• 专题一：双清论坛“低空经济信息系统与安全管理理论及关键技术” •
DOI: 10.3724/BNSFC-2025-0219

脑启发的“低慢小”目标视觉感知研究现状与展望*

王 刚[1†] 杨 欣[1,2†] 高 晋[3**] 李 椋[1] 朱鹏飞[4] 胡卫明[3] 王以政[1]

1. 军事科学院 军事认知与脑科学研究所，北京 100850
2. 电子科技大学 生命科学与技术学院，成都 611731
3. 中国科学院自动化研究所，北京 100190
4. 天津大学 智能与计算学部，天津 300350

[**摘 要**] 移动速度慢、雷达反射面小的“低慢小”目标需要引入光电传感器(红外/可见光等)，协同雷达应对杂波干扰和电磁压制等，进而基于深度学习等目标视觉感知技术实现智能化光电信息处理。然而，“低慢小”目标外观弱隐、背景复杂、场景多变，当前主流的智能感知技术仍面临虚警/漏检率高、计算能效比差、可信可靠性不足等问题。如何借鉴人类视觉系统(Human Visual System，HVS)获取、感知和认知加工视觉信息的优势特性，提高“低慢小”目标视觉感知的关键性能和智能化水平，已成为业内新的研究热点。本文将在介绍相关技术发展现状的同时，进一步对该领域研究的未来发展趋势进行讨论和展望。

[**关键词**] 视觉感知；人类视觉系统；“低慢小”目标；深度学习；表观—运动建模

现代无人机技术快速发展的同时，也为总体国家安全带来严峻挑战。例如，近年来在机场和要地等敏感地域的无人机黑飞事件频频引发社会关注；我国南海维权行动多次遭遇非合作无人机袭扰威胁；部分邻国和外域国家持续使用无人机对我国管辖区域实施非法侦察监视等。面向国防和公共安全等重大现实需求，我国亟需加强对飞行高度低、运动速度慢/多变、雷达反射面小(相对于战斗机、导弹等飞行器)的“低慢小”目标风险防范能力，加快提升复杂背景“低慢小”目标视觉感知技术及装备的智能化水平[1]。尽管以YOLO、Transformer等为骨干网络的深度学习算法已经显著提高了图像小目标检测识别的任务性能，但受成像信噪比和背景干扰等因素影响，当前主流的“低慢小”目标视觉感知技术仍面临虚警/漏检率高、计算能效比差、可靠性不足等问题。例如，在2025年IEEE国际计算机视觉与模式识别会议(IEEE Conference on Computer Vision and Pattern Recognition，CVPR)举办的Anti-UAV挑战赛中，目标检测赛道的绝大多数参赛算法准确率不足60%，与高复杂动态应用场景的实际需求差距较大。

相比于机器视觉，人类视觉系统(Human Visual System，HVS)在获取、感知和认知加工视觉信息方面具有独特优势。例如，经过简短培训的人员等能够在复杂光电画面中综合利用外观特征、移动方向、运动模态等时空上下文信息快速辨识飞鸟、无人机等极小飞行目标。借鉴生物甚至人类视觉系统获取、感知和认知加工视觉信息的优势特性，提高“低慢小”目标视觉感知的关键性能和智能化水平，已成为业内新的研究热点。围绕“低慢小”目标视觉感知问题，本文重点关注“脑科学—人工智能”跨学科交叉融合的创新技术路线，梳理国内外脑启发视觉计算特别是“低慢小”目标感知方面的前沿进展，探讨现有技术在实际应用中尚存在的瓶颈性问题，并对发展趋势进行展望。

收稿日期：2025-07-30；修回日期：2026-01-09
† 共同第一作者。
* 本文根据国家自然科学基金委员会第407期“双清论坛”讨论的内容整理。
** 通信作者，Email：jin.gao@nlpr.ia.ac.cn
本文受到国家自然科学基金项目(62422317)、北京市科技新星项目(20240484703)的资助。

引用格式： 王刚，杨欣，高晋，等. 脑启发的“低慢小”目标视觉感知研究现状与展望. 中国科学基金，2026，40(1)：65–72.
Wang G，Yang X，Gao J，et al. Brain-inspired visual perception of low-altitude slow and small targets：Current research and future perspectives. Bulletin of National Natural Science Foundation of China，2026，40(1)：65–72.(in Chinese)

1 “低慢小”目标视觉感知瓶颈问题分析

“低慢小”目标易隐藏于楼宇、山坳或树林等复杂背景中，传统雷达探测面临杂波干扰、电磁压制和低空盲区等问题[2]。光电探测相比于雷达探测具有隐蔽性好、可辨性高、抗干扰性强等特点，可有效弥补雷达探测手段的短板，近年来已逐渐成为低空安全探测的必要手段。特别是在“包以德循环(发现—定向—决策—执行)”的各个流程，均需要开发基于图像/视频的视觉感知技术，实现高度自动化的高通量光电数据处理分析。然而，远距离微/小型无人机等飞行目标在光电视频中的像素尺寸极小(≤10×10像素)[3]，“低慢小”目标视觉感知在现实应用中仍然存在诸多瓶颈问题。

