SEARCH WORD: TRACKING

[['TSBA: A two-stage poison-only backdoor attack on visual object tracking',
'https://www.sciencedirect.com/science/article/pii/S0031320325008830',
'',
'Yilang Zhang, Yanjun Pu, Jingzheng Li, Shuxin Zhao, Bo Lang',
'2',
'2026',
'3'],
['Light field collaborative perception for visual object tracking',
'https://www.sciencedirect.com/science/article/pii/S003132032500857X',
'',
'Mianzhao Wang, Fan Shi, Xu Cheng, Meng Zhao',
'2',
'2026',
'3'],
['BlackboxBench: A Comprehensive Benchmark of Black-Box Adversarial Attacks',
'http://ieeexplore.ieee.org/document/11016822',
'Adversarial examples are well-known tools to evaluate the vulnerability of deep neural networks (DNNs). Although lots of adversarial attack algorithms have been developed, it's still challenging in the practical scenario that the model's parameters and architectures are inaccessible to the attacker/evaluator, i.e., black-box adversarial attacks. Due to the practical importance, there has been rapid progress from recent algorithms, reflected by the quick increase in attack success rate and quick decrease in query numbers to the target model. However, there lacks thorough evaluations and comparisons among these algorithms, causing difficulties in tracking the real progress, analyzing advantages and disadvantages of different technical routes, as well as designing future development roadmap of this field. Thus, we aim at building a comprehensive benchmark of black-box adversarial attacks, called BlackboxBench. It mainly provides: 1) a unified, extensible and modular-based codebase, implementing 29 query-based attack algorithms and 30 transfer-based attack algorithms; 2) comprehensive evaluations: we evaluate the implemented algorithms against several mainstreaming model architectures on 2 widely used datasets (CIFAR-10 and a subset of ImageNet), leading to 14,950 evaluations^{1}1 in total; 3) thorough analysis and new insights, as well analytical tools.',
'Meixi Zheng, Xuanchen Yan, Zihao Zhu, Hongrui Chen, Baoyuan Wu',
'3',
'2025',
'8'],
['A Learning Paradigm for Selecting Few Discriminative Stimuli in Eye-Tracking Research',
'http://ieeexplore.ieee.org/document/11015567',
'Eye-tracking is a reliable method for quantifying visual information processing and holds significant potential for group recognition, such as identifying autism spectrum disorder (ASD). However, eye-tracking research typically faces the heterogeneity of stimuli and is time-consuming due to the large number of observed stimuli. To address these issues, we first mathematically define the stimulus selection problem and introduce the concept of stimulus discrimination ability to reduce the computational complexity of the solution. Then, we construct a scanpath-based recognition model to mine the stimulus discrimination ability. Specifically, we propose cross-subject entropy and cross-subject divergence scores for quantitatively evaluating stimulus discrimination ability, effectively capturing differences in intra-group collective trends and inter-subject consistency within a group. Furthermore, we propose an iterative learning mechanism that employs stimulus-wise attention to focus on discriminative stimuli for discrimination purification. In the experiment, we construct an ASD eye-tracking dataset with diverse stimulus types and conduct extensive tests on three representative models to validate our approach. Remarkably, our method demonstrates superior performance using only 10 selected stimuli compared to models utilizing 220 stimuli. Additionally, we perform experiments on another eye-tracking task, gender prediction, to further validate our method. We believe that our approach is both simple and flexible for integration into existing models, promoting large-scale ASD screening and extending to other eye-tracking research domains.',
'Wenqi Zhong, Chen Xia, Linzhi Yu, Kuan Li, Zhongyu Li, Dingwen Zhang, Junwei Han',
'3',
'2025',
'8'],
['Revisiting Siamese-Based 3D Single Object Tracking With a Versatile Transformer',
'http://ieeexplore.ieee.org/document/11045222',
'3D Single Object Tracking (SOT) plays an important role in real-world visual applications such as autonomous driving and planning. How to realize effective 3D SOT is still a valuable challenge due to its

carrier-sparse point clouds and its role-complex influencing factors. Inspired by the remote modeling of popular transformers, we further propose a Versatile Point Tracking Transformer (VPTT) method for 3D SOT, with object guidance from the template point cloud to the search area point cloud under the siamese-based tracking paradigm. Specifically, VPTT employs self- and cross- attention mechanisms and extends four matching operations, resulting in leveraging the contextual information of consecutive frames to improve the tracking results. By constructing a deep network VerFormer consisting of four successive transformer layers, which performs matching operations involving fusional transformation, separative discrimination, intersectional interaction, and unidirectional propagation from shallow to deep. Considering that the tracking task involves multiple processes, VPTT further learns how to forecast intermediate outputs including mask probability, trailing distance, and heading angle at each stage. Such a specialized design allows our VPTT to revisit the end-to-end training paradigm used for 3D tracking while developing a versatile transformer that is a perfect fit for the 3D SOT task. Experiments on three benchmarks, KITTI, nuScenes, and Waymo, show that VPTT achieves state-of-the-art tracking performance on siamese-based tracking running at $\\sim$62 FPS.',
  'Jiaming Liu, Yue Wu, Qiguang Miao, Maoguo Gong, Linghe Kong',
  '3',
  '2025',
  '8'],
 ['AFTER: Attention-Based Fusion Router for RGBT Tracking',
  'http://ieeexplore.ieee.org/document/11079869',
  'Multi-modal feature fusion as a core investigative component of RGBT tracking emerges numerous fusion studies in recent years. However, existing RGBT tracking methods widely adopt fixed fusion structures to integrate multi-modal feature, which are hard to handle various challenges in dynamic scenarios. To address this problem, this work presents a novel Attention-based Fusion router called AFTER, which optimizes the fusion structure to adapt to the dynamic challenging scenarios, for robust RGBT tracking. In particular, we design a fusion structure space based on the hierarchical attention network, each attention-based fusion unit corresponding to a fusion operation and a combination of these attention units corresponding to a fusion structure. Through optimizing the combination of attention-based fusion units, we can dynamically select the fusion structure to adapt to various challenging scenarios. Unlike complex search of different structures in neural architecture search algorithms, we develop a dynamic routing algorithm, which equips each attention-based fusion unit with a router, to predict the combination weights for efficient optimization of the fusion structure. Extensive experiments on five mainstream RGBT tracking datasets demonstrate the superior performance of the proposed AFTER against state-of-the-art RGBT trackers. We release the code in https://github.com/Alexadlu/AFter',
  'Andong Lu, Wanyu Wang, Chenglong Li, Jin Tang, Bin Luo',
  '4',
  '2025',
  '8'],
 ['Associate Everything Detected: Facilitating Tracking-by-Detection to the Unknown',
  'http://ieeexplore.ieee.org/document/11105000',
  'Multi-object tracking (MOT) emerges as a pivotal and highly promising branch in the field of computer vision. Classical closed-vocabulary MOT (CV-MOT) methods aim to track objects of predefined categories. Recently, some open-vocabulary MOT (OV-MOT) methods have successfully addressed the problem of tracking unknown categories. However, we found that the CV-MOT and OV-MOT methods each struggle to excel in the tasks of the other. In this paper, we present a unified framework, Associate Everything Detected (AED), that simultaneously tackles CV-MOT and OV-MOT by integrating with any off-the-shelf detector and supports unknown categories. Different from existing tracking-by-detection MOT methods, AED gets rid of prior knowledge (e.g., motion cues) and relies solely on highly robust feature learning to handle complex trajectories in OV-MOT tasks while keeping excellent performance in CV-MOT tasks. Specifically, we model the association task as a similarity decoding problem and propose a sim-decoder with an association-centric learning mechanism. The sim-decoder calculates similarities in three aspects: spatial, temporal, and cross-clip. Subsequently, association-centric learning leverages these threefold similarities to ensure that the extracted features are appropriate for continuous tracking and robust enough to generalize to unknown categories. Compared with existing powerful OV-MOT and CV-MOT methods, AED achieves superior performance on TAO, SportsMOT, and DanceTrack without any prior knowledge. Our code is available at https://github.com/balabooooo/AED',
  'Zimeng Fang, Chao Liang, Xue Zhou, Shuyuan Zhu, Xi Li',
  '4',
  '2025',
  '8'],
 ['Diffusion Model-Based Path Follower for a Salamander-Like Robot',
  'http://ieeexplore.ieee.org/document/10934035',
  'Salamander-like robots, renowned for their versatile locomotion, present unique challenges in the

development of effective path-following controllers due to their distinctive movement patterns and complex body structures. Conventional path-following controllers, while effective for various bionic robots, struggle with the intricate modeling for salamander-like robots and often require laborious manual tuning. Conversely, learning-based methods offer promising alternatives but face issues such as reliance on environmental interactions, short-sighted prediction, and irrational design of state space and reward function. To overcome these limitations, this article proposes a diffusion model-based hierarchical control framework that treats path tracking as a sequence generation problem. The diffusion model's capability to model joint distributions of state, action, and reward sequences enables it to outperform other learning-based approaches in efficient data utilization, stable training, and long-horizon dependency modeling. Our framework integrates a high-level policy driven by guided diffusion with a low-level controller for parsing commands into executable movements via inverse kinematics, reducing the action space and improving learning efficiency. In addition, we design a more reasonable state space and reward function tailored to the path-following task, addressing shortcomings in prior learning-based controllers. Furthermore, we optimize the diffusion model (DM) by developing lightweight network architectures and incorporating advanced attention mechanisms, to ensure its practical deployment on physical robots with limited computational resources, without compromising performance. Extensive simulations and real-world experiments demonstrate the framework's effectiveness, efficiency, and robustness in diverse path-following tasks for salamander-like robots, marking a significant advancement in the control of biomimetic robots.',
  'Zhiang Liu, Yang Liu, Yongchun Fang',
  '5',
  '2025',
  '8'],
 ['Neuroadaptive Control With Enhanced Stability and Reliability',
  'http://ieeexplore.ieee.org/document/10931795',
  'The performance of neural network (NN)-driven control systems hinges on the reliability and functionality of the NN unit in the controller. Maintaining the compact set condition for NN training signals (inputs) during operation is crucial for preserving the NN's universal learning and approximation capabilities, yet this requirement is often overlooked in existing studies. This article introduces a constraint transformation-based design method that ensures excitation signals always originate from a fixed region, regardless of initial conditions. By meeting the compactness condition required by the universal approximation theorem, this approach safeguards the functionality of the NN-driven control unit. Additionally, a decaying damping rate is employed to enable the tracking error to asymptotically converge to zero, rather than being ultimately uniformly bounded (UUB). To further ensure robust operation even if the NN underperforms due to an insufficient number of neurons or violation of the compact set condition, a new control strategy is developed based on the worst case behavior of NNs. This "fail-secure" mechanism significantly enhances the reliability of the NN-based control scheme. The effectiveness and benefits of the proposed method are confirmed through numerical simulations, demonstrating its potential to substantially improve the robustness and performance of NN-driven control systems.',
  'Kaili Xiang, Ruotong Ming, Siyu Chen, Frank L. Lewis',
  '5',
  '2025',
  '8'],
 ['Exploring Dynamic Transformer for Efficient Object Tracking',
  'http://ieeexplore.ieee.org/document/10947615',
  'The speed-precision tradeoff is a critical problem in visual object tracking, as it typically requires low latency and is deployed on resource-constrained platforms. Existing solutions for efficient tracking primarily focus on lightweight backbones or modules, which, however, come at a sacrifice in precision. In this article, inspired by dynamic network routing, we propose DyTrack, a dynamic transformer framework for efficient tracking. Real-world tracking scenarios exhibit varying levels of complexity. We argue that a simple network is sufficient for easy video frames, while more computational resources should be assigned to difficult ones. DyTrack automatically learns to configure proper reasoning routes for different inputs, thereby improving the utilization of the available computational budget and achieving higher performance at the same running speed. We formulate instance-specific tracking as a sequential decision problem and incorporate terminating branches to intermediate layers of the model. Furthermore, we propose a feature recycling mechanism to maximize computational efficiency by reusing the outputs of predecessors. Additionally, a target-aware self-distillation strategy is designed to enhance the discriminating capabilities of early-stage predictions by mimicking the representation patterns of the deep model. Extensive experiments demonstrate that DyTrack achieves promising speed-precision tradeoffs with only a single model. For instance, DyTrack obtains 64.9% area under the curve (AUC) on LaSOT with a speed of 256 fps.',
  'Jiawen Zhu, Xin Chen, Haiwen Diao, Shuai Li, Jun-Yan He, Chenyang Li, Bin Luo, Dong Wang, Huchuan Lu',
  '5',
  '2025',

'8'],
['A Review of Multi-Object Tracking in Recent Times',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cvi2.70010?af=R',
'This paper discusses many recent deep-learning MOT methods. Moreover, to highlight their contributions, these methods are categorised into four main groups: detection-based, SOT-based, and segmentation-based methods according to the integrated core technologies.\n\n\n\n\n\nABSTRACT\nMulti-object tracking (MOT) is a fundamental problem in computer vision that involves tracing the trajectories of foreground targets throughout a video sequence while establishing correspondences for identical objects across frames. With the advancement of deep learning techniques, methods based on deep learning have significantly improved accuracy and efficiency in MOT. This paper reviews several recent deep learning-based MOT methods and categorises them into three main groups: detection-based, single-object tracking (SOT)-based, and segmentation-based methods, according to their core technologies. Additionally, this paper discusses the metrics and datasets used for evaluating MOT performance, the challenges faced in the field, and future directions for research.',
'Suya Li, Hengyi Ren, Xin Xie, Ying Cao',
'6',
'2025',
'3'],
['A New Large-Scale Dataset for Marine Vessel Re-Identification Based on Swin Transformer Network in Ocean Surveillance Scenario',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cvi2.70007?af=R',
'A new large-scale marine vessel dataset with well-annotated vessel orientation, vessel colour, and vessel type labels has been collected and created in a real marine environment for vessel Re-ID research. (2)A side information embedding module is introduced through the learnable embedding layer to encode more kinds of information, including marine vessel orientation, type and colour. (3)A deep neural network framework based on Swin Transformer for marine vessel Re-ID task is proposed to learn and extract discriminative features, and achieves SOTA performance on vessel, vehicle and person Re-ID benchmark datasets.\n\n\n\n\n\nABSTRACT\nIn recent years, there has been an upward trend that marine vessels, an important object category in marine monitoring, have gradually become a research focal point in the field of computer vision, such as detection, tracking, and classification. Among them, marine vessel re-identification (Re-ID) emerges as a significant frontier research topics, which not only faces the dual challenge of huge intra-class and small inter-class differences, but also has complex environmental interference in the port monitoring scenarios. To propel advancements in marine vessel Re-ID technology, SwinTransReID, a framework grounded in the Swin Transformer for marine vessel Re-ID, is introduced. Specifically, the project initially encodes the triplet images separately as a sequence of blocks and construct a baseline model leveraging the Swin Transformer, achieving better performance on the Re-ID benchmark dataset in comparison to convolution neural network (CNN)-based approaches. And it introduces side information embedding (SIE) to further enhance the robust feature-learning capabilities of Swin Transformer, thus, integrating non-visual cues (orientation and type of vessel) and other auxiliary information (hull colour) through the insertion of learnable embedding modules. Additionally, the project presents VesselReID-1656, the first annotated large-scale benchmark dataset for vessel Re-ID in real-world ocean surveillance, comprising 135,866 images of 1656 vessels along with 5 orientations, 12 types, and 17 colours. The proposed method achieves 87.1% $\\%$ mAP and 96.1% $\\%$ Rank-1 accuracy on the newly-labelled challenging dataset, which surpasses the state-of-the-art (SOTA) method by 1.9% $\\%$ mAP regarding to performance. Moreover, extensive empirical results demonstrate the superiority of the proposed SwinTransReID on the person Market-1501 dataset, vehicle VeRi-776 dataset, and Boat Re-ID vessel dataset.',
'Zhi Lu, Liguo Sun, Pin Lv, Jiuwu Hao, Bo Tang, Xuanzhen Chen',
'6',
'2025',
'3'],
['Unlocking the power of multi-modal fusion in 3D object tracking',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cvi2.12335?af=R',
'3D Single Object Tracking plays a vital role in autonomous driving and robotics, yet traditional approaches have predominantly focused on using pure LiDAR-based point cloud data, often neglecting the benefits of integrating image modalities. To address this gap, we propose a novel Multi-modal Image-LiDAR Tracker (MILT) designed to overcome the limitations of single-modality methods by effectively combining RGB and point cloud data. Our key contribution is a dual-branch architecture that separately extracts geometric features from LiDAR and texture features from images. These features are then fused in a BEV perspective to achieve a comprehensive representation of the tracked object. A significant innovation in our approach is the Image-to-LiDAR Adapter module, which transfers the rich feature representation capabilities of the image modality to the 3D tracking task, and the BEV-Fusion module, which facilitates the interactive fusion of geometry and texture features. By validating MILT on public datasets, we demonstrate substantial

