation for medical x-ray image segmentation suitable for small datasets, in: B. Gimi, A. Krol (Eds.), Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging, volume 10953, International Society for Optics and Photonics, SPIE, 2019, pp. 453 – 463. URL: https://doi.org/10.1117/12.2512451. doi:10.1117/12.2512451.

[38] Z. Dai, H. Liu, Q. V. Le, M. Tan, Coatnet: Marrying convolution and attention for all data sizes, Advances in neural information processing systems 34 (2021) 3965–3977.

[39] X. Wen, A. Zhang, C. Lin, X. Pang, CRNet: Cascaded Refinement Network for polyp segmentation, Journal of King Saud University-Computer and Information Sciences 36 (2024) 102250.

[40] J. Yang, C. Li, X. Dai, J. Gao, Focal modulation networks, Advances in Neural Information Processing Systems 35 (2022) 4203–4217.

[41] K. Sun, B. Xiao, D. Liu, J. Wang, Deep High-Resolution Representation Learning for Human Pose Estimation, in: CVPR, 2019.

[42] W. Yu, P. Zhou, S. Yan, X. Wang, Inceptionnext: When inception meets convnext, in: Proceedings of the IEEE/cvf conference on computer vision and pattern recognition, 2024, pp. 5672–5683.

[43] Y. Liu, H. Zhu, M. Liu, H. Yu, Z. Chen, J. Gao, Rolling-unet: Revitalizing mlp's ability to efficiently extract long-distance dependencies for medical image segmentation, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, 2024, pp. 3819–3827.

[44] H. Huang, X. Zhou, J. Cao, R. He, T. Tan, Vision Transformer with Super Token Sampling, arXiv:2211.11167 (2022).

[45] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo, SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers, in: Neural Information Processing Systems (NeurIPS), 2021.

[46] M. Cicconet, D. Geiger, K. C. Gunsalus, M. Werman, Mirror symmetry histograms for capturing geometric properties in images, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 2981–2986.

[47] R. Sapkota, R. H. Cheppally, A. Sharda, M. Karkee, YOLO26: key architectural enhancements and performance benchmarking for real-time object detection, arXiv preprint arXiv:2509.25164 (2025).

[48] A. Nibali, Z. He, S. Morgan, L. Prendergast, Numerical coordinate regression with convolutional neural networks, arXiv preprint arXiv:1801.07372 (2018).

[49] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (1986) 679–698.

[50] R. O. Duda, P. E. Hart, Use of the hough transformation to detect lines and curves in pictures, Communications of the ACM 15 (1972) 11–15.

[51] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE transactions on pattern analysis and machine intelligence 40 (2017) 834–848.