Optics & Laser Technology 192 (2025) 113429

Contents lists available at ScienceDirect

Optics and Laser Technology

journal homepage: www.elsevier.com/locate/optlastec

Joint intensity-spectral polarization hierarchical fusion guided efficient transparent object detection

Xueqiang Fan, Longyu Qiao, Bing Lin, Zhongyi Guo *

School of Computer and Information, Hefei University of Technology, Hefei 230009, China

ARTICLE INFO

Keyword:
Spectral polarization imaging
Transparent object detection
Hierarchical multimodal fusion
2D/3D convolutional neural networks
FuseISP

ABSTRACT

The field of object detection has emerged as a critical and valuable research frontier. Nevertheless, the detection of transparent objects remains an unresolved and challenging problem, primarily due to their limited texture and color information. Towards being able to address this situation, we propose a novel intensity-spectral polarization fusion framework, termed as FuseISP, specifically designed for transparent object discrimination. FuseISP starts by utilizing hierarchical feature extractor for each feature source, *i.e.*, trichromatic intensities or trichromatic linear polarization cues, to produce abundant high- and low-frequency features. Subsequently, we implement an intensity-spectral polarization mixed modulator to enhance interactions between intensity and spectral polarization information. Additionally, FuseISP introduces a new hierarchical feature fusion module to establish connections among different levels for modelling the shared information. Lastly, a multi-level decoder module based on the integration of 2D convolutional neural networks (CNNs) and 3D CNNs, which can simultaneously capture inter- and intra-polarization relationships, is designed to construct the transparent object detector in a deeply supervised manner. Experimental results show our proposed method outperforms other advanced approaches in the real-world scenes.

1. Introduction

With the rise of deep learning (DL) technologies, *e.g.*, CNNs and Vision Transformers, significant progress has been made in scene parsing including object detection. The existing object detection paradigm has made remarkable progress, establishing themselves as fundamental solutions across various applications, including visual tracking [1], robotic navigation [2], and autonomous driving [3,4], and industrial quality inspection [5].

As a subfield of object detection, transparent object detection (TOD) still faces greater challenges than traditional object detection tasks [6,7]. The transparent objects (TO), *e.g.*, glass guardrails and window panes, which are very common in daily life scenes, exhibit the following typical characteristics: *i*) they do not have their own visual appearances; *ii*) they easily confuse the vision systems due to their inherently special properties [8,9], which only transmit/reflect the appearances of their surroundings; *iii*) they mostly share an extremely thin separation boundary with the background region [10]. As shows in Fig. 1, we display two common but challenging scenarios. These challenges are particularly acute for the segmentation tasks that support autonomous driving and underwater submarine navigation, which have very high safety requirements. For addressing this task, TOD has gradually received attention and intensive research.

Most of the current TOD solutions, *e.g.*, MirrorNet [11] and EBLNet [12], dependent on trichromatic (RGB) textures for feature extraction. Nonetheless, the TO is not visible in RGB cameras due to the absence of both texture and colour information. This can result in the model's inability to effectively extract key features of the TO from RGB images, thereby these methods often fail to distinguish the TO from the background.

To remedy this issue, recent noteworthy advancements have explored richer modalities especially polarization information for more robust identification of TO. Unlike the intensity information used in traditional vision systems, the property of polarization provides intriguing physical characteristics of light, enabling the extraction of distinctive information about an object, *e.g.*, its surface smoothness and material composition [13–16]. Fig. 1 illustrates these advantages of polarization imaging. Recently, polarization imaging system becomes commercially available, spawning a wide array of various applications such as polarization scattering imaging [17–23]. A number of

* Corresponding author.
E-mail address: guozhongyi@hfut.edu.cn (Z. Guo).

https://doi.org/10.1016/j.optlastec.2025.113429
Received 15 April 2025; Received in revised form 30 May 2025; Accepted 20 June 2025
Available online 28 June 2025

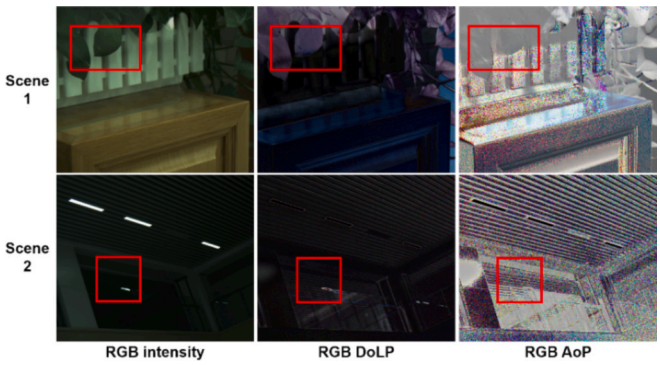

Fig. 1. Typical challenges in transparent object detection. Scene 1: the transparent object area blends into the background environment and lacks texture. Scene 2: low-light results in poor visibility in the transparent object area.

polarization-aided TOD (PTOD) studies, *e.g.*, Polarized CNN [8], have been carried out, and the results report that PTOD exhibits its distinct advantages, especially in strongly jamming environments. These methods mainly exploit Deep Convolutional Neural Networks (DCNNs) that have powerful feature extraction and feature learning capabilities and RGB linear polarization cues, *e.g.*, degree of linear polarization (DoLP) and angle of polarization (AoP). For example, a polarization-based glass segmentation model *i.e.*, PGSNet [24], is developed using RGB intensity, DoLP, and AoP cues extracted from a single image acquired by an RGB polarizing camera as the input for the model.

Nevertheless, these polarization-based approaches still remain a huge challenge for reliable TOD due to the following two-fold obstacles. *Firstly*, the utilization of RGB intensity and linear polarization cues allows neural networks to obtain plentiful features. However, a new question that deserves asking is how to seamlessly merge the coarse- to fine-level features with varying polarization states across different scales and depths. *Secondly*, the joint analysis among 3D polarization characteristics have not been considered effectively, hindering the effective modelling for TOD. Additionally, 2D convolutions primarily focus on extracting local spatial features (*e.g.*, edge and texture) within a single polarization component. In contrast, 3D convolutions operate on volumetric polarization data, considering not only the spatial dimensions but also an extra dimension related to the correlation among polarization parameters like S_0, DoP, and AoP. Hence, we integrate 2D and 3D convolutions to process multi-polarization information. This interaction effectively expands the model's ability to handle complex data structures and tasks, as it enables the model to learn features at diverse scales and dimensions.