performance improvements over traditional methods, effectively showcasing the advantages of our multi-modal fusion strategy. This work advances the state-of-the-art in SOT by integrating complementary information from RGB and LiDAR modalities, resulting in enhanced tracking accuracy and robustness.',
  'Yue Hu',
  '6',
  '2025',
  '2'],
 ['Representation alignment contrastive regularisation for multi-object tracking',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cvi2.12331?af=R',
  "Achieving high-performance in multi-object tracking algorithms heavily relies on modelling spatial-temporal relationships during the data association stage. Mainstream approaches encompass rule-based and deep learning-based methods for spatial-temporal relationship modelling. While the former relies on physical motion laws, offering wider applicability but yielding suboptimal results for complex object movements, the latter, though achieving high-performance, lacks interpretability and involves complex module designs. This work aims to simplify deep learning-based spatial-temporal relationship models and introduce interpretability into features for data association. Specifically, a lightweight single-layer transformer encoder is utilised to model spatial-temporal relationships. To make features more interpretative, two contrastive regularisation losses based on representation alignment are proposed, derived from spatial-temporal consistency rules. By applying weighted summation to affinity matrices, the aligned features can seamlessly integrate into the data association stage of the original tracking workflow. Experimental results showcase that our model enhances the majority of existing tracking networks' performance without excessive complexity, with minimal increase in training overhead and nearly negligible computational and storage costs.",
  'Shujie Chen, Zhonglin Liu, Jianfeng Dong, Xun Wang, Di Zhou',
  '6',
  '2025',
  '2'],
 ['Real-Time Underwater Vision Sensing System for AUV Tracking',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70183?af=R',
  'In this article, we developed an underwater real-time visual sensing system for AUV tracking based on the edge embedded-device NVIDIA Jetson Xavier NX. After experimental verification, compared with other algorithms, the UW-Siam tracking algorithm we proposed not only has a higher success rate and accuracy, but also has the best speed performance, and can easily cope with underwater object tracking tasks in different scenarios.\n\n\n\n\n\n\n\nABSTRACT\nWith the continuous development of emerging technologies such as big data and artificial intelligence, the related technologies of perception tasks on which underwater object tracking relies have made great progress. However, a significant barrier still exists in implementing high-performance real-time underwater object tracking on low-power edge devices. To achieve real-time tracking of underwater objects for edge devices, this article develops an underwater real-time visual sensing system applied to AUV tracking. First, an underwater object tracking device is designed in this article employing a stereo binocular camera, an edge embedded NVIDIA Jetson Xavier NX and an STM32 control board. Then, after preprocessing, the input image with the effective USM algorithm, we propose a quick approach for detecting underwater objects based on SIoU-YOLOv8n, which enables automatic object recognition and selection. At the same time, this article proposes a twin network UW-Siam for continuous tracking of underwater objects, which achieves more accurate underwater object tracking. Finally, the algorithm is deployed to the designed real-time underwater vision sensing system and tested in real-world scenarios. The tracking accuracy reached 0.652, and the detection mAP reached 0.97. The results indicate that the system can rapidly detect and continuously monitor objects, performing well in real-world scenarios with high accuracy and\xa0robustness.',
  'Canrong Chen, Tingzhuang Liu, Linglu He, Yi Zhu, Fei Yuan',
  '7',
  '2025',
  '8'],
 ['Spatiotemporal Context Adapting Framework for Visual Object Tracking',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70150?af=R',
  'Due to noise interference, existing methods often partially exploit auxiliary data, leading to underutilisation of spatiotemporal information. To address these issues, we propose a novel and concise tracking framework, uniformly encoding all auxiliary data and mitigating the unstable factors.\n\n\n\n\n\n\n\nABSTRACT\nVisual object tracking is widely applied in intelligent transportation systems and visual surveillance systems that serve smart cities, as well as in autonomous vehicles. Existing methods usually utilise a relation-modelling framework to model the visual object tracking problem, with auxiliary spatial context and temporal information. The spatial context is often extracted by enlarging the target template, which can introduce more background and positional information. The temporal correlation is obtained by associating the search image with previous images. However, due to noise interference, existing methods often partially exploit auxiliary data, leading to underutilisation of spatiotemporal information. To address these issues, we propose a novel

and concise tracking framework, uniformly encoding all auxiliary data, including the enlarged target template, previous images, and corresponding target bounding boxes. Specifically, to mitigate the unstable factors introduced by these raw inputs, we propose a spatiotemporal context adaptive encoder, which can adaptively select appropriate information in noisy data. Extensive experiments show that the proposed method achieves state-of-the-art performance on various benchmarks, demonstrating its\xa0superiority.',
  'Kunlong Zhao, Dawei Zhao, Xu Wang, Liang Xiao, Yulong Huang, Yiming Nie, Yonggang Zhang, Bin Dai',
  '7',
  '2025',
  '7'],
 ['A Two-Stage Homography Matrix Prediction Approach for Trajectory Generation in Multi-Object Tracking on Sports Fields',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70136?af=R',
  'To overcome the limitations of existing homography matrix prediction methods, we developed a two-stage prediction approach by optimizing the pixel position feature extraction structure. In the first stage, 4-key-point coordinate deviation is leveraged to indirectly estimate the macroscopic values of the homography matrix, followed by direct fine-tuning of the microscopic 8-DOF numerical components in the second stage. We validated the effectiveness of this approach in three competitive sports scenarios: basketball, ice hockey, and handball.\n\n\n\n\n\n\n\nABSTRACT\nHomography estimation is a fundamental topic in computer vision, especially in scenarios that require perspective changes for intelligent analysis of sports fields, where it plays a crucial role. Existing methods predict the homography matrix either indirectly by evaluating the 4-key-point coordinate deviation in paired images with the same visual content or directly by fine-tuning the 8 degrees of freedom numerical values that define the matrix. However, these approaches often fail to effectively incorporate coordinate positional information and overlook optimal application scenarios, leading to significant accuracy bottlenecks, particularly for paired images with differing visual content. To address these issues, we propose an approach that integrates both methods in a staged manner, leveraging their respective advantages. In the first stage, positional information is embedded to enhance convolutional computations, replacing serial concatenation in traditional feature fusion with parallel concatenation, while using 4-key-point coordinate deviation to predict the macroscopic homography matrix. In the second stage, positional information is further integrated into the input images to refine the direct 8 degrees of freedom numerical predictions, improving matrix fine-tuning accuracy. Comparative experiments with state-of-the-art methods demonstrate that our approach achieves superior performance, yielding a root mean square error as low as 1.25 and an average corner errror as low as 14.1 in homography transformation of competitive sports image\xa0pairs.',
  'Pan Zhang, Jiangtao Luo, Guoliang Xu, Xupeng Liang',
  '7',
  '2025',
  '6'],
 ['Vorticity Transport Equation-Based Shadow Removal Approach for Image Inpainting',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70114?af=R',
  '(1) A new image inpainting algorithm using the VTE is developed; (2) FMM is introduced as an alternative to the iterative numerical solution of the VTE to improve the speed of image in painting; and (3) the VTE equation is transformed into a weighted-average equation to be used in conjunction with FMM.\n\n\n\n\n\n\n\nABSTRACT\nShadows are common in many types of images, causing information loss or disturbance. Shadow removal can help improve the quality of the digital image. If there is no effective information available to restore the original image in the shaded area, the interpolation-based inpainting technique can be used to remove the shadow from the digital image. This image inpainting technique typically involves establishing and solving partial differential equations (PDEs), an iterative solving process that is very time-consuming. To solve the time-consuming problem, a method that introduces the fast marching method (FMM) into the vorticity transport equation (VTE) is demonstrated. VTE is a type of partial differential equation describing two-dimensional fluids. FMM is a numerical scheme for tracking the evolution of monotonically advancing interfaces via finite difference solution of the eikonal equation. The proposed method contains three main steps: (a) by investigating the relationship between VTE and the traditional PDE-based image inpainting method, a new image inpainting model using VTE is developed;(b) the area to be inpainted is divided into boundaries that shrink in layers from the outside inwards using FMM; and (c) the VTE image inpainting model is converted into a weighted average form to coordinate with FMM. The visual and quantitative evaluation of the experimental results of shadow removal shows that the proposed method outperforms PDE-based and state-of-the-art methods in terms of shadow-removal effect and running time. The results also show that our method excels at inpainting images with near-smooth textures and simple geometric structures and where the pixels to be inpainted are continuous with neighbouring pixels.',
  'Xiaoying Ti, Li Yu, Quanhua Zhao',
  '7',
  '2025',
  '5'],

['Structure-Aware Transformer for Shadow Detection',
 'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70031?af=R',
 'We propose a structure-aware transformer network for robust shadow detection. An edge-guided multi-task learning framework is designed to predict shadow maps with rich structures. An auxiliary semantic-aware learning is introduced to overcome the interference from complex\xa0scenes.\n\n\n\n\n\n\n\nABSTRACT\nShadow detection helps reduce ambiguity in object detection and tracking. However, existing shadow detection methods tend to misidentify complex shadows and their similar patterns, such as soft shadow regions and shadow-like regions, since they treat all cases equally, leading to an incomplete structure of the detected shadow regions. To alleviate this issue, we propose a structure-aware transformer network (STNet) for robust shadow detection. Specifically, we first develop a transformer-based shadow detection network to learn significant contextual information interactions. To this end, a context-aware enhancement (CaE) block is also introduced into the backbone to expand the receptive field, thus enhancing semantic interaction. Then, we design an edge-guided multi-task learning framework to produce intermediate and main predictions with a rich structure. By fusing these two complementary predictions, we can obtain an edge-preserving refined shadow map. Finally, we introduce an auxiliary semantic-aware learning to overcome the interference from complex scenes, which facilitates the model to perceive shadow and non-shadow regions using a semantic affinity loss. By doing these, we can predict high-quality shadow maps in different scenarios. Experimental results demonstrate that our method reduces the balance error rate (BER) by 4.53%, 2.54%, and 3.49% compared to state-of-the-art (SOTA) methods on the benchmark datasets SBU, ISTD, and UCF,\xa0respectively.',
 'Wanlu Sun, Liyun Xiang, Wei Zhao',
 '7',
 '2025',
 '3'],
['An intelligent retrievable object-tracking system with real-time edge inference capability',
 'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13297?af=R',
 "An intelligent retrievable object-tracking system assists users in quickly and accurately locating lost objects. However, challenges such as real-time processing on edge devices, low image resolution, and small-object detection significantly impact the accuracy and efficiency of video-stream-based systems, especially in indoor home environments. To overcome these limitations, a novel real-time intelligent retrievable object-tracking system is designed. The system incorporates a retrievable object-tracking algorithm that combines DeepSORT and sliding window techniques to enhance tracking capabilities. Additionally, the YOLOv7-small-scale model is proposed for small-object detection, integrating a specialized detection layer and the convolutional batch normalization LeakyReLU spatial-depth convolution module to enhance feature capture for small objects. TensorRT and INT8 quantization are used for inference acceleration on edge devices, doubling the frames per second. Experiments on a Jetson Nano (4 GB) using YOLOv7-small-scale show an 8.9% improvement in recognition accuracy over YOLOv7-tiny in video stream processing. This advancement significantly boosts the system's performance in efficiently and accurately locating lost objects in indoor home\xa0settings.",
 'Yujie Li, Yifu Wang, Zihang Ma, Xinghe Wang, Benying Tan, Shuxue Ding',
 '7',
 '2024',
 '12'],
['Research on Extended Target Tracking Algorithm Based on RHM and RMM Fusion',
 'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70371?af=R',
 'This paper addresses the extended target tracking problem with irregular shapes. A fusion algorithm combining random hypersurface model (RHM) and random matrix model (RMM) for extended target tracking is proposed. By leveraging the fast convergence and strong robustness of RMM, the tracking efficiency of RHM is improved. Through simulation experiments, the effectiveness of the algorithm proposed in this paper is verified. The extended target fusion method proposed in this paper can provide a more accurate means for precise tracking and state prediction in practical engineering.\n\n\n\n\n\n\n\nABSTRACT\nAiming at the extended target tracking problem with irregular shapes, an extended target tracking algorithm that fuses random hypersurface model (RHM) and random matrix model (RMM) is proposed in the case where the priori shape of the target is unknown. First, the algorithm adopts the RMM method to track the extended target and then takes the reference points on the elliptical shape contour and constructs an ordered sequence. Second, the discrete Fourier transform (DFT) is performed on the sequence to find the coefficients of the harmonic components to obtain the elliptical shape parameters. Then, the ellipsoidal shape parameter is used as the a priori shape parameter of the star-convex RHM. Finally, the effectiveness of the proposed algorithm is verified by simulation experiments. The proposed extended target fusion method provides a more accurate means for precise tracking and state prediction of extended targets.',
 'Guoqing Qi, Zixuan Xu, Yinya Li, Andong Sheng',
 '8',
 '2025',

'7'],
['Longitudinal and Feedforward Controller Design Based on Linear Matrix Inequalities for a Mini Aerial Vehicle',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70365?af=R',
'In this paper, the design of a MIMO optimal PID controller and a feedforward controller to enhance disturbance rejection and reference tracking for an MAV system is proposed. The feedback controller employs LMI constraints for iterative variable tuning, while a feedforward strategy is developed to further optimise performance. \n\n\n\n\n\n\n\nABSTRACT\nThis paper presents the design of a multi-input multi-output (MIMO) optimal proportional–integral–derivative (PID) controller and a feedforward controller to enhance disturbance rejection and reference tracking in a mini aerial vehicle (MAV). The feedback controller is developed using linear matrix inequality (LMI) constraints within an iterative framework for tuning the PID parameters. In addition, a feedforward control strategy is proposed to further improve performance in reference tracking and disturbance rejection. The feedforward controller is designed under LMI constraints to minimise the H∞ norm of the transfer function matrix relating disturbances (or set-points) to the output (error). This optimisation problem is challenging due to its non-convex and nonlinear nature. Therefore, the main contribution of this paper is to propose a method based on utilising frequency sampling techniques, iterative algorithms and convex optimisation formulation to guarantee optimal solutions for reference tracking and disturbance rejection within specified constraints, which is crucial for preventing irreversible damage in control systems and also enhances the manoeuvrability of the MAV.',
'Saeedreza Tofighi',
'8',
'2025',
'7'],
['FPGA Implementation of a Chopper Controller in PV Systems for IoT Applications',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70364?af=R',
'A digital control system based on FPGA is developed to generate a chopping frequency ranging from 12.5\xa0kHz to 100\xa0kHz, adjusting its output voltage, frequency, and duty cycle accordingly. The proposed digital controller, with optimisations in architecture and process technology, will improve chip area, decrease delay, and enhance the power efficiency of the converter compared to past researches.\n\n\n\n\n\n\n\nABSTRACT\nThis work presents the design and simulation of an Field-Programmable Gate Array (FPGA) based digital controller for a micro-power photovoltaic (PV) energy harvesting system tailored to Internet of Things (IoT) applications. The proposed system features a DC-DC boost converter with an H-bridge topology, designed to interface effectively with low-power PV sources and deliver regulated output to IoT loads. A constant voltage maximum power point tracking (CV-MPPT) algorithm is implemented to maintain optimal energy extraction under variable irradiance. The controller dynamically adjusts the chopping frequency, duty cycle, and output voltage. Leveraging the parallelism and reconfigurability of FPGAs, a high-speed digital control architecture is developed and synthesised on a Cyclone IV FPGA platform. Compared to conventional microcontroller or DSP-based solutions, this approach offers significantly improved response time, power efficiency, and scalability. The final implementation achieves a chip area of 0.92 mm2, a control loop delay of 106\xa0ms, and an overall efficiency of 92.36%, demonstrating its advantages over existing designs in terms of compactness, performance, and energy efficiency.',
'Keyvan Barati, Abdolsamad Hamidi',
'8',
'2025',
'7'],
['High-Impedance Dipoles for Auto-Receiver Tracking Wireless Power Transfer Systems',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70339?af=R',
'This study presents a self-tuning multi-transmitter (multi-Tx) wireless power transfer (WPT) system designed for applications requiring freely positioned receivers (Rxs).\n\n\n\n\n\n\n\nABSTRACT\nThis study presents a self-tuning multi-transmitter (multi-Tx) wireless power transfer (WPT) system designed for applications requiring freely positioned receivers (Rxs). The system leverages high-impedance dipole (HID) as Txs to dynamically activate the nearest Tx to a simple dipole antenna as the Rx while deactivating other Txs without additional control circuits. The Txs design incorporates a modified coaxial cable dipole antenna featuring symmetrical gaps over the shield and inductive tuning elements between the central wire and the shield to optimise the resonance frequency and ensure effective power transfer. The feeding point is connected to the central wire via a small cut in the shield. This autonomous behaviour simplifies system operation, enhances flexibility, and ensures acceptable efficient power transfer. Experimental evaluation demonstrates that the system achieves a power transfer efficiency of approximately 65% while maintaining consistent performance as the Rx rotates over two orthogonally located Txs. The dynamic self-tuning mechanism adapts to the position of the Rx, proving particularly advantageous on curved surfaces.',
'Ali Sharifian Mazraeh Mollaei, Ilkka Laakso, Masoud Sharifian Mazraeh Mollaei',
'8',

'2025',
'7'],
['Fine-Grained Temporal Encoding and Decoding-Based Underwater Object Tracking',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70346?af=R',
'This study focuses on this domain and proposes an innovative fine-grained temporal encoding and decoding-based underwater object tracking method.\n\n\n\n\n\n\n\nABSTRACT\nUnderwater object tracking is a highly challenging task in the field of computer vision. This study focuses on this domain and proposes an innovative fine-grained temporal encoding and decoding-based underwater object tracking method. Due to the complex and dynamic underwater environment, such as uneven lighting, turbid water quality, and complex target motion patterns, existing underwater object tracking methods face significant limitations in accuracy and stability. By carefully designing a refinement module that combines fine-grained consistency and candidate elimination, this method can accurately extract fine-grained features of the target and effectively mitigate the interference of various underwater complexities during feature extraction, thereby improving the precision of target features. Furthermore, leveraging the temporal encoding–decoding module, the target features are continuously propagated along the temporal sequence, allowing full utilization of the relational information between frames, which further enhances tracking stability. Experiments were conducted on the UVOT400 dataset, which is large-scale and rich in attributes with diverse target categories. The results demonstrate that, compared to existing methods, this approach significantly outperforms in both accuracy and stability of underwater object tracking, providing new insights and effective solutions for the advancement of underwater object tracking\xa0technology.',
'Zhen Sun, Zhenggang Guan, Qinghua Li, Mengyang Yuan, Haonan Sun',
'8',
'2025',
'6'],
['Full-Space Scanning Leaky-Wave Antenna Utilising Spoof Surface Plasmon Polaritons',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70328?af=R',
'In this paper, we present a novel spoof surface plasmon polariton leaky-wave antenna (SSPP-LWA) capable of full-space beam scanning. The proposed antenna achieves full-space beam scanning from -90° to +90° across the frequency band of 9.8–13.3\xa0GHz, with a high scanning rate of 5.94° per 1% frequency variation. Additionally, the design features a simplified structure conducive to practical fabrication. \n\n\n\n\n\n\n\nABSTRACT\nA novel spoof surface plasmon polariton leaky-wave antenna (SSPP-LWA) capable of full-space beam scanning is proposed in this letter. First, the surface plasmon polariton (SPPs) unit and its corresponding transmission line are carefully designed, and the underlying mechanism to achieve full-space scanning is analysed based on a monopole array structure. Specifically, a spoof surface plasmon polariton (SSPP) transmission line combined with a semicircular, three-dimensionally loaded radiating patch is introduced and equivalently modelled as a monopole radiating element. Subsequently, the proposed LWA is realised by employing a monopole array composed of 12 identical radiating elements. Simulation results verify that the antenna achieves continuous beam scanning from -90° to +90° within the frequency range of 9.8–13.3\xa0GHz, exhibiting a high scanning rate of up to 5.94°/%, with an average gain of approximately 9 dBi. Benefiting from its excellent scanning capability and stable radiation characteristics, this antenna demonstrates significant potential for target detection and tracking applications in radar systems.',
'Ran Yu, Leilei Liu',
'8',
'2025',
'6'],
['Implementation and Performance Analysis of RTK-GNSS in Wearable Devices for Athletes in Harsh Environments',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70289?af=R',
'This study investigates the implementation of real-time kinematic global navigation satellite system (RTK-GNSS) technique in wearable electronic performance and tracking system for football players to obtain highly accurate positioning data, even in harsh environments. The experiment results demonstrate that the utilization of RTK-GNSS can significantly improve positioning accuracy in various stadium environments, which is crucial for evaluating player performance.\n\n\n\n\n\n\n\nABSTRACT\nThis study presents the performance analysis of RTK-GNSS (real-time kinematic global navigation satellite system) mounted on wearable devices for athletes in harsh environments. GNSS signal can deteriorate by obstacles, i.e. stadium roofs, due to the high GDOP (geometric dilution of precision). RTK-GNSS is used to overcome harsh environment problems in this study. Hardware experiments are performed to analyse the performance of RTK-GNSS in sports wearable devices in harsh environments. The full pitch tracking test results of a wearable device equipped with RTK-GNSS demonstrate the best performance among three different wearable devices in a harsh environment; the maximum positioning error of 1.56 m$\\text{m}$ is approximately 3–30 times smaller than that of wearable devices equipped with stand-alone GNSS. The distance test results of the wearable device equipped with RTK-GNSS validate the most accurate

