# Exploring Window-Level Drone Delivery in Urban Air Logistics

Yiwen Wang *
School of Mechanical and Electrical Engineering, Soochow University, Suzhou, China
* Corresponding Author Email: 2329403033@stu.suda.edu.cn

**Abstract.** Window-level UAVs delivery applies to the emerging field of urban air logistics, enabling autonomous UAVs to deliver packages directly to building windows. However, unprecedented levels of precision, perception, and adaptability in complex, GPS-denied environments are required to achieve reliable window-level operation. This paper provides a systematic review of the key technological advances making this challenging scenario possible. How the fusion of visual, LiDAR, and inertial sensing, which is enhanced by deep learning and semantic SLAM, enables centimeter-level localization among urban canyons. Robust window detection is addressed through hybrid methods that combine geometric efficiency with deep semantic understanding, improving accuracy under architectural repetition and occlusion. Furthermore, intelligent navigation strategies have been developed to support real-time obstacle avoidance and smooth trajectory generation in cluttered airspace. These incorporate reinforcement learning with dynamic window constraints and factor graph optimization. Critical challenges remain in sensor resilience, energy efficiency during vertical operation, generalization across diverse urban forms, and scalable multi-UAV coordination despite progress. This review integrates current advancements and highlights future research needs, thereby providing a foundational reference for the development of reliable, precise, scalable and adaptable window-level UAV delivery systems in smart cities.

**Keywords:** Window-level UAVs delivery; sensor fusion; urban air logistics.

## 1. Introduction

The growing demand for fast and efficient urban logistics has driven significant interest in unmanned aerial vehicles (UAVs) as an innovative solution for last-mile delivery. Window-level UAVs delivery is recognized as one of the most ambitious frontiers in this domain, with drones autonomously navigating dense urban environments to deliver packages directly to specific building windows. By avoiding congested road networks, this method allows UAVs to reach upper floors of tall buildings autonomously and minimizes the dependence on manual delivery personnel. However, it poses unprecedented technical challenges, including centimeter-level positioning accuracy, robust window detection under occlusion and architectural diversity, and real-time path planning in dynamic, GPS-denied environments. In contrast to rooftop or ground-level delivery, window docking requires joint operation of high-precision perception, semantic understanding, and intelligent decision-making, making it a critical testbed for advanced autonomous systems.

Recent research has led to significant advances in addressing various components of urban drone logistics. Vision-language navigation frameworks such as LogisticsVLN enable UAVs to interpret natural language instructions for window-level terminal delivery and navigate complex urban scenarios without predefined maps, leveraging foundation models to bridge semantic goals with physical actions [1]. This agentic capability demonstrates the potential for goal-driven navigation in unstructured environments. Meanwhile, optimization-based approaches in this study have focused on energy-efficient routing, multi-drone coordination, and battery-aware scheduling, proposing reinforcement learning and heuristic algorithms to minimize flight time and collision risk in cluttered airspace [2]. Further system-level modeling efforts have introduced risk-aware navigation and genetic algorithm-based task planning, which coordinates delivery and recharge tasks to enhance delivery reliability under dynamic constraints such as payload limits, UAV energy or speed restrictions, and random task deadlines [3].

Building upon these advances, this paper discusses three core technical sectors for window-level UAV delivery.

First, high-precision positioning is achieved through the integration of visual localization, deep learning-enhanced SLAM (Simultaneous Localization and Mapping), and LiDAR-based relative navigation. By fusing absolute georeferenced data with relative motion estimation, the system suppresses drift and maintains robustness in GPS-denied areas.

Second, window recognition is addressed through a hybrid approach combining geometry-driven point cloud slicing for efficiency and deep learning-based instance segmentation for accuracy, while keypoint-heatmap fusion improves grouping robustness in repetitive facade layouts.

Third, path planning leverages reinforcement learning integrated with dynamic window fusion for real-time obstacle avoidance, and with positioning-coordinated formation control for multi-UAV collaboration in complex urban environments. Together, these components form a cohesive architecture that advances the feasibility of safe, precise, and scalable window-level drone delivery.