Motivated by these analyses and the inherent shortcomings of existing methods, to detect TO, we propose a new intensity-spectral polarization fusion framework, termed as FuseISP. FuseISP proposes a new intensity-spectral polarization mixed modulator (ISPMM) and a hierarchical feature fusion (HFF) to fuse features across multi-scale and multi-level. FuseISP also introduces a multi-level decoder (MLD) based on the joint 2D/3D CNNs to model transparent regions by analysing comprehensively multiple polarization characteristics, 2D plane, and 3D space. We have conducted extensive ablation experiments to validate the effectiveness of our proposed framework.

To sum up, there are three-fold contributions to this work:

- We proposed two feature fusion modules, *i.e.*, ISPMM and HFF, where ISPMM aggregates information at the modality-level while HFF fuses information between features at different levels and scales.
- Our introduced the MLD module simultaneously capture inter- and intra-spectro-polarimetric relationships.
- Through extensive experiments, results show that our approach is more efficient for TOD than several state-of-the-art methods with superior generalization and robustness.

The remainder of this article is organized as follows. Section 2 describes several typical polarization object detection (POD) methods. Section 3 introduces the proposed FuseISP, including overall pipeline and network structures. Then, experiments on real-world scenes are performed to verify our method in Section 4. Finally, the conclusion is given in Section 5.

2. Related work

This section offers a concise review of POD based on their *application scenarios*, and focuses on the specific methods that are most pertinent to our work.

Polarization-Based Vehicle-Road Scene Detection. Automatic and accurate detection of road traffic objects is an important task in traffic safety and intelligent transportation system, which mainly includes road and vehicle detection. Casselgren *et al.* [25] first propose a method of using polarized short-wave infrared light to classify the four road conditions of dry, water, ice, and snow in winter, and the discrimination effect is obvious. Aiming at the problem of vehicle–road environment perception in low visibility, Wang *et al.* [26] fuse polarization features and intensity as the input feature of U-Net network to segment the vehicle–road scene. Although this scheme can alleviate the problem of vehicle–road scene detection in low visibility to some extent, it may not work at night. To overcome the above problem, Wang *et al.* [27] proposed a multi-branch input end-to-end network model, *i.e.*, MBIBEDN, based on polarized and infrared images to deal with vehicle-road scene detection in daytime, night and low visibility. Recently, Dong *et al.* [28] propose a special method for vehicle detection, which uses spectral polarization characteristics as additional clues to eliminate the challenges brought by the changes of lighting/weather conditions and vehicle density in the scene. This vehicle detection model dynamically integrates the complementary characteristics of RGB and polarization to learn the inherent material characteristics of the vehicle and realize vehicle detection.

Polarization-Based Camouflaged Object Detection. Camouflaged object detection (COD) has important application value in many fields, such as medicine and military. Because of the similarity between the camouflage object and its background, the accuracy of object identification is reduced. Polarization can provide valuable insights for understanding the characteristics of objects with different material properties and surface roughness. It reflects the difference of polarization information between the camouflage object and its background, increases the contrast between them, and can improve the accuracy of object detection even in complex scenes. Using polarization information to assist COD has attracted the attention of some researchers, and a series of excellent work has emerged. For instance, Fu *et al.* [29] propose a multi-modal COD method based on gating fusion from the perspective of combining polarization and intensity characteristics. Wang et al. tried to 'amplify' the difference between the object and the surrounding environment by using polarization information, and successively proposed two network models for COD, *i.e.*, PolarNet [30] and IPNet [31].

Polarization-Based Transparent Object Detection. The study of TOD from polarization is still in its infancy. To our best knowledge, there are a few methods (*i.e.*, Polarized Mask R-CNN [8], IEEETOSNet [9], and PGSNet [24]) available for TOD. For example, PGSNet [24] dynamically fuses and weights both the RGB colour and polarization cues using a novel global-guidance and multiscale self-attention module, and leverages global cross domain contextual information to achieve robust segmentation. Unlike PGSNet, IEEETOSNet [9] introduces an edge-enhanced and input extensible TO segmentation network that is capable of selecting the most optimal combination of inputs from 12 polarization parameters. Further, IEEETOSNet constructs a dataset containing multi-polarization cues captured with a monochromatic polarization imaging system. Nevertheless, it still remains a huge challenge for reliable TOD from polarization. we design a new solution to explore TOD in this paper.

Other Application Scenarios. In addition to the wide application of polarization imaging in the above-mentioned fields, the polarization-based target detection technology is also used in other fields, including material identification [32] and agricultural product quality inspection [5]. Recently, Yang *et al.* [5] combined polarization imaging with ResNet-18 and ghost bottleneck for nectarine damage detection, and the detection accuracy reached more than 96%, showing excellent early nectarine damage detection performance.

3. Proposed method

3.1. Description for polarization optics

To facilitate network training, researchers usually use a commercial DOFP polarization camera to collect a real-world polarization dataset. In DOFP polarization imagers, a micro-polarizer array (MPA) typically comprises four distinct polarization orientations, *i.e.*, 0°, 45°, 90°, and 135°, dividing the light into four different orientations. Hence, for one-shot, it can photograph simultaneously four polarized images, *i.e.*, $I_{0°}$, $I_{45°}$, $I_{90°}$, and $I_{135°}$. Fig. 2 exhibits the DOFP polarization camera and its corresponding MPA pattern.

