SCIENCE CHINA
Information Sciences

• LETTER •

November 2026, Vol. 69, Iss. 11, 210307:1–210307:2
https://doi.org/10.1007/s11432-026-5217-4

Special Topic: Communication, Sensing, and Target Recognition for Low-Altitude Economy

Towards low-altitude UAV detection in urban radar clutter: a benchmark and spatio-temporal channel transfer network

Jianhong ZHOU[1], Feng KE[1*], Yikui ZHAI[2*], Xueqiang ZHENG[3], Ziyi JIANG[1], Haolin LV[4] & Xiu-Yin ZHANG[1]

[1]*School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China*
[2]*School of Electronics and Information Engineering, Wuyi University, Jiangmen 529020, China*
[3]*College of Communications Engineering, Army Engineering University of PLA, Nanjing 210042, China*
[4]*Future Aviation Technology (Jiangmen) Co., Ltd., Jiangmen 529000, China*

Received 26 March 2026/Revised 24 August 2026/Accepted 15 October 2026/Published online 28 October 2026

Citation Zhou J H, Ke F, Zhai Y K, et al. Towards low-altitude UAV detection in urban radar clutter: a benchmark and spatio-temporal channel transfer network. Sci China Inf Sci, 2026, 69(11): 210307, https://doi.org/10.1007/s11432-026-5217-4

The rapid proliferation of unmanned aerial vehicles (UAVs) is accelerating the development of low-altitude intelligent transportation, urban logistics, and emergency response applications, while simultaneously placing greater demands on airspace safety and security. Reliable UAV detection is therefore essential for downstream tasks such as target recognition, risk warning, and intervention. Compared with passive sensing modalities, radar supports long-range, all-weather operation and directly measures target range and radial velocity [1]. In particular, multiple-input multiple-output frequency-modulated continuous-wave (MIMO-FMCW) radar offers enhanced angular resolution and clutter suppression, making it a promising technology for urban UAV surveillance [2]. Nevertheless, detecting small, slow-moving UAVs remains challenging because of their low radar cross-section (RCS), which results in weak radar echoes, as well as strong environmental clutter and multipath interference. Conventional threshold-based methods, such as constant false alarm rate (CFAR) detectors, typically rely on statistical assumptions that may be violated in non-stationary urban environments, thereby increasing the likelihood of missed detections for weak targets.

Deep learning offers a promising alternative by learning discriminative representations directly from radar measurements. However, two key challenges remain. First, multi-chirp range-azimuth (RA) measurements contain fine-grained temporal information encoded in phase variations and micro-motion characteristics. Existing methods commonly reduce computational complexity through chirp sampling or temporal aggregation [3], which may discard weak but informative dynamic cues. Therefore, effectively preserving inter-chirp information while maintaining computational efficiency is essential for reliable detection of weak UAV targets. Second, publicly available radar datasets with accurate UAV position annotations collected in realistic urban environments remain scarce [4, 5]. The low signal-to-noise ratio (SNR) of UAV radar returns, compounded by strong environmental clutter and multipath propagation, and complicates reliable ground-truth annotation, thereby limiting rigorous model evaluation.

To address these challenges, we construct UAVRadar, an real-time kinematic (RTK)-annotated benchmark comprising 23700 frames of high-resolution radar measurements collected under diverse flight modes, sensing conditions, and environmental disturbances, along with a reproducible annotation and evaluation pipeline. For a more detailed description of UAVRadar, please refer to Appendix A. Building on UAVRadar, we further propose the spatio-temporal channel transfer network (STCT-Net) for accurate and efficient detection of weak UAV targets. Its STCT module transfers temporal dynamics from the chirp dimension to the channel space, thereby preserving motion- and phase-related cues that may otherwise be suppressed by conventional temporal aggregation. A principal component analysis (PCA)-based chirp sampling strategy prioritizes informative components while reducing redundant information, noise, and computational overhead. In addition, task-adapted feature extraction and a Soft-BCE loss tailored to Gaussian heatmap supervision are incorporated to improve weak-target representation and localization accuracy.