一是安全任务条件下的虚警率与灵敏度的平衡问题。区别于通用目标检测任务，在真实应用场景条件下“低慢小”目标出现的概率非常低，不仅会造成目标特性数据匮乏的矛盾，还会导致训练数据与真实应用场景数据的分布差异；大气湍流扰动或背景遮挡会进一步导致目标非连续显现。在低空安全任务中，如何保证系统以低虚警率和低漏检率来输出低概率显现目标的检测结果是当前面临的一个关键瓶颈问题。

其次，远距离“低慢小”目标的光电成像尺寸极小，形状和纹理等关键视觉特征严重缺失，天空背景中云层和强光的动态变化会干扰目标的显著性特征提取，传感器变焦或失焦也会引入显著的尺度变化和图像退化问题，在极低目标信噪比的条件下很难解算估计目标类别置信度，从而导致漏检率高和虚警率高等问题。

此外，“低慢小”目标的型号、颜色和外形特征复杂多样，飞鸟、风筝、气球、民航客机、楼宇边角、塔吊等动/静态干扰物与“低慢小”无人机在视觉特征上具有高度相似性。因此，系统在输出潜在目标时很容易将各种背景干扰物与“低慢小”目标混淆，导致较高的虚警率[4]。

业内基于稀疏编码[5]、超分辨率[6]、深度学习[7]等技术已经提出较多光电目标感知方案[8-11]，在技术层面基本实现了对面目标和部分小目标(≤30×30像素)特征增强/学习和对复杂背景特征刻画等。针对数据匮乏问题，近期若干研究[1,12-14]通过构建专门数据集来助力深度网络模型训练和性能评估(表1)。然而，目前“低慢小”目标视觉感知的性能仍然难以满足现实需求，特别是与人脑的智能化程度相比还存在较大差距[15]。在结构和功能等多维度借鉴人脑发现、识别、跟踪小目标的优势特性，进而研发高度智能化、无人化、自主化的脑启发“低慢小”目标视觉感知系统，逐渐受到学术界和产业界的广泛关注。

2 脑启发的“低慢小”目标感知

2.1 脑视觉感知机制概述

脑视觉系统既具备多尺度静态信息处理能力，还具有多时空尺度动态信息加工的机制[16]。特别地，如图1所示，在瞬时阶段，视网膜—外膝体等区域的大细胞通路等能够快速捕捉局部动态信息[17]；然后，视皮层背侧通路(dorsal pathway)加工光流等信息；在短时阶段，视皮层背侧通路与腹侧通路(ventral pathway)交互投射并调制融合多尺度特征，判别目标类别[18]；在长时认知阶段，前额叶等高级脑区轨迹推理与上下文关联能力可支撑高阶语义判别[19]等。具体到视觉运动目标检测任务上，生物视觉在静态信息加工的基础上，还呈现出“注意引导—定位预测—模态辨识”等环节的递进链路。

人类视觉系统对动态信息的加工始于视网膜[25]。视网膜中央凹区域高空间分辨率的细胞主要负责精细采样，而周边区域则更多地负责运动检测和整体结构感知，这种选择性机制通过非均匀特征采样和编码有效平衡了感知性能和资源利用率。视网膜双极细胞具备“ON/OFF—中心”拮抗感受野的细胞类型，可以有效处理亮/暗变化，使得视觉系统能够精确感知亮暗边缘与颜

表1 业内主要“低慢小”目标相关数据集示例

Table 1 Datasets Containing Low-altitude Slow and Small Targets

数据库	视频数	模态	分辨率/帧率	帧数	平均目标大小(像素)	背景类型
Anti-UAV[20]	318	RGB-T	640×512 25FPS	29.7万	125×63(RGB)52×30(IR)	建筑物、云层、树木
AntiUAV410[1]	410	红外	640×512 25FPS	43.9万	30×30	建筑物、山脉、森林、城市区域、云层、水面
SIRSTD[21]	39	红外	640×512 25FPS	5万	7×7	建筑物、山脉、森林、城市区域、云层
Drone-vs-Bird[22]	77	可见光	720×576、3840×2160	10.5万	34×23	天空或植被、不同天气类型(多云、晴)、直射阳光眩光
NPS-Drones[23]	50	可见光	1920×1080、1280×960 30FPS	7万	16×12	天空、草地
ARD-MAV[24]	60	可见光	1920×1080 30FPS	10.6万	20×20	复杂背景、非平面场景、遮挡、快速移动

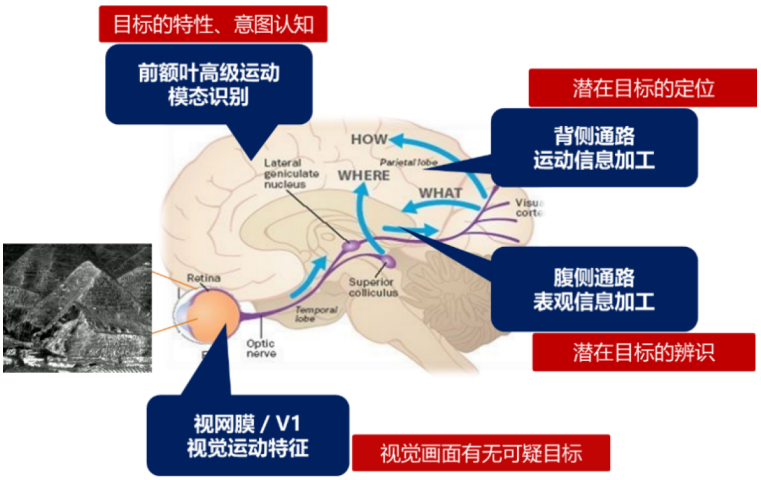

图1　人类视觉系统感知“低慢小”目标的信息加工过程示意图

Fig.1　Illustration of Human Visual Systems on the Perception of Low-slow-small Targets

