BDNet:Bio-Inspired Dual-Backbone Small Object Detection Network

Wenchao Guan[1] , Chuan Lin[1*] , Sihan Huang[1] , Xiongzhen Wang[1] , Xintao Pang[2*]
[1]School of Automation, Guangxi University of Science and Technology
[2]Faculty of Applied Sciences, Macao Polytechnic University

Abstract

In remote sensing images, small objects often exhibit low color contrast and blurred edges, leading to suboptimal feature extraction. Physiological studies indicate that the LGN/V1–V2–V4 pathway provides color-opponent sensitivity and hierarchical enhancement for color information extraction, whereas the V1–V4 pathway exhibits strong orientation selectivity for edge extraction. Integrating these complementary visual signals in the V4 region can substantially improve target discrimination. Motivated by these findings, we propose a dual-backbone network (BDNet) to enhance feature extraction for small objects. BDNet adopts a parallel architecture to capture fine-grained features from color and edge cues. Specifically, the color-extraction backbone simulates the color-opponent mechanism in LGN/V1 via a Color Antagonism Module (CAM) to amplify color differences, and further mimics the chromatic processing hierarchy in V2 using a Visual Cortex Hue Enhancement Module (VCHM) to enrich hue representations. Together, these two modules alleviate low color contrast. The edge-extraction backbone simulates the orientation selectivity of receptive fields in V1 through an Orientation Selective Module (OrSM) to select and enhance salient edges, thereby reducing edge blurring from fragmented edge responses. Finally, the two feature types are fused via a Feature Fusion Module (FFM) that emulates integration in V4, yielding a comprehensive feature representation. Experiments demonstrate that BDNet outperforms state-of-the-art methods on the VisDrone2019, NWPU VHR-10, and AI-TODv2 datasets, providing a bio-inspired solution for small-object detection in remote sensing images.

1. Introduction

Remote Sensing Object Detection (RSOD) aims to localize and recognize targets in high-resolution remote sensing imagery. However, small objects in remote sensing images often exhibit limited visual cues, such as low color contrast and blurred edges, and their features are further degraded by downsampling in feature-extraction networks, which severely limits detection accuracy [4].

To enhance the representation of fine-grained cues like color and edges, existing feature-enhancement methods have achieved notable gains [23]. However, they are often limited because they focus on a single type of feature. For example, COSE [22] improves color consistency in low-contrast regions through color-shift correction, DCFL [49] enhances edge textures by combining super-resolution and detail compensation, and SET [42] strengthens high-frequency details via spectral enhancement. Multi-backbone networks typically aim to complement different characteristics within the same feature space. For instance, TransFuse [53] combines parallel CNN and Transformer backbones to integrate local and global cues, while DSOD++ [41] employs dual paths to enhance information from different receptive fields, and DCAL [55] introduces bidirectional guidance to improve cross-attention complementarity. However, research on multi-branch networks that explicitly target multi-dimensional low-level cues (e.g., color and edges) remains scarce. In contrast, studies of biological vision provide a well-established theoretical basis for understanding low-level visual processing.

Physiological studies suggest the visual system integrates multiple features through hierarchical processing. The Lateral Geniculate Nucleus (LGN), Primary Visual Cortex (V1), and Secondary Visual Cortex (V2) process color and luminance [39, 40], while orientation-selective neurons in V1 extract edges [18]. These two streams converge in visual area V4, improving object representation [14, 36]. Motivated by these observations, we propose a bio-inspired dual-backbone detection network, termed BDNet. BDNet simulates the LGN/V1–V2–V4 and V1–V4 pathways to enhance and complement color and edge cues, resulting in a dual-backbone architecture with a color-enhancement branch and an edge-reinforcement branch. This design enables multi-dimensional enhancement and hierarchical fusion of small-object features, improving the detectability of weak targets in remote sensing imagery. Our main contributions are:

*Corresponding authors. Chuan Lin (chuanlin@gxust.edu.cn)
Xintao Pang (p2424471@mpu.edu.mo)

- We propose a "color enhancement–edge reinforcement–hierarchical fusion" detection framework that leverages the biological color-processing pathway (LGN/V1–V2–V4) and edge-processing pathway (V1–V4) to achieve complementary enhancement across a color-extraction backbone and an edge-extraction backbone for remote-sensing small-object detection.
- In the color-extraction backbone, we design a Color Antagonism Module (CAM) and a Visual Cortex Hue Enhancement Module (VCHM) to synergistically enhance color cues and improve hue representation. In the edge-extraction backbone, we introduce an Orientation Selective Module (OrSM) to strengthen edge and contour details. Finally, a Feature Fusion Module (FFM) integrates these features to yield a more comprehensive representation, thereby addressing insufficient feature extraction.
- Extensive experiments on the VisDrone2019, NWPU VHR-10, and AI-TODv2 remote-sensing small-object datasets show that BDNet achieves state-of-the-art performance, demonstrating the effectiveness of the proposed bio-inspired design.

2. Related Work

2.1. Remote Sensing Image Small Object Detection

The core challenge in detecting small objects in remote sensing images lies in their small pixel footprint, which can cause feature attenuation and information loss during downsampling in deep networks. In recent years, research has gradually shifted from adapting general detection frameworks to designing architectures that explicitly account for small-object characteristics. For instance, Xu et al. [49] addressed scale mismatch in tiny-object detection through multi-stage feature refinement. Wu et al. [48] improved localization accuracy across multiple remote-sensing categories by integrating spatial- and frequency-domain features. Liu et al. [24] proposed a progressive context-reasoning framework with multi-branch structures, and Hou et al. [15] employed multi-path designs to cope with the complexity of remote-sensing imagery. Although these methods demonstrate notable performance, most optimize features in a holistic manner without explicitly decoupling and enhancing the low-level cues that are critical for small objects. Consequently, fine-grained information can gradually diminish as features propagate through deep networks. Furthermore, many architectures are not explicitly grounded in visual information-processing mechanisms and therefore struggle to establish structured, prior-guided pathways akin to those in biological vision.

2.2. Bio-Inspired Small Object Detection Models

Bio-inspired models provide valuable insights for small-object detection in aerial imagery. Several representative approaches address these challenges. ClusterNet [20] improves localization for low-pixel targets by simulating a two-stage strategy that combines wide-field perception with foveal focus. Brstd [17], inspired by visual modulation, introduces antagonistic receptive fields and feedback inhibition to strengthen small-object features while suppressing background interference. Magno-VTOD [44] draws on the retinal magnocellular pathway to improve detection of tiny moving targets in infrared scenes. VSTDet [32] mimics hierarchical processing in the ventral pathway to preserve fine details with lightweight computation, and EVMNet [3] emulates dual foveal and tectal pathways to progressively extract semantics and enhance contextual modeling. Nevertheless, these works primarily emphasize single perceptual mechanisms or functional separation. Building upon these studies, we systematically simulate the LGN/V1–V2–V4 pathway to support collaborative processing of color and edge cues, and construct a dual-backbone network that more directly addresses feature degradation in small objects.