In order to describe the polarization of light and the interaction between light and objects systematically and scientifically, the researchers put forward some mathematical characterization methods of polarized light, such as Poincare sphere [33], Jones vector [34] and Stokes vector S=[S_0, S_1, S_2, S_3] [35,36]. Among them, Stokes vector method can represent polarized light with arbitrary polarization state, and can fully characterize the polarization characteristics of incident light waves and light waves after interaction with substances. Therefore, Stokes vector description is the most commonly used method to express polarization characteristics in the field of polarization detection. In general, circularly polarized light is rarely available in the natural environment in visible band [37], so S_3 component is not considered in this paper. The polarization parameters Stokes vector S can be expressed as:

$$S = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} I_{0°} + I_{90°} \\ I_{0°} - I_{90°} \\ I_{45°} - I_{135°} \end{bmatrix}, \tag{1}$$

where S_0 refers to the total intensity received by the camera; S_1 represents the intensity difference between the vertical and horizontal components; S_2 denotes the intensity difference between the 45° and 135° components. Based on the Stokes vector S, DoLP and AoP are defined as

$$\mathrm{DoLP} = \frac{\sqrt{S_1^2 + S_2^2}}{S_0}, \tag{2}$$

$$\mathrm{AoP} = \frac{1}{2} \times \arctan\left(\frac{S_2}{S_1}\right). \tag{3}$$

3.2. Overview of our approach

The overall pipeline of our proposed FuseISP is illustrated in Fig. 3. Given the input features RGB S_0, DoLP, and AoP, they are first separately processed by three parallel stem layers consisting of one ConvNeXt [38] backbone-based HFE, to encode multi-scale features at {1/4, 1/8, 1/16, 1/32} of the original image resolution. Here, the ConvNeXt backbone is

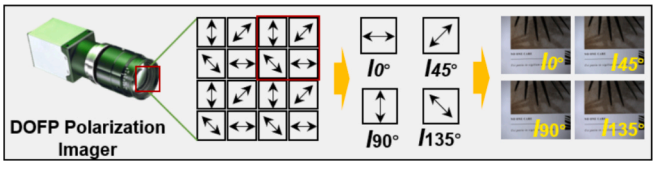

Fig. 2. The DoFP polarization camera, along with its corresponding MPA pattern. The direction of the arrow aligns with the transmission direction.

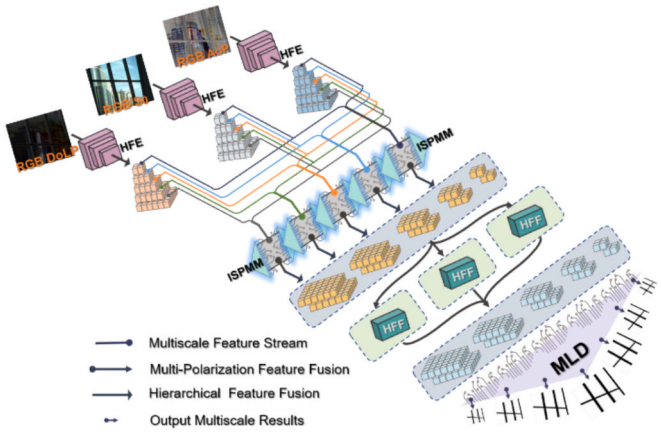

Fig. 3. Pipeline of our proposed **FuseISP**.

initialized with a model pre-trained on ImageNet. Next, the multi-level features are executed with multiple ISPMMs to produce the fused information at the modality-level. Then, to fully exploit complementary information from hierarchical features, the HFF is employed to generate the fused information on different levels. Finally, we sent all fused features to the MLD to produce predictions. We will describe the ISPMM, HFE, and MLD in detail below.

3.3. Intensity-spectral polarization mixed modulator

Diagrammatic details on how to produce fused features between S_0 and polarization features are shown in Fig. 4. Specifically, for each input feature map x ,*e.g.*, x∈[S_0, DoLP, AoP], is first partitioned into J groups along the channel dimension. Then, the ISPMM concatenates all grouped feature maps from different modalities to obtain J groups of mixed features. Finally, to obtain final mixed feature M, the ISPMM aggregates the J groups of fused features along the channel dimension and adopts J groups *DWConv* to integrate the polarization information from different receptive field perspectives. The above process can be formulated as:

$$[x^1, \cdots, x^j] = GELU(Conv_{1\times1}(DWConv_{3\times3}(x))), j = 1, \cdots, J \tag{4}$$

$$m^{j,k} = GELU\left(Conv_{1\times1}\left(DWConv_{k\times k}\left(x^j\right)\right)\right), k = 1, \cdots, K \tag{5}$$

$$M = Conv_{1\times1}(Concat(m^{1,5}, \cdots, m^{j,k}, \cdots, m^{j,k})) \tag{6}$$

where $DWConv_{3\times3}(\bullet)$ is a depth-wise convolution [39] with kernel size 3×3, $DWConv_{k\times k}(\bullet)$ is a depth-wise convolution with kernel size k×k and $k = j$×2+3; $Concat(\bullet)$ denotes feature concatenation operation; $Conv_{1\times1}(\bullet)$ is a 2D CNN layer with kernel size 1×1. Note that the hyper-parameter J is empirically set to 3.

Overall, our ISPMM offers following advantages: 1) It assists our model in capturing the relationships across different modalities; 2) It allows each spatial location to observe the local environment in different scale spaces, further expanding the receptive field of the whole network.

3.4. Hierarchical feature fusion

The progressive down-sampling operations in the feature extractor stage cause the feature information loss problem. To address this limitation, we introduce the HFE to capture connections between features at different levels, following [40]. Fig. 5 shows the detailed network structure of the HFE. HFE consists of two consecutive stages: fine- to coarse-level fusion (FCF) and coarse- to fine-level fusion (CFF).

In the FCF stage: *step 1*, feature $\mathscr{T}_2$ from the intermediate branch is executed with a GELU activation function and a 3×3 convolution to obtain the enhanced feature $\mathscr{T}_2^1$; *step 2*, the difference between $\mathscr{T}_2^1$ and

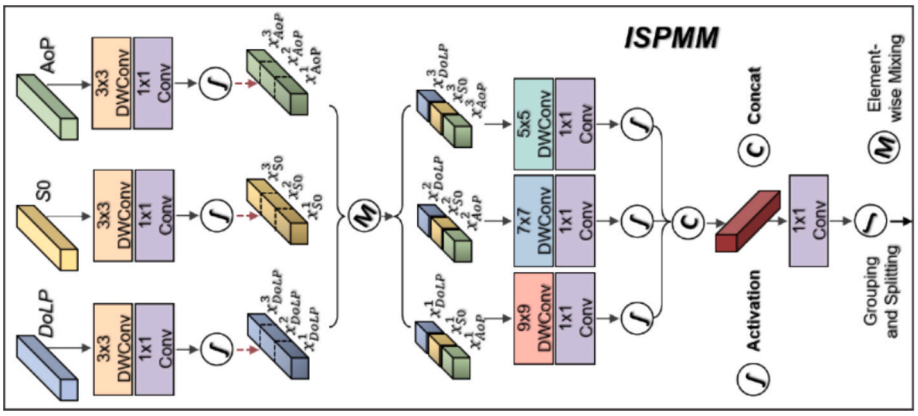

Fig. 4. Illustration of our proposed **ISPMM**.