色对比。神经节细胞各向同性感受野强化了侧抑制机制的结构基础,提升了对边缘、局部对比度的感知能力[26]。视觉信号进入大脑皮层后,视觉信息主要沿腹侧通路进行处理。初级视觉皮层(V1)中的视觉神经元具有方位选择性、边缘检测和空间频率选择性[27],可有效提取图像内像素间的几何结构特征。接着,视觉信号投射传输至V2、V4、IT等视觉区域,逐层整合出与形状、颜色、纹理等有关的高层次语义信息[28]。高层次视觉处理机制还存在自上而下的动态调节,使得视觉系统即使在干扰复杂的环境中,也能保持稳定而精准的识别能力。

人类视觉系统对动态信息的加工同样始于视网膜,具有信息时延能力的无长突细胞能够与神经节细胞配合快速提取运动显著性信息,辅助生成视觉注意。灵长类初级视觉皮层(V1)神经元具有运动方向选择性(即当动态刺激方向恰好与其偏好方向一致时,会引发强烈的响应),在光流特征提取过程中起到重要作用。进一步的动态信息由视皮层背侧通路的脑区完成,例如,视皮层V5区在整合运动信息方面具备更强的全局动态建模能力,该区域的神经元不再只关注某条边缘的运动,而是对整个物体的运动方向形成统一的响应。更高级的运动模态信息(例如移动轨迹、姿态变化规律等)则可能涉及工作记忆等机制的参与,一般认为由前额叶等高级脑区负责实现。

2.2　脑启发的“低慢小”目标感知技术

2.2.1　基于静态特征的技术

在人类视觉系统中,静态视觉信息的处理呈现出从浅层感知到深层认知的层级处理模式。受此启发,近年来“低慢小”目标检测领域在静态特征建模方面逐渐形成了模仿人类视觉分层加工机制的多层次技术路径,从浅层的纹理增强,到中层的结构整合,再到高层的显著性建模。

低层视觉方面,部分研究聚焦于模拟人类早期视觉系统对边缘、纹理和局部对比度[29]等信息的敏感性,尤其是视网膜及V1区对高空间频率信号的响应特征。例如,LCM(Local Contrast Measure)模型[30]借鉴了视网膜水平细胞及V1皮层中侧抑制机制,提出了有效抑制强背景干扰并增强小目标区域的方法。在此基础上,SALCM(Size-Aware Local Contrast Measure)模型[31]进一步引入尺寸感知机制,通过融合多种局部特征分量与Canny边缘检测器估计目标形状,增强了模型对不同尺度目标的检测能力。中级视觉的特征建模则更多关注空间频率整合与结构感知机制。HS-FPN(High-frequency Spatial-aware Feature Pyramid Network)模型[32]通过高通滤波提取高频响应以增强小目标特征表达,并引入空间依赖性感知模块以捕捉特征间的空间结构关系。该设计与人类视觉系统中V2、V4区域在边界整合与结构识别方面的功能相一致,有效模拟了局部特征向全局结构的集成过程。

在更高层级的视觉建模方面,研究者进一步引入自上而下的注意机制来增强目标显著性[33]。例如,BRSTD(Bio-inspired Remote Sensing Tiny object Detection)模型[34]通过构建具有X、Y、W细胞拮抗感受野特性的空间通道注意模块,并引入自上而下的抑制机制,模拟高级皮层对低层次区域的调控,以深层语义信息引导低层特征的选择性增强,从而提高目标区域的判别能力。此外,针对小目标本身语义信息稀缺的特点,ILNet(Infrared Low-level Network)模型[35]认为其区域显著但语义薄弱,强调对浅层特征的增强与重建,提升了浅层特征在深层语义建模中的表达能力。而IAANet(Interior Attention-Aware Network)模型[36]则进一步引入Transformer架构以建模像素之间的全局依赖,通过区域候选与注意力感知机制,

从粗到细实现显著区域的精确定位。

2.2.2 基于动态特征的技术

考虑到“低慢小”目标的静态特征相对有限，业内开始重视挖掘和利用其动态特征[37]，意图通过时空特征类脑融合等方式提升“低慢小”目标检测能力。

针对“低慢小”目标定位困难的问题，STDMANet (Spatio-Temporal Difference Multi-scale Attention Network)模型[38]提出通过时间多尺度特征提取器获取不同时域下的时空表示，并引入空间多尺度细化模块以增强小目标的局部显著性。Magno模型[37]借鉴生物视网膜运动信息加工特性，利用时空域多帧图像信息增强持续动态变化区域的能量，通过“目标—背景”的运动特征解耦与融合来降低目标位置的不确定性，由此提出的脑启发视频小目标检测算法(MVTOD)显著提升了复杂背景目标检测能力(表2、图2)。OTHR(On-off Time-delay with Hassenstein-Reichardt Correlator)模型[39]则模拟初级视皮层运动方向选择性机制，提出了面向运动小目标的光流计算模型，显著降低小目标光流刻画的误差，进而增强“低慢小”目标区域的特异性，提高算法抑制复杂背景的能力。