3. Methods

3.1. Network Architecture

As shown in Fig. 1, BDNet simulates the hierarchical processing mechanism of the biological visual system for color and edge information, constructing a "dual-path feature extraction - hierarchical fusion" detection framework to address the core issue of ineffective feature extraction for small objects in remote sensing images. The network consists of three components: the Color Information Path (CIP), the Edge Information Path (EIP), and the Feature Fusion Module (FFM).

The CIP focuses on enhancing color contrast. It first employs the Color Antagonism Module (CAM) and then the Visual Cortex Hue-enhancement Module (VCHM). CAM mimics the color antagonism mechanism of the LGN/V1 areas to enhance color representation, while VCHM simulates the hierarchical chromatic processing structure of the V2 area to adjust hue levels.

The EIP focuses on preserving edge direction information for small targets. It consists of the Enhanced Learnable Laplacian Operator Module (ELLOM) and the Orientation Selection Module (OrSM). ELLOM first extracts edge information from the RGB image, which is then passed to OrSM. OrSM simulates the orientation selectivity of neurons in V1/V2 areas to reinforce the contour integrity of small targets.

Features from the CIP and EIP paths undergo hierarchical fusion across different dimensions via the Feature Fusion Module (FFM), which simulates the information integration mechanism of the V4 area. A feature pyramid network then aggregates multi-scale features, and the detection head predicts target categories and locations.

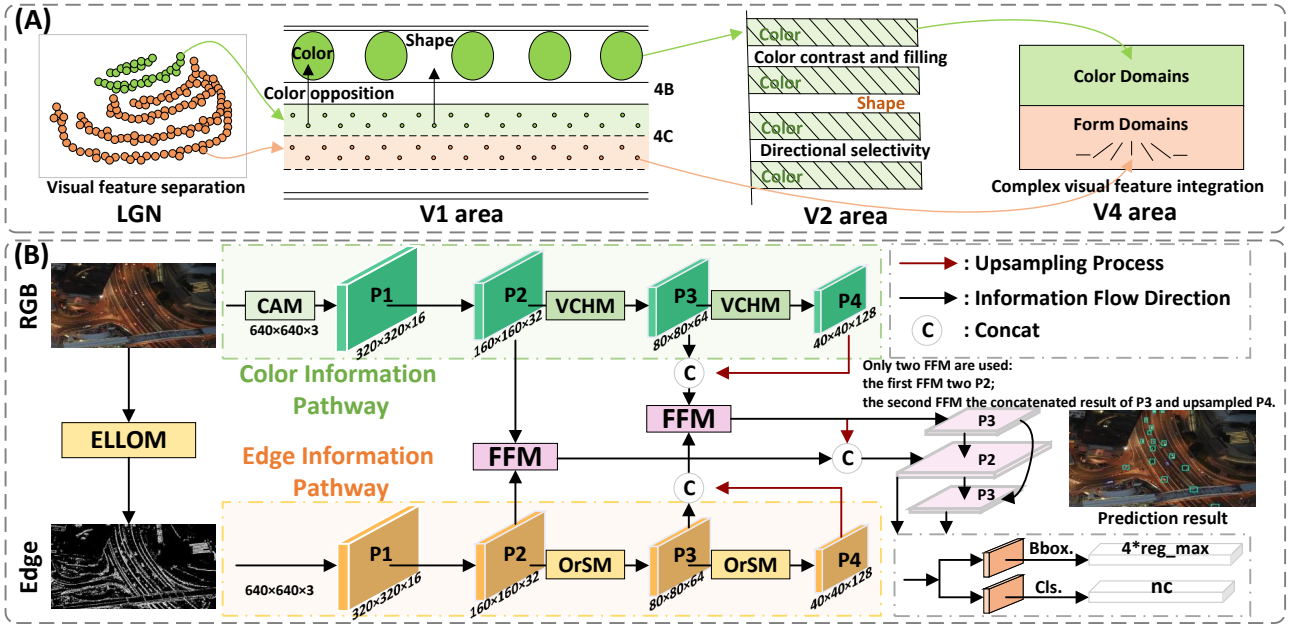

Figure 1. (A) Schematic diagram of biological visual information processing: Color information is processed via the LGN/V1 → V2 → V4 pathway, while form (edge, shape) information is processed via the LGN/V1 → V4 pathway. Color and form information interact and integrate in the V4 area. (B) Block diagram of the proposed BDNet architecture: After the RGB image is input to the network, the Color Information Path (CIP, comprising CAM + VCHM) and the Edge Information Path (EIP, comprising ELLOM + OrSM) extract color and edge information, respectively. FFM then interactively fuses the dual-path features.

3.2. Color Information Pathway

Color Antagonism Module. The LGN and V1 areas of the biological visual system contain three types of cone cells—L (long-wave sensitive), M (medium-wave sensitive), and S (short-wave sensitive)—which perform preliminary processing of color information through an "excitation-inhibition" antagonistic mechanism [6, 8, 35], as illustrated in Fig. 2(A1). Based on this mechanism, CAM implements the following specific steps:

The R, G, and B channels of the input RGB image are mapped to channels simulating the L, M, and S cone cells, respectively, forming the six "excitation-inhibition" antagonistic pairs shown in Fig. 2(A2). Inspired by opponent color theory [21], which reveals that the visual system enhances color contrast via antagonistic effects of complementary color pairs (e.g., red-green, blue-yellow), an Adaptive Selection Enhancement (ASE) mechanism is designed. Specifically, ASE introduces a learnable weight vector to weight the RGB channels and modulate their contributions in opponent color combinations. Based on these antagonistic pairs, ASE processes the R, G, and B channels and generates six types of "excitation channel-inhibition channel" feature pairs. Finally, features of the same type of excitation channel are aggregated, and feature fusion is achieved via convolution, thereby completing the bio-inspired enhancement of color features.

Visual Cortex Hue Enhancement Module. After CAM processing, while remote sensing image hue is enhanced via color differences, hue features still deviate from human perception. Physiological studies show the cortical hue map's matching degree with the perceptual color space improves significantly with increasing cortical hierarchy [30], as shown in Fig. 2(B1).