performance, 98.97%$\\%$, compared to the true value in harsh\xa0environment.',
  'Mingu Kim, Bit Kim, Chulwoo Park, Jinsung Yoon',
  '8',
  '2025',
  '5'],
 ['Robust Beam Tracking for 3D Manoeuvrable UAV in DFRC Systems',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70211?af=R',
  "This letter presents a novel 3D robust manoeuvrable unmanned aerial vehicle (UAV) beam tracking algorithm within the dual-functional radar-communication framework. Specifically, we propose the interacting multiple model algorithm to accommodate multiple potential UAV motion models and improve the unscented Kalman filter algorithm to enhance its stability.\n\n\n\n\n\n\n\nABSTRACT\nDual-functional radar-communication (DFRC) will be a key technology in future sixth-generation (6G) network. Specifically, an integrated radar and communication platform (IRCP) equipped with full-dimensional antenna arrays can execute 3D beamforming, which can effectively minimize interference for communicating with unmanned aerial vehicles (UAVs). However, a significant challenge in fully exploiting 3D beamforming gain is the IRCP's ability to precisely track manoeuvrable UAVs. In this letter, we propose a novel interacting multiple model with enhanced unscented Kalman filter algorithm to realize 3D beam tracking for one manoeuvrable UAV within the framework of DFRC. To be specific, we build multiple state transition models to adopt the manoeuvrability of the UAV. Meanwhile, the stability of the unscented Kalman filter is improved by using the singular value decomposition. Simulation results show that the proposed algorithm has higher tracking accuracy and better transmission performance for one manoeuvrable UAV than the traditional single-motion model based beam tracking\xa0algorithm.",
  'Yuhang Tang, Wei Liu, Jinkun Zhu, Jing Lei, Haoying Mo',
  '8',
  '2025',
  '3'],
 ['Optimal Performance Analysis for Networked System with Quantitative Control Input Over Feedback Channel',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70183?af=R',
  "This paper focuses on the performance analysis of networked control systems, specifically addressing the optimal tracking performance with quantitative control inputs over feedback channels. A dual-degree-of-freedom controller is designed using the Youla parameterization method, and the system's performance is assessed through numerical simulations. The findings highlight the impact of unstable poles, non-minimum phase zeros, and quantization errors on tracking performance, offering insights for practical control system design.\n\n\n\n\n\n\n\nABSTRACT\nNetworked control systems (NCS) have become an emerging and important research area given the rapidly developing network communication technology. Numerous studies in this area concentrate on the stability of control systems, while little has been done on its performance analysis, especially the index of an optimal control system. We fill this gap here by analysing the optimal tracking performance for an NCS that has quantitative control inputs over feedback channels. We design a dual-degree-of-freedom controller along with an optimal control system, by applying the Youla parameterization method based on the time domain and coprime decompositions. The proposed methodology is assessed by numerical simulations. The system's tracking performance is found to be deteriorated by the unstable poles and non-minimum phase zeros of the controlled object, as well as by quantization signal errors. These findings are beneficial for analysing and designing practical control\xa0systems.",
  'Fang Han, Xiaowei Jiang',
  '8',
  '2025',
  '3'],
 ['Applying Adaptive Wavelet Neural Network and Sliding Mode Control for Tracking Control of MEMS Gyroscope',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70187?af=R',
  'In this letter, an algorithm applying an adaptive wavelet neural network (AWNN) and sliding-mode control (SMC) is proposed, investigated and exploited for tracking control of micro-electromechanical system (MEMS) gyroscope. Such an AWNN model can be regarded as a special radius basis function neural network and utilizes Mexican hat function as activation function. Besides, Taylor expansion is used for analyzing activation radius, which is considered as an adaptive variable. The parameters of the MEMS gyroscope model are hard to obtain in engineering applications; thus, AWNN is designed to approximate the uncertain function of MEMS gyroscope and the unknown asymmetrical dead zone in the control scheme. The weights updating laws and the activation radius adaptive laws in AWNN are derived from the Lyapunov stability analysis, which results in the control error converging to the desired value and the weights and activation radius converging to its real value. To achieve the effect of error acceleration, a power function is used to design a sliding mode function. Computer simulation results substantiate the theoretical analysis and

further demonstrate the efficacy of such an algorithm combined with AWNN and SMC for MEMS gyroscope control.',
  'Guo Luo, Bingling Chen',
  '8',
  '2025',
  '3'],
 ['A GGIW PMBM Smoother for Multiple Extended Object Tracking',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70188?af=R',
  'A gamma Gaussian inverse Wishart Poisson multi-Bernoulli mixture (PMBM) filter that uses multiple time scans for smoothing is proposed. The results of continuous tracking are smoothed and corrected at regular intervals, resulting in more accurate tracking performance. Simulation results show that the proposed algorithm effectively reduces generalized optimal sub-pattern assignment errors compared to the original PMBM filter. \n\n\n\n\n\n\n\nABSTRACT\nPoisson multi-Bernoulli mixture (PMBM) filtering is an effective method for multi-extended target tracking. However, it has a low utilization of information from multiple time scans, leading to potential errors such as jumps and discontinuities in the target trajectories. To address this, propose a gamma Gaussian inverse Wishart PMBM filter that uses multiple time scans for smoothing. The results of continuous tracking are smoothed and corrected at regular intervals, resulting in more accurate tracking performance. Simulation results show that the proposed algorithm effectively reduces generalized optimal sub-pattern assignment errors compared to the original PMBM filter.',
  'Wenhui Wang, Ye Xu, Kejie Zhang, YouPeng Sun, Peng Li',
  '8',
  '2025',
  '2'],
 ['A Novel D-InSAR and Offset Tracking Fusion Algorithm Based on Variance-Masked Mean Filtering',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70173?af=R',
  'This paper proposes a novel fusion algorithm based on variance-masked mean filtering to combine the strengths of both methods. The algorithm filters noise in gentle regions of the offset map and fuses refined results with D-InSAR to enhance accuracy. Applied to the 2023 Türkiye earthquake, the method effectively reduces noise, corrects unwrapping errors, and achieves superior accuracy in high-gradient deformation scenarios such as earthquakes and landslides.\n\n\n\n\n\n\n\nABSTRACT\nAccurate ground deformation monitoring is crucial for understanding geophysical activities and mitigating disaster risks. D-InSAR and offset tracking are widely used but face limitations: D-InSAR struggles with phase unwrapping errors in large-scale deformation regions, while offset tracking lacks precision in small-scale deformation. This paper proposes a fusion algorithm based on variance-masked mean filtering to combine the strengths of both methods. The algorithm filters noise in gentle regions of the offset map and fuses refined results with D-InSAR to enhance accuracy. Applied to the 2023 Türkiye earthquake, the method effectively reduces noise, corrects unwrapping errors, and achieves superior accuracy in large-scale deformation scenarios such as earthquakes and landslides.',
  'Yifan Wang, Gen Li, Zihan Hu, Chuanxu Sun, Zhe Li',
  '8',
  '2025',
  '2'],
 ['Extended target tracking using neural network and Gaussian process',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70151?af=R',
  "In extended target tracking, Gaussian Process (GP) is utilized to model unknown contour functions based on the model-predicted target center and contour measurements. However, model prediction relies on accurate prior knowledge. When the model-predicted target center is inaccurate, it will affect the modelling of the measurement model. To address issue, this letter introduces a hybrid-driven approach that combines extended Kalman filter using GP with neural network; proposes an extended target tracking algorithm using neural network and GP. The algorithm predicts the target center according to the neural network and the target's kinematic model, and takes the prediction center and the contour measurements at the current moment as the input of the neural network, which in turn provides real-time estimates for the predicted center compensation. The simulation results show that the algorithm has a significant improvement in tracking performance and better accuracy in estimating the center position and extent state of the target.",
  'Hao Wang, Liping Song',
  '8',
  '2025',
  '1']]

SEARCH WORD: SMALL TARGET

[['MTMLNet: Multi-Task Mutual Learning Network for Infrared Small Target Detection and Segmentation',

'http://ieeexplore.ieee.org/document/11080263',
'Infrared small target detection has been extensively studied due to its wide range of applications. Most studies treat infrared small target detection as an independent task, either as a detection-based or a segmentation-based, failing to fully leverage the supervisory information from different annotation forms. To address this issue, we propose a multi-task mutual learning network (MTMLNet) specifically designed for infrared small targets, aiming to enhance both detection and segmentation performance by effectively utilizing various forms of supervisory information. Specifically, we design a multi-stage feature aggregation (MFA) module capable of capturing features with varying gradients and receptive fields simultaneously. Additionally, a hybrid pooling down-sampling (HPDown) module is proposed to mitigate information loss during the down-sampling process of infrared small targets. Finally, the hierarchical feature fusion (HFF) module is designed to adaptively select and fuse features from different semantic layers, learning the optimal way to fuse features across semantic layers. The results on IRSTD-1k and SIRST-V2 datasets show that our proposed MTMLNet achieves state-of-the-art (SOTA) performance in both detection-based and segmentation-based methods. The codes are available at https://github.com/YangBo0411/MTMLNet',
'Bo Yang, Fengqian Li, Songliang Zhao, Wei Wang, Jun Luo, Huayan Pu, Mingliang Zhou, Yangjun Pi',
'4',
'2025',
'8'],
['Computational Fluid Dynamic Network for Infrared Small Target Detection',
'http://ieeexplore.ieee.org/document/10949663',
'Infrared small target detection (IRSTD) aims to identify and locate small targets amidst background noise. It is highly valuable in various practical application domains, such as maritime rescue and early warning systems deployed in challenging conditions such as harsh weather, low illumination, and long imaging distances. Different from existing works that either adopt well-designed backbone networks or devise specific modules to improve them from different aspects, in this article, we formulate the learning process of IRSTD from a novel perspective, i.e., the mechanism of pixel movement. Considering that the movement of pixels passing through the layers of the network for IRSTD can be analogized to the flow of particles in a fluid dynamic system, we propose a computational fluid dynamic network (CFD-Net) derived from computational fluid dynamics. Technically, we leverage the superiority of the unilateral difference equation with third-order accuracy and devise a unilateral differential residual structure as the backbone of CFD-Net. This design ensures that the pixel stream only flows in the forward direction. In addition, a switch-controlled multidirectional treatment tank (SMTT) is introduced to CFD-Net to dynamically guide the pixel stream to the appropriate path for different targets with varying shapes and orientations, facilitating learning robust target representation and improving detection performance. The proposed CFD-Net is evaluated on the IRSTD-1k and SIRST datasets and is found to outperform existing state-of-the-art (SOTA) methods.',
'Mingjin Zhang, Ke Yue, Jie Guo, Qiming Zhang, Jing Zhang, Xinbo Gao',
'5',
'2025',
'8'],
['NBCDC-YOLOv8: A new framework to improve blood cell detection and classification based on YOLOv8',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cvi2.12341?af=R',
'In recent years, computer technology has successfully permeated all areas of medicine and its management, and it now offers doctors an accurate and rapid means of diagnosis. Existing blood cell detection methods suffer from low accuracy, which is caused by the uneven distribution, high density, and mutual occlusion of different blood cell types in blood microscope images, this article introduces NBCDC-YOLOv8: a new framework to improve blood cell detection and classification based on YOLOv8. Our framework innovates on several fronts: it uses Mosaic data augmentation to enrich the dataset and add small targets, incorporates a space to depth convolution (SPD-Conv) tailored for cells that are small and have low resolution, and introduces the Multi-Separated and Enhancement Attention Module (MultiSEAM) to enhance feature map resolution. Additionally, it integrates a bidirectional feature pyramid network (BiFPN) for effective multi-scale feature fusion and includes four detection heads to improve recognition accuracy of various cell sizes, especially small target platelets. Evaluated on the Blood Cell Classification Dataset (BCCD), NBCDC-YOLOv8 obtains a mean average precision (mAP) of 94.7%, and thus surpasses the original YOLOv8n by 2.3%.',
'Xuan Chen, Linxuan Li, Xiaoyu Liu, Fengjuan Yin, Xue Liu, Xiaoxiao Zhu, Yufeng Wang, Fanbin Meng',
'6',
'2025',
'2'],
['PAP-YOLOv8: Hierarchical Feature Optimization With Dual-Stage Attention for Object Detection in Remote Sensing Scenes',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70177?af=R',
'This study addresses the challenges of high missed detection rates for dense small targets and insufficient multi-scale feature interaction in remote sensing images by proposing the PAP-YOLOv8 framework, an

improved YOLOv8-based architecture. By integrating a pixel dynamic convolutional attention (PDCA) module, an adaptive channel attention network (ACAN), and a pyramidal wavelet attention fusion block (PWAFB), the framework achieves precise multi-scale object detection in complex backgrounds, attaining an mAP@0.5 of 86.9% on the DIOR dataset—a 5.3% improvement over the baseline—while maintaining a lightweight architecture, offering an efficient solution for real-time satellite monitoring.\n\n\n\n\n\n\n\n\nABSTRACT\nRemote sensing image target recognition plays an important role in environmental monitoring, disaster management, urban planning, and traffic systems. To address issues such as high missing detection rates of dense small targets and insufficient multi-scale feature interaction in remote sensing image target detection, an improved PAP-YOLOv8 network architecture was proposed in this study based on YOLOv8. This framework included three core innovations: (1) the pixel dynamic convolution attention module (PDCA), which enhanced small target detection capabilities in complex backgrounds through direction-sensitive convolution and a dual-stage attention mechanism; (2) the adaptive channel attention network (ACAN), which optimized multi-scale feature fusion and suppressed background interference via dynamic weight allocation and channel recalibration; (3) the pyramidal wavelet attention fusion block (PWAFB), which synergistically integrated wavelet multi-resolution analysis with composite attention mechanisms to model spatial details and channel relationships in remote sensing images through a collaborative decomposition and reconstruction workflow. Experiments on the Detection in Optical Remote Sensing (DIOR) dataset showed that the mAP@0.5 of PAP-YOLOv8 reached 86.9%, an increase of 5.3% over the baseline YOLOv8, while demonstrating enhanced detection capability for small and multi-scale targets. Compared with other detection models, PAP-YOLOv8 achieves a balanced trade-off between network size and detection capability, demonstrating superior overall performance with an F1 score of 83.1%. This study confirms the potential of PAP-YOLOv8 for satellite monitoring applications and provides a valuable reference for improving target detection performance in optical remote sensing satellite imagery.',
  'Pengfei Zhang, Jian Liu, Jianqiang Zhang, Xingda Li',
  '7',
  '2025',
  '8'],
 ['A Coarse-to-Fine Detection Framework for Automated Lung Tumour Detection From 3D PET/CT Images',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70146?af=R',
  'The limitations in positron emission tomography (PET) imaging resolution and the complexity of cancer characteristics make detecting small lesions particularly challenging. A novel coarse-to-fine detection framework is proposed to reduce missed diagnoses of small lung lesions in positron emission tomography/computed tomography (PET/CT) images. The proposed method integrates a stacked detection structure with a multi-attention guidance mechanism, effectively leveraging spatial and contextual information from small lesions to enhance lesion localisation. \n\n\n\n\n\n\n\nABSTRACT\nLung cancer remains the leading cause of cancer-related mortality worldwide. Early detection is critical to improving treatment outcomes and survival rates. Positron emission tomography/computed tomography (PET/CT) is a widely used imaging modality for identifying lung tumours. However, limitations in imaging resolution and the complexity of cancer characteristics make detecting small lesions particularly challenging. To address this issue, we propose a novel coarse-to-fine detection framework to reduce missed diagnoses of small lung lesions in PET/CT images. Our method integrates a stacked detection structure with a multi-attention guidance mechanism, effectively leveraging spatial and contextual information from small lesions to enhance lesion localisation. Experimental evaluations on a PET/CT dataset of 225 patients demonstrate the effectiveness of our method, achieving remarkable results with a precision of 81.74%, a recall of 76.64%, and an mAP of 84.72%. The proposed framework not only improves the detection accuracy of small target lesions in the lung but also provides a more reliable solution for early diagnosis.',
  'Yunlong Zhao, Qiang Lin, Junfeng Mao, Jingjun Wei, Yongchun Cao, Zhengxing Man, Caihong Liu, Jingyan Ma, Xiaodi Huang',
  '7',
  '2025',
  '6'],
 ['SEM-YOLO: A Small Target Defect Detection Model for Photovoltaic Modules',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70134?af=R',
  'This paper proposes a PV module defect detection model, SEM-YOLO, based on YOLOv8.The experimental results show that the mAP of the improved model reaches 93.8%, among which the mAP of small target defects reaches 83%, compared with the mainstream models (RE-DETR, YOLOv9s, YOLOv10n, and YOLOv11), the detection accuracies in terms of overall and small-target defects are significantly improved, which validates the effectiveness of the model.\n\n\n\n\n\n\n\nABSTRACT\nDefect detection is key to extending the lifetime of PV modules. However, existing methods still face significant challenges in detecting small and ambiguous targets. To this end, this paper proposes a PV module defect detection model, SEM-YOLO, based on YOLOv8. The model improves the performance through the following improvements: first, the SPD-Conv module is introduced to replace the traditional convolution in the backbone and neck sections to reduce the information loss caused by excessive down-sampling, thus