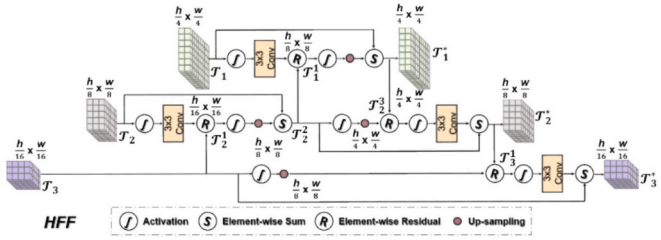

Fig. 5. The overall architecture of the **HFF**.

$\mathscr{T}_3$ is calculated and update the $\mathscr{T}_2$; *step 3*, $\mathscr{T}_1$ performs the same operation as the *step 1*. Meanwhile, the difference between the output of the *step 2* and $\mathscr{T}_1$ is computed and obtain the updated feature $\mathscr{T}_1^*$. The FCF stage aims to propagate information hierarchically from the high-level to the low-level features. This process is formulated as

$$\mathscr{T}_2^2 = Upsampling(GELU(\mathscr{T}_3 - Conv_{3\times3}(GELU(\mathscr{T}_2)))) + \mathscr{T}_2 \tag{7}$$

$$\mathscr{T}_1^* = Upsampling\big(GELU\big(\mathscr{T}_2^2 - Conv_{3\times3}(GELU(\mathscr{T}_1))\big)\big) + \mathscr{T}_1 \tag{8}$$

where $Conv_{3\times3}(\bullet)$ denotes a 3×3 2D CNN layer with a stride of 2; *Upsampling* $(\bullet)$ refers to bilinear interpolation operation.

Conversely, in the CFF stage, it is designed to spread information hierarchically from the low-level to the high-level features. The process of the CFF can be presented as

$$\mathscr{T}_2^* = \mathscr{T}_2^2 + Conv_{3\times3}\big(GELU\big(\mathscr{T}_1^* - Upsampling\big(GELU\big(\mathscr{T}_2^2\big)\big)\big)\big) \tag{9}$$

$$\mathscr{T}_3^* = \mathscr{T}_3 + Conv_{3\times3}\big(GELU\big(\mathscr{T}_2^* - Upsampling(GELU(\mathscr{T}_3))\big)\big) \tag{10}$$

3.5. Multi-level decoder

Previous methods lack the joint analysis among multiple polarization characteristics, specifically inter- and intra-polarization connections hindering the effective detection of many TO areas. To alleviate the above problem, we design the MLD to consider the correlations among multiple polarization characteristics inspired by [4,15,41], and Fig. 6A displays the architecture of the MLD. Concretely, MLD is constructed by three basic neural cells, as shown in Fig. 6 (A.1), (A.2), and (A.3). The modules in Fig. 6 (A.1) and (A.2) have a single 2D CNNs block and 3D CNNs block, respectively, while the module in Fig. 6 (A.3) contains both the 2D and 3D CNNs blocks. The linear interpolation operation is adopted to gradually recover the spatial resolution of feature maps. MLD's structure is a standard triangle, which consists of two stems, neural cells for cross-space information transfer, and skip connection (*i.e.*, [*a*, *b*, ⋯, *j*]). Both stems primarily transmit 2D and 3D information,

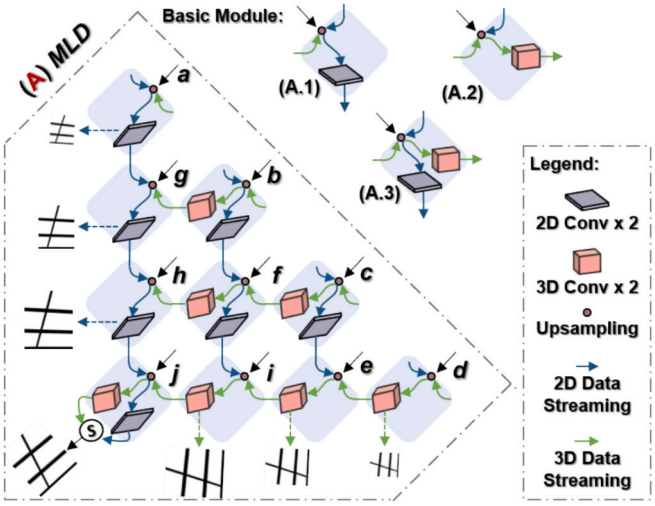

Fig. 6. The overall architecture of the **MLD**.

respectively. The interaction between 2D and 3D information of the branches on the two stems through a joint 2D/3D CNN, which then propagates to more distant branches. It aims at distilling more discriminative features.

3.6. Deep supervision loss

Based on MLD, we use deep supervision to enhance the robustness of the learning process for hidden layers. Specifically, for the outputs of the MLD's two stems, we sequentially apply a 3D CNNs with a single output channel and the sigmoid activation function to derive several predictions $\{P_i, i = 1, 2, \cdots, 7\}$. We adopt a binary cross-entropy (BCE) loss $\mathscr{L}_{bce}$ and a *IoU* loss $\mathscr{L}_{iou}$ for training our model, which can be formulated as follows:

$$\mathscr{L} = \mathscr{L}_{bce}(P_1, G) + \mathscr{L}_{iou}(P_1, G) + \lambda \sum_{i=2}^{7} [\mathscr{L}_{bce}(P_i, G) + \mathscr{L}_{iou}(P_i, G)] \tag{11}$$

where G denotes the ground-truth. λ denotes the weighting scalar for loss balance, and we empirically set it to 0.4.