围绕“低慢小”目标类别辨识能力差的问题，业内人员尝试采用更多维度的特征进行模式判别。EV-UAV方法[40]基于事件相机运动小目标脉冲数据集，挖掘时空事件点云的特征表达，对稀疏事件点云中移动目标生成的事件进行分段，进而实现目标与背景的高置信度区分。多尺度空间运动检测法(Multi-Scale Space Kinematic, MUSAK)[41]提出了基于运动特征的多尺度空间运动学检测方法，在三个尺度空间提取三维、伪三维和二维运动学参数，构建三个基于门控循环单元的无人机识别分支，相比于二维图像识别方法显著提升了目标识别能力。

为了进一步提高对目标运动模态和属性特征的刻画能力，JTD-UAV方法[42]采用视觉语言多模态大模型模拟人类视觉系统前额叶的功能，将“低慢小”目标感知与视觉问答相结合，在实现“低慢小”目标精准定位的同时，生成对其行为与意图的自然语言描述，持续判断并刻画其运动状态与行为意图，从而促进更加全面的目标监控体系的构建。

2.3 脑启发“低慢小”目标感知的瓶颈问题

需要指出，目前仍然没有任何技术和系统能够在核心指标和综合表现上接近人脑的感知水平，脑启发的“低慢小”目标感知技术仍存在相当数量的瓶颈问题。

(1)对深层次脑结构和脑原理的借鉴程度不足。类脑目标感知是深度借鉴脑科学所阐释的生物大脑视觉机制来研发算法软件和架构系统的研究领域，脑工作原理是突破传统智能计算领域瓶颈局限的关键。当前，人类对大脑神经环路等多尺度机制的研究还远远不够，例如缺乏对想象、意识、决策等高级功能的机制理解。这种局限性在源头上限制了类脑研究的深度、广度和粒度。

表2 不同算法在主流数据集上的典型指标评测结果

Table 2 Results in Terms of Typical Metrics Obtained on Widely Used Datasets

Method	AntiUAV410[1]			Anti-UAV2021[20]		
	AP_{50}	AP_t	AP_s	AP_{50}	AP_t	AP_s
YOLOv6[43]	87.4	64.1	81.0	85.6	53.4	81.2
YOLOv7[44]	84.3	57.6	73.6	82.7	44.3	75.6
YOLOv8[45]	84.3	59.7	76.4	82.8	50.3	76.8
YOLOv9[46]	86.2	61.3	79.8	84.7	50.8	80.0
YOLOv10[47]	84.6	60.6	76.4	83.3	51.4	76.9
YOLOv11[48]	84.5	59.4	78.5	83.0	50.6	77.9
YOLOv12[49]	84.3	59.6	75.9	82.4	47.9	75.8
Deformable DETR[50]	79.2	45.2	67.0	78.0	38.3	69.4
Conditional DETR[51]	81.7	54.0	71.5	79.5	41.7	72.6
Dab DETR[52]	81.8	52.2	69.8	79.5	40.5	69.9
Dino DETR[53]	84.6	60.3	75.8	82.9	49.3	76.5
Dyhead[54]	81.3	41.3	65.0	79.2	33.3	66.7
CFINet[55]	78.3	36.6	66.0	76.1	31.5	65.6
RFLA[56]	82.2	55.5	71.7	80.4	44.6	73.0
SSPNet[57]	82.9	55.9	72.5	81.0	45.6	73.0
RT DETR[58]	86.5	64.3	81.0	84.5	52.9	80.2
Mamba YOLO[59]	84.0	62.8	78.5	82.5	50.1	78.6
MVTOD[37]	**89.4**	**67.4**	**85.3**	**90.0**	**64.5**	**86.3**