enhancing the detection of small targets. Second, the neck section C2f-EMA module is introduced, in which the efficient multiscale attention module (EMA) enhances feature extraction by redistributing weights and prioritizing relevant features to improve the perception and recognition of small target defects (hot spots). Finally, we add a small target detection layer and increase the MultiSEAM detection header, so that the model can capture and detect small targets more efficiently at the output stage. The experimental results show that the mAP of the improved model reaches 93.8%, among which the mAP of small target defects reaches 83%, which is an improvement of 2.23% and 7.62% compared with YOLOv8. In addition, compared with the mainstream models (RT-DETR, YOLOv9s, YOLOv10n, and YOLOv11), the detection accuracies in terms of overall and small-target defects are significantly improved, which further validates the effectiveness of the model.',
'Wang Yun, Yin Wang, Gang Xie, Zhicheng Zhao',
'7',
'2025',
'6'],
['AMFE-YOLO: A Small Object Detection Model for Drone Images',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70110?af=R',
'To address the issue of poor detection performance in drone aerial images, we innovatively propose AMFE-YOLO. Solutions are introduced for challenges such as occlusion and targets under varying lighting conditions. In the end, our model shows significant improvement compared to current popular algorithms. \n\n\n\n\n\n\n\nABSTRACT\nDrones, due to their high efficiency and flexibility, have been widely applied. However, small objects captured by drones are easily affected by various conditions, resulting in suboptimal surveying performance. While the YOLO series has achieved significant success in detecting large targets, it still faces challenges in small target detection. To address this, we propose an innovative model, AMFE-YOLO, aimed at overcoming the bottlenecks in small target detection. Firstly, we introduce the AMFE module to focus on occluded targets, thereby improving detection capabilities in complex environments. Secondly, we design the SFSM module to merge shallow spatial information from the input features with deep semantic information obtained from the neck, enhancing the representation ability of small target features and reducing noise. Additionally, we implement a novel detection strategy that introduces an auxiliary detection head to identify very small targets. Finally, we reconfigured the detection head, effectively addressing the issue of false positives in small-object detection and improving the precision of small object detection. AMFE-YOLO outperforms methods like YOLOv10 and YOLOv11 in terms of mAP on the VisDrone2019 public dataset. Compared to the original YOLOv8s, the average precision improved by 5.5%, while the model parameter size was reduced by 0.7\xa0M.',
'Qi Wang, Chengxin Yu',
'7',
'2025',
'5'],
['A High-Accuracy YOLOv8-ResAttNet Framework for Maritime Vessel Detection Using Residual Attention',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70085?af=R',
"A new model YOLOv8 ResAttNet based on YOLOv8 algorithm combined with residual and attention mechanism (ResAttNet) is proposed to improve the accuracy of ship detection in remote sensing images. \n\n\n\n\n\n\n\nABSTRACT\nAgainst the backdrop of constantly upgrading maritime security requirements and dynamic marine environments, satellite based ship detection has become a key technology for national maritime surveillance, resource management, and environmental protection. However, existing methods often struggle to address ongoing challenges, including insufficient sensitivity to small vessels and susceptibility to errors or missed detections in complex ocean backgrounds caused by wave reflections, cloud cover, and lighting changes. To address these limitations, this study proposes YOLOv8 ResAttNet, an enhanced model that integrates residual learning and attention mechanisms into the YOLOv8 framework. The core innovation lies in a custom designed backbone network that combines multi-scale feature aggregation with an improved ICBAM attention module to achieve precise localization of ship targets while suppressing irrelevant background noise. This architecture dynamically recalibrates feature channel weights through residual attention blocks, enhancing the model's ability to distinguish subtle ship features (such as hull contours and superstructures) in different maritime scenarios. Extensive experiments on high-resolution HRSID datasets have demonstrated the superiority of this model: the average accuracy (mAP50) of YOLOv8 ResAttNet is 95.2%, which is 4.9% higher than the original YOLOv8 and over 4% higher than state-of-the-art models such as YOLO SENet and YOLO11. These improvements highlight its robustness in handling scale changes and complex background interference. The research results emphasize the effectiveness of combining residual connectivity with attention driven feature refinement for maritime target detection, especially in small target scenes. This work not only advances the technological frontier of remote sensing image analysis, but also provides a scalable framework for real-world applications such as illegal fishing monitoring, maritime traffic management, and disaster response. Future research directions include extending the model to multimodal satellite data fusion, optimizing the computational efficiency of edge device deployment, and further bridging the gap between theoretical innovation and maritime surveillance

systems.",
 'Peixue Liu, Mingze Sun, Xinyue Han, Shu Liu, Yujie Chen, Han Zhang',
 '7',
 '2025',
 '4'],
 ['An Improved Object Detection Algorithm for UAV Images Based on Orthogonal Channel Attention Mechanism and Triple Feature Encoder',
 'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70061?af=R',
 "The OATF-YOLO algorithm addresses Unmanned Aerial Vehicle (UAV) image challenges like complex scenes, dense small targets and scale variations in object detection. It incorporates an orthogonal channel attention mechanism for enhanced feature extraction, and integrates a triple feature encoder and a scale sequence feature fusion module to improve multi - scale feature fusion. Also, an inner factor is introduced into the loss function to boost the algorithm's robustness and accuracy. The experimental results on the VisDrone2019 and NWPU-VHR10 datasets show that the improved algorithm significantly improves on Precision, Recall, mAP50%, and mAP50:90% metrics in both the validation and testing sets, indicating the superiority of the proposed algorithm.\n\n\n\n\n\n\n\n\nABSTRACT\nObject detection in Unmanned Aerial Vehicle (UAV) imagery plays an important role in many fields. However, UAV images usually exhibit characteristics different from those of natural images, such as complex scenes, dense small targets, and significant variations in target scales, which pose considerable challenges for object detection tasks. To address these issues, this paper presents a novel object detection algorithm for UAV images based on YOLOv8 (referred to as OATF-YOLO). First, an orthogonal channel attention mechanism is added to the backbone network to imporve the algorithm's ability to extract features and clear up any confusion between features in the foreground and background. Second, a triple feature encoder and a scale sequence feature fusion module are integrated into the neck network to bolster the algorithm's multi-scale feature fusion capability, thereby mitigating the impact of substantial differences in target scales. Finally, an inner factor is introduced into the loss function to further upgrade the robustness and detection accuracy of the algorithm. Experimental results on the VisDrone2019-DET dataset indicate that the proposed algorithm significantly outperforms the baseline model. On the validation set, the OATF-YOLO algorithm achieves a precision of 59.1%, a recall of 40.5%, an mAP50 of 42.5%, and an mAP50:95 of 25.8%. These values represent improvements of 3.8%, 3.0%, 4.1%, and 3.3%, respectively. Similarly, on the test set, the OATF-YOLO algorithm achieves a precision of 52.3%, a recall of 34.7%, an mAP50 of 33.4%, and an mAP50:95 of 19.1%, reflecting enhancements of 4.0%, 3.3%, 4.0%, and 2.6%, respectively. To further validate the model's robustness and scalability, experiments are conducted on the NWPU-VHR10 dataset, and OATF-YOLO also achieves excellent performance. Furthermore, compared to several classical object detection algorithms, OATF-YOLO demonstrates superior detection performance on both datasets and indicates that it is better suited for UAV image object detection scenarios.",
 'Wenfeng Wang, Chaomin Wang, Sheng Lei, Min Xie, Binbin Gui, Fang Dong',
 '7',
 '2025',
 '4'],
 ['MFF-YOLOv8: Small Object Detection Based on Multi-Scale Feature Fusion for UAV Remote Sensing Images',
 'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70066?af=R',
 "In this work, we propose a novel network for small target detection in UAV aerial photography scenarios to address the challenges often encountered in UAV aerial photography scenarios where small target features are difficult to extract and are susceptible to reception light, scale variations and occlusion. First, we design a pyramid network with high-resolution feature fusion, which reconstructs the high-resolution feature map for feature selection, allowing the network to fully exploit the high-resolution feature information after removing a large amount of noise. Second, we design a multi-scale hybrid attention mechanism that maximizes the retention and extraction of small target feature information by fully capturing multi-scale features in the channel and space. Finally, a regression loss function that combines the joint auxiliary bounding box and the dynamic focusing bounding box is designed to improve the processing capability for a large number of low IoU samples in UAV aerial photography scenes. Based on the experiments, we compare the proposed model with other state-of-the-art methods on the VisDrone2019 dataset. The experimental results show that the proposed model has higher detection accuracy and excellent robustness.\n\n\n\n\n\n\n\n\nABSTRACT\nAs a popular task in drone-captured scenes, object detection involves images with a large number of small objects, but current networks often suffer missed and false detections. To address this problem, we propose a YOLO algorithm MFF-YOLOv8 based on multi-scale feature fusion for small target detection in UAV aerial images. First, a high-fesolution feature fusion pyramid (HFFP) is designed, which utilizes high-resolution feature maps containing much information about small objects to guide the feature fusion module, weighting and fusing feature maps to enhance the network's ability to represent small targets. Meanwhile, a reconstruction feature selection (RFS) module is employed to remove the large amounts of noise produced by high-resolution feature maps. Second, a hybrid efficient multi-scale attention (HEMA) mechanism is

designed in the backbone network to maximize the retention and extraction of feature information related to small objects while simultaneously suppressing background noise interference. Finally, an Inner-Wise IoU loss function (Inner-WIoU) is designed for joint auxiliary bounding box and dynamic focal bounding box regression, which enhances the accuracy of network regression results, thus improving the detection precision of the model for small objects. MFF-YOLOv8 was experimented on the VisDrone2019 dataset, achieving a 47.9% mAP50, 9.3% up compared with that of the baseline network YOLOv8s. Also, in order to verify the generalization of the overall network, it was evaluated on the DOTA and UAVDT datasets, and the mAP50 was improved by 3.7% and 1.8%, respectively. The results demonstrate that MFF-YOLOv8 significantly enhances detection precision for small objects in UAV aerial\xa0scenes.",
  'Kun Hu, Jinzheng Lu, Chaoquan Zheng, Qiang Xiang, Ling Miao',
  '7',
  '2025',
  '4'],
 ['LCM-YOLO: A Small Object Detection Method for UAV Imagery Based on YOLOv5',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70051?af=R',
  "The LCM-YOLO model, an enhancement of the YOLOv5, improves small target detection and adaptability to scale variations in UAV aerial images by incorporating a local fusion mechanism, cross-scale feature fusion and multi-head attention, resulting in significant performance improvements over the original YOLOv5 on the VisDrone2019 dataset. Specifically, LCM-YOLO achieves a 7.2% increase in mAP50 and a 5.1% increase in mAP50-95, with final scores of 40.7% and 22.5%, respectively.
\n\n\n\n\n\n\n\nABSTRACT\nThis study addresses the challenges of detecting small targets and targets with significant scale variations in UAV aerial images. We propose an improved YOLOv5 model, named LCM-YOLO, to tackle these challenges. Initially, a local fusion mechanism is introduced into the C3 module, forming the C3-LFM module to enhance feature information acquisition during feature extraction. Subsequently, the CCFM is employed as the neck structure of the network, leveraging its lightweight convolution and cross-scale feature fusion characteristics to effectively improve the model's ability to integrate target features at different levels, thereby enhancing its adaptability to scale variations and detection performance for small targets. Additionally, a multi-head attention mechanism is integrated at the front end of the detection head, allowing the model to focus more on the detailed information of small targets through weight distribution. Experiments on the VisDrone2019 dataset show that LCM-YOLO has excellent detection capabilities. Compared to the original YOLOv5 model, its mAP50 and mAP50-95 metrics are improved by 7.2% and 5.1%, respectively, reaching 40.7% and 22.5%. This validates the effectiveness of the LCM-YOLO model for detecting small and multi-scale targets in complex backgrounds.",
  'Shaodong Liu, Faming Shao, Weijun Chu, Heng Zhang, Dewei Zhao, Jinhong Xue, Qing Liu',
  '7',
  '2025',
  '3'],
 ['Efficient method for detecting targets from remote sensing images based on global attention mechanism',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70012?af=R',
  "Remote sensing image target detection provides an effective and accurate data analysis tool for many application areas. Due to complex backgrounds, large differences in target scales, and missed detection of small targets, remote sensing image target detection is challenging. In order to enhance the model's understanding of the global information of remote sensing images, this paper proposes the GFA module. This module can establish the global contextual connection of remote sensing images to provide rich context to help understand the complex scene and background in which the target is located, without being limited to local information. Additionally, it focuses on channel information for enhanced target feature extraction. For the purpose of alleviating the serious imbalance in foreground–background samples that is present in single-level target detection models. The loss function is reconstructed based on focal loss by redefining the balance factor α and focus factor γ, so that it can be dynamically adjusted during network training. Meanwhile, EIoU is used to further enhance the bounding box regression capability. Affine transformations were also used to augment the dataset in order to assist the model in adjusting to real-world situations. The proposed method is experimentally validated on the publicly available HRRSD dataset. In comparison with YOLO v5, the mAP of the detection results improved by 2.7%. Compared with YOLO v8 and YOLO v10, the mAP improved by 3.2% and 3.3%. The model achieves an FPS of 40.1, an optimal balance between speed and accuracy. Further, experiments are conducted using the NWPU VHR-10 dataset and the RSOD dataset, both of which demonstrated that the proposed method outperforms other target detection methods and improves remote sensing target detection performance.",
  'Zijun Gao, Jingwen Su, Bo Li, Jue Wang, Zhankui Song',
  '7',
  '2025',
  '2'],
 ['Modified You Only Look Once Network Model for Enhanced Traffic Scene Detection Performance for Small Targets',

'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70014?af=R',
"We propose a model based on you only look once version 8X (Yolov8X) network model, which has been combined with receptive fields block (RFB) and multidimensional collaborative attention (MCA). The validation results on the Argoverse 1.1 autonomous driving dataset demonstrate that the enhanced network model outperforms the prevailing detectors, achieving an F1 score of 78.6, an average precision of 55.1, and an average recall of 72.4.\n\n\n\n\n\n\nABSTRACT\nIn order to address the challenge of small target recognition in traffic scenes, we propose a model based on you only look once version 8X (Yolov8X) network model, which has been combined with receptive fields block (RFB) and multidimensional collaborative attention (MCA). First, the model employs the RFB to extract reliable and distinctive features, thereby enhancing the precision of small target identification. Furthermore, the MCA structure is introduced to simulate multidimensional attention through three parallel branches, thereby enhancing the feature expression ability of the model. This fragment describes a compression transformation and an excitation transformation that captures the differentiated feature representation of the command. These transformations facilitate the network's ability to locate and predict the location of small objects more accurately. Utilizing these transformations enhances the expressiveness and diversity of features, thereby improving the detection performance of small objects. Furthermore, data augmentation and hyperparameter optimization techniques are employed to enhance the model's generalisability. The validation results on the Argoverse 1.1 autonomous driving dataset demonstrate that the enhanced network model outperforms the prevailing detectors, achieving an F1 score of 78.6, an average precision of 55.1, and an average recall of 72.4. The algorithm's excellent performance for small target detection was demonstrated through visual analysis, proving its high application value and potential for promotion in fields such as autonomous\xa0driving.",
'Lei Shi, Shuai Ren, Xing Fan, Ke Wang, Shan Lin, Zhanwen Liu',
'7',
'2025',
'2'],
['MCFNet: Research on small target detection of RGB-infrared under UAV perspective with multi-scale complementary feature fusion',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13320?af=R',
"The task of single-modal image target detection from a drone's perspective faces multiple challenges such as large image sizes, small and dense targets, insufficient lighting conditions, and hardware resource constraints, all of which affect the accuracy and real-time performance of algorithms. In response, a small target detection algorithm that fuses multi-scale complementary features, named MCFNet is proposed. Firstly, in order to independently extract the unimodal features of visible light and infrared images and effectively enhance the complementary information between images, a dual-stream backbone network with a terminal fusion mechanism is proposed. Secondly, the complementary feature information residual module is utilized to optimize the integration process of residual features. Subsequently, by designing a multi-scale feature enhancement module the network's capability to capture multi-scale features is enhanced. Finally, for targets of varying sizes, a lightweight Transformer feature extraction module is proposed to improve the detection accuracy of small targets from a drone's perspective. Test results on the drone-vehicle dataset show that this method achieved an average detection accuracy of 67.92%, while the detection accuracy on the self-constructed UAV-data dataset reached 96.4%. Additionally, a series of ablation experiments validated the effectiveness of the different modules.",
'Jing Jing, Jian Feng Hu, Zuo Peng Zhao, Ying Liu',
'7',
'2025',
'1'],
['Channel selection and local attention transformer model for semantic segmentation on UAV remote sensing scene',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13298?af=R',
'Compared with common urban landscape semantic segmentation, unmanned aerial vehicle (UAV) image semantic segmentation is more challenging because small targets have very low pixel percentages and multi-scale features due to the influence of flight altitude. Yet, the commonly used successive grid downsampling strategy in the current transformer-based methods omits some important features of small targets. Furthermore, due to the complex background interference, it can lead to even worse results. In reaction to this, existing strategies aim to maintain superior resolution. Nevertheless, the application of this method incurs considerable computational costs, which brings challenges for the practical applications of UAVs. So it is significant to design a novel framework to balance retaining more pixels representing small objects during downsampling and reducing computational costs. For this, the Channel Selection and the Local Attention Transformer Model (CSLFormer) are proposed. During the overlap patch embedding process of feature maps, the model allocates half of the important channels to global attention and local attention. These two types of attention focus on different aspects: one learns the relationships and importance among various patches, while the other emphasizes the features of individual patches. The method shows superior performance on two public datasets: AeroScapes and Vaihingen, achieving mean intersection over union