Based on this, VCHM leverages the hierarchical characteristics of cortical hue processing and optimizes color feature representation via a three-step operation: *grouped convolution – channel coupling – feature embedding*, as illustrated in Fig. 2(B2). Specifically, given the input feature $X = (x_0, x_1, \cdots, x_{c-1})$, grouped convolution with C groups is applied to obtain the intermediate feature $X' = (x'_0, x'_1, \cdots, x'_{c-1})$. To generate new hue features, X' undergoes pairwise channel coupling via Eq.(1). Subsequently, Eq.(2) embeds the newly generated hue features into the original channel sequence based on their generation order, further optimizing the color features.

$$Y = Conv\left(\begin{bmatrix}[1,1,0,\cdots,0,0]\\ [0,1,1,\cdots,0,0]\\ \vdots \\ [0,0,0,\cdots,1,1]\end{bmatrix} \cdot W, \begin{bmatrix}x'_0\\ x'_1\\ \vdots \\ x'_{c-1}\end{bmatrix}\right) \quad (1)$$

$$E(X,Y) = (x_0, w_0 \cdot x'_0 + w_1 \cdot x'_1, x_1, \cdots, x_{c-1}) \quad (2)$$

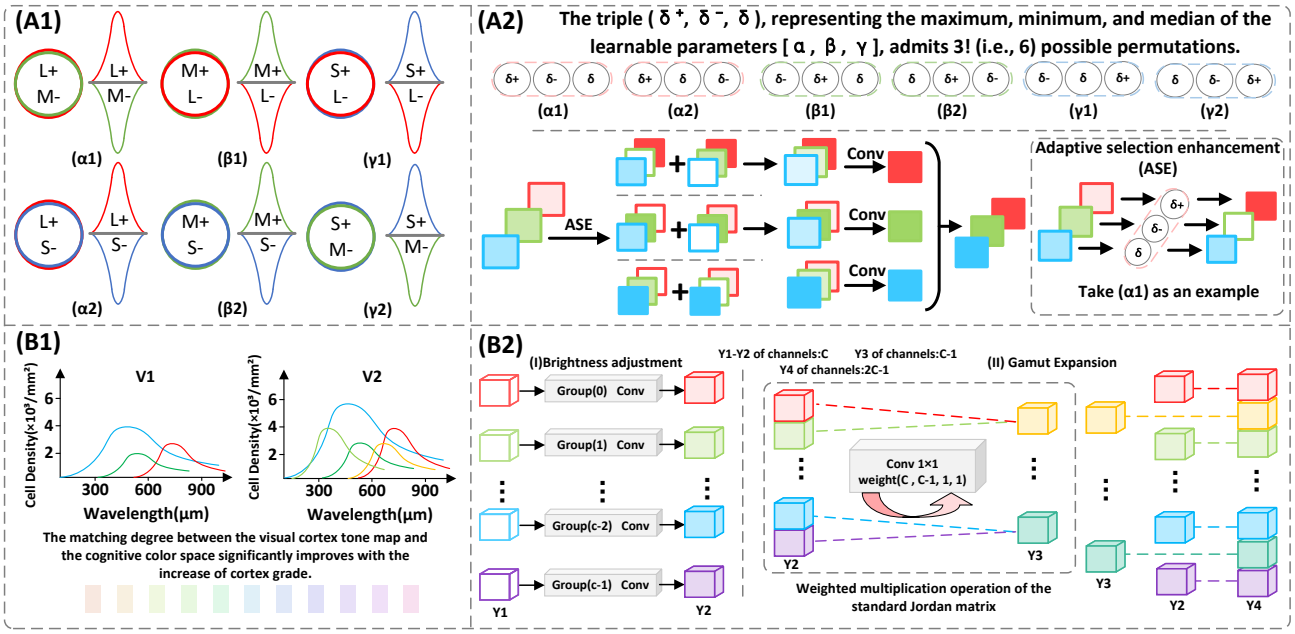

Figure 2. Architecture design of CIP. (A1) illustrates the color-opponent receptive fields in the LGN/V1 region, where three types of cone cells (L, M, S) respond to spectral stimuli through different excitation (+)/inhibition (-) patterns. (A2) presents the designed color channel combination scheme demonstrating antagonistic receptive fields. (B1) reflects hue differences in the visual cortex. (B2) displays the proposed method for enhancing hue representation.

where W represents the original convolution kernel weights of the 1×1 convolution, which are multiplied by the Jordan-form kernel to form the new convolution kernel; $E(X,Y)$ embeds all channels of X and Y into the odd and even positions of the new feature, respectively, following the specified sequence.

3.3. Edge Information Pathway

The edge contours of small targets serve as the core carriers of their shape features. To enhance the model's learnability, the ELLOM adaptive method is employed to generate edge images. The traditional Laplacian operator [7] is composed of the first two second-order differential operators in Eq.(3), while ELLOM consists of the latter two. Using learnable weights $(w_1, w_2, \cdots, w_9)$, it achieves adaptivity, where Σ represents the sum of the surrounding weights.

$$\begin{bmatrix}1&0&1\\0&-4&0\\1&0&1\end{bmatrix}\begin{bmatrix}0&1&0\\1&-4&1\\0&1&0\end{bmatrix}\begin{bmatrix}w_1&0&w_3\\0&-\Sigma&0\\w_7&0&w_9\end{bmatrix}\begin{bmatrix}0&w_2&0\\w_4&-\Sigma&w_6\\0&w_8&0\end{bmatrix} \quad (3)$$

Neurons in biological V1/V2 areas respond selectively to specific orientation stimuli—activating those aligned with the target direction while suppressing responses to other orientations [1, 2, 34], as illustrated in Fig. 3(A). To simulate this mechanism, OrSM implements the Enhancement-Suppression Convolution Kernel (ESCK) and an adaptive orientation selection method.

The construction of ESCK adopts a design concept that shifts from pixel-based differences to weight-based differences in differential convolution. Detailed design specifications are provided in the Supplementary Material (see Sec. A), and the corresponding ESCK are shown in Fig. 3(B).

To address kernel interference from direct superposition in existing multi-directional feature fusion methods, this paper proposes an adaptive directional kernel selection method (illustrated Fig. 3(C)). After ESCK generates eight distinct directional salient features $X^1, X^2, \ldots, X^8$ from the input feature X, a sub-network encodes the input feature map to produce a directional index map with values ranging from 0 to 7. This index map is then converted into eight binary mask maps $\{M_k\}_{k=0}^{7}$ (where M_k is 1 only at index k), which eliminate feature interference from non-salient directions. Finally, the filtered features are multiplied and fused with the original eight salient features, achieving adaptive selection of the most salient convolution kernel in the spatial domain. This approach fundamentally avoids mutual interference from direct superposition of multiple convolution kernels and enhances feature discriminability.