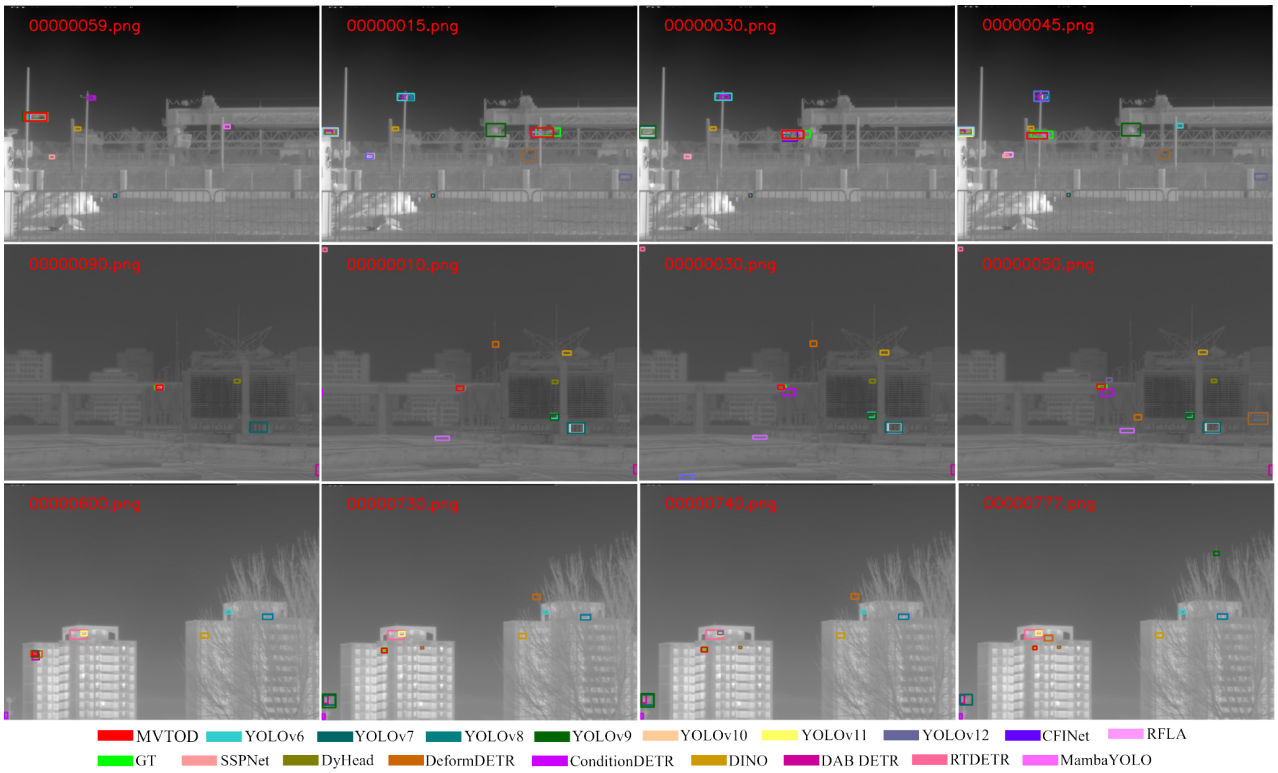

图2　脑启发算法与对比算法在主流数据集上的目标检测结果示例

Fig.2　Examples of Detection Results Obtained by the Brain-inspired Method and the Advanced Computer Vision Methods

(2)算法模型的生物合理性、可解释性和可靠性不足。类脑智能以借鉴脑工作原理为主要技术路线,因此生物合理性、可解释性和可靠性被广泛认为是类脑技术应具备的重要特色。然而,现有的类脑技术受限于模型和硬件的技术瓶颈,与人脑相比在低能耗、高容错、多尺度等方面的生物合理性明显不足。例如,基于脉冲神经网络(Spiking Neural Network,SNN)算法或类脑计算芯片来模拟突触可塑性动态加工机制的技术较为初步和浅显,通常忽略生物神经元的多样性与动态交互特性;很多智能感知技术及系统的内在过程不可追溯、响应不符常识、应用不合预期;类脑算法模型在优化阶段,仍然存在梯度消失、过拟合、灾难性遗忘、长尾数据、对抗攻击等诸多问题。

(3)难以满足多样化任务需求和多类型场景。人脑能够并行处理多个高级任务,能够根据实际需要进行任务切换并快速适应环境,但目前的脑启发视觉感知算法更多的是面向单一或有限任务。例如,当成像载体剧烈运动时,其视觉特征提取和表征会严重降质,导致模型算法性能大幅下降;主流视觉感知系统主要是被动响应,缺乏自主监控、目标重规划、主动探测等高级功能;传统算法缺乏脑的动态记忆与持续学习能力,难以在“记什么”和“何时记”间智能抉择,在应对目标消失重现、长期跟踪等方面性能有限。

(4)应用生态及产业化需要进一步发展。我国类脑技术产业的发展现状仍存在基础能力偏弱、要素协同不畅、治理体系滞后、标准规范缺位等系统性短板,需树立系统思维和生态理念,以生态视角优化升级类脑计算发展的组织框架与推进逻辑。例如,未来对标脑参数容量开发多模态目标感知类脑大模型,在部署到算力、功耗受限的边缘端智能体时存在一定困难;此外,类脑脉冲相机、神经形态芯片、存算一体传感器等新型硬件尚未成熟化、规模化,与传统系统架构的融合存在障碍。

3　总结与展望

通过对脑启发“低慢小”目标视觉感知系统的梳理,可以发现现有算法虽然已初步实现“低慢小”目标的自动化检测,但在复杂场景下同时满足低虚警率、低漏检率和强鲁棒性的需求仍面临重大挑战。针对基于脑启发的“低慢小”深度学习视觉感知系统存在的瓶颈性问题,未来工作在以下几个方面值得深入研究。

(1)数据层面,现有数据集在场景多样性、目标类型/尺寸多样性和标注质量等方面与实际需求还有较大差距,业内应着力构建更具代表性的大规模基准数据集,同时应探索自监督学习、弱监督学习、小样本学习、持续学习、脑机智能融合等新型范式,以缓解数据稀缺、数据分布未知等问题。

(2)目标定位层面,复杂场景下的“低慢小”目标静态特征可挖掘能力有限,即使人眼在凝视状态下也很难发现定位单帧图像中的低信噪比弱隐目标。应借鉴人类视觉系统的多时空尺度时空特征解耦机理,提出更有效的目标运动信息表征方法,从而在更多信息维度层面突破“低慢小”目标发现识别能力的瓶颈。