4. Experiments

4.1. Experimental settings

A) ***Dataset***: To experimentally demonstrate the effectiveness of our

proposed FuseISP, we train and verify it on a large-scale polarization glass segmentation dataset, called RGBP-Glass, which is currently the publicly available high-quality dataset. The dataset comprises 4510 RGB intensity along with pixel-aligned RGB AoP and DoLP images. Each image features manually annotated accurate ground-truth at the pixel-level. Notably, each sample in RGBP-Glass contains at least one real-world glass object [24]. The types of transparent objects in this study include window panes, mirrors, glass guards, automotive window panes, etc. These targets represent common transparent objects encountered in daily life. For more detailed information on the dataset construction, please refer to ref. [24]. For the input polarization images, we first employ a BM3D [42] and a polarization difference model [43] to denoise and de-mosaic respectively. Finally, we randomly selected 3157 images and 451 images as the training set (denoted as TR3157) and validation set (denoted as VA451), respectively. The remaining 902 images constitute the testing set (denoted as TE902).

B) ***Implementation Details***: The proposed FuseISP is implemented in PyTorch [44] library and is trained for 150 epochs with a batch size of 15 on a NVIDIA A800-80 GB GPU using the Adam optimizer to optimize the model parameters with momentum term (β_1=0.5 and β_2=0.999), weight decay of 10^{-4}. The initial learning rate is set to 0.001, and the learning rate is reduced to half of the original every 15 epochs. Due to computational resource constraints, the resolution of all images was adjusted from 612×512 to 416×416. FuseISP exhibits a computational complexity of 107.59G FLOPs, and achieves 12fps during real-time inference.

C) ***Evaluation Criteria***: Five widely recognized metrics are employed to comprehensively evaluate the performance of our method for TOD: overall accuracy (*OA*), intersection over Union (*IoU*), balanced error rate (*BER*) [45], weighted F-measure (F_β) [46], mean absolute error (*MAE*). For *OA*, *IoU*, and F_β, higher is better, while for *MAE* and *BER*, lower is better.

4.2. Comparison with the advanced methods

We validate the effectiveness of our method by comparing it with several advanced object detection methods, including UNet3+ [47], PSPNet [48], DeepLabv3+ [49], FastSCNN [50], LRSR-net [51], MirrorNet [11], PolarNet [30], and IPNet [31]. For a fair comparison, all methods are retrained on the training dataset TR3157 and evaluated on the testing dataset TE902.

A) ***Quantitative Performance Comparison***

Table 1 shows the evaluation results of our FuseISP compared with previous methods. The results clearly show that FuseISP achieves much better performance, especially in terms of *OA*, *MAE*, and F_β, compared to other methods. Particularly, FuseISP achieves 10.20%, 3.01%, and 5.57% improvements in *MAE*, *IoU*, and *BER*, respectively, compared with the recent best polarization object detection method, *i.e.*, IPNet. Furthermore, our FuseISP achieves 0.68%, 1.65%, and 0.43% performance improvement than the previous advanced CNNs-based method FastSCNN on *OA*, *IoU* and F_β, respectively. The difference between our method and competitors in the *IoU* values is statistically significant, with a *p*-value<0.01 in the Student's *t*-test. This result demonstrates that FuseISP has a stronger ability to rectify and fuse multi-modal information features than the advanced methods.

Table 1
Objective evaluation results of different methods on testing dataset. The best and second-best results are in red and blue, respectively. ↑&↓ denote larger and smaller is better, respectively.

Method	*OA* (%)↑	*MAE*↓	*IoU* (%)↑	*BER* (%)↓	F_β(%)↑	*P-value*$^{(IoU)}$
FastSCNN	91.76	0.115	82.47	9.34	90.66	<0.01**
UNet3+	89.64	0.131	79.70	10.48	89.12	<0.0001*
DeepLabv3+	90.96	0.109	80.98	9.56	89.96	<0.0001*
PSPNet	89.61	12.68	78.46	12.45	86.55	<0.0001*
MirrorNet	89.44	0.121	79.22	11.30	87.02	<0.0001*
IPNet	91.08	0.098	81.38	8.97	90.15	<0.0001*
LRSR-net	81.58	0.236	63.75	21.82	75.80	<0.0001*
PolarNet	91.15	0.099	83.11	9.66	89.84	<0.0001*
FuseISP (Ours)	92.39	0.088	83.83	8.47	91.05	–

P-value$^{(IoU)}$ denotes that significance (*P*-value) in the *IoU* values between FuseISP and other methods are assessed by Student's *t*-test. **P*-value < 0.0001, ***P*-value < 0.01.

B) ***Qualitative Performance Comparison***

Moreover, the visual comparison between FuseISP and other comparison methods in Fig. 7 further demonstrate the advantage of our method. Here we select four cases to illustrate. Each case is associated with different properties, including complex geometry, low contrast, low-light, and complex surroundings. Our goal is to demonstrate that our approach can work more robustly and better under different circumstances. From Fig. 7, it can be easily seen that our method not only discriminates the right objects but also maintains their sharp boundaries in almost all circumstances. However, the other methods sometimes fail when dealing with complex scenes, especially when the objects are with complex surroundings (the 3th row in Fig. 7) and low-light environment (the 2th row in Fig. 7). It is noteworthy that our FuseISP exhibits a stable generalization capability across dark and light environments.

4.3. Model analysis

Here, we present extensive ablation studies of our model to explore its performance on both validation dataset VA451 and testing dataset TE902, including the feature sources and the module parts.

A) **Features ablation:** This section examines the extent to which the introduced spectral polarization characteristics can help FuseISP discriminate transparent regions. Table 2 reports the discriminative performance comparison of these feature combinations and *p*-values in the Student's *t*-test, including S_0, S_0+DoLP, S_0+AoP, and S_0+DoLP+AoP. By carefully observing Table 2, the performance of S_0+DoLP+AoP consistently outperforms other features. The introduction of the DoLP and AoP achieves 30.26%, 30.93%, 6.26%, 18.02%, and 2.16% average improvements of *OA*, *MAE*, *IoU*, *BER*, and F_β, respectively compared with S_0 on both datasets VA451 and TE902, with *p*-value<10^{-3}. These comparison results can demonstrate that the detection performance of FuseISP is indeed improved after the combination of intensity and spectral polarization is adopted.