3.4. Feature Fusion Module

Color and edge information exhibit complementarity in small object detection: color distinguishes surface attributes, while edges localize spatial boundaries. Physiolog-

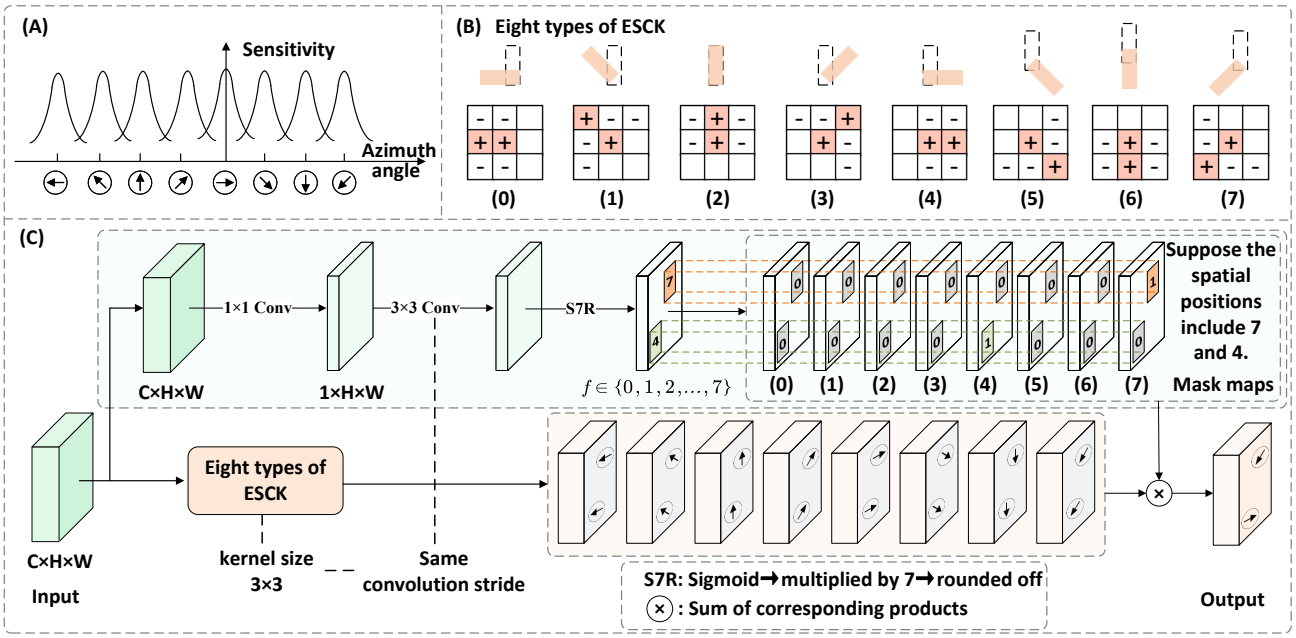

Figure 3. Design of OrSM in EIP. (A) Sensitivity variation of V1/V2 neurons under different stimulus angles. The horizontal axis denotes stimulus angle (0°–180°), and the vertical axis denotes neuronal response strength, illustrating response characteristics to grating stimuli of different orientations. (C) Specific implementation of OrSM.

ical studies have shown that the visual cortex area V4 simultaneously contains neuronal populations sensitive to color and those sensitive to form (edges, orientations, etc.) [37], demonstrating functional partitioning and feature interactions at the cortical level [43], as shown in Fig. 4(A). To simulate the interactive characteristics of the V4 area, FFM employs an information injection strategy to achieve hierarchical integration of dual-path features, with design details illustrated in Fig. 4(B).

Specifically, taking a single sample as an example, for features $Z_1 \in \mathbb{R}^{C\times H\times W}$ and $Z_2 \in \mathbb{R}^{C\times H\times W}$ from the color and edge extraction backbones, global average pooling and 1×1 convolution are first applied to learn the importance scores of each channel. Assume that each channel of feature Z_1 is represented as $\{p_1, p_2, \ldots, p_c\}$, while each channel of feature Z_2 is represented as $\{q_1, q_2, \ldots, q_c\}$. Then, after dimension adjustment, an outer product operation is performed to obtain the matrix $R \in \mathbb{R}^{1\times C\times C}$, referring to Eq.(4).

$$R[1, C, C] = \begin{bmatrix} p_1q_1 & p_1q_2 & \cdots & p_1q_c \\ p_2q_1 & p_2q_2 & \cdots & p_2q_c \\ \vdots & \vdots & \ddots & \vdots \\ p_cq_1 & p_cq_2 & \cdots & p_cq_c \end{bmatrix} \tag{4}$$

Each element of the matrix R quantifies the semantic association strength between channel α_i of feature Z_1 and channel α_j of feature Z_2. To reconstruct Z_2 by injecting information from Z_1, we find that using the matrix R as a 1×1 convolution kernel and convolving it with Z_2 can cleverly align with the semantic structure of feature Z_2. The modified 1×1 convolution kernel $K[C, C, 1, 1]$ is defined referring to Eq.(5).

If each channel of feature Z_2 is represented as $Z_2 = [Q_1, Q_2, \cdots, Q_c]^T$, where $Q_c \in \mathbb{R}^{H\times W}$ is the spatial feature of the c-th channel, and it is convolved with K, then the k-th output channel is computed using the k-th row of the convolution kernel, as shown in Eq.(6). It can be observed that this convolution result on Z_2 incorporates information from every channel of Z_1.

$$K[C, C, 1, 1] = \begin{bmatrix} [p_1q_1, p_1q_2, \cdots, p_1q_c] \\ [p_2q_1, p_2q_2, \cdots, p_2q_c] \\ \vdots \\ [p_cq_1, p_cq_2, \cdots, p_cq_c] \end{bmatrix} \tag{5}$$

$$\begin{aligned} Output_k &= [p_kq_1, p_kq_2, \ldots, p_kq_c] \cdot [Q_1, Q_2, \ldots, Q_c]^T \\ &= \sum_{i=1}^{c} (p_kq_i \cdot Q_i) \end{aligned} \tag{6}$$

It is worth noting that the matrix R, obtained from the outer product of two one-dimensional channel weight vectors, is mathematically a rank-1 matrix. This rank-1 property endows the matrix with a strong regularization effect,