(3)目标辨识层面,低空自然鸟类、风筝、气球等动/静态干扰物与“低慢小”无人机目标在视觉特征上具有高度相似性。借鉴人脑腹背侧通路交互融合机制,有效聚合目标静态表观特征和运动特征,综合利用时空上下文信息进行无人机目标判别,或将是突破现有模型检测性能的关键所在。

(4)目标认知层面,“低慢小”目标感知是目标预警发现和监视管控的起始阶段,后续还涉及敌我识别、目标处置等任务。借鉴人脑高级脑区目标价值判断机理,在认知层面进一步实现目标属性刻画、意图判别等,对于发展自主式无人作业系统、提高相关装备的智能化水平具有重要意义。

参 考 文 献

[1] Huang B,Li JN,Chen JJ,et al. Anti-UAV410:A thermal infrared benchmark and customized scheme for tracking drones in the wild. IEEE Transactions on Pattern Analysis and Machine Intelligence,2024,46(5):2852—2865.

[2] 杨欣,王刚,李椋,等. 基于深度卷积神经网络的小型民用无人机检测研究进展. 红外技术,2022,44(11):1119—1131.
Yang X,Wang G,Li L,et al. Civil drone detection based on deep convolutional neural networks:A survey. Infrared Technology,2022,44(11):1119—1131. (in Chinese)

[3] Zhao MJ,Li W,Li L,et al. Single-frame infrared small-target detection:A survey. IEEE Geoscience and Remote Sensing Magazine,2022,10(2):87—119.

[4] 陈唯实,牛红闯,王鑫,等. 机场净空区飞鸟与无人机多源探测技术综述. 航空学报,2025,46(10):233—259.
Chen WS,Niu HC,Wang X,et al. Review on multi-source detection technologies for birds and drones in airport clearance area. Acta Aeronautica et Astronautica Sinica,2025,46(10):233—259. (in Chinese)

[5] Liu T,Yin Q,Yang JG,et al. Combining deep denoiser and low-rank priors for infrared small target detection. Pattern Recognition,2023,135:109184.

[6] Ma TL,Wang H,Liang J,et al. MSMA-net:An infrared small target detection network by multiscale super-resolution enhancement and multilevel attention fusion. IEEE Transactions on Geoscience and Remote Sensing,2024,62:5602620.

[7] 李文博,王琦,高尚. 基于深度学习的红外小目标检测算法综述. 激光与红外,2023,53(10):1476—1484.
Li WB,Wang Q,Gao S. A review of infrared small target detection algorithms based on deep learning. Laser & Infrared,2023,53(10):1476—1484. (in Chinese)

[8] 朱逸轩,陈慈发,蔡腾. 结合交叉注意力的红外小目标检测. 红外技术,2025,47(12):1559—1566,1575.
Zhu YX,Chen CF,Cai T. Infrared small object detection with cross-attention. Infrared Technology,2025,47(12):1559—1566,1575. (in Chinese)

[9] 张毅,张焱,张宇,等. 基于多级特征增强融合的红外飞机目标检测方法. 航空学报,2023,44(22):48—62.
Zhang Y,Zhang Y,Zhang Y,et al. Infrared aircraft target detection method based on multi-level feature enhancement fusion. Acta Aeronautica et Astronautica Sinica,2023,44(22):48—62. (in Chinese)

[10] 谷雨,张宏宇,孙仕成. 融合多尺度分形注意力的红外小目标检测模型. 电子与信息学报,2023,45(8):3002—3011.
Gu Y,Zhang HY,Sun SC. Infrared small target detection model with multi-scale fractal attention. Journal of Electronics & Information Technology,2023,45(8):3002—3011. (in Chinese)

[11] 薛雅丽,俞潼安,崔闪,等. 基于级联嵌套U-Net的红外小目标检测. 吉林大学学报(工学版),2025,55(5):1714—1721.
Xue YL,Yu TA,Cui S,et al. Infrared small target detection based on cascaded nested U-Net. Journal of Jilin University (Engineering and Technology Edition),2025,55(5):1714—1721. (in Chinese)

[12] Zhu PF,Wen LY,Du DW,et al. Detection and tracking meet drones challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence,2022,44(11):7380—7399.

[13] Cao B,Yao HY,Zhu PF,et al. Visible andClear:Finding tiny objects inDifference map. Computer Vision – ECCV 2024. Cham:Springer,2025:1—18.

[14] Zhang J,Jiang Z,Zhang J,et al. Tracking tiny drones against clutter:Large-scale infrared benchmark with motion-centric adaptive algorithm//Proceedings of the IEEE International Conference on Computer Vision. 2025.

[15] Kou Y,Gao J,Li B,et al. ZoomTrack:Target-aware non-uniform resizing for efficient visual tracking. Advances in Neural Information Processing Systems,2023,36:50959—50977.

[16] 寿天德. 视觉信息处理的脑机制. 2版. 合肥:中国科学技术大学出版社,2010.
Shou TD. The brain mechanism of visual information processing. 2nd edition. Hefei:China Science and Technology University Press,2010. (in Chinese)

[17] Miura SK,Scanziani M. Distinguishing externally from saccade-induced motion in visual cortex. Nature,2022,610(7930):135—142.