## 2. High-Precision Positioning Technology in UAV Window-Level Delivery Scenarios

High-precision positioning is the core prerequisite for unmanned aerial vehicles (UAVs) in window-level delivery, where sub-meter or even centimeter-level accuracy is required to ensure safe and precise navigation in dense urban environments. Traditional GPS-based methods face significant limitations due to signal blockage, multipath effects, and dynamic obstacles, making them inadequate for such stringent requirements. As a result, alternative high-precision localization technologies have emerged as critical solutions, including visual localization, SLAM enhancement, and LiDAR-based systems. This section evaluates these approaches in terms of accuracy, robustness, and real-time applicability for UAV self-localization in complex urban settings.

### 2.1. Visual Localization Technology: From Feature Matching to Deep Learning

Traditional visual localization mainly relies on template or feature-based matching. These methods often fail to meet the precision requirements for window-level delivery: specifically, template matching suffers from high computational costs and sensitivity to changes in georeferenced data, and feature-based matching is prone to mismatches and drift accumulation, especially at varying UAV altitudes.

Deep learning has significantly advanced aerial visual localization by utilizing end-to-end feature extraction. Convolutional Neural Networks (CNN)-based models, such as a pre-trained VGG16 employed by Amer et al, have demonstrated improved recognition of urban structures like building facades and roads, achieving up to 91.2% matching accuracy by leveraging pre-trained "deep urban signatures" [4]. While such models enhance robustness, architectural efficiency requires optimization. For instance, although the CNN architecture PoseNet combined with Long Short-Term Memory (LSTM) was designed to improve localization continuity, experimental results indicate that the simplified CompactPN model achieve comparable accuracy with faster inference, demonstrating the real-time requirements of UAVs achieved by deep learning without losing accuracy.

Visual localization systems are broadly categorized into absolute AVL and relative RVL methods. AVL uses georeferenced data (e.g., satellite imagery) to estimate the global coordinate, effectively avoiding cumulative errors in RVL. Integrating both of them by combining visual odometry (VO) with satellite image matching enables joint optimization of relative motion and absolute position errors. This fusion leverages the continuity of VO and the global reference of map matching, effectively suppressing drift and improving localization accuracy, thus offering a viable multi-source strategy for UAV navigation in GPS-denied urban canyons.

### 2.2. SLAM Enhancement: Robustness in Dynamic Urban Environments

SLAM is essential in unknown or constructing environments where prior maps are unavailable. However, traditional SLAM frameworks, such as ORB-SLAM3, rely on static scene assumptions and are susceptible to dynamic elements (e.g., pedestrians, vehicles, or moving objects), which might reduce localization accuracy.

To address this, deep learning-enhanced SLAM systems integrate real-time object detection model (e.g., YOLOv5) to identify dynamic targets [5]. By excluding dynamic feature points from pose estimation, these systems preserve the static environmental features such as building edges and road structures. Optical flow is further employed to track feature trajectories, improving the discrimination between moving objects and static backgrounds. Experimental results show that such fusion reduces Absolute Trajectory Error (APE) by 43.47% and Relative Pose Error (RPE) by 26.47% compared to ORB-SLAM3 in dynamic scenes, while maintaining average tracking time in 34 ms per frame, meeting real-time requirements for UAV window-level delivery.

Semantic segmentation further enhances map stability by extracting key static features like building contours. These semantically segmented features are refined and matched through a Semantic Shape Matching (SSM) process, providing robust geometric and contextual constraints that improve homography estimation and pose accuracy in urban GPS-denied environments.

### 2.3. LiDAR Localization: High-Precision Relative Positioning and Adaptive Control

LiDAR offers high-accuracy 3D point cloud data that is unaffected by lighting conditions, making it well-suited for close-range UAV operations. In window-level delivery, where precise relative positioning between the UAV and the target window is vital, LiDAR-based methods perform exceptionally well [6]. By clustering point cloud data (e.g., via k-means) and extracting center coordinates of static structures, such as walls and architectural fixtures, these systems compute real-time relative poses with reported accuracy as high as 0.02 m, significantly outperforming visual localization in proximity UAV tasks.

To enhance UAV self-motion states in complex environments, deep reinforcement learning has been integrated into LiDAR localization [6]. Using algorithms like Deep Deterministic Policy Gradient (DDPG), UAVs learn to optimize their localization and motion strategies using collision risk assessment maps. Collision risk assessment maps enable autonomous posture adjustments, mitigating errors caused by occlusions or sparse features. Additionally, dynamic prioritized experience replay accelerates training convergence by emphasizing high-impact samples, improving the adaptive regulation capability of UAV motion states (velocity, attitude) in complex scenarios.