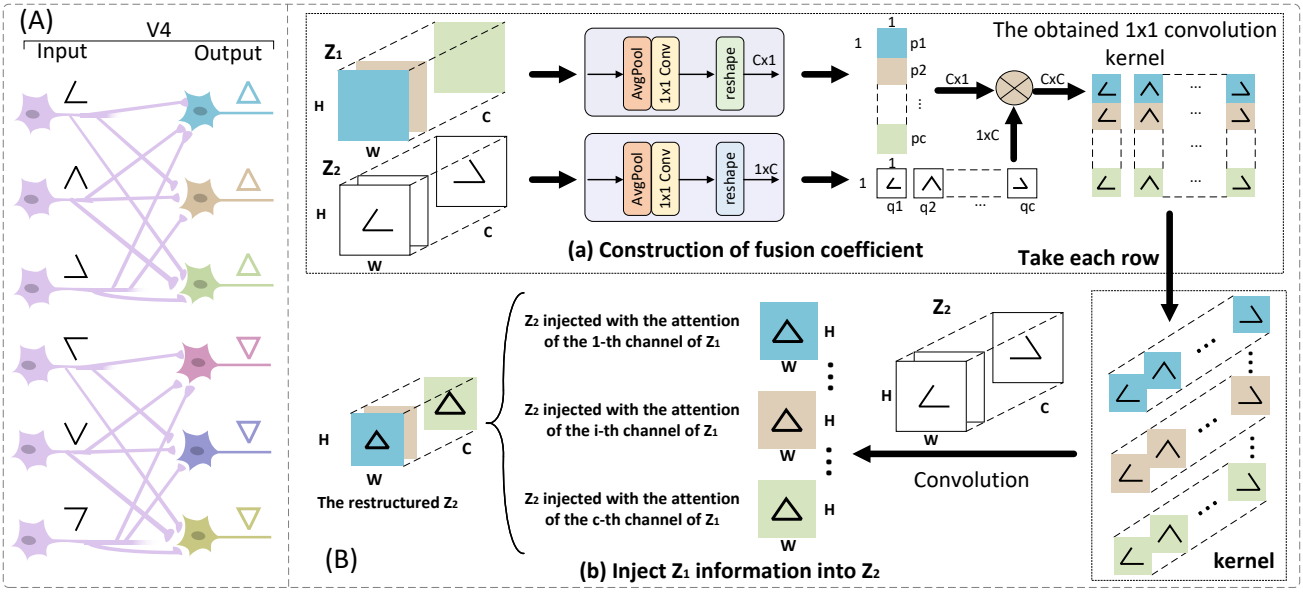

Figure 4. Design of the Color and Edge Feature Fusion Module (FFM). (A) Visual representation of the interaction between the color domain and form domain in the V4 area. (B) Illustrates the specific construction of FFM.

which is particularly beneficial for small objects with high noise levels. To mitigate the feature expression limitations imposed by the rank-1 structure, we preserve the diversity of channel features through the residual and concatenation operations of Eq.(7), thereby enhancing the model's detection robustness in complex remote sensing scenarios.

$$Output = Z_1 \| \left(Output_1 \| Output_2 \| \ldots \| Output_c + Z_2\right) \quad (7)$$

where $\|$ represents the channel-wise concatenation operation.

4. Experiments

4.1. Experimental Key Support

This study employs an optimized YOLO12 as the baseline model, without loading any pre-trained weights, to validate the model's generalization capability on three remote sensing small object datasets: VisDrone2019 [4], NWPU VHR-10 [5], and AI-TODv2 [9]. Systematic ablation experiments are primarily conducted on the VisDrone2019 dataset. Performance is evaluated using standard object detection metrics, including mAP_{50} and $mAP_{50\text{-}95}$ in YOLO format, AP, AP_{50}, and AP_{75} under COCO standard, as well as AP_{vt}, AP_t, AP_s, and AP_m for different target scales. The number of parameters (Params/M) and computational complexity (GFLOPs) are also reported to assess model efficiency (see Supplementary Material Sec. B for details).

4.2. Comparative Experiments and Analysis

We conducted comprehensive comparisons between BDNet and various current SOTA methods. All comparative experimental results with citations are sourced from the relevant literature. To ensure a fair comparison, we retrained high-performance detectors YOLO11, YOLO12, and YOLO26 under the same experimental conditions as BDNet.

Evaluation results on NWPU VHR-10. As shown in Tab. 1, although the proposed method achieves optimal performance in only some categories (BD, TC, BC, HA), its performance distribution across all categories is more balanced with the smallest variance, ultimately achieving the highest mean average precision (mAP_{50}) of 94.1%.This demonstrates that the color extraction backbone effectively enhances the recognition capability for targets with low color contrast (such as TC and BC, which have similar tones to the background), the edge extraction backbone accurately captures structured targets (such as the problem of blurred edges and contours in BD and HA), resolving the issue of edge blurring.

Evaluation results on VisDrone2019. In Tab. 2, the proposed method achieves the best results on the validation set for both mAP_{50} (50.5%) and $mAP_{50\text{-}95}$ (31.2%), while requiring significantly fewer parameters and lower computational cost. This demonstrates that BDNet maintains advantages in both accuracy and efficiency, validating the superiority of its dual-backbone architecture. For object detection in complex UAV imagery, the model efficiently extracts and integrates key color and edge features with a

Table 1. Performance comparison across categories on NWPU VHR-10 test set. Abbreviations represent: airplane (AI), ship (SH), storage tank (ST), baseball diamond (BD), tennis court (TC), basketball court (BC), ground track field (GT), harbor (HA), bridge (BR), vehicle (VE). **bold** indicates the best performance.

Method	AI	SH	ST	BD	TC	BC	GT	HA	BR	VE	mAP_{50}
DINO(4scale)[50]	98.9	93.4	97.9	97.9	94.1	95.4	95.8	91.6	71.2	92.5	92.7
MaskFormer[45]	**100**	93.4	92.6	96.9	94.6	95.6	**100**	79.5	**94.1**	91.3	93.8
YOLOv8[29]	99.5	90.7	97.9	96.6	**96.9**	90.3	99.5	85.8	77.8	91.4	92.6
RS-YOLO[12]	99.9	**94.1**	96.6	97.8	93.8	92.7	97.6	89.5	84.7	92.3	93.9
YOLO-RDNet[29]	99.0	88.2	97.7	97.2	94.2	93.4	99.5	90.0	83.3	**93.5**	93.6
HWANet[19]	99.4	92.3	**99.0**	98.9	95.4	78.2	96.2	97.3	84.9	89.7	93.1
Ours	99.5	86.0	97.6	**99.5**	**96.9**	**99.5**	99.5	**99.0**	76.2	87.6	**94.1**

Table 2. Performance comparison on VisDrone2019 validation set. "-" indicates the metric was not adopted, **bold** indicates the best performance.

Method	mAP_{50}	$mAP_{50\text{-}95}$	Params	GFLOPs
Dab-DETR[28]	41.1	21.8	43.45	101
RTMDet-L[31]	46.8	29.3	52.26	79.97
EMAattention[33]	49.7	30.4	91.2	315.0
RT-DETR[54]	47.5	29.2	19.88	57.0
VMC-DETR[11]	45.9	27.9	-	70.5
UAV-DETR[51]	50.0	30.9	21.26	72.5
AOD-YOLO-S[46]	44.9	26.5	10.7	124.0
YOLOv8s-FSD [13]	46.7	28.5	13.0	43.2
RT-DETR-R50[52]	49.8	30.8	41.9	129.6
LUFE-Net[52]	50.2	30.9	9.7	**33.1**
YOLO11-I	46.1	28.7	25.3	87.28
YOLO12-I	47.3	29.3	29.2	89.4
YOLO26-I	45.2	27.7	26.3	93.8
Ours	**50.5**	**31.2**	**2.59**	52.44

Table 3. Performance comparison on the AI-TODv2 test set under the COCO evaluation standard. **Bold** indicates the best performance in the current comparison. LTDNet* is a high-precision variant of LTDNet.