Method. For UAV radar target detection, this study proposes a UAV detection framework termed STCT-Net (see Figure 1). The framework comprises three components: a data preprocessing module, an STCT module, and an optimized feature extractor. Among these, the data preprocessing module and the STCT module are specifically designed for input feature enhancement and spatio-temporal information modeling, respectively. Specifically, during the data preprocessing stage, a bin-based PCA dimensionality reduction strategy is introduced along the chirp dimension of the RA spectrum to replace the conventional uniform sampling approach. Compared with conventional methods, this strategy more effectively highlights the critical dynamic information of weak and small targets, thereby enhancing feature saliency and providing a more compact, information-dense input representation for subsequent network learning. In the STCT module, a chirp merging mechanism is further introduced. This mechanism explicitly de-

* Corresponding author (email: fengke@scut.edu.cn, yikuizhai@163.com)

© Science China Press 2026

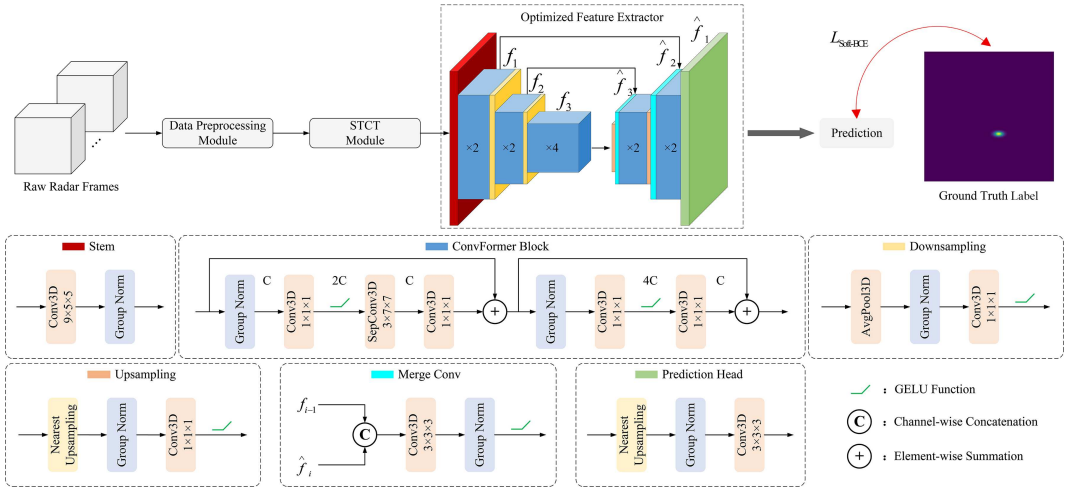

Figure 1 (Color online) The overall architecture of STCT-Net comprises a data preprocessing module, an STCT module, and an optimized feature extractor. The raw radar frame sequence is first processed by the data preprocessing and STCT modules to extract micro-temporal features and construct an input representation suitable for subsequent network learning. All components shown in the lower portion of the figure, including the stem, ConvFormer block, downsampling and upsampling modules, Merge Conv, and prediction head, constitute the optimized feature extractor. The resulting features are then fed into the optimized feature extractor, which adopts a U-shaped encoder-decoder architecture. In the encoder, the input features are progressively downsampled to generate multi-scale feature maps, denoted as f_1, f_2, and f_3. In the decoder, these feature maps are progressively upsampled and fused, through skip connections, with the high-resolution features from the corresponding encoder stages, yielding refined feature maps $\hat{f}_1$, $\hat{f}_2$, and $\hat{f}_3$. Finally, the decoder output is passed through a prediction head to produce the final detection results, and the proposed Soft-BCE loss function is employed for supervised training using the ground-truth labels.