(mIoU) of 75.57% and 78.93%, respectively. The proposed CSLFormer has been released on GitHub: https://github.com/leoda1/CSLFormer.',
  'Da Liu, Hao Long, Zhenbao Liu',
  '7',
  '2024',
  '12'],
 ['YOLO-Tiny: A lightweight small object detection algorithm for UAV aerial imagery',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13314?af=R',
  "In unmanned aerial vehicle (UAV) aerial target detection tasks, two main challenges exist: the limited computational resources of UAV terminals, which are not conducive to running complex models, and the prevalence of small targets, which can easily lead to missed detections and false positives. To address these issues, this study proposes a lightweight and high-accuracy small-object detection algorithm for UAV aerial imagery that is based on YOLOv5s. First, the network structure is optimized by removing the layers in YOLOv5s primarily used for detecting large targets (P4 and P5) and adding layers primarily used for detecting small targets (P2 and P3). This enables the model to focus more on extracting small-object features. A lightweight dynamic convolution is subsequently introduced in the C3 module, and the lightweight LW_C3 and LW_downsampling modules are designed for feature extraction and downsampling operations. This enhances the model's feature extraction capability while achieving a lightweight design. Finally, the adaptive multi-scale spatial feature fusion (AMSFF) module is designed to adaptively learn the spatial weights of the feature maps at different levels, thereby further strengthening the effective fusion of multi-scale features. Experimental results show that the improved YOLO-Tiny model has higher accuracy and lower complexity, hence validating its excellent performance.",
  'Fei Feng, Lu Yang, Quanxing Zhou, Weipeng Li',
  '7',
  '2024',
  '12'],
 ['MOE-IR: Infrared Dim Small Target Detection Method With Mixture of Experts Feature Extraction',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70359?af=R',
  'A multi expert model-based infrared small target detection method mixture of experts infrared (MOE-IR) has been proposed to solve the problem of robust detection of infrared small targets in complex environments. In the MOE-IR, the target feature extraction expert model and the background suppression expert model are proposed to suppress background clutter interference and extract rich infrared small target features, respectively. Meanwhile, an adaptive gate controlled network is designed to adaptively allocate the weights of the two expert models according to the input infrared image, so as to achieve effective detection of infrared small targets in different complex scenes. Comprehensive experiments demonstrate the detection performance of the proposed MOE-IR in the public dataset SIRST is better than the existing advanced infrared small target detection methods, with Pd increased by 3% and Fa reduced by 73%.\n\n\n\n\n\n\n\nABSTRACT\nInfrared small target detection holds significant importance across various domains, including military and security applications. Nevertheless, detecting small targets is highly challenging due to their minimal pixel presence in images, indistinct features, and complex backgrounds. Existing detection methods are often interfered by complex backgrounds, resulting in unsatisfactory detection results. To overcome this challenge, this study introduces the mixture of experts infrared small target detection (MOE-IR) method. The core idea of this method is to construct a mixture of experts feature extraction network to perform rich feature extraction and complex background suppression on small targets, respectively, so as to achieve robust infrared small target detection in complex backgrounds. Specifically, the MOE-IR comprises a target feature extraction expert and a background suppression expert. The target feature extraction expert focuses on enhancing the features of infrared small targets, while the background suppression expert aims to mitigate background clutter. Additionally, an adaptive gate controlled network is incorporated to dynamically assign weights to the two experts based on the input infrared image, ensuring effective detection of infrared small targets across diverse and complex scenarios. Extensive experiments demonstrate that the proposed algorithm surpasses existing infrared small target detection methods in terms of detection accuracy and false alarm rate. It can reliably and stably identify small targets within infrared images, thus offering an effective solution for practical infrared small target detection\xa0applications.',
  'Zhengkui Weng, Xinjie Fu, Xu Zhang, Siyuan Sun',
  '8',
  '2025',
  '7'],
 ['Global Context Attention in YOLO for Broadband Signal Identification',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70358?af=R',
  'Accurate detection of broadband communication signals is critical for information restoration in complex electromagnetic environments, yet traditional methods struggle with modern demands. This study proposes NLYOLO, a novel detection model that optimises anchor configurations for signal shapes, integrates non-local and SimAM attention modules, and improves the loss function to enhance small-target feature

extraction in YOLOv11s. Experimental results demonstrate significant improvements in precision and recall over conventional energy detection and baseline models, effectively addressing signal detection challenges in intricate scenarios.\n\n\n\n\n\n\n\n\nABSTRACT\nDetecting broadband communication signals is crucial for signal processing, as its accuracy directly impacts subsequent processing such as information restoration. However, due to the ever-growing complexity of the electromagnetic environment, traditional broadband detection algorithms are no longer suitable for current application needs. This paper proposes a broadband signal detection model, NLYOLO. We improve the precision of the image by analysing the characteristics of the target signal by selecting suitable parameters. We also select anchors adapted to the special shape of the target signals and introduce attention modules Non-local and SimAM in the feature extraction of NLYOLO and improve the loss function. This enhances the ability of high-level semantic information feature extraction and improves the precision rate and recall of small target signals. The experimental results show that compared with the traditional energy detection algorithm and the baseline model, the detection precision and recall rate are significantly improved, which effectively achieves the target detection task in the field of signal processing in complex scenes.',
  'Xiaoya Wang, Songlin Sun, Haiying Zhang',
  '8',
  '2025',
  '7'],
 ['UAV-Based Real-Time Object Detection Network Using Structured Pruning Strategy',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70206?af=R',
  "We propose an effective aerial object detection algorithm, AIR-YOLO, which achieves the balance between accuracy and computational cost. We propose a structured pruning strategy to remove the redundant parts in AIR-YOLO, significantly reducing the model's complexity. Compared with the model before pruning, the pruned model's accuracy decreases slightly. Experiments show that our model is superior to state-of-the-art works in both model complexity and detection accuracy.\n\n\n\n\n\n\n\n\nABSTRACT\nReal-time object detection networks based on UAV have been used in various fields. However, some challenges need to be solved: (1) Conventional detection algorithms are not suitable for small targets; (2) The computational capacity of the UAV platform is limited; (3) The sample distribution in the aerial dataset shows the characteristics of long-tail distribution. Categories at the tail end often need to be better learned. To address these challenges, we propose the AIR-YOLO-pruned method, a lightweight UAV-based object detection method built on the YOLOv8. In this paper, we propose the AIR-YOLO which is suitable for small object detection. We introduce the gradient adaptive allocation loss to enhance the model's learning ability for tail categories. To eliminate redundant components in AIR-YOLO, we design a kind of structured pruning strategy. Experiment results indicate that our AIR-YOLOn-pruned method, with competitive computational cost, achieves a 17% improvement in accuracy compared to YOLOv8n.",
  'Donghui Zhao, Bo Mo',
  '8',
  '2025',
  '3'],
 ['Small target detection with decoupling and decorrelation of confusion features',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70158?af=R',
  'Abstract\nSmall target detection has important application value in military, cigarette defect detection, monitoring and remote sensing fields. However, due to the complex background interference of infrared images, the background and target features are confused with each other, which makes it difficult for existing detection methods to effectively distinguish the features of the background and the target, resulting in poor detection effect. Therefore, to solve this problem, this paper proposes a small target detection method based on confusion feature decoupling and decorrelation, aiming to achieve accurate detection of small targets in complex environments by extracting robust small target features. Specifically, in confusion feature decoupling and decorrelation, we propose a confusion feature decoupling and decorrelation module. By introducing a feature decoupling mechanism, the features of the input image are decomposed into independent background features and target features, and feature decorrelation is used to achieve independence between the target and the background. The target features after decoupling and decorrelation are purer, which helps to reduce background interference and thus improve detection performance. Systematic experimental results show that the detection performance of the proposed method on public small target detection datasets is much better than that of existing advanced detection\xa0methods.',
  'Yonghua Zhang, Hui Wang, He Tang, Xingze Liu, Benxue Liu, Siyuan Sun',
  '8',
  '2025',
  '2']]

[['Object Detection Based on CNN and Vision-Transformer: A Survey',

'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cvi2.70028?af=R',
'Analyse the most significant advances in CNN-based and transformer-based object detectors. The development of object detectors is used as a timeline to summarise important techniques such as loss function and multi-scale features. Suggest recent research developments on difficult tasks such as small object detection, object detection with few shots and accelerated detection.\n\n\n\n\n\n\nABSTRACT\nObject detection is the most crucial and challenging task of computer vision and has been used in various fields in recent years, such as autonomous driving and industrial inspection. Traditional object detection methods are mainly based on the sliding windows and the handcrafted features, which have problems such as insufficient understanding of image features and low accuracy of detection. With the rapid advancements in deep learning, convolutional neural networks (CNNs) and vision transformers have become fundamental components in object detection models. These components are capable of learning more advanced and deeper image properties, leading to a transformational breakthrough in the performance of object detection. In this review, we comprehensively review the representative object detection models from deep learning periods, tracing their architectural shifts and technological breakthroughs. Furthermore, we discuss key challenges and promising research directions in the object detection. This review aims to provide a comprehensive foundation for practitioners to enhance their understanding of object detection technologies.',
'Jinfeng Cao, Bo Peng, Mingzhong Gao, Haichun Hao, Xinfang Li, Hongwei Mou',
'6',
'2025',
'5'],
['MCD-YOLOv10n: A Small Object Detection Algorithm for UAVs',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70145?af=R',
"In this paper, an improved algorithm, MCD-YOLOv10n, is proposed to address the small object detection challenge in UAV data acquisition. The MEMAttention module combines EMAttention and multi-scale convolution to capture image features at different scales.C2f combines with MEMAttention and DSConv to enhance the flexibility of the network. The algorithm outperforms models such as YOLOv10n on three datasets, VisDrone-DET2019, Exdark, and DOTA-v1.5, providing an efficient solution for small object detection.\n\n\n\n\n\n\n\nABSTRACT\nDeep neural networks deployed on UAVs have made significant progress in data acquisition in recent years. However, traditional algorithms and deep learning models still face challenges in small and unevenly distributed object detection tasks. To address this problem, we propose the MCD-YOLOv10n model by introducing the MEMAttention module, which combines EMAttention with multiscale convolution, uses Softmax and AdaptiveAvgPool2d to adaptively compute feature weights, dynamically adjusts the region of interest, and captures cross-scale features. In addition, the C2f_MEMAttention and C2f_DSConv modules are formed by the fusion of C2f with MEMAttention and DSConv, which enhances the model's ability of extracting and adapting to irregular target features. Experiments on three datasets, VisDrone-DET2019, Exdark and DOTA-v1.5, show that the evaluation metric mAP50 achieves the best detection accuracy of 32.9%, 52.9% and 68.2% when the number of holdout parameters is at the minimum value of 2.24M. Moreover, the mAP50-95 metrics (19.5% for VisDrone-DET2019 and 45.0% for DOTA-v1.5) are 1.1 and 1.2 percentage points ahead of the second place, respectively. In terms of Recall, the VisDrone-DET2019 and DOTA-v1.5 datasets improved by 1.0% and 0.7% over the baseline model. These results validate that MCD-YOLOv10n has strong adaptability and generalization ability for small object detection in complex\xa0scenes.",
'Jinshuo Shi, Xitai Na, Shiji Hai, Qingbin Sun, Zhihui Feng, Xinyang Zhu',
'7',
'2025',
'6'],
['Scale-Wise Interaction Fusion Network for Land Cover Classification of Urban Scene Imagery',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70139?af=R',
'This study proposes the scale-wise interaction fusion network (SIFN) to improve land cover classification in urban aerial imagery, addressing challenges related to small objects and similar features. The SIFN utilises multi-scale feature extraction, scale-wise interaction, feature shuffle-fusion and adaptive mask generation to enhance classification accuracy, particularly in delineating boundaries and reducing misclassification in transitional zones. Experimental results demonstrate significant improvements in feature representation and overall classification performance.\n\n\n\n\n\n\n\nABSTRACT\nAccurate land cover classification of urban aerial imagery presents significant challenges, particularly in recognising small objects and similar-appearing features (e.g., flat land, prepared land for cultivation, crop growing areas and built-up regions along with ground water resource areas). These challenges arise due to the irregular scaling of extracted features at various rates from complex urban scenes and the mismatch in feature information flow across channels, ultimately affecting the overall accuracy (OA) of the network. To address these issues, we propose the scale-wise interaction fusion network (SIFN) for land cover classification of urban scene imagery. The SIFN comprises four key modules: multi-scale feature extraction, scale-wise interaction, feature shuffle-fusion and adaptive mask generation. The multi-scale feature extraction module captures contextual information across different dilation rates of convolutional layers, effectively handling varying object sizes. The scale-wise

interaction module enhances the learning of multi-scale contextual features, while the feature shuffle-fusion module facilitates cross-scale information exchange, improving feature representation. Lastly, adaptive mask generation ensures precise boundary delineation and reduces misclassification in transitional zones. The proposed network significantly improves boundary masking accuracy for small and similar objects, thereby enhancing the overall land cover classification performance.',
  'Muhammad Shafiq, Waeal J. Obidallah, Quanrun Fan, Anas Bilal, Yousef A. Alduraywish',
  '7',
  '2025',
  '6'],
 ['Enhanced Small-Object Detection in UAV Images Using Modified YOLOv5 Model',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70121?af=R',
  'ABSTRACT\nThis study presents a modified YOLOv5 algorithm specifically designed to enhance small-object detection in unmanned aerial vehicle (UAV) images. Traditional object detection in UAV images is particularly challenging due to the high altitude of the cameras, which results in small object sizes and varying viewing angles. To address these challenges, the algorithm incorporates an additional prediction head to detect objects across a wide range of scales, a channel feature fusion with involution (CFFI) block to minimize information loss, a convolutional block attention module (CBAM) to highlight the crucial spatial and channel features, and a C3 structure with a Transformer block (C3TR) to capture contextual information. The algorithm additionally employs soft non-maximum suppression to enhance the bounding box scoring of overlapping objects in dense scenes. Extensive experiments were conducted on the VisDrone-DET2019 dataset, which demonstrated the effectiveness of the proposed algorithm. The results showed improvements with precision scores of 55.0%, recall scores of 44.6%, mean average precision scores of mAP50 = 50.9% and mAP50:95 = 33.0% on the VisDrone-DET2019 validation set, and precision of 50.8%, recall of 37.3%, mAP50 = 44.2%, and mAP50:95 = 27.3% on the VisDrone-DET2019 testing set. The improved performance is due to the incorporation of attention mechanisms, which allow the proposed model to stay lightweight while still extracting the features needed to detect small objects.',
  'Bach-Thanh Lieu, Chi-Khang Nguyen, Huynh-Lam Nguyen, Thanh-Hai Le',
  '7',
  '2025',
  '6'],
 ['ACDR-CRAFF Net: A Multi-Scale Network Based on Adaptive Channel and Coordinate Relational Attention Network for Remote Sensing Scene Classification',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70112?af=R',
  "The adaptive channel dimensionality reduction (ACDR) module dynamically adjusts channel representations to retain key information and reduce redundancy, improving the efficiency of feature fusion. Meanwhile, the coordinate relational attention multi-scale feature fusion (CRAFF) module utilizes a coordinate relation attention mechanism to capture spatial dependencies between different feature levels, optimizing the multi-scale feature fusion process. The collaborative interaction of these two modules facilitates the gradual fusion of local to global features, enhancing the model's robustness and generalization ability.\n\n\n\n\n\n\n\n\nABSTRACT\nAccurate classification of remote sensing scene images is crucial for diverse applications, from environmental monitoring to urban planning. While convolutional neural networks (CNNs) have dramatically improved classification accuracy, challenges remain due to the complex distribution of small objects, varied spatial configurations, and intra-class multimodality in remote sensing images. In this work, we make three key contributions to address these challenges. (1) We propose the adaptive channel and coordinate relational attention network (ACDR-CRAFF), a novel multi-scale feature fusion framework designed to enhance feature representation across scales. (2) We introduce two innovative modules: the adaptive channel dimensionality reduction (ACDR) module, which dynamically adjusts channel representations to retain essential low-dimensional features, and the coordinate relational attention multi-scale feature fusion (CRAFF) module, which effectively captures and transfers spatial information between feature levels. (3) By integrating ACDR and CRAFF, our model achieves a progressive fusion of local to global features, ensuring robust feature expressiveness at multiple scales. Experimental results on four widely used benchmark datasets demonstrate that ACDR-CRAFF consistently outperforms several state-of-the-art methods, achieving significant improvements in classification accuracy and setting a new benchmark for complex remote sensing scene classification tasks. These results underscore the effectiveness of our approach in addressing the limitations of existing methods and advancing the state of the art in remote sensing image analysis.",
  'Wei Dai, Haixia Xu, Furong Shi, Liming Yuan, Xinyu Wang, Xianbin Wen',
  '7',
  '2025',
  '5'],
 ['AMFE-YOLO: A Small Object Detection Model for Drone Images',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70110?af=R',
  'To address the issue of poor detection performance in drone aerial images, we innovatively propose AMFE-

YOLO. Solutions are introduced for challenges such as occlusion and targets under varying lighting conditions. In the end, our model shows significant improvement compared to current popular algorithms. \n\n\n\n\n\n\n\nABSTRACT\nDrones, due to their high efficiency and flexibility, have been widely applied. However, small objects captured by drones are easily affected by various conditions, resulting in suboptimal surveying performance. While the YOLO series has achieved significant success in detecting large targets, it still faces challenges in small target detection. To address this, we propose an innovative model, AMFE-YOLO, aimed at overcoming the bottlenecks in small target detection. Firstly, we introduce the AMFE module to focus on occluded targets, thereby improving detection capabilities in complex environments. Secondly, we design the SFSM module to merge shallow spatial information from the input features with deep semantic information obtained from the neck, enhancing the representation ability of small target features and reducing noise. Additionally, we implement a novel detection strategy that introduces an auxiliary detection head to identify very small targets. Finally, we reconfigured the detection head, effectively addressing the issue of false positives in small-object detection and improving the precision of small object detection. AMFE-YOLO outperforms methods like YOLOv10 and YOLOv11 in terms of mAP on the VisDrone2019 public dataset. Compared to the original YOLOv8s, the average precision improved by 5.5%, while the model parameter size was reduced by 0.7\xa0M.',
  'Qi Wang, Chengxin Yu',
  '7',
  '2025',
  '5'],
 ['MFF-YOLOv8: Small Object Detection Based on Multi-Scale Feature Fusion for UAV Remote Sensing Images',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70066?af=R',
  "In this work, we propose a novel network for small target detection in UAV aerial photography scenarios to address the challenges often encountered in UAV aerial photography scenarios where small target features are difficult to extract and are susceptible to reception light, scale variations and occlusion. First, we design a pyramid network with high-resolution feature fusion, which reconstructs the high-resolution feature map for feature selection, allowing the network to fully exploit the high-resolution feature information after removing a large amount of noise. Second, we design a multi-scale hybrid attention mechanism that maximizes the retention and extraction of small target feature information by fully capturing multi-scale features in the channel and space. Finally, a regression loss function that combines the joint auxiliary bounding box and the dynamic focusing bounding box is designed to improve the processing capability for a large number of low IoU samples in UAV aerial photography scenes. Based on the experiments, we compare the proposed model with other state-of-the-art methods on the VisDrone2019 dataset. The experimental results show that the proposed model has higher detection accuracy and excellent robustness.\n\n\n\n\n\n\n\nABSTRACT\nAs a popular task in drone-captured scenes, object detection involves images with a large number of small objects, but current networks often suffer missed and false detections. To address this problem, we propose a YOLO algorithm MFF-YOLOv8 based on multi-scale feature fusion for small target detection in UAV aerial images. First, a high-fesolution feature fusion pyramid (HFFP) is designed, which utilizes high-resolution feature maps containing much information about small objects to guide the feature fusion module, weighting and fusing feature maps to enhance the network's ability to represent small targets. Meanwhile, a reconstruction feature selection (RFS) module is employed to remove the large amounts of noise produced by high-resolution feature maps. Second, a hybrid efficient multi-scale attention (HEMA) mechanism is designed in the backbone network to maximize the retention and extraction of feature information related to small objects while simultaneously suppressing background noise interference. Finally, an Inner-Wise IoU loss function (Inner-WIoU) is designed for joint auxiliary bounding box and dynamic focal bounding box regression, which enhances the accuracy of network regression results, thus improving the detection precision of the model for small objects. MFF-YOLOv8 was experimented on the VisDrone2019 dataset, achieving a 47.9% mAP50, 9.3% up compared with that of the baseline network YOLOv8s. Also, in order to verify the generalization of the overall network, it was evaluated on the DOTA and UAVDT datasets, and the mAP50 was improved by 3.7% and 1.8%, respectively. The results demonstrate that MFF-YOLOv8 significantly enhances detection precision for small objects in UAV aerial\xa0scenes.",
  'Kun Hu, Jinzheng Lu, Chaoquan Zheng, Qiang Xiang, Ling Miao',
  '7',
  '2025',
  '4'],
 ['LCM-YOLO: A Small Object Detection Method for UAV Imagery Based on YOLOv5',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70051?af=R',
  "The LCM-YOLO model, an enhancement of the YOLOv5, improves small target detection and adaptability to scale variations in UAV aerial images by incorporating a local fusion mechanism, cross-scale feature fusion and multi-head attention, resulting in significant performance improvements over the original YOLOv5 on the VisDrone2019 dataset. Specifically, LCM-YOLO achieves a 7.2% increase in mAP50 and a 5.1% increase in mAP50-95, with final scores of 40.7% and 22.5%, respectively.