In multi-UAV delivery scenarios, LiDAR also enables high-precision inter-drone relative positioning. Through formation control algorithms, it maintains the relative distance and attitude of the drone swarm, providing support for collaborative perception of the coordinates and motion states of multiple UAVs in dense building areas.

## 3. Window Recognition in UAV Window-Level Delivery Scenarios

In UAV-based window-level delivery, precise window detection bridges high-precision localization and final delivery execution. After reaching the target building, the UAV must locate the correct window from visual or point cloud data, assessing its geometric boundaries and environmental status (e.g., occlusion, obstacles). This task is complicated by urban architectural diversity, sensor variability, and limited annotated data, particularly from UAV perspectives. Current approaches fall into three categories: geometry feature-driven, deep learning-driven, and keypoint-heatmap fusion methods, which will be discussed below.

### 3.1. Geometry-Driven Detection: Efficiency vs. Robustness in Point Cloud Processing

Image-domain window detection leverages visual sensors on UAVs and deep learning models for window localization. Traditional approaches, such as Façade Delaunay (FD), employ Delaunay triangulation to detect window gaps by identifying long edges in point cloud meshes [7]. While effective for regular rectangular windows (98.5% accuracy), FD suffers from high computational complexity (requiring up to 16 hours to process 2.6 million points) and fails on non-rectangular or curved windows. It also demands high point cloud density (>175 pts/m²), limiting its practicality in real-world UAV operations.

In contrast, point cloud slicing methods significantly improve efficiency by transferring 3D point cloud analysis to 1D distance calculation. By aligning a local coordinate system with the building facade and slicing horizontally and vertically, these methods project points onto a single axis and identify gaps where inter-point distances exceed twice the median. This dimensionality reduction enables processing in 2.5 seconds, representing a 2300 times speedup over FD. Moreover, it maintains 93–96% accuracy on complex facades, including historical buildings with curved or protruding windows. It remains functional even at low point densities (31 pts/m²), enhancing robustness in sparse data conditions.

However, slicing methods are vulnerable to occlusion. Air conditioners or vegetation can create artificial gaps, leading to false detections and boundary displacement. While post-processing compensation can improve accuracy (e.g., from 94.8% to 98.1%), such limitations highlight the need for occlusion-aware preprocessing, especially in cluttered urban environments.

### 3.2. Deep Learning-Driven Detection: Balancing Annotation Cost and Precision

Deep learning approaches for extracting rich visual features are divided into two categories of methods: "few-shot cross-view recognition" and "pixel-level instance segmentation". Shengming Li proposed the Hybrid Convolutional-Transformer (HCT) framework, addresses the challenge of limited labeled data [8]. Fig. 1 is the architecture of the proposed module, including SWTB, MPCA, and SFA. The left, middle and right panels indicate the SWTB, MPCA, and SFA, respectively. The $\oplus$ is the element-wise addition and $\otimes$ is the element-wise multiplication. It uses satellite and street-view images as support sets to assist UAV-view window localization, requiring few image-level labels. By integrating Vision Transformers with CNNs through the HCT framework, ViT captures global facade patterns across views in shallow layers, while deep CNNs extract local window features such as frame edges and glass reflection textures. HCT integrates global and local features via Multi-Path Cross Attention (MPCA) or Spatial Fusion Attention (SFA), achieving strong few-shot building detection on drone images with minimal supervision.