[18] Chariker L,Shapley R,Hawken M,et al. A theory of direction selectivity for macaque primary visual cortex. Proceedings of the National Academy of Sciences of the United States of America,2021,118(32):e2105062118.

[19] Hiser J,Koenigs M. The multifaceted role of the ventromedial prefrontal cortex in emotion,decision making,social cognition,and psychopathology. Biological Psychiatry,2018,83(8):638—647.

[20] Zhao J,Zhang J,Li D,et al. Vision-based anti-uav detection and tracking. IEEE Transactions on Intelligent Transportation Systems,2022,23(12):25323—25334.

[21] Tong X,Zuo Z,Su S,et al. ST-trans:Spatial-temporal transformer for infrared small target detection in sequential images. IEEE Transactions

on Geoscience and Remote Sensing，2024，62：1—19.

[22] Coluccia A，Fascista A，Schumann A，et al. Drone-vs-bird detection challenge at IEEE AVSS2021//2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE，2021：1—8.

[23] Li J，Ye D H，Chung T，et al. Multi-target detection and tracking from a single camera in Unmanned Aerial Vehicles (UAVs)//2016 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE，2016：4992—4997.

[24] Guo H，Zheng Y，Zhang Y，et al. Global-local MAV detection under challenging conditions based on appearance and motion. IEEE Transactions on Intelligent Transportation Systems，2024，25(9)：12005—12017.

[25] Boyun VP，Voznenko LO，Malkush IF. Principles of organization of the human eye retina and their use in computer vision systems. Cybernetics and Systems Analysis，2019，55(5)：701—713.

[26] Charbel Issa P，Troeger E，Finger R，et al. Structure-function correlation of the human central retina. PLoS One，2010，5(9)：e12864.

[27] Ptito M，Bleau M，Bouskila J. The retina：A window into the brain. Cells，2021，10(12)：3269.

[28] Wässle H. Parallel processing in the mammalian retina. Nature Reviews Neuroscience，2004，5(10)：747—757.

[29] Sun Y，Cao B，Zhu P，et al. Dynamic brightness adaptation for robust multi-modal image fusion//Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. 2024：1317—1325.

[30] Philip Chen CL，Li H，Wei YT，et al. A local contrast method for small infrared target detection. IEEE Transactions on Geoscience and Remote Sensing，2014，52(1)：574—581.

[31] Ye LH，Liu J，Zhang JT，et al. A novel size-aware local contrast measure for tiny infrared target detection. IEEE Geoscience and Remote Sensing Letters，2025，22：8001005.

[32] Shi ZC，Hu J，Ren J，et al. HS-FPN：High frequency and spatial perception FPN for tiny object detection. Proceedings of the AAAI Conference on Artificial Intelligence，2025，39(7)：6896—6904.

[33] Sun YM，Cao B，Zhu PF，et al. DetFusion：A detection-driven infrared and visible image fusion network//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa，Portugal. ACM，2022：4003—4011.

[34] Huang SH，Lin C，Jiang XT，et al. BRSTD：Bio-inspired remote sensing tiny object detection. IEEE Transactions on Geoscience and Remote Sensing，2024，62：5643115.

[35] Li HQ，Yang JF，Wang RS，et al. ILNet：Low-level matters for salient infrared small target detection. IEEE Transactions on Aerospace and Electronic Systems，2025，61(4)：8306—8318.

[36] Wang KW，Du SY，Liu CX，et al. Interior attention-aware network for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing，2022，60：5002013.

[37] Wang G，Yang X，Li L，et al. Tiny drone object detection in videos guided by the bio-inspired magnocellular computation model. Applied Soft Computing，2024，163：111892.

[38] Yan PT，Hou RZ，Duan XG，et al. STDMANet：Spatio-temporal differential multiscale attention network for small moving infrared target detection. IEEE Transactions on Geoscience and Remote Sensing，2023，61：5602516.

[39] Ji XY，Wang G，Wang YZ. Optical flow estimation for tiny objects：New problem，specialized benchmark，and bioinspired scheme//Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence. Montreal，Canada. International Joint Conferences on Artificial Intelligence Organization，2025：1215—1223.

[40] Chen N，Xiao C，Dai Y，et al. Event-based tiny object detection：A benchmark dataset and baseline//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2025：7209—7218.

[41] Liu S，Li GT，Zhan YF，et al. MUSAK：A multi-scale space kinematic method for drone detection. Remote Sensing，2022，14(6)：1434.

[42] Wang YF，Zhao J，Fan ZX，et al. JTD-UAV：MLLM-enhanced joint tracking and description framework for anti-UAV systems//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville，TN，USA. IEEE，2025：1633—1644.

[43] Chuyi Li，Lulu Li，Hongliang Jiang，et al. “YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications，”Published on arXiv.org, September 7，2022. [Online]. Available：https：//arxiv.org/abs/2209.02976. [Accessed：March 5，2026].

[44] Wang CY，Bochkovskiy A，Liao HYM. YOLOv7：Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023：7464—7475.

[45] Jocher G，Chaurasia A，Qiu J. YOLO by Ultralytics. (2023-01-01) [2026-02-03]. https：//github.com/ultralytics/ultralytics.