Method	AP	AP_{50}	AP_{75}	AP_{vt}	AP_t	AP_s	AP_m
DetectoRS[38]	16.1	35.5	12.5	0.1	12.6	28.3	40.0
ESANet[47]	17.6	45.0	10.5	5.4	15.8	22.9	33.8
ORFENet[25]	18.9	44.4	12.7	6.9	18.4	23.4	30.3
DNTR[27]	23.1	51.9	16.8	9.7	22.7	27.2	36.1
DCENet[16]	23.5	53.9	16.8	8.5	**24.1**	28.1	37.1
LTDNet*[26]	23.0	54.6	15.5	8.9	23.6	27.2	33.1
Ours	**24.7**	**54.9**	**18.2**	**10.2**	23.5	**31.7**	**41.8**

streamlined parameter count.

AI-TODv2 Evaluation Results. Tab. 3 demonstrates that the proposed method achieves optimal performance across all metrics, including AP, AP_{50}, AP_{75}, AP_{vt}, AP_s, and AP_m, fully showcasing its superior detection performance. This comprehensive result demonstrates the unique value of the dual-backbone architecture in handling tiny objects: the color extraction pathway enhances recognition capability for low-contrast small targets, while the edge extraction pathway improves precision in capturing blurred contours. Ultimately, through the feature fusion mechanism, the model achieves accurate detection of targets of various sizes, particularly ultra-small objects in UAV imagery.

4.3. Ablation Experiments and Analysis

To evaluate the contributions of each module in BDNet, we conducted ablation experiments on the VisDrone2019 validation set and performed feature visualization analysis. We further provide additional supplemental analysis and verification (see Supplementary Material Sec. C for details). to validate the effectiveness of the model design.

CIP Internal Components. As illustrated in Table 4, CAM elevates mAP_{50} from the baseline 47.8% to 48.0%, capturing finer color contrasts in low-illumination scenarios. Subsequently, VCHM further enriches feature discriminability through hierarchical hue processing, pushing performance to 48.3%. When integrated, these two modules achieve a combined mAP_{50} of 48.8%, underscoring their complementary functionality.

EIP Internal Components. According to Table 4, ELLOM component raises mAP_{50} to 48.1% by emphasizing semantically salient contours, while OrSM further improves it to 48.3% through biologically-inspired directional filtering. Together, they attain 48.8%, demonstrating a clear synergistic interaction. OrSM proves particularly critical by modeling precise edge orientation and spatial relationships, effectively mitigating motion-induced blurring artifacts and reinforcing structural coherence in small object detection.

Complete Model. As shown in Table 5, staged integration of CIP, EIP, and FFM systematically elevates mAP_{50} from 47.8% to 50.5%, with a manageable increase in parameters and computational load. This structured enhancement confirms each module's indispensability and illustrates their cumulative effect—color and edge pathways resolve key challenges in low-contrast and blurred-boundary conditions, while FFM enables cross-dimensional feature

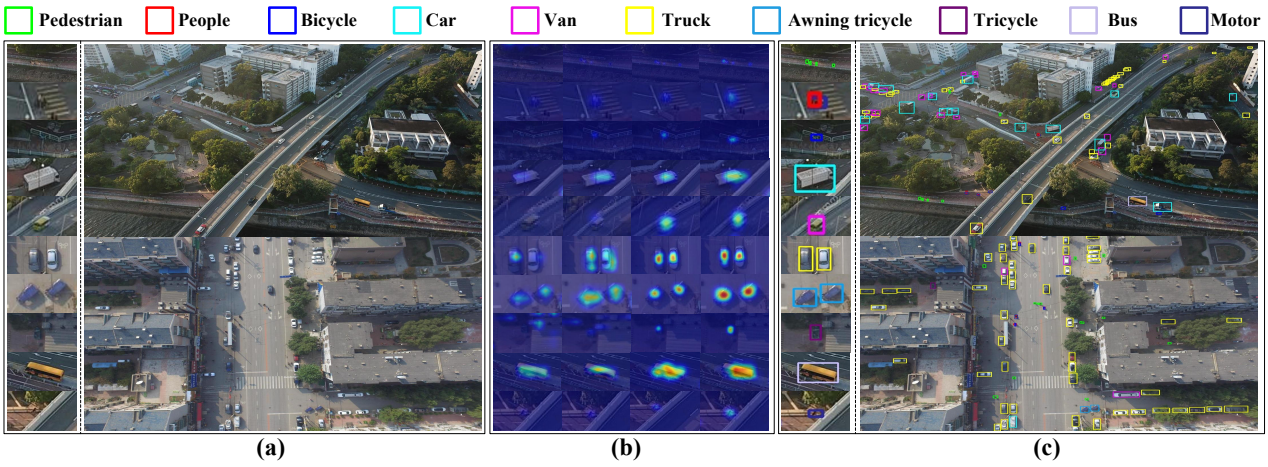

Figure 5. Visualization. (a) displays the input image containing ten categories of small objects with blurred edges and low contrast, intuitively illustrating the complexity of the detection task; (b) presents a four-column layout showing heatmaps of features from different layers for the ten categories of small objects (starting from the baseline model, sequentially adding CIP, EIP, and FFM), where higher brightness indicates stronger feature response and target discriminability; (c) shows the output image and detection results.

consolidation, culminating in a balanced, powerful detection network.

Table 4. Ablation Study on CIP and EIP Internal Components

CAM	VCHM	mAP_{50}	Params	GFLOPs
×	×	47.8	1.98	45.98
✓	×	48.0	1.98	46.39
×	✓	48.3	2.04	47.15
✓	✓	48.8	2.04	47.56
ELLOM	OrSM	mAP_{50}	Params	GFLOPs
×	×	47.8	1.98	45.98
✓	×	48.1	1.98	46.38
×	✓	48.3	2.04	47.13
✓	✓	48.8	2.04	47.52

Table 5. Ablation Study on the Entire Model

CIP	EIP	FFM	mAP_{50}	Params	GFLOPs
×	×	×	47.8	1.98	45.98
✓	×	×	48.8	2.04	47.56
×	✓	×	48.8	2.04	47.52
×	×	✓	49.4	2.48	49.20
✓	✓	×	49.8	2.52	52.31
×	✓	✓	50.3	2.53	50.78
✓	×	✓	49.7	2.53	50.75
✓	✓	✓	50.5	2.59	52.44

Visualization Analysis. To qualitatively assess our approach, we selected a challenging scene from VisDrone2019 containing ten distinct object categories and compared the feature heatmaps and final detection outputs generated by the baseline and BDNet (Fig. 5). The baseline model yields indistinct heatmaps, exhibiting weak activation for both low-contrast targets (e.g., People, Bicycle) and objects with blurred contours (e.g., Pedestrian), causing foreground-background ambiguity. Incorporating CIP noticeably intensifies the feature responses; however, certain edge details remain blurred, and the discriminability between adjacent instances is still limited. The subsequent integration of the EIP branch effectively suppresses background noise and further sharpens the feature representations. Ultimately, FFM integration achieves target representation through color-edge feature fusion, yielding strong and accurate feature responses.