\n\n\n\n\n\n\n\nABSTRACT\nThis study addresses the challenges of detecting small targets and targets with significant scale variations in UAV aerial images. We propose an improved YOLOv5 model, named LCM-YOLO, to tackle these challenges. Initially, a local fusion mechanism is introduced into the C3 module, forming the C3-LFM module to enhance feature information acquisition during feature extraction. Subsequently, the CCFM is employed as the neck structure of the network, leveraging its lightweight convolution and cross-scale feature fusion characteristics to effectively improve the model's ability to integrate target features at different levels, thereby enhancing its adaptability to scale variations and detection performance for small targets. Additionally, a multi-head attention mechanism is integrated at the front end of the detection head, allowing the model to focus more on the detailed information of small targets through weight distribution. Experiments on the VisDrone2019 dataset show that LCM-YOLO has excellent detection capabilities. Compared to the original YOLOv5 model, its mAP50 and mAP50-95 metrics are improved by 7.2% and 5.1%, respectively, reaching 40.7% and 22.5%. This validates the effectiveness of the LCM-YOLO model for detecting small and multi-scale targets in complex backgrounds.",
  'Shaodong Liu, Faming Shao, Weijun Chu, Heng Zhang, Dewei Zhao, Jinhong Xue, Qing Liu',
  '7',
  '2025',
  '3'],
 ['Multi-Shadow Scenarios Tennis Ball Detection by an Improved RTMdet-Light Model',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70054?af=R',
  'Our paper considers a lightweight object detection model named improved RTMDet-light. Specially, it has compatible capacities in the backbone and neck, constructed by a basic building block that consists of large-kernel depth-wise convolutions. Furthermore, soft labels when calculating matching costs in the dynamic label assignment was introduced to improve accuracy. \n\n\n\n\n\n\n\nABSTRACT\nThe real-time and rapid recording of sport sensor data related to tennis ball trajectories facilitates the analysis of this information and the development of intelligent training regimes. However, there are three essential challenges in the task of tennis ball recognition using sport vision sensors: the small size of the ball, its high speed, and the complex match scenarios. As a result, this paper considers a lightweight object detection model named improved RTMDet-light to deal with these challenges. Specifically, it has compatible capacities in the backbone and neck, constructed by a basic building block that consists of large-kernel depth-wise convolutions. Furthermore, GhosNet and ShuffleNet are used to replace the CSPLayers which reduce the parameters of our model. The lightweight model proposed addresses the inherent challenges of detecting small objects and muti scenarios in the match. After training, the proposed model performed better on four scenarios with different shades of tennis ball match, with results visualized through heatmaps and performance metrics tabulated for detailed analysis. The recall, FLOPs and number of parameters of the improved RTMDet-light are 71.4%, 12.543G, and 4.874M, respectively. The results demonstrate robustness and effectiveness of our model in accurate tennis ball detecting across various scales. In conclusion, our model for real-time detection in tennis ball detection offers a lightweight and faster solution for sport sensors.',
  'Yukun Zhu, Yanxia Peng, Cong Yu',
  '7',
  '2025',
  '3'],
 ['SF-YOLO: A Novel YOLO Framework for Small Object Detection in Aerial Scenes',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70027?af=R',
  'The authors propose a new supervised contrastive learning framework (FSSCL-OSC) that can achieve open-set classification of hyperspectral images in scenarios with very few sample sizes. Experimental results on three classical HSI datasets show that FSSCL-OSC provides a significant improvement over existing methods, under a sample size of only 10%, the overall accuracy reached 82.38% and 90.76% and 84.70%, respectively. \n\n\n\n\n\n\n\nABSTRACT\nObject detection models are widely applied in the fields such as video surveillance and unmanned aerial vehicles to enable the identification and monitoring of various objects on a diversity of backgrounds. The general CNN-based object detectors primarily rely on downsampling and pooling operations, often struggling with small objects that have low resolution and failing to fully leverage contextual information that can differentiate objects from complex background. To address the problems, we propose a novel YOLO framework called SF-YOLO for small object detection. Firstly, we present a spatial information perception (SIP) module to extract contextual features for different objects through the integration of space to depth operation and large selective kernel module, which dynamically adjusts receptive field of the backbone and obtains the enhanced features for richer understanding of differentiation between objects and background. Furthermore, we design a novel multi-scale feature weighted fusion strategy, which performs weighted fusion on feature maps by combining fast normalized fusion method and CARAFE operation, accurately assessing the importance of each feature and enhancing the representation of small objects. The extensive experiments conducted on VisDrone2019, Tiny-Person and PESMOD datasets demonstrate that our proposed method enables comparable detection performance to state-of-the-art detectors.',
  'Meng Sun, Le Wang, Wangyu Jiang, Fayaz Ali Dharejo, Guojun Mao, Radu Timofte',

'7',
'2025',
'3'],
['Modified You Only Look Once Network Model for Enhanced Traffic Scene Detection Performance for Small Targets',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70014?af=R',
"We propose a model based on you only look once version 8X (Yolov8X) network model, which has been combined with receptive fields block (RFB) and multidimensional collaborative attention (MCA). The validation results on the Argoverse 1.1 autonomous driving dataset demonstrate that the enhanced network model outperforms the prevailing detectors, achieving an F1 score of 78.6, an average precision of 55.1, and an average recall of 72.4.\n\n\n\n\n\n\nABSTRACT\nIn order to address the challenge of small target recognition in traffic scenes, we propose a model based on you only look once version 8X (Yolov8X) network model, which has been combined with receptive fields block (RFB) and multidimensional collaborative attention (MCA). First, the model employs the RFB to extract reliable and distinctive features, thereby enhancing the precision of small target identification. Furthermore, the MCA structure is introduced to simulate multidimensional attention through three parallel branches, thereby enhancing the feature expression ability of the model. This fragment describes a compression transformation and an excitation transformation that captures the differentiated feature representation of the command. These transformations facilitate the network's ability to locate and predict the location of small objects more accurately. Utilizing these transformations enhances the expressiveness and diversity of features, thereby improving the detection performance of small objects. Furthermore, data augmentation and hyperparameter optimization techniques are employed to enhance the model's generalisability. The validation results on the Argoverse 1.1 autonomous driving dataset demonstrate that the enhanced network model outperforms the prevailing detectors, achieving an F1 score of 78.6, an average precision of 55.1, and an average recall of 72.4. The algorithm's excellent performance for small target detection was demonstrated through visual analysis, proving its high application value and potential for promotion in fields such as autonomous\xa0driving.",
'Lei Shi, Shuai Ren, Xing Fan, Ke Wang, Shan Lin, Zhanwen Liu',
'7',
'2025',
'2'],
['An intelligent retrievable object-tracking system with real-time edge inference capability',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13297?af=R',
"An intelligent retrievable object-tracking system assists users in quickly and accurately locating lost objects. However, challenges such as real-time processing on edge devices, low image resolution, and small-object detection significantly impact the accuracy and efficiency of video-stream-based systems, especially in indoor home environments. To overcome these limitations, a novel real-time intelligent retrievable object-tracking system is designed. The system incorporates a retrievable object-tracking algorithm that combines DeepSORT and sliding window techniques to enhance tracking capabilities. Additionally, the YOLOv7-small-scale model is proposed for small-object detection, integrating a specialized detection layer and the convolutional batch normalization LeakyReLU spatial-depth convolution module to enhance feature capture for small objects. TensorRT and INT8 quantization are used for inference acceleration on edge devices, doubling the frames per second. Experiments on a Jetson Nano (4 GB) using YOLOv7-small-scale show an 8.9% improvement in recognition accuracy over YOLOv7-tiny in video stream processing. This advancement significantly boosts the system's performance in efficiently and accurately locating lost objects in indoor home\xa0settings.",
'Yujie Li, Yifu Wang, Zihang Ma, Xinghe Wang, Benying Tan, Shuxue Ding',
'7',
'2024',
'12'],
['Channel selection and local attention transformer model for semantic segmentation on UAV remote sensing scene',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13298?af=R',
'Compared with common urban landscape semantic segmentation, unmanned aerial vehicle (UAV) image semantic segmentation is more challenging because small targets have very low pixel percentages and multi-scale features due to the influence of flight altitude. Yet, the commonly used successive grid downsampling strategy in the current transformer-based methods omits some important features of small targets. Furthermore, due to the complex background interference, it can lead to even worse results. In reaction to this, existing strategies aim to maintain superior resolution. Nevertheless, the application of this method incurs considerable computational costs, which brings challenges for the practical applications of UAVs. So it is significant to design a novel framework to balance retaining more pixels representing small objects during downsampling and reducing computational costs. For this, the Channel Selection and the Local Attention Transformer Model (CSLFormer) are proposed. During the overlap patch embedding process of feature maps, the model allocates half of the important channels to global attention and local attention.

These two types of attention focus on different aspects: one learns the relationships and importance among various patches, while the other emphasizes the features of individual patches. The method shows superior performance on two public datasets: AeroScapes and Vaihingen, achieving mean intersection over union (mIoU) of 75.57% and 78.93%, respectively. The proposed CSLFormer has been released on GitHub: https://github.com/leoda1/CSLFormer.',
  'Da Liu, Hao Long, Zhenbao Liu',
  '7',
  '2024',
  '12'],
 ['YOLO-Tiny: A lightweight small object detection algorithm for UAV aerial imagery',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13314?af=R',
  "In unmanned aerial vehicle (UAV) aerial target detection tasks, two main challenges exist: the limited computational resources of UAV terminals, which are not conducive to running complex models, and the prevalence of small targets, which can easily lead to missed detections and false positives. To address these issues, this study proposes a lightweight and high-accuracy small-object detection algorithm for UAV aerial imagery that is based on YOLOv5s. First, the network structure is optimized by removing the layers in YOLOv5s primarily used for detecting large targets (P4 and P5) and adding layers primarily used for detecting small targets (P2 and P3). This enables the model to focus more on extracting small-object features. A lightweight dynamic convolution is subsequently introduced in the C3 module, and the lightweight LW_C3 and LW_downsampling modules are designed for feature extraction and downsampling operations. This enhances the model's feature extraction capability while achieving a lightweight design. Finally, the adaptive multi-scale spatial feature fusion (AMSFF) module is designed to adaptively learn the spatial weights of the feature maps at different levels, thereby further strengthening the effective fusion of multi-scale features. Experimental results show that the improved YOLO-Tiny model has higher accuracy and lower complexity, hence validating its excellent performance.",
  'Fei Feng, Lu Yang, Quanxing Zhou, Weipeng Li',
  '7',
  '2024',
  '12'],
 ['UAV-Based Real-Time Object Detection Network Using Structured Pruning Strategy',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70206?af=R',
  "We propose an effective aerial object detection algorithm, AIR-YOLO, which achieves the balance between accuracy and computational cost. We propose a structured pruning strategy to remove the redundant parts in AIR-YOLO, significantly reducing the model's complexity. Compared with the model before pruning, the pruned model's accuracy decreases slightly. Experiments show that our model is superior to state-of-the-art works in both model complexity and detection accuracy.\n\n\n\n\n\n\n\n\nABSTRACT\nReal-time object detection networks based on UAV have been used in various fields. However, some challenges need to be solved: (1) Conventional detection algorithms are not suitable for small targets; (2) The computational capacity of the UAV platform is limited; (3) The sample distribution in the aerial dataset shows the characteristics of long-tail distribution. Categories at the tail end often need to be better learned. To address these challenges, we propose the AIR-YOLO-pruned method, a lightweight UAV-based object detection method built on the YOLOv8. In this paper, we propose the AIR-YOLO which is suitable for small object detection. We introduce the gradient adaptive allocation loss to enhance the model's learning ability for tail categories. To eliminate redundant components in AIR-YOLO, we design a kind of structured pruning strategy. Experiment results indicate that our AIR-YOLOn-pruned method, with competitive computational cost, achieves a 17% improvement in accuracy compared to YOLOv8n.",
  'Donghui Zhao, Bo Mo',
  '8',
  '2025',
  '3']]

SEARCH WORD: TINY

[['Dual-Level Cross-Modality Neural Architecture Search for Guided Image Super-Resolution',
  'http://ieeexplore.ieee.org/document/11029618',
  'Guided image super-resolution (GISR) aims to reconstruct a high-resolution (HR) target image from its low-resolution (LR) counterpart with the guidance of a HR image from another modality. Existing learning-based methods typically employ symmetric two-stream networks to extract features from both the guidance and target images, and then fuse these features at either an early or late stage through manually designed modules to facilitate joint inference. Despite significant performance, these methods still face several issues: i) the symmetric architectures treat images from different modalities equally, which may overlook the inherent differences between them; ii) lower-level features contain detailed information while higher-level features

capture semantic structures. However, determining which layers should be fused and which fusion operations should be selected remain unresolved; iii) most methods achieve performance gains at the cost of increased computational complexity, so balancing the trade-off between computational complexity and model performance remains a critical issue. To address these issues, we propose a Dual-level Cross-modality Neural Architecture Search (DCNAS) framework to automatically design efficient GISR models. Specifically, we propose a dual-level search space that enables the NAS algorithm to identify effective architectures and optimal fusion strategies. Moreover, we propose a supernet training strategy that employs a pairwise ranking loss trained performance predictor to guide the supernet training process. To the best of our knowledge, this is the first attempt to introduce the NAS algorithm into GISR tasks. Extensive experiments demonstrate that the discovered model family, DCNAS-Tiny and DCNAS, achieve significant improvements on several GISR tasks, including guided depth map super-resolution, guided saliency map super-resolution, guided thermal image super-resolution, and pan-sharpening. Furthermore, we analyze the architectures searched by our method and provide some new insights for future research.',
  'Zhiwei Zhong, Xianming Liu, Junjun Jiang, Debin Zhao, Shiqi Wang',
  '3',
  '2025',
  '8'],
 ['SF-YOLO: A Novel YOLO Framework for Small Object Detection in Aerial Scenes',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70027?af=R',
  'The authors propose a new supervised contrastive learning framework (FSSCL-OSC) that can achieve open-set classification of hyperspectral images in scenarios with very few sample sizes. Experimental results on three classical HSI datasets show that FSSCL-OSC provides a significant improvement over existing methods, under a sample size of only 10%, the overall accuracy reached 82.38% and 90.76% and 84.70%, respectively. \n\n\n\n\n\n\n\nABSTRACT\nObject detection models are widely applied in the fields such as video surveillance and unmanned aerial vehicles to enable the identification and monitoring of various objects on a diversity of backgrounds. The general CNN-based object detectors primarily rely on downsampling and pooling operations, often struggling with small objects that have low resolution and failing to fully leverage contextual information that can differentiate objects from complex background. To address the problems, we propose a novel YOLO framework called SF-YOLO for small object detection. Firstly, we present a spatial information perception (SIP) module to extract contextual features for different objects through the integration of space to depth operation and large selective kernel module, which dynamically adjusts receptive field of the backbone and obtains the enhanced features for richer understanding of differentiation between objects and background. Furthermore, we design a novel multi-scale feature weighted fusion strategy, which performs weighted fusion on feature maps by combining fast normalized fusion method and CARAFE operation, accurately assessing the importance of each feature and enhancing the representation of small objects. The extensive experiments conducted on VisDrone2019, Tiny-Person and PESMOD datasets demonstrate that our proposed method enables comparable detection performance to state-of-the-art detectors.',
  'Meng Sun, Le Wang, Wangyu Jiang, Fayaz Ali Dharejo, Guojun Mao, Radu Timofte',
  '7',
  '2025',
  '3'],
 ['An intelligent retrievable object-tracking system with real-time edge inference capability',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13297?af=R',
  "An intelligent retrievable object-tracking system assists users in quickly and accurately locating lost objects. However, challenges such as real-time processing on edge devices, low image resolution, and small-object detection significantly impact the accuracy and efficiency of video-stream-based systems, especially in indoor home environments. To overcome these limitations, a novel real-time intelligent retrievable object-tracking system is designed. The system incorporates a retrievable object-tracking algorithm that combines DeepSORT and sliding window techniques to enhance tracking capabilities. Additionally, the YOLOv7-small-scale model is proposed for small-object detection, integrating a specialized detection layer and the convolutional batch normalization LeakyReLU spatial-depth convolution module to enhance feature capture for small objects. TensorRT and INT8 quantization are used for inference acceleration on edge devices, doubling the frames per second. Experiments on a Jetson Nano (4 GB) using YOLOv7-small-scale show an 8.9% improvement in recognition accuracy over YOLOv7-tiny in video stream processing. This advancement significantly boosts the system's performance in efficiently and accurately locating lost objects in indoor home\xa0settings.",
  'Yujie Li, Yifu Wang, Zihang Ma, Xinghe Wang, Benying Tan, Shuxue Ding',
  '7',
  '2024',
  '12'],
 ['YOLO-Tiny: A lightweight small object detection algorithm for UAV aerial imagery',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13314?af=R',