[46] Wang CY，Yeh IH，Mark LHY. Yolov9：Learning what you want to learn using programmable gradient information//European Conference on Computer Vision. Cham：Springer Nature Switzerland，2024：1—21.

[47] Wang A，Chen H，Liu L，et al. Yolov10：Real-time end-to-end object detection. Advances in Neural Information Processing Systems，2024，37：107984—108011.

[48] He L，Zhou Y，Liu L，et al. Research on object detection and recognition in remote sensing images based on YOLOv11. Scientific Reports，2025，15(1)：14032.

[49] Tian Y，Ye Q，Doermann D. Yolov12：Attention-centric real-time object detectors//Proceedings of the Thirty-ninth Annual Conference on Neural Information Processing Systems. 2025.

[50] Zhu X，Su W，Lu L，et al. Deformable DETR：Deformable transformers for end-to-end object detection//Proceedings of the International Conference on Learning Representations.2021.

[51] Meng D，Chen X，Fan Z，et al. Conditional detr for fast training convergence//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021：3651—3660.

[52] Liu S，Li F，Zhang H，et al. DAB-DETR：Dynamic anchor boxes are better queries for DETR//Proceedings of the International Conference on Learning Representations. 2022.

[53] Zhang H，Li F，Liu S，et al. DINO：DETR with improved denoising anchor boxes for end-to-end object detection//Proceedings of the International Conference on Learning Representations. 2023.

[54] Dai X，Chen Y，Xiao B，et al. Dynamic head：Unifying object detection heads with attentions//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021：7373—7382.

[55] Yuan X，Cheng G，Yan K，et al. Small object detection via coarse-to-fine proposal generation and imitation learning//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023：6317

—6327.

[56] Xu C,Wang J,Yang W,et al. RFLA:Gaussian receptive field based label assignment for tiny object detection//European Conference on Computer Vision. Cham:Springer Nature Switzerland,2022:526—543.

[57] Hong M,Li S,Yang Y,et al. SSPNet:Scale selection pyramid network for tiny person detection from UAV images. IEEE Geoscience and Remote Sensing Letters,2021,19:1—5.

[58] Zhao Y,Lv W,Xu S,et al. Detrs beat yolos on real-time object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024:16965—16974.

[59] Wang Z,Li C,Xu H,et al. Mamba yolo:A simple baseline for object detection with state space model//Proceedings of the AAAI Conference on Artificial Intelligence. 2025,39(8):8205—8213.

Brain-inspired Visual Perception of Low-altitude Slow and Small Targets: Current Research and Future Perspectives

Gang Wang[1†] Xin Yang[1, 2†] Jin Gao[3*] Liang Li[1] Pengfei Zhu[4] Weiming Hu[3] Yizheng Wang[1]

1. *Institute of Military Cognition and Brain Sciences, Academy of Military Sciences, Beijing 100850, China*
2. *School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 611731, China*
3. *Institute of Automation Chinese Academy of Sciences, Beijing 100190, China*
4. *College of Intelligence and Computing, Tianjin University, Tianjin 300350, China*

Abstract Low-altitude Slow and Small (LSS) targets pose significant detection challenges for conventional radar systems due to their low flight altitude,slow velocity,small radar cross-section,and ability to hide within complex backgrounds (e.g.,urban structures,mountainous terrain,or forests). Photoelectric sensors (including infrared and visible spectrum) coupled with intelligent visual perception technologies,particularly deep learning,offer promising solutions to these limitations. However,the diversity of LSS target signatures,the complexity of detection environments,and the openness of application scenarios collectively constrain the transferability and generalization capability of trained deep learning models. Current research focuses on drawing inspiration from the Human Visual System (HVS),specifically its capacity for integrated appearance-motion analysis during visual information acquisition,processing,and interpretation. This biomimetic approach has emerged as a key direction for advancing LSS target perception. This paper systematically reviews the state-of-the-art technological developments while providing critical insights into future research in this field.

Keywords brain-inspired visual perception;Human Visual System;low-altitude slow and small target;deep learning;appearance-motion modeling

高 晋 中国科学院自动化研究所研究员,博士研究生导师。主持国家自然科学基金青年科学基金项目(B类)、北京市杰出青年科学基金项目、国家自然科学基金联合基金项目等。目前研究方向为视觉目标自主 感知与理解。

王 刚 军事科学院副研究员,现任中国神经科学学会类脑智能分会副秘书长、中国人工智能学会青工委委员等,国家重大科技专项青年科学家,入选国防领域高层次青年人才、北京市科技新星和北脑青年学者等;在*Nature Communications*、*IEEE Transactions on Pattern Analysis and Machine Intelligence*等发表论文40余篇;获*IEEE Conference on Computer Vision and Pattern Recognition*反无人机挑战赛冠军/最佳学生论文提名奖5项;研究方向为类脑/脑机智能。

杨 欣 电子科技大学生物医学工程专业博士研究生,参与承担"脑科学与类脑研究"国家科技重大专项、国家自然科学基金青年科学基金(C类)等课题5项;曾获*IEEE Conference on Computer Vision and Pattern Recognition*反无人机挑战赛冠军2项。研究方向包括深度学习与脑启发目标检测。

(责任编辑 王 磊 张 强)

† Contributed equally as co-first authors.
* Corresponding Author,Email:jin.gao@nlpr.ia.ac.cn