These visualization results systematically validate that the proposed BDNet effectively addresses the feature extraction challenges arising from low color contrast and blurred edges in remote sensing small targets, demonstrating robust generalization.

5. Conclusion

This paper proposes a biologically-inspired dual-backbone network for small object detection (BDNet), designed to address poor feature extraction in remote sensing images caused by small objects' low color contrast and blurred edges. BDNet simulates the processing mechanism of biological vision systems by constructing a color information pathway (CIP), an edge information pathway (EIP), and a feature fusion module, strengthening color and edge feature extraction. Extensive experiments on three datasets validate its effectiveness. This research provides a biologically plausible solution for small object detection in complex remote sensing scenarios.

6. Acknowledgment

This work is supported by the National Natural Science Foundation of China (Grant No. 62266006) and the Guangxi Natural Science Foundation (Grant Nos. 2025GXNSFAA069690 and 2020GXNSFDA297006).

References

[1] Matteo Carandini and David J Heeger. Normalization as a canonical neural computation. *Nature reviews neuroscience*, 13(1):51–62, 2012. 4

[2] James R Cavanaugh, Wyeth Bair, and J Anthony Movshon. Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons. *Journal of neurophysiology*, 88(5):2530–2546, 2002. 4

[3] Xi Chen and Chuan Lin. Evmnet: Eagle visual mechanism-inspired lightweight network for small object detection in uav aerial images. *Digital Signal Processing*, 158:104957, 2025. 2

[4] Gong Cheng and Junwei Han. A survey on object detection in optical remote sensing images. *ISPRS journal of photogrammetry and remote sensing*, 117:11–28, 2016. 1

[5] Gong Cheng, Junwei Han, Peicheng Zhou, and Lei Guo. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 98: 119–132, 2014. 6

[6] Dennis M Dacey. Circuitry for color coding in the primate retina. *Proceedings of the National Academy of Sciences*, 93 (2):582–588, 1996. 3

[7] Pierre Simon de Laplace. *Théorie analytique des probabilités*. Courcier, 1820. 4

[8] Andrew Derrington. The lateral geniculate nucleus. *Current Biology*, 11(16):R635–R637, 2001. 3

[9] Dawei Du, Yuankai Qi, Hongyang Yu, Yifan Yang, Kaiwen Duan, Guorong Li, Weigang Zhang, Qingming Huang, and Qi Tian. The unmanned aerial vehicle benchmark: Object detection and tracking. In *Proceedings of the European conference on computer vision (ECCV)*, pages 370–386, 2018. 6

[4] Dawei Du, Pengfei Zhu, Longyin Wen, Xiao Bian, Haibin Lin, Qinghua Hu, Tao Peng, Jiayu Zheng, Xinyao Wang, Yue Zhang, et al. Visdrone-det2019: The vision meets drone object detection in image challenge results. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pages 0–0, 2019. 6, 2

[11] Xin Ge, Liping Qi, Qingsen Yan, Jinqiu Sun, Yu Zhu, and Yanning Zhang. Enhancing real-time aerial image object detection with high-frequency feature learning and context-aware fusion. *Remote Sensing*, 17(12):1994, 2025. 7

[12] Dongen Guo, Zhuoke Zhou, Fengshuo Guo, Chaoxin Jia, Xiaohong Huang, Jiangfan Feng, and Zhen Shen. A remote sensing target detection model based on lightweight feature enhancement and feature refinement extraction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:9569–9581, 2024. 7

[13] Zihan Guo, Xingyu Mu, Chao Chang, Weijie Cheng, and Xincheng Tian. An enhanced framework for small object detection with middle-order interaction and adaptive cross-scale aggregation. *Engineering Applications of Artificial Intelligence*, 159:111730, 2025. 7

[14] CA Heywood, A Gadotti, and A Cowey. Cortical area v4 and its role in the perception of color. *Journal of Neuroscience*, 12(10):4056–4065, 1992. 1

[15] Liping Hou, Ke Lu, and Jian Xue. Refined one-stage oriented object detection method for remote sensing images. *IEEE Transactions on Image Processing*, 31:1545–1558, 2022. 2

[16] Xinkai Hu, Zhida Ren, Uzair Aslam Bhatti, Mengxing Huang, and Yirong Wu. Dcedet: Tiny object detection in remote sensing images based on dual-contrast feature enhancement and dynamic distance measurement. *Remote Sensing*, 17(16):2876, 2025. 7

[17] Sihan Huang, Chuan Lin, Xintong Jiang, and Zhenshen Qu. Brstd: Bio-inspired remote sensing tiny object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 2

[18] David H Hubel and Torsten N Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243, 1968. 1

[19] Baohua Jin, Fukang Yin, Wenpeng Cai, Hongchan Li, Haodong Zhu, Wei Huang, Qinggang Wu, Hui Chen, and Zhongchuan Sun. Hwanet: A haar wavelet-based attention network for remote sensing object detection. *PLoS One*, 20 (9):e0330759, 2025. 7

[20] Rodney LaLonde, Dong Zhang, and Mubarak Shah. Clusternet: Detecting small objects in large scenes by exploiting spatio-temporal information. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4003–4012, 2018. 2

[21] Peichao Li, Anupam K Garg, Li A Zhang, Mohammad S Rashid, and Edward M Callaway. Cone opponent functional domains in primary visual cortex combine signals for color appearance mechanisms. *Nature communications*, 13 (1):6344, 2022. 3

[22] Yiyu Li, Ke Xu, Gerhard Petrus Hancke, and Rynson WH Lau. Color shift estimation-and-correction for image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25389–25398, 2024. 1

[23] Yanfeng Li, Kahou Chan, Yue Sun, Chantong Lam, Tong Tong, Zitong Yu, Keren Fu, Xiaohong Liu, and Tao Tan. Moedit: On learning quantity perception for multi-object image editing. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2683–2693, 2025. 1

[24] Binhui Liu, Chunyan Xu, Zhen Cui, and Jian Yang. Progressive context-dependent inference for object detection in remote sensing imagery. *IEEE Transactions on Image Processing*, 32:580–590, 2022. 2

[25] Dongyang Liu, Junping Zhang, Yunxiao Qi, Yinhu Wu, and Ye Zhang. Tiny object detection in remote sensing images based on object reconstruction and multiple receptive field adaptive feature enhancement. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024. 7

[26] Dongyang Liu, Junping Zhang, Yunxiao Qi, Yunqiao Xi, and Jing Jin. Exploring lightweight structures for tiny object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2025. 7