"In unmanned aerial vehicle (UAV) aerial target detection tasks, two main challenges exist: the limited computational resources of UAV terminals, which are not conducive to running complex models, and the prevalence of small targets, which can easily lead to missed detections and false positives. To address these issues, this study proposes a lightweight and high-accuracy small-object detection algorithm for UAV aerial imagery that is based on YOLOv5s. First, the network structure is optimized by removing the layers in YOLOv5s primarily used for detecting large targets (P4 and P5) and adding layers primarily used for detecting small targets (P2 and P3). This enables the model to focus more on extracting small-object features. A lightweight dynamic convolution is subsequently introduced in the C3 module, and the lightweight LW_C3 and LW_downsampling modules are designed for feature extraction and downsampling operations. This enhances the model's feature extraction capability while achieving a lightweight design. Finally, the adaptive multi-scale spatial feature fusion (AMSFF) module is designed to adaptively learn the spatial weights of the feature maps at different levels, thereby further strengthening the effective fusion of multi-scale features. Experimental results show that the improved YOLO-Tiny model has higher accuracy and lower complexity, hence validating its excellent performance.",
'Fei Feng, Lu Yang, Quanxing Zhou, Weipeng Li',
'7',
'2024',
'12']]

SEARCH WORD: NAVIGATION

[['Toward Unified 3D Object Detection via Algorithm and Data Unification',
'http://ieeexplore.ieee.org/document/11016195',
'Realizing unified 3D object detection, including both indoor and outdoor scenes, holds great importance in applications like robot navigation. However, involving various scenarios of data to train models poses challenges due to their significantly distinct characteristics, e.g., diverse geometry properties and heterogeneous domain distributions. In this work, we propose to address the challenges from two perspectives, the algorithm perspective and data perspective. In terms of the algorithm perspective, we first build a monocular 3D object detector based on the bird’s-eye-view (BEV) detection paradigm, where the explicit feature projection is beneficial to addressing the geometry learning ambiguity. In this detector, we split the classical BEV detection architecture into two stages and propose an uneven BEV grid design to handle the convergence instability caused by geometry difference between scenarios. Besides, we develop a sparse BEV feature projection strategy to reduce the computational cost and a unified domain alignment method to handle heterogeneous domains. From the data perspective, we propose to incorporate depth information to improve training robustness. Specifically, we build the first unified multi-modal 3D object detection benchmark MM-Omni3D and extend the aforementioned monocular detector to its multi-modal version, which is the first unified multi-modal 3D object detector. We name the designed monocular and multi-modal detectors as UniMODE and MM-UniMODE, respectively. The experimental results reveal several insightful findings highlighting the benefits of multi-modal data and confirm the effectiveness of all the proposed strategies.',
'Zhuoling Li, Xiaogang Xu, Ser-Nam Lim, Hengshuang Zhao',
'3',
'2025',
'8'],
['Inverse RL Scene Dynamics Learning for Nonlinear Predictive Control in Autonomous Vehicles',
'http://ieeexplore.ieee.org/document/10944500',
'This article introduces the deep learning-based nonlinear model predictive controller with scene dynamics (DL-NMPC-SD) method for autonomous navigation. DL-NMPC-SD uses an a priori nominal vehicle model in combination with a scene dynamics model learned from temporal range sensing information. The scene dynamics model is responsible for estimating the desired vehicle trajectory, as well as to adjust the true system model used by the underlying model predictive controller. We propose to encode the scene dynamics model within the layers of a deep neural network, which acts as a nonlinear approximator for the high-order state space of the operating conditions. The model is learned based on temporal sequences of range-sensing observations and system states, both integrated by an Augmented Memory component. We use inverse reinforcement learning (IRL) and the Bellman optimality principle to train our learning controller with a modified version of the deep Q-learning (DQL) algorithm, enabling us to estimate the desired state trajectory as an optimal action-value function. We have evaluated DL-NMPC-SD against the baseline dynamic window approach (DWA), as well as against two state-of-the-art End2End and RL methods, respectively. The performance has been measured in three experiments: 1) in our GridSim virtual environment; 2) on indoor and outdoor navigation tasks using our RovisLab autonomous mobile test unit (AMTU) platform; and 3) on a full-scale autonomous test vehicle driving on public roads.',

'Sorin M. Grigorescu, Mihai V. Zaha',
'5',
'2025',
'8'],
['Point Cloud Registration Based on Multiple Neighborhood Feature Difference',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70097?af=R',
'This paper proposes a point cloud registration method based on multiple neighborhood feature difference (MNFD) that employs a coarse-to-fine strategy to effectively enhance both registration efficiency and accuracy. In the coarse registration stage, a novel feature point extraction approach based on MNFD is introduced, capable of identifying highly stable and distinctive feature points in the point cloud. These feature points are then utilized in combination with the fast point feature histogram (FPFH) algorithm to achieve an initial alignment between the target and template point clouds. In the fine registration stage, the results from the coarse alignment are refined using algorithms such as iterative closest point (ICP) to ensure both efficiency and precision during the registration process.\n\n\n\n\n\n\n\nABSTRACT\nDense point cloud registration is a critical problem in computer vision and 3D reconstruction, with widespread applications in scenarios such as robotic navigation, autonomous driving, and 3D measurement. However, dense point cloud registration faces significant challenges, including high computational complexity and prolonged processing times. To address these issues, this paper proposes a point cloud registration method based on multiple neighborhood feature difference (MNFD) that employs a coarse-to-fine strategy to effectively enhance both registration efficiency and accuracy. The proposed method consists of two stages: coarse registration and fine registration. In the coarse registration stage, a novel feature point extraction approach based on MNFD is introduced, capable of identifying highly stable and distinctive feature points in the point cloud. These feature points are then utilized in combination with the fast point feature histogram (FPFH) algorithm to achieve an initial alignment between the target and template point clouds. In the fine registration stage, the results from the coarse alignment are refined using algorithms such as iterative closest point (ICP) to ensure both efficiency and precision during the registration process. Experiments conducted on publicly available datasets demonstrate the superiority of the proposed method compared to existing\xa0approaches.',
'Haixia Wang, Teng Wang, Zhiguo Zhang, Xiao Lu, Qiaoqiao Sun, Shibin Song, Jun Nie',
'7',
'2025',
'5'],
['Fisheye image rectification and restoration based on Swin Transformer',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13294?af=R',
"Fisheye cameras are widely used in surveillance, automotive systems, virtual reality, and panoramic photography due to their wide-angle perspective. However, images captured by fisheye cameras suffer from significant geometric distortions, affecting image analysis and understanding. This distortion bends straight lines into curves, resulting in a barrel-shaped appearance of the image. To mitigate these effects and transform fisheye images into a regular perspective, fisheye image correction is necessary, enabling more accurate and reliable performance in applications like object recognition, 3D reconstruction, and visual navigation. While convolutional neural networks based fisheye image correction has progressed, it has not fully utilized the global distribution and local symmetry of distortions due to the limitations of fixed receptive fields. This paper introduces a new model based on the Swin Transformer that effectively utilizes both global and local distortion features to adapt automatically to fisheye lens distortions. It also incorporates image restoration functionality to enhance texture details in the corrected images. A novel approach to synthetic dataset generation is proposed to improve the network's generalization capabilities.",
'Jian Xu, Dewei Han, Kang Li, Junjie Li, Zhaoyuan Ma',
'7',
'2024',
'12'],
['Nullifying Jammer Effects in RIS-Assisted GNSS-Free Drone Localisation',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70368?af=R',
'This letter explores a double reconfigurable intelligent surfaces (RIS)-aided global navigation satellite system-free drone localisation setup to optimise jammer nullification while enhancing the desired signal. We propose an alternating projection algorithm to optimise the RIS phase shifts, achieving efficient jammer suppression and improved 3D drone positioning. Simulation results demonstrate that the proposed approach significantly enhances localisation accuracy and mitigates jamming effects.\n\n\n\n\n\n\n\nABSTRACT\nA general challenge in drone localisation and navigation is the strong dependence on satellite positioning. Reconfigurable intelligent surfaces (RIS) offer a promising\xa0technology for global navigation satellite system (GNSS)-free drone localisation and for the alleviation of the jammer effect by nulling the jamming signal. In this letter, we investigate how to optimise the nullification of the jammer while maximising the desired signal in a double RIS-aided GNSS-free drone localisation setup. The main objectives of the presented procedure are the maximisation of the desired signal received by the drone, minimisation of the

jamming signal effect, accurate estimation of the angle of departure (AOD) from RIS to drone, and drone positioning in 3D. To address this challenge, we propose an alternating projection algorithm to perform jamming nullification by optimising the phase shift at each RIS element. Simulations show that by proper RIS element phase optimisation, the effect of the jammer is efficiently reduced, the AODs from RIS surfaces to drone are computed more accurately, and the drone 3D localisation accuracy is\xa0improved.',
  'Mehari Meles, Akash Rajasekaran, Reino Virrankoski, Riku Jäntti',
  '8',
  '2025',
  '7'],
 ['A Combined Navigation Fault Detection Method Based on Wavelet-ResNet',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70316?af=R',
  'A combined navigation dataset, using wavelet transform is constructed and a residual network structure is trained for fault detection. The results show that this method can effectively improve the training efficiency and the accuracy of slowly varying faults.\n\n\n\n\n\n\n\nABSTRACT\nWith the goal of realizing fault detection of airborne combined navigation system, this letter proposes a fault detection method combining wavelet transform and residual network. Firstly, the model equation\xa0of the system is established, and the combined navigation data set is constructed by combining the characteristics of the UAV flight trajectory. At the same time, the normal data of the UAV is added with common faults to obtain the fault data set. Then, the navigation data is preprocessed by discrete wavelet transform, and the residual network model is trained according to the preprocessed navigation data set to improve the training efficiency of the model. Finally, the trained model is tested and evaluated. The simulation results show that the accuracy of detecting mutation faults, slope faults and precision level reduction faults is more than 95%$95\\%$.',
  'Yunhua Liu, Hao Wang, Yongqiang Liu, Qing Chang',
  '8',
  '2025',
  '6'],
 ['Implementation and Performance Analysis of RTK-GNSS in Wearable Devices for Athletes in Harsh Environments',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70289?af=R',
  'This study investigates the implementation of real-time kinematic global navigation satellite system (RTK-GNSS) technique in wearable electronic performance and tracking system for football players to obtain highly accurate positioning data, even in harsh environments. The experiment results demonstrate that the utilization of RTK-GNSS can significantly improve positioning accuracy in various stadium environments, which is crucial for evaluating player performance.\n\n\n\n\n\n\n\nABSTRACT\nThis study presents the performance analysis of RTK-GNSS (real-time kinematic global navigation satellite system) mounted on wearable devices for athletes in harsh environments. GNSS signal can deteriorate by obstacles, i.e. stadium roofs, due to the high GDOP (geometric dilution of precision). RTK-GNSS is used to overcome harsh environment problems in this study. Hardware experiments are performed to analyse the performance of RTK-GNSS in sports wearable devices in harsh environments. The full pitch tracking test results of a wearable device equipped with RTK-GNSS demonstrate the best performance among three different wearable devices in a harsh environment; the maximum positioning error of 1.56 m$\\text{m}$ is approximately 3–30 times smaller than that of wearable devices equipped with stand-alone GNSS. The distance test results of the wearable device equipped with RTK-GNSS validate the most accurate performance, 98.97%$\\%$, compared to the true value in harsh\xa0environment.',
  'Mingu Kim, Bit Kim, Chulwoo Park, Jinsung Yoon',
  '8',
  '2025',
  '5']]

SEARCH WORD: INFRA

[['Self cycle strategy for unpaired visible-to-infrared image translation',
  'https://www.sciencedirect.com/science/article/pii/S0031320325009148',
  '',
  'Decao Ma, Juan Su, Bing Li, Yong Xian, Shaopeng Li, Yao Ding',
  '2',
  '2026',
  '3'],
 ['Reliable and Balanced Transfer Learning for Generalized Multimodal Face Anti-Spoofing',
  'http://ieeexplore.ieee.org/document/11015656',
  'Face Anti-Spoofing (FAS) is essential for securing face recognition systems against presentation attacks.

Recent advances in sensor technology and multimodal learning have enabled the development of multimodal FAS systems. However, existing methods often struggle to generalize to unseen attacks and diverse environments due to two key challenges: (1) Modality unreliability, where sensors such as depth and infrared suffer from severe domain shifts, impairing the reliability of cross-modal fusion; and (2) Modality imbalance, where over-reliance on a dominant modality weakens the model's robustness against attacks that affect other modalities. To overcome these issues, we propose MMDG++, a multimodal domain-generalized FAS framework built upon the vision-language model CLIP. In MMDG++, we design the Uncertainty-Guided Cross-Adapter++ (U-Adapter++) to filter out unreliable regions within each modality, enabling more reliable multimodal interactions. Additionally, we introduce Rebalanced Modality Gradient Modulation (ReGrad) for adaptive gradient modulation to balance modality convergence. To further enhance generalization, propose Asymmetric Domain Prompts (ADPs) that leverage CLIP's language priors to learn generalized decision boundaries across modalities. We also develop a novel multimodal FAS benchmark to evaluate generalizability under various deployment conditions. Extensive experiments across this benchmark show our method outperforms state-of-the-art FAS methods, demonstrating superior generalization capability.',
  'Xun Lin, Ajian Liu, Zitong Yu, Rizhao Cai, Shuai Wang, Yi Yu, Jun Wan, Zhen Lei, Xiaochun Cao, Alex Kot',
  '3',
  '2025',
  '8'],
 ['FreeFusion: Infrared and Visible Image Fusion via Cross Reconstruction Learning',
  'http://ieeexplore.ieee.org/document/11010882',
  "Existing fusion methods empirically design elaborate fusion losses to retain the specific features from source images. Since image fusion has no ground truth, the hand-crafted losses may not make the fused images cover all the vital features, and then affect the performance of the high-level tasks. Here, there are two main challenges: domain discrepancy among source images and semantic mismatch at different-level tasks. This paper proposes an infrared and visible image fusion via cross reconstruction learning, which doesn't using any hand-crafted fusion losses, but prompts the network to adaptively fuse complementary information of source images. Firstly, we design a cross reconstruction learning model that decouples the fusion features to reconstruct another-modality source image. Thus, the fusion network is forced to learn the domain-adaptive representations of two modal features, which enables their domain alignment in a latent space. Secondly, we propose a dynamic interactive fusion strategy that builds a correlation matrix between fusion features and object semantic features to overcome the semantic mismatch. Further, we enhance the strong correlation features and suppress the weak correlation features to improve the interactive ability. Extensive experiments on three datasets demonstrate the superior fusion performance compared to the state-of-the-art methods, concurrently facilitating the segmentation accuracy.",
  'Wenda Zhao, Hengshuai Cui, Haipeng Wang, You He, Huchuan Lu',
  '3',
  '2025',
  '8'],
 ['MTMLNet: Multi-Task Mutual Learning Network for Infrared Small Target Detection and Segmentation',
  'http://ieeexplore.ieee.org/document/11080263',
  'Infrared small target detection has been extensively studied due to its wide range of applications. Most studies treat infrared small target detection as an independent task, either as a detection-based or a segmentation-based, failing to fully leverage the supervisory information from different annotation forms. To address this issue, we propose a multi-task mutual learning network (MTMLNet) specifically designed for infrared small targets, aiming to enhance both detection and segmentation performance by effectively utilizing various forms of supervisory information. Specifically, we design a multi-stage feature aggregation (MFA) module capable of capturing features with varying gradients and receptive fields simultaneously. Additionally, a hybrid pooling down-sampling (HPDown) module is proposed to mitigate information loss during the down-sampling process of infrared small targets. Finally, the hierarchical feature fusion (HFF) module is designed to adaptively select and fuse features from different semantic layers, learning the optimal way to fuse features across semantic layers. The results on IRSTD-1k and SIRST-V2 datasets show that our proposed MTMLNet achieves state-of-the-art (SOTA) performance in both detection-based and segmentation-based methods. The codes are available at https://github.com/YangBo0411/MTMLNet',
  'Bo Yang, Fengqian Li, Songliang Zhao, Wei Wang, Jun Luo, Huayan Pu, Mingliang Zhou, Yangjun Pi',
  '4',
  '2025',
  '8'],
 ['Computational Fluid Dynamic Network for Infrared Small Target Detection',
  'http://ieeexplore.ieee.org/document/10949663',
  'Infrared small target detection (IRSTD) aims to identify and locate small targets amidst background noise. It is highly valuable in various practical application domains, such as maritime rescue and early warning systems deployed in challenging conditions such as harsh weather, low illumination, and long imaging distances. Different from existing works that either adopt well-designed backbone networks or devise specific

modules to improve them from different aspects, in this article, we formulate the learning process of IRSTD from a novel perspective, i.e., the mechanism of pixel movement. Considering that the movement of pixels passing through the layers of the network for IRSTD can be analogized to the flow of particles in a fluid dynamic system, we propose a computational fluid dynamic network (CFD-Net) derived from computational fluid dynamics. Technically, we leverage the superiority of the unilateral difference equation with third-order accuracy and devise a unilateral differential residual structure as the backbone of CFD-Net. This design ensures that the pixel stream only flows in the forward direction. In addition, a switch-controlled multidirectional treatment tank (SMTT) is introduced to CFD-Net to dynamically guide the pixel stream to the appropriate path for different targets with varying shapes and orientations, facilitating learning robust target representation and improving detection performance. The proposed CFD-Net is evaluated on the IRSTD-1k and SIRST datasets and is found to outperform existing state-of-the-art (SOTA) methods.',
  'Mingjin Zhang, Ke Yue, Jie Guo, Qiming Zhang, Jing Zhang, Xinbo Gao',
  '5',
  '2025',
  '8'],
 ['Feature-Level Compensation and Alignment for Visible-Infrared Person Re-Identification',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cvi2.70005?af=R',
  'In this paper, the authors present a novel feature-level compensation and alignment network (FCA-Net) to address the challenging visible-infrared person re-identification task. In FCA-Net, a channel feature compensation (CFC) module is proposed, which enforces the network to learn consistent distribution patterns of channel features, and thereby the cross-modality discrepancy is narrowed. To perform spatial alignment, a novel pairwise relation module is proposed to incorporate human structural information, and a cross-modality part alignment loss is designed to encourage accurate matching.\n\n\n\n\n\n\nABSTRACT\nVisible-infrared person re-identification (VI-ReID) aims to match pedestrian images captured by nonoverlapping visible and infrared cameras. Most existing compensation-based methods try to generate images of missing modality from the other ones. However, the generated images often fail to possess enough quality due to severe discrepancies between different modalities. Moreover, it is generally assumed that person images are roughly aligned during the extraction of part-based local features. However, this does not always hold true, typically when they are cropped via inaccurate pedestrian detectors. To alleviate such problems, the authors propose a novel feature-level compensation and alignment network (FCA-Net) for VI-ReID in this paper, which tries to compensate for the missing modality information on the channel-level and align part-based local features. Specifically, the visible and infrared features of low-level subnetworks are first processed by a channel feature compensation (CFC) module, which enforces the network to learn consistent distribution patterns of channel features, and thereby the cross-modality discrepancy is narrowed. To address spatial misalignment, a pairwise relation module (PRM) is introduced to incorporate human structural information into part-based local features, which can significantly enhance the feature discrimination power. Besides, a cross-modality part alignment loss (CPAL) is designed on the basis of a dynamic part matching algorithm, which can promote more accurate local matching. Extensive experiments on three standard VI-ReID datasets are conducted to validate the effectiveness of the proposed method, and the results show that state-of-the-art performance is achieved.',
  'Husheng Dong, Ping Lu, Yuanfeng Yang, Xun Sun',
  '6',
  '2025',
  '2'],
 ['Category-instance distillation based on visual-language models for rehearsal-free class incremental learning',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cvi2.12327?af=R',
  'Recently, visual-language models (VLMs) have displayed potent capabilities in the field of computer vision. Their emerging trend as the backbone of visual tasks necessitates studying class incremental learning (CIL) issues within the VLM architecture. However, the pre-training data for many VLMs is proprietary, and during the incremental phase, old task data may also raise privacy issues. Moreover, replay-based methods can introduce new problems like class imbalance, the selection of data for replay and a trade-off between replay cost and performance. Therefore, the authors choose the more challenging rehearsal-free settings. In this paper, the authors study class-incremental tasks based on the large pre-trained vision-language models like CLIP model. Initially, at the category level, the authors combine traditional optimisation and distillation techniques, utilising both pre-trained models and models trained in previous incremental stages to jointly guide the training of the new model. This paradigm effectively balances the stability and plasticity of the new model, mitigating the issue of catastrophic forgetting. Moreover, utilising the VLM infrastructure, the authors redefine the relationship between instances. This allows us to glean fine-grained instance relational information from the a priori knowledge provided during pre-training. The authors supplement this approach with an entropy-balancing method that allows the model to adaptively distribute optimisation weights across training samples. The authors' experimental results validate that their method, within the framework of VLMs, outperforms traditional CIL methods.',