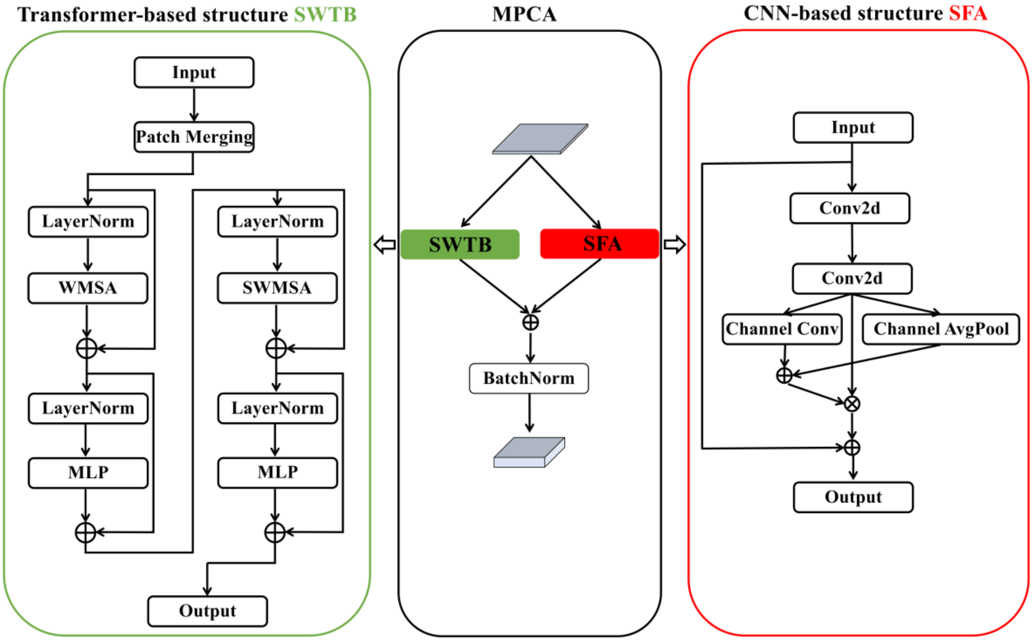


**Fig 1.** Architectures of the proposed module

Nils Nordmark and Mola Ayenew proposed a Mask R-CNN method for window instance segmentation, using ResNet101-FPN as the backbone [9]. The Region Proposal Network (RPN) generates 9 anchors per location, retaining high-probability window proposals. A mask head after RoIAlign outputs 28×28 soft masks, enabling pixel-accurate contour extraction aligned with bounding boxes. Trained on a self-collected dataset of 100 images (1,540 windows) with precise outer-frame, the model leverages the transference of COCO dataset pre-training and the combination of data augmentation strategies. It achieves 94.76% accuracy on the eTRIMS dataset, outperforming WD-Net (~90%) and DeepFacade (91%). However, it requires pixel-level annotations, making labeling labor-intensive and limiting adaptability to new areas.

In contrast, HCT needs only image-level labels and fuses cross-view data, including satellite rooftops and street-view details, to assist UAV-view detection. It runs at approximately 40 FPS, suitable for high-speed flight up to 30 m/s, while Mask R-CNN processes at 10–15 FPS due to heavy pixel-level computation, favoring high-precision, low-speed scenarios. HCT is ideal for unannotated areas; Mask R-CNN excels where accuracy is prioritized. Future integration could use HCT to pre-label data and reduce Mask R-CNN’s annotation burden.

### 3.3. Keypoint-Heatmap Fusion: Robust Grouping in Repetitive Layouts

Keypoint-based methods detect window corners and centers, then group them into instances [10]. Traditional Part Affinity Fields (PAFs) learn vector fields between keypoints but struggle in dense, repetitive window arrays. Adjacent windows with similar geometries often lead to incorrect associations (e.g., linking corners across instances), reducing grouping accuracy ($AP_{50}$: 86.4% in facades with >30 windows).

The heatmap fusion approach improves robustness by decoupling detection from geometric association. It outputs a 5-channel heatmap (four corners + center) and a tagmap assigning embedding labels to keypoints. A "pull" loss minimizes intra-window tag variance, while a "push" loss maximizes inter-window differences ($\Delta$=1), enabling clustering-free grouping. A center verification module further filters invalid proposals by checking the center heatmap response (threshold: 0.23), reducing false positives to below 4.6%.

Evaluated on a dataset of 3,418 images, this method achieves 91.4% precision and 91.0% recall, outperforming PAFs (89.4%, 90.7%). Its tag-based grouping is less sensitive to window layout regularity, achieving 92.1% $AP_{50}$ in dense scenarios. This makes it particularly suitable for urban high-rises with uniform window patterns.

Future enhancements could integrate LiDAR-derived 3D geometry to improve keypoint accuracy under occlusion. Additionally, adaptive push-loss thresholds could better distinguish complex shapes (e.g., arched vs. rectangular windows), expanding shape generalization.