couples the modeling of spatial and temporal features, enabling the network to preserve key temporal characteristics in target echoes more effectively while extracting spatial information, thereby improving its ability to characterize the dynamic properties of UAV targets. For feature extraction, this study adopts the ConvFormer paradigm, which offers effective feature representation and stable optimization during training. To accommodate the characteristics of radar target detection, it is extended from a 2D architecture to a 3D one, and its key components are selectively replaced and optimized to improve its suitability for radar data representation. In addition, given the weak-target characteristics of UAVs in radar sensing, a task-adapted training loss function, Soft-BCE, is proposed. During training, this loss function places greater emphasis on modeling the responses of target regions while enhancing the suppression of background noise, thereby improving weak-target detection performance. Owing to the coordinated design of these modules, the proposed STCT-Net achieves excellent performance in the UAV detection task. For a more detailed description of our method, please refer to Appendix B.

Experiments. To verify the superiority of STCT-Net, experiments were conducted from three aspects. First, we compared STCT-Net with representative SOTA radar target detection methods and conventional CFAR detectors on UAVRadar to evaluate its detection accuracy and computational efficiency. Second, extensive ablation studies were performed to verify the effectiveness of the PCA-based sampling strategy, STCT module, optimized feature extractor, and Soft-BCE loss. Finally, we further analyzed the robustness of STCT-Net under different SNR and target-range conditions, together with the receiver operating characteristic (ROC) and visualization results, demonstrating its advantages in weak-target detection and false-alarm suppression. Detailed experimental results are provided in Appendix C.

Conclusion. This study presented UAVRadar, a large-scale benchmark for low-altitude UAV detection in complex urban radar-clutter environments, and proposed STCT-Net for detecting weak and small UAV targets. STCT-Net integrates PCA-based chirp sampling, spatio-temporal channel transfer, backbone optimization, and Soft-BCE supervision to improve target representation and suppress background interference. Experiments on UAVRadar show that STCT-Net achieves 91.16% mAP_{50}, 82.41% mAP_{70}, 76.55% $\text{mAP}_{50:90}$, and 81.14% $\text{mAR}_{50:90}$, outperforming the best competing methods while maintaining 6.33M parameters and 191.20 GFLOPs. These results demonstrate its effectiveness and efficiency for urban UAV radar perception. However, UAVRadar currently focuses mainly on single-UAV scenarios and provides limited coverage of mixed maneuvers and adverse weather conditions. Future work will extend the benchmark to multiple UAV types, multi-UAV scenarios, mixed motion patterns, and broader weather conditions to improve model generalization and practical robustness.

Acknowledgements This work was supported in part by Guangdong Science and Technology (S&T) Program (Grant No. 2025B010120004), Guangxi Key Research and Development (R&D) Program (Grant No. 2024AB08139), and Guangzhou Science and Technology (S&T) Program (Grant No. 2025B03J0064).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

Availability of data Our code and dataset can be found at https://github.com/jackychouLab/STCT-Net.

References

1 Sun Y, Abeywickrama S, Jayasinghe L, et al. Micro-Doppler signature-based detection, classification, and localization of small UAV with long short-term memory neural network. IEEE Trans Geosci Remote Sens, 2021, 59: 6285–6300
2 Yu X, Alhujaili K, Cui G, et al. MIMO radar waveform design in the presence of multiple targets and practical constraints. IEEE Trans Signal Process, 2020, 68: 1974–1989
3 Wang Y, Jiang Z, Li Y, et al. RODNet: a real-time radar object detection network cross-supervised by camera-radar fused object 3D localization. IEEE J Sel Top Signal Process, 2021, 15: 954–967
4 Yuan S, Yang Y, Hoang T, et al. MMAUD: a comprehensive multi-modal anti-UAV dataset for modern miniature drone threats. In: Proceedings of the 2024 IEEE International Conference on Robotics and Automation, 2024. 2745–2751
5 Han S-K, Jung Y H. A time-synchronized multi-sensor drone dataset acquired from multiple radars and RF receiver. Scientific Data, 2026, 13: 1–18