B) **Modules ablation:** We progressively delete the modules of FuseISP. As shown in Table 3, removing each component of FuseISP will cause a performance decrease, demonstrating the necessity and importance of the modules we designed. The removal of ISPMM causes a significant loss of performance, indicating that our ISPMM module is crucial for TO prediction. However, the removal of HFF does not cause a significant loss of performance, while the removal of the MLD will result in a noticeable decrease. Because the inter- and intra-polarization relationships can help the model directly, adding the shared information among different levels may not be helpful when down-sampling operations in the feature extractor stage cause the feature information loss. Additionally, we also show the impact of removing multiple modules to assess interdependencies. If we remove all the components and only feed the splicing features into the HFE and its matching decoder, the performance will be much poorer, thus, brutally combining embeddings without a suitable model is impracticable.

4.4. Failure case analysis

Like all the other TOD methods, our FuseISP cannot detect TO with the strong or extensive reflections due to limited context information as shown in Fig. 8. The highlight in the reflection layer obscures the invisible background, making it easy to miss detections and cause misclassifications. To be concrete, the *MAE* and *OA* of FuseISP are 0.112 and 90.1%, respectively. Similarly, large-area reflective obstacles disrupt the

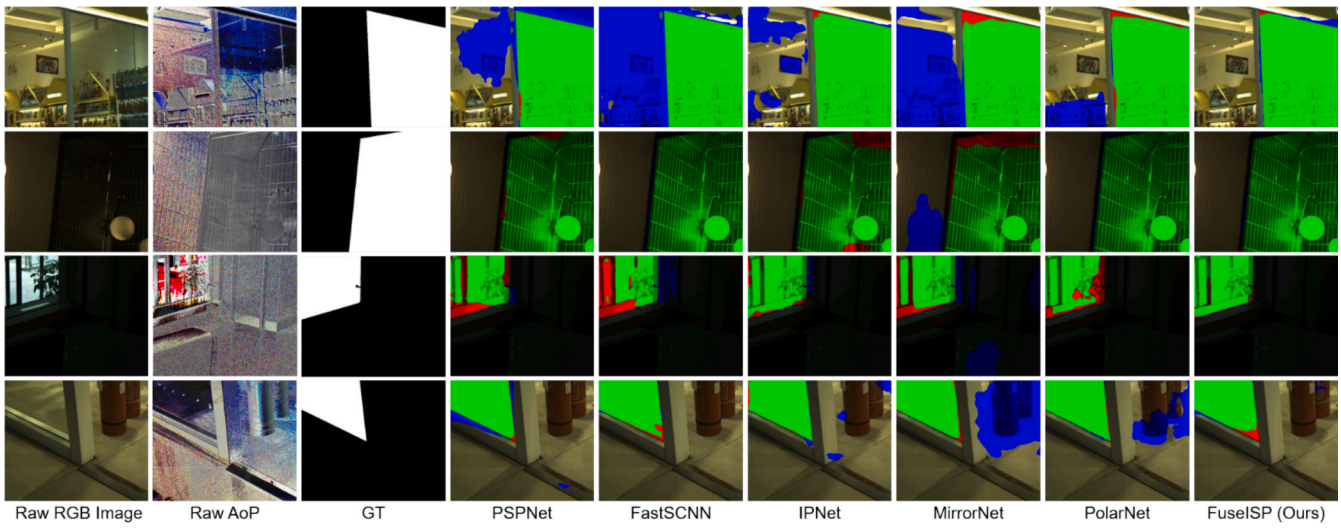

Fig. 7. Detection results of different comparison methods. Green regions, blue regions, and red regions represent true positive, false positive, and false negative, respectively.

Table 2
Objective evaluation results of different features. ↑&↓ denote larger and smaller is better, respectively. Numbers highlight with red indicate the best results.

Dataset	Data Type	*OA* (%)↑	*MAE*↓	*IoU* (%)↑	*BER* (%)↓	F_β (%)↑	*P-value* (*IoU*)
VA451	S_0	89.99	0.124	79.61	9.88	89.72	<0.001*
	S_0 + DoLP	92.41	0.907	83.44	8.26	90.57	<0.01**
	S_0 + AoP	91.21	0.101	81.57	9.43	90.42	<0.01**
	S_0 + DoLP + AoP	92.31	0.087	83.52	8.20	91.05	–
TE902	S_0	89.51	0.128	78.48	10.50	89.10	<0.001*
	S_0 + DoLP	92.40	0.906	83.63	8.38	90.88	<0.01**
	S0 + AoP	91.56	0.098	82.07	9.37	90.63	<0.01**
	S_0 + DoLP + AoP	92.62	0.087	84.46	8.50	91.63	–

***P-value*$^{(IoU)}$** denotes that significance (*P*-value) in the *IoU* values between S_0+DoLP+AoP and other features are assessed by Student's *t*-test. *P-value < 0.001, **P-value < 0.01.

true color distribution of both the TO and their surroundings, further complicating TO detection. The main reason for this phenomenon should be that there are only few images containing highly reflective transparent objects in the training dataset. Thus, FuseISP is unable to learn sufficient knowledge for this case, leading to less effectiveness in this case.