## 4. Path Planning in UAV Window-Level Delivery Scenarios

Effective path planning is critical for UAVs in window-level delivery, where navigation must balance safety, efficiency, and real-time responsiveness in complex urban environments. Key challenges include dynamic obstacle avoidance, formation control in GPS-denied environments, and error accumulation. Two dominant technologies have emerged: reinforcement learning combined with dynamic window fusion, and reinforcement learning integrated with positioning-coordinated formation control. These approaches address distinct operational demands—single-agent navigation in cluttered airspace and multi-UAV coordination in signal-denied zones, respectively.

### 4.1. Reinforcement Learning with Dynamic Window Fusion: Enhancing Local Obstacle Avoidance

Chuanbo Wu, at el. Proposed a method combining deep reinforcement learning (DRL) with the dynamic window algorithm (DWA) to form the DWA-PPO method [11]. The DWA-PPO method integrates the sampling efficiency of the Dynamic Window Approach (DWA) with the policy optimization power of Proximal Policy Optimization (PPO), targeting real-time obstacle avoidance in dynamic urban airspace. DWA first constrains the feasible velocity space by considering vehicle dynamics, sensor range, and acceleration limits. This structured action space is then used to guide

PPO’s exploration, significantly reducing inefficient sampling caused by sparse rewards in pure reinforcement learning.

Compared to traditional DWA, which relies on greedy local optimization and often results in suboptimal global paths due to excessive heading changes (e.g., 4.2 rad), DWA-PPO generates smoother trajectories with reduced cumulative orientation shifts (1.89 rad). Against pure PPO, it achieves faster convergence—reaching stable performance in 100 iterations versus 300—by leveraging DWA’s heuristic constraints to bias exploration toward physically plausible actions. This structural guidance mitigates the "reward sparsity" problem to enable more efficient learning of safe and efficient flight policies.

In scenarios with multiple dynamic obstacles, the DWA-PPO algorithm achieves superior performance with a 0% collision rate and computation time of 2.5 seconds for real-time decision making. By comparison, traditional DWA and pure PPO require 3.8 s and 4.1 s, respectively, with collision rates of 5.3% and 2.1%. Notably, DWA-PPO’s single-action inference time of 0.04 seconds enables control frequencies exceeding 10 Hz, meeting real-time flight requirements. Fig. 2 demonstrates the integrated architecture, where DWA's velocity space sampling provides heuristic constraints for PPO's policy optimisation. The success of the proposed method is attributed to a multi-objective reward structure that balances safety (via obstacle distance penalties), speed (through forward velocity incentives), and task completion (with terminal rewards), enabling stable policy convergence without oscillatory behavior.