[27] Hou-I Liu, Yu-Wen Tseng, Kai-Cheng Chang, Pin-Jyun Wang, Hong-Han Shuai, and Wen-Huang Cheng. A denoising fpn with transformer r-cnn for tiny object detection. *IEEE transactions on geoscience and remote sensing*, 62:1–15, 2024. 7

[28] Shilong Liu, Feng Li, Hao Zhang, Xiao Yang, Xianbiao Qi, Hang Su, Jun Zhu, and Lei Zhang. Dab-detr: Dynamic anchor boxes are better queries for detr. *arXiv preprint arXiv:2201.12329*, 2022. 7

[29] XiaoDong Liu, Hao Zhang, Wenyin Gong, and Xiang Li. Yolo-pdnet: Small target recognition improvement for remote sensing image based on yolov8. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE, 2024. 7

[30] Ye Liu, Ming Li, Xian Zhang, Yiliang Lu, Hongliang Gong, Jiapeng Yin, Zheyuan Chen, Liling Qian, Yupeng Yang, Ian Max Andolina, et al. Hierarchical representation for chromatic processing across macaque v1, v2, and v4. *Neuron*, 108(3):538–550, 2020. 3

[31] Chengqi Lyu, Wenwei Zhang, Haian Huang, Yue Zhou, Yudong Wang, Yanyi Liu, Shilong Zhang, and Kai Chen. Rtmdet: An empirical study of designing real-time object detectors. *arXiv preprint arXiv:2212.07784*, 2022. 7

[32] Yansong Niu, Chuan Lin, Xintong Jiang, and Zhenshen Qu. Vstdet: A lightweight small object detection network inspired by the ventral visual pathway. *Applied Soft Computing*, 171:112775, 2025. 2

[33] Daliang Ouyang, Su He, Guozhong Zhang, Mingzhu Luo, Huaiyong Guo, Jian Zhan, and Zhijie Huang. Efficient multi-scale attention module with cross-spatial learning. In *ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1–5. IEEE, 2023. 7

[34] Xintao Pang, Chuan Lin, Fuzhang Li, and Yongcai Pan. Bio-inspired xyw parallel pathway edge detection network. *Expert Systems with Applications*, 237:121649, 2024. 4

[35] Xintao Pang, Fengjuan Yao, Yanming Zhang, Yue Sun, Edmundo Patricio Lopes Lao, Chuan Lin, Patrick Cheong-Iao Pang, Wei Wang, Wei Li, Zhifan Gao, et al. Blenet: a bio-inspired lightweight and efficient network for left ventricle segmentation in echocardiography. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025. 3

[36] Anitha Pasupathy and Charles E Connor. Shape representation in area v4: position-specific tuning for boundary conformation. *Journal of neurophysiology*, 86(5):2505–2519, 2001. 1

[37] Anitha Pasupathy, Dina V Popovkina, and Taekjun Kim. Visual functions of primate area v4. *Annual review of vision science*, 6(1):363–385, 2020. 5

[38] Siyuan Qiao, Liang-Chieh Chen, and Alan Yuille. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10213–10224, 2021. 7

[39] Anna W Roe, Leonardo Chelazzi, Charles E Connor, Bevil R Conway, Ichiro Fujita, Jack L Gallant, Haidong Lu, and Wim Vanduffel. Toward a unified theory of visual area v4. *Neuron*, 74(1):12–29, 2012. 1

[40] Robert Shapley and Michael J Hawken. Color in the cortex: single-and double-opponent cells. *Vision research*, 51 (7):701–717, 2011. 1

[41] Zhiqiang Shen, Zhuang Liu, Jianguo Li, Yu-Gang Jiang, Yurong Chen, and Xiangyang Xue. Dsod: Learning deeply supervised object detectors from scratch. In *Proceedings of the IEEE international conference on computer vision*, pages 1919–1927, 2017. 1

[42] Huixin Sun, Runqi Wang, Yanjing Li, Linlin Yang, Shaohui Lin, Xianbin Cao, and Baochang Zhang. Set: Spectral enhancement for tiny object detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 4713–4723, 2025. 1

[43] Hisashi Tanigawa, Haidong D Lu, and Anna W Roe. Functional organization for color and orientation in macaque v4. *Nature neuroscience*, 13(12):1542–1548, 2010. 5

[44] Gang Wang, Xin Yang, Liang Li, Kai Gao, Jin Gao, Jia-yi Zhang, Da-jun Xing, and Yi-zheng Wang. Tiny drone object detection in videos guided by the bio-inspired magnocellular computation model. *Applied Soft Computing*, 163:111892, 2024. 2

[45] Keyan Wang, Feiyu Bai, Jiaojiao Li, Yajing Liu, and Yunsong Li. Mashformer: A novel multiscale aware hybrid detector for remote sensing object detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:2753–2763, 2023. 7

[46] Yingqing Wang, Weili Zeng, Ziyu Zhao, Baogeng Li, and Zhibin Quan. Aod-yolo: A self-modulating multi-scale feature aggregation mechanism for small object detection in airport surface scenes. *Applied Soft Computing*, page 113849, 2025. 7

[47] Jixiang Wu, Zongxu Pan, Bin Lei, and Yuxin Hu. Fsanet: Feature-and-spatial-aligned network for tiny object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022. 7

[48] Xin Wu, Danfeng Hong, Jiaojiao Tian, Jocelyn Chanussot, Wei Li, and Ran Tao. Orsim detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7):5146–5158, 2019. 2

[49] Chang Xu, Jian Ding, Jinwang Wang, Wen Yang, Huai Yu, Lei Yu, and Gui-Song Xia. Dynamic coarse-to-fine learning for oriented tiny object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7318–7328, 2023. 1, 2

[50] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M Ni, and Heung-Yeung Shum. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*, 2022. 7

[51] Huaxiang Zhang, Kai Liu, Zhongxue Gan, and Guo-Niu Zhu. Uav-detr: efficient end-to-end object detection for unmanned aerial vehicle imagery. *arXiv preprint arXiv:2501.01855*, 2025. 7

[52] Xi Zhang, Huicheng Lai, Tingting Yuan, Guxue Gao, and Di Jiang. Lufe-net: A lightweight uav aerial object feature enhancement network for complex scenarios. *Optics & Laser Technology*, 195:114543, 2026. 7

[53] Yundong Zhang, Huiye Liu, and Qiang Hu. Transfuse: Fusing transformers and cnns for medical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 14–24. Springer, 2021. 1

[54] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, and Jie Chen. Detrs beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16965–16974, 2024. 7

[55] Haowei Zhu, Wenjing Ke, Dong Li, Ji Liu, Lu Tian, and Yi Shan. Dual cross-attention learning for fine-grained visual categorization and object re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4692–4702, 2022. 1