5. Conclusion

In this paper, we propose a novel approach, FuseISP, designed from the ground up specifically for identifying transparent regions by integrating neural networks and spectral polarization imaging. FuseISP leverages intensity information, DoLP and AoP, along with three novel modules, *i.e.*, ISPMM, HFF, and MLD, to establish the correlations among polarization, 2D planes, 3D spaces, and transparent regions. The results demonstrate significant improvements and are consistently better than the existing several advanced methods. Despite its strengths, FuseISP faces limitations. It still occasionally struggles with poor performance when dealing with highly reflective transparent objects. Additionally, FuseISP suffers from a lack of interpretability. The underlying mechanisms of the DL model remain opaque, presenting a black-box. However, challenges remain in effectively utilizing multiple

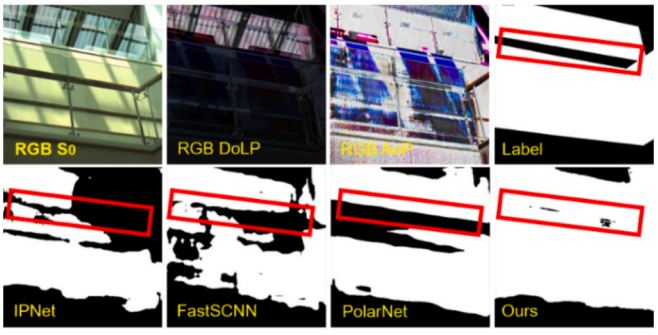

Fig. 8. The strong or extensive reflections cause transparent object detection to fail.

Table 3
Ablation Analysis for the Design Choices of FuseISP. ↑&↓ denote larger and smaller is better, respectively. Numbers highlight with red indicate the best results.

Configuration for FuseISP	VAL451					TE902				
	OA(%)↑	*MAE*↓	*IoU*(%)↑	*BER*(%)↓	F_β (%)↑	*OA*(%)↑	*MAE*↓	*IoU*(%)↑	*BER*(%)↓	F_β(%)↑
FuseISP	92.31	0.087	83.52	8.20	91.05	92.62	0.087	84.46	8.50	91.63
– ISPMM	91.21	0.107	81.23	9.30	89.11	91.33	0.106	81.44	9.38	89.39
– HFF	92.38	0.092	83.20	9.08	91.13	92.42	0.093	83.40	9.16	91.24
– MLD	91.22	0.104	81.53	9.62	89.60	91.46	0.103	82.04	9.47	89.89
– HFF-MLD	89.08	0.171	78.55	10.88	87.71	88.98	0.170	78.60	11.20	87.85
– ISPMM-MLD	90.76	0.142	80.87	9.55	88.90	90.81	0.143	80.71	9.22	88.86

polarization characteristics to enhance TOD, and further advancements are needed in characterizing polarization information and developing a lightweight network model. Nonetheless, significant improvement has been achieved by our proposed FuseISP. We expect the technique described here to be useful for TOD.

CRediT authorship contribution statement

Xueqiang Fan: Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation, Conceptualization. **Longyu Qiao:** Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Bing Lin:** Validation, Resources, Investigation, Formal analysis. **Zhongyi Guo:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization.

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.

Acknowledgments

This research is funded by the National Natural Science Foundation of China under Grant 61775050. The computation of this research is supported by the HPC Platform of Hefei University of Technology.