'Weilong Jin, Zilei Wang, Yixin Zhang',
'6',
'2025',
'2'],
['Enhancing Quantum Image Encryption With QLSTM and Chaos Synchronisation Control: A Deep Neural Network Approach',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70091?af=R',
'This paper presents a quantum image encryption method using QLSTM to enhance chaotic sequences and achieve synchronisation control. The approach improves encryption security and reliability, maintaining high information entropy and complexity under attacks. \n\n\n\n\n\n\n\nABSTRACT\nQuantum image encryption is crucial for data protection, but current methods lack attack resistance and have complex encryption processes. This paper proposes a quantum long short - term memory (QLSTM)-based quantum image encryption method to enhance chaotic sequences and achieve chaos synchronisation control. The QLSTM network improves the Lorenz chaotic sequence, increasing its unpredictability. An adaptive synchronisation control algorithm, using the enhanced chaotic sequence from QLSTM, ensures sender-receiver synchronization. Optimised through deep neural networks, the system maintains stable synchronization under interference. New cryptographic quantum infrastructure (NCQI) was constructed, and images were encrypted using third-order radial diffusion, quantum generalised Arnold transform, and quantum W transform. The QLSTM-improved chaotic sequence showed excellent LLE and 0–1 test results. Information entropy was near 8, with R, G and B channels exceeding 7.999. Anti-attack analysis revealed high information entropy, strong attack resistance, and number of pixels change rate/unified average changing intensity (NPCR/UACI) values of 99.698% and 33.460%, respectively, indicating significant pixel-level changes. Combining quantum chaotic system prediction with the QLSTM model enhanced quantum communication stability and anti-interference ability. This QLSTM-based quantum encryption method, with chaos synchronisation control, significantly improves encryption security and reliability, maintaining high information entropy and complexity under attacks, proving its effectiveness in image encryption.',
'Yuebo Wu, Duansong Wang, Jian Zhou, Huifang Bao',
'7',
'2025',
'5'],
['Infrared and Visible Image Fusion Based on Autoencoder Network',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.70086?af=R',
'(1) An infrared and visible image fusion network is proposed which skilfully combines information decomposition, feature saliency measurement and extraction of significant targets. (2) The common features are extracted by the two scene information encoders with shared weights, and the unique features are extracted by the two salient information encoders with unshared weights for infrared and visible images, respectively. Also, the constraints are added to the loss function in order to ensure the ability of the salient information encoders to extract representative features and the cross-modality feature extraction ability of the scene information encoder. (3) The ability of fusion network to distinguish between targets and backgrounds is further enhanced by introducing a pre-trained semantic segmentation network to guide fusion network training and constructing a feature saliency-based fusion strategy.\n\n\n\n\n\n\n\nABSTRACT\nTo overcome the problems of texture information loss and insufficiently prominent targets in existing fusion networks, an information decomposition-based autoencoder fusion network for infrared and visible images is proposed in this paper. Two salient information encoders with unshared weights and two scene information encoders with shared weights are designed to extract different features from infrared and visible images, respectively. The constraint is added to the loss function in order to ensure the ability of the salient information encoders to extract representative features and the scene information encoder to extract the cross-modality feature. In addition, by introducing the pre-trained semantic segmentation networks to guide the network training and constructing a feature saliency-based fusion strategy, the ability of the fusion network is further enhanced to distinguish between targets and backgrounds. Extensive experiments are carried out on five datasets. Comparison experiments with state-of-the-art fusion networks and ablation experiments indicate that the proposed method can obtain fused images with richer and more comprehensive information and is more robust to challenging factors, such as strong and weak light smoke and fog environments. At the same time, the fused images by our proposed method are more beneficial for downstream tasks such as target detection.',
'Hongmei Wang, Xuanyu Lu, Zhuofan Wu, Ruolin Li, Jingyu Wang',
'7',
'2025',
'4'],
['Research on the registration of infrared and visible images based on phase consistency and edge extreme points',
'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13317?af=R',
'Due to the significant differences in the imaging principles of infrared images and visible images, the

feature points and descriptors between the two cannot be effectively matched directly by traditional feature extraction methods such as SIFT. To solve this problem, this study proposes a registration algorithm for infrared and visible images based on phase information and edge information. The algorithm extracts the feature points of the infrared image and the visible image through the principle of phase agreement and the edge binary map of the image and then calculates the descriptors of the gradient images of the infrared image and the visible image, and the descriptor calculation draws on some SIFT principles. Finally, the cosine similarity was used to match the feature points, and the improved random sample consensus algorithm was used to screen out the correct registration points. Experiments show that this method can effectively register between infrared and visible images and is also suitable for the registration of infrared and visible images with different rotation angles and similar structures.',
  'Jie Li, Rougang Zhou, Zhenchao Ruan, Chou Jay Tsai Chien, Junjie Zhu',
  '7',
  '2025',
  '1'],
 ['Image inpainting with aggregated convolution progressive network',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13318?af=R',
  'Images can be corrupted during capture or transmission due to clouds, overlaps, and other interferences, deviating from their original state. Image inpainting techniques restore such images, but different types—Synthetic Aperture Radar (SAR), RGB, and infrared—require varying field-of-view sizes. SAR and infrared images, with less information, need a larger field of view, leading to uncorrelated interference in distant areas. RGB images, richer in information, are constrained by a limited local field of view, hindering access to full semantic details. To address these challenges, an aggregated convolution progressive network is proposed. This model employs a coarse-grained inpainting module for initial restoration, enhanced by an aggregated convolution module to capture contextual information. Local and global details are then used to refine the output, improving restoration quality. Additionally, existing datasets predominantly focus on RGB images, lacking diversity. To bridge this gap, a comprehensive dataset covering SAR, RGB, and infrared images under cloud, overlap, and corruption conditions is constructed. This method achieves superior performance, with MAE of 0.05, SSIM of 0.95, and PSNR of 36.68 within a 20–30% mask size range, outperforming state-of-the-art techniques across diverse image types and size ranges. Experimental results validate its effectiveness in advancing image inpainting.',
  'Yang Li, Jia Zhai, Wen Lu, Haipeng Guo, JiaZheng Wen, Huanyu Liu, Junbao Li',
  '7',
  '2025',
  '1'],
 ['MCFNet: Research on small target detection of RGB-infrared under UAV perspective with multi-scale complementary feature fusion',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13320?af=R',
  "The task of single-modal image target detection from a drone's perspective faces multiple challenges such as large image sizes, small and dense targets, insufficient lighting conditions, and hardware resource constraints, all of which affect the accuracy and real-time performance of algorithms. In response, a small target detection algorithm that fuses multi-scale complementary features, named MCFNet is proposed. Firstly, in order to independently extract the unimodal features of visible light and infrared images and effectively enhance the complementary information between images, a dual-stream backbone network with a terminal fusion mechanism is proposed. Secondly, the complementary feature information residual module is utilized to optimize the integration process of residual features. Subsequently, by designing a multi-scale feature enhancement module the network's capability to capture multi-scale features is enhanced. Finally, for targets of varying sizes, a lightweight Transformer feature extraction module is proposed to improve the detection accuracy of small targets from a drone's perspective. Test results on the drone-vehicle dataset show that this method achieved an average detection accuracy of 67.92%, while the detection accuracy on the self-constructed UAV-data dataset reached 96.4%. Additionally, a series of ablation experiments validated the effectiveness of the different modules.",
  'Jing Jing, Jian Feng Hu, Zuo Peng Zhao, Ying Liu',
  '7',
  '2025',
  '1'],
 ['MOE-IR: Infrared Dim Small Target Detection Method With Mixture of Experts Feature Extraction',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70359?af=R',
  'A multi expert model-based infrared small target detection method mixture of experts infrared (MOE-IR) has been proposed to solve the problem of robust detection of infrared small targets in complex environments. In the MOE-IR, the target feature extraction expert model and the background suppression expert model are proposed to suppress background clutter interference and extract rich infrared small target features, respectively. Meanwhile, an adaptive gate controlled network is designed to adaptively allocate the weights of the two expert models according to the input infrared image, so as to achieve effective detection of

infrared small targets in different complex scenes. Comprehensive experiments demonstrate the detection performance of the proposed MOE-IR in the public dataset SIRST is better than the existing advanced infrared small target detection methods, with Pd increased by 3% and Fa reduced by 73%.\n\n\n\n\n\n\n\n\nABSTRACT\nInfrared small target detection holds significant importance across various domains, including military and security applications. Nevertheless, detecting small targets is highly challenging due to their minimal pixel presence in images, indistinct features, and complex backgrounds. Existing detection methods are often interfered by complex backgrounds, resulting in unsatisfactory detection results. To overcome this challenge, this study introduces the mixture of experts infrared small target detection (MOE-IR) method. The core idea of this method is to construct a mixture of experts feature extraction network to perform rich feature extraction and complex background suppression on small targets, respectively, so as to achieve robust infrared small target detection in complex backgrounds. Specifically, the MOE-IR comprises a target feature extraction expert and a background suppression expert. The target feature extraction expert focuses on enhancing the features of infrared small targets, while the background suppression expert aims to mitigate background clutter. Additionally, an adaptive gate controlled network is incorporated to dynamically assign weights to the two experts based on the input infrared image, ensuring effective detection of infrared small targets across diverse and complex scenarios. Extensive experiments demonstrate that the proposed algorithm surpasses existing infrared small target detection methods in terms of detection accuracy and false alarm rate. It can reliably and stably identify small targets within infrared images, thus offering an effective solution for practical infrared small target detection\xa0applications.',
  'Zhengkui Weng, Xinjie Fu, Xu Zhang, Siyuan Sun',
  '8',
  '2025',
  '7'],
 ['Vehicle Target Detection Based on Cross-Modality Projective-Invariant Features Extracted from Unpaired SAR and Infrared Images',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70336?af=R',
  'A novel dual-channel context-guided feature-alignment network (CGFAN) that is capable of fusing the cross-modality projective-invariant features extracted from unpaired SAR and IR images is proposed. First, region of interest matching between SAR and IR images are realized based on special landmarks exhibiting consistent cross-modality features. After that, generative models trained with historical SAR and IR images are used to synthesize SAR images based on the IR images collected in real time for the current mission.\n\n\n\n\n\n\n\n\nABSTRACT\nSynthetic aperture radar (SAR) automatic target recognition (ATR) is remarkably challenging since the SAR image defies the foundation for human and computer vision, i.e., the Gestalt perceptual principles. We propose to address this problem by fusing the target features reflected in SAR and infrared (IR) images via a novel dual-channel context-guided feature-alignment network (CGFAN) that is capable of fusing the cross-modality projective-invariant features extracted from unpaired SAR and IR images. First, region of interest (ROI) matching between SAR and IR images is realized based on special landmarks exhibiting consistent cross-modality features. After that, generative models trained with historical SAR and IR images are used to synthesize SAR images based on the IR images collected in real time for the current mission. Since SAR imaging takes more time than IR imaging, by using these synthesized SAR images as auxiliary data, the spatial-coverage rate in a typical collaborative SAR/IR ATR mission carried out by drone swarms is effectively improved. The proposed CGFAN is tested against the proprietary monostatic-bistatic circular SAR and IR dataset constructed by the researchers at our institution, which consists of nine types of military vehicles. Experimental results show that the proposed CGFAN offers better ATR performance than the baseline networks.A novel dual-channel CGFAN that is capable of fusing the cross-modality projective-invariant features extracted from unpaired SAR and IR images is proposed. First, ROI matching between SAR and IR images are realized based on special landmarks exhibiting consistent cross-modality features. After that, generative models trained with historical SAR and IR images are used to synthesize SAR images based on the IR images collected in real time for the current mission.',
  'Zhe Geng, Chongqi Xu, Chen Xin, Xiang Yu, Daiyin Zhu',
  '8',
  '2025',
  '6'],
 ['An Optimal Ultra-Thin Broadband Polarization-Independent Metamaterial Absorber for Visible and Infrared Spectrum Applications',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70314?af=R',
  "The proposed MMA as a strong candidate for environmental monitoring, IR detection, solar energy harvesting, and sensor technologies.\n\n\n\n\n\n\n\n\nABSTRACT\nThis paper presents the design and study of an extremely thin, wideband, and polarization-insensitive metamaterial absorber (MMA) tailored for applications in the visible and infrared (IR) spectral ranges. The proposed MMA introduces a diamond-shaped resonator setup, achieves high absorption efficiency beyond 90% across a wide wavelength range from 331.64 to 2163.1\xa0nm. The average bandwidth of absorption is 95.64%, having the highest absorption rate of 99.039% witnessed at 511.47\xa0nm. The unit cell of the absorber is condensed and

designed to cover the visible optical range, along with the near and short infrared optical windows. Aluminium nitride (AlN) is used as the substrate, while nickel (Ni) serves as both the ground and resonator layer material, contributing to the MMA's durability and thermal stability. As the absorber is symmetric in design, it ensures polarization independence, with steady absorption performance maintained at oblique angles of incidence up to 75°. Comprehensive parametric simulations were performed, examining elements such as surface current distribution, layer thicknesses, different substrate materials and variations in permittivity and permeability, all of which demonstrate the MMA's robustness and adaptability for various applications and by covering a wide spectral range and exhibiting flexibility to polarization and incident angle variations, this absorber offers an effective platform for environmental monitoring, IR detection, and solar energy harvesting.",
  'Md. Murad Kabir Nipun, Md. Jahedul Islam, Md. Moniruzzaman',
  '8',
  '2025',
  '5'],
 ['Deep Reinforcement Learning-Based Beam Hopping Scheduling Algorithm for Satellite Internet System',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70250?af=R',
  "The authors proposed a Deep Reinforcement Learning (DRL)-based downlink Beam Hopping Scheduling (BHS) algorithm, which leverages DRL's advantage to make dynamic decisions according to varying states.\n\n\n\n\n\n\n\nABSTRACT\nSatellite internet is a new network system that combines satellite communication and internet technology. The main space-based infrastructure of satellite internet is a low-Earth-orbit (LEO) satellite constellation. The system uses flexible control of satellite antenna pointing to provide worldwide on-demand coverage with finite satellites' resource. beam hopping (BH) is a promising technology to improve resource allocation's flexibility, which typically uses spot beams to illuminate different beam positions in different time slots. However, considering the uneven and dynamic traffic demand of every cell, it is difficult to design a dynamic beam hopping scheduling (BHS) algorithm that can provide high resource utilisation and good Quality of Experience (QoE). We propose a deep reinforcement learning (DRL)-based downlink BHS algorithm, which leverages DRL to make dynamic decisions according to varying states. The experimental results show that compared with other algorithms, the proposed algorithm performs better in terms of average queuing delay, capacity utilisation ratio, fairness between different cells and the whole throughput.",
  'Dong Zhang, Baoqiang Jiang, Chengcheng Li, Peiying Zhang, Huaxi Gu',
  '8',
  '2025',
  '4'],
 ['Small target detection with decoupling and decorrelation of confusion features',
  'https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ell2.70158?af=R',
  'Abstract\nSmall target detection has important application value in military, cigarette defect detection, monitoring and remote sensing fields. However, due to the complex background interference of infrared images, the background and target features are confused with each other, which makes it difficult for existing detection methods to effectively distinguish the features of the background and the target, resulting in poor detection effect. Therefore, to solve this problem, this paper proposes a small target detection method based on confusion feature decoupling and decorrelation, aiming to achieve accurate detection of small targets in complex environments by extracting robust small target features. Specifically, in confusion feature decoupling and decorrelation, we propose a confusion feature decoupling and decorrelation module. By introducing a feature decoupling mechanism, the features of the input image are decomposed into independent background features and target features, and feature decorrelation is used to achieve independence between the target and the background. The target features after decoupling and decorrelation are purer, which helps to reduce background interference and thus improve detection performance. Systematic experimental results show that the detection performance of the proposed method on public small target detection datasets is much better than that of existing advanced detection\xa0methods.',
  'Yonghua Zhang, Hui Wang, He Tang, Xingze Liu, Benxue Liu, Siyuan Sun',
  '8',
  '2025',
  '2']]