References

[1] N. Li, Y. Zhao, Q. Pan, et al., Illumination-invariant road detection and tracking using LWIR polarization characteristics, ISPRS J. Photogramm. Remote Sens. 180 (2021) 357–369.
[2] T. Wang, B. Chen, Z. Zhang, et al., Applications of machine vision in agricultural robot navigation: a review, Comput. Electron. Agric. 198 (2022) 107085.
[3] J. Xie, J. Dou, L. Zhong, et al., A dual-mode intensity and polarized imaging system for assisting autonomous driving, IEEE Trans. Instrum. Meas. 73 (2024) 1–13.
[4] X. Fan, B. Lin, Z. Guo, Infrared polarization-empowered full-time road detection via lightweight multi-pathway collaborative 2D/3D convolutional networks, IEEE Trans. Intell. Transp. Syst. 25 (9) (2024) 12762–12775.
[5] Y. Yang, L. Wang, M. Huang, et al., Polarization imaging based bruise detection of nectarine by using ResNet-18 and ghost bottleneck, Postharvest Biol. Technol. 189 (2022) 111916.
[6] A. Li, Y. Mao, J. Zhang, et al., Mutual information regularization for weakly-supervised RGB-D salient object detection, IEEE Trans. Circuits Syst. Video Technol. 34 (1) (2023) 397–410.
[7] J.-J. Liu, Q. Hou, Z.-A. Liu, et al., Poolnet+: exploring the potential of pooling for salient object detection, IEEE Trans. Pattern Anal. Mach. Intell. 45 (1) (2022) 887–904.
[8] A. Kalra, V. Taamazyan, S.K. Rao, et al., Deep polarization cues for transparent object segmentation, Proc. IEEE/CVF Conf. Comput. Vis. Patt. Recognit. (2020) 8602–8611.
[9] R. Yu, W. Ren, M. Zhao, et al., Transparent objects segmentation based on polarization imaging and deep learning, Opt. Commun. 555 (2024) 130246.
[10] K. Fan, C. Wang, Y. Wang, et al., Rfenet: Towards reciprocal feature evolution for glass segmentation. arXiv preprint arXiv:2307.06099, 2023.
[11] X. Yang, H. Mei, K. Xu, et al., Where is my mirror? Proc. IEEE Int. Conf. Comput. Vis. (2019) 8809–8818.
[12] H. He, X. Li, G. Cheng, et al., Enhanced boundary learning for glass-like object segmentation, Proc. IEEE Int. Conf. Comput. Vis. (2021) 15859–15868.
[13] Q. Kang, K. Guo, X. Zhang, et al., Dynamically manipulating long-wave infrared polarized thermal radiation by a vanadium dioxide metasurface, Opt. Lett. 49 (9) (2024) 2485–2488.
[14] G. Wang, J. Gao, Y. Xiang, et al., Deep learning-driven underwater polarimetric target detection based on the dispersion of polarization characteristics, Opt. Laser Technol. 174 (2024) 110549.
[15] X. Tian, R. Liu, Z. Wang, et al., High quality 3D reconstruction based on fusion of polarization imaging and binocular stereo vision, Inf. Fusion 77 (2022) 19–28.
[16] J. Liu, S. Li, R. Dian, et al., DT-F Transformer: dual transpose fusion transformer for polarization image fusion, Inf. Fusion 106 (2024) 102274.
[17] B. Lin, W. Chen, X. Fan, et al., Transformer-based improved U-net for high-performance underwater polarization imaging, Opt. Laser Technol. 181 (2025) 111664.
[18] B. Lin, X. Fan, P. Peng, et al., Dynamic polarization fusion network (DPFN) for imaging in different scattering systems, Opt. Express 32 (1) (2024) 511–525.
[19] X. Fan, B. Lin, K. Guo, et al., TSMPN-PSI: high-performance polarization scattering imaging based on three-stage multi-pipeline networks, Opt. Express 31 (23) (2023) 38097–38113.
[20] X. Fan, W. Chen, B. Lin, et al., Improved polarization scattering imaging using local-global context polarization feature learning framework, Opt. Lasers Eng. 178 (2024) 108194.
[21] L. Shen, H. Xia, X. Zhang, et al., U^{2}PNet: an unsupervised underwater image-restoration network using polarization, IEEE Trans. Cybern., 2024.
[22] X. Fan, M. Ding, T. Lv, et al., Meta-DNET-UPI: Efficient underwater polarization imaging combining deformable convolutional networks and meta-learning, Opt. Laser Technol. 187 (2025) 112900.
[23] B. Lin, L. Qiao, X. Fan, et al., Large-range polarization scattering imaging with an unsupervised multi-task dynamic-modulated framework, Opt. Lett. 50 (10) (2025) 3413–3416.
[24] H. Mei, B. Dong, W. Dong, et al., Glass segmentation using intensity and spectral polarization cues, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition2022, pp. 12622-31.
[25] J. Casselgren, M. Sjödahl, Polarization resolved classification of winter road condition in the near-infrared region, Appl. Opt. 51 (15) (2012) 3036–3045.
[26] H.-F. Wang, Y.-H. Shan, T. Hao, et al., Vehicle-road environment perception under low-visibility condition based on polarization features via deep learning, IEEE Trans. Intell. Transp. Syst. 23 (10) (2022) 17873–17886.
[27] H.-F. Wang, Y.-M. Jiao, T. Hao, et al., Low-visibility vehicle-road environment perception based on the multi-modal visual features fusion of polarization and infrared, IEEE Trans. Intell. Transp. Syst. 2023.
[28] W. Dong, H. Mei, Z. Wei, et al., Exploiting polarized material cues for robust car detection, Proc. AAAI Conf. Artif. Intell. (2024) 1564–1572.
[29] B. Fu, T. Cao, Y. Zheng, et al., Polarization-driven camouflaged object segmentation via gated fusion, Appl. Opt. 61 (27) (2022) 8017–8027.
[30] X. Wang, Z. Zhang, J. Gao, Polarization-based camouflaged object detection, Pattern Recognit. Lett. 174 (2023) 106–111.
[31] X. Wang, J. Ding, Z. Zhang, et al., IPNet: polarization-based camouflaged object detection via dual-flow network, Eng. Appl. Artif. Intell. 127 (2024) 107303.
[32] H. Chen, L.B. Wolff, Polarization phase-based method for material classification and object recognition in computer vision, in: Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 1996, pp. 128–35.
[33] G. Deschamps, P. Mast, Poincaré sphere representation of partially polarized fields, IEEE Trans. Antennas Propag. 21 (4) (1973) 474–478.
[34] C.J. Sheppard, Jones and Stokes parameters for polarization in three dimensions, Phys. Rev. A 90 (2) (2014) 023809.
[35] H.G. Berry, G. Gabrielse, A. Livingston, Measurement of the Stokes parameters of light, Appl. Opt. 16 (12) (1977) 3200–3325.
[36] W.S. Bickel, W.M. Bailey, Stokes vectors, mueller matrices, and polarized scattered light, Am. J. Phys 53 (5) (1985) 468–478.
[37] Z. Dong, D. Zheng, Y. Huang, et al., A polarization-based image restoration method for both haze and underwater scattering environment, Sci. Rep. 12 (1) (2022) 1–11.
[38] Z. Liu, H. Mao, C.-Y. Wu, et al., A convnet for the 2020s, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 11976-86.
[39] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2017, pp. 1251–58.
[40] T. Wang, G. Tao, W. Lu, et al., Restoring vision in hazy weather with hierarchical contrastive learning, Pattern Recognit. 145 (2024) 109956.
[41] Z. Dong, Y.H e, X. Qi, et al., MNet: rethinking 2D/3D networks for anisotropic medical image segmentation. arXiv preprint arXiv:2205.04846, 2022.
[42] A. Abubakar, X. Zhao, S. Li, et al., A block-matching and 3-D filtering algorithm for Gaussian noise in DoFP polarization images, IEEE Sens. J. 18 (18) (2018) 7429–7435.
[43] N. Li, Y. Zhao, Q. Pan, et al., Demosaicking DoFP images using Newton's polynomial interpolation and polarization difference model, Opt. Express 27 (2) (2019) 1376–1391.
[44] A. Pajankar, A. Joshi, Neural Network and PyTorch Basics, Hands-on Machine Learning with Python, Springer, 2022, pp. 215–26.
[45] V. Nguyen, T.F. Yago Vicente, M. Zhao, et al., Shadow detection with conditional generative adversarial networks, Proc. IEEE Int. Conf. Comput. Vis. (2017) 4510–4518.
[46] R. Margolin, L. Zelnik-Manor, A. Tal, How to evaluate foreground maps? Proc. IEEE Conf. Comput. Vis Pattern Recognit. (2014) 248–255.
[47] H. Huang, L. Lin, R. Tong, et al., UNet 3+: A full-scale connected UNet for medical image segmentation, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)2020, pp. 1055-59.
[48] H. Zhao, J. Shi, X. Qi, et al., Pyramid scene parsing network, Proc. IEEE Conf. Comput. Vis Pattern Recognit. (2017) 2881–2890.
[49] L.-C. Chen, Y. Zhu, G. Papandreou, et al., Encoder-decoder with atrous separable convolution for semantic image segmentation, Proc. Eur. Conf. Comput. vis. (2018) 801–818.
[50] R.P. Poudel, S. Liwicki, R. Cipolla, Fast-scnn: Fast semantic segmentation network. arXiv preprint arXiv:1902.04502 2019.
[51] S. Sun, Z. Yang, T. Ma, Lightweight remote sensing road detection network, IEEE Geosci. Remote Sens. Lett. 19 (2022) 1–5.