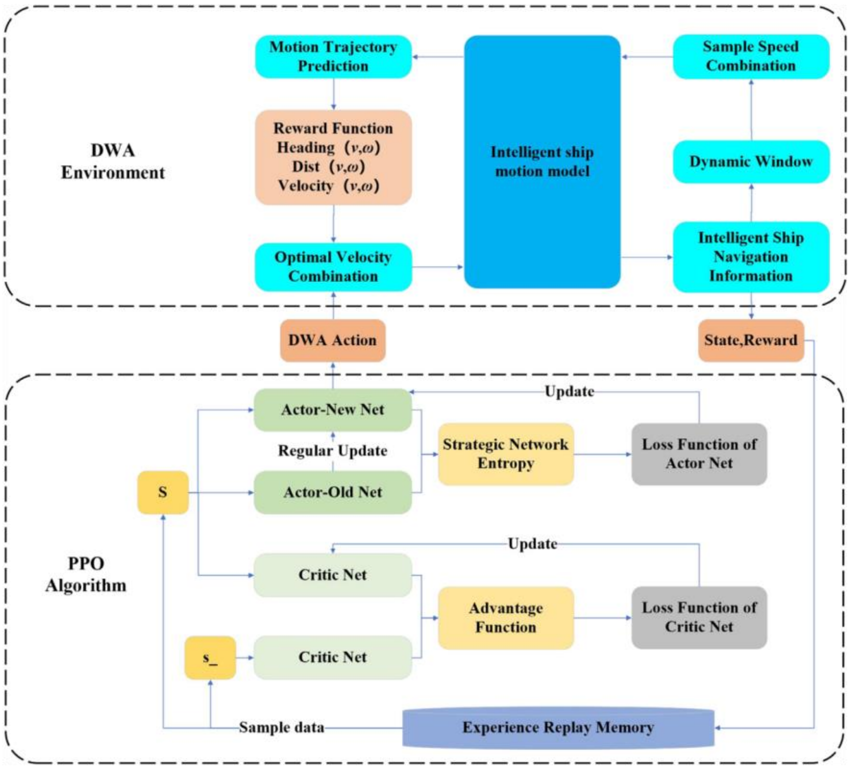


**Fig 2.** Improved algorithm structure combining PPO and DWA

However, DWA-PPO remains dependent on perceiving environment accurately. Obstacle prediction degrades under limited sensor range or poor texture conditions. Integration with monocular visual SLAM could enhance environmental modeling by supplementing geometric data with semantic texture cues. Additionally, incorporating energy-aware terms, such as penalties on vertical climb rate or aggressive acceleration, could enhance flight efficiency while maintaining safety.

### 4.2. Reinforcement Learning with Positioning-Coordinated Formation Control: Robustness in GPS-Denied Environments

Motivated by recent progress in UAV navigation, sophisticated approaches have been developed to address urban operational challenges using probabilistic optimization and integrated control frameworks. The approach proposed by Runqi Chai establishes a convex optimization framework that systematically handles sensor uncertainties by transforming chance constraints into deterministic convex approximations [12]. This formulation uses the bounding function $\Psi(g(u,\xi)) = (g(u,\xi)+1)^{+}$ to maintain convexity while ensuring 95% confidence in constraint satisfaction.

Additionally, the optimized trajectories generated by the proposed framework are compared with those produced by the nonlinear pseudospectral algorithm, demonstrating that the convex approach achieves superior smoothness in both state and control profiles and effectively eliminates the high-frequency oscillations present in NPS-derived control signals, a characteristic that is vital for real-world actuator performance. Experimental results confirm its efficiency: it completes trajectory generation in 2.5 seconds per decision cycle with no constraint violations, outperforming conventional methods that take 3.8 seconds and have a 5.3% violation rate under similar urban conditions.

Building on this, Peiwen Yang and Weisong Wen's Joint Positioning-Control Method (JPCM) rethinks the sensor-to-actuation pipeline via factor graph optimization [13]. It unifies GNSS/LiDAR measurements (as positioning factors) and model predictive control objectives (as dynamic constraints) into a cohesive network. Fig. 3 is the paths of Nominal MPC and JPCM-GI. The black dashed line, red solid line, and blue solid line represent the reference path, the tracking path based on nominal MPC, and the tracking path based on JPCM-GI, respectively. This figure illustrates JPCM's advantage: the blue solid line (JPCM-GI path) fits the reference circular trajectory (black dashed line) far better than the red solid line (nominal MPC path, with severe jitter). Quantitatively, JPCM reduces x-axis positioning error by 40% (from 0.0434m to 0.026m RMSE) and improves trajectory smoothness by 55% (lowering accumulated heading changes from 4.2 rad to 1.89 rad).

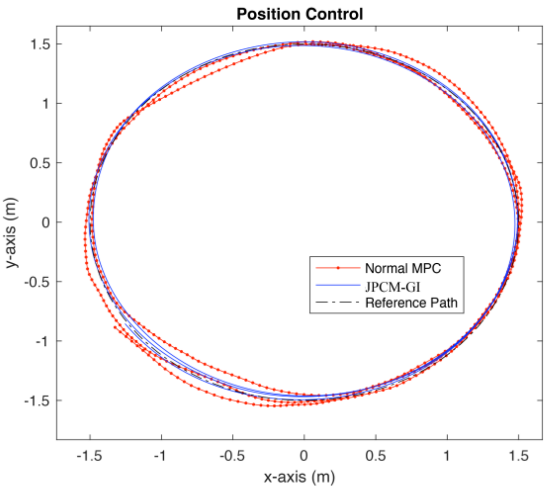


**Fig 3.** The paths of Nominal MPC and JPCM-GI

Practical urban evaluations confirm JPCM's viability, with 0.35m relative positioning accuracy during complex maneuvers and 3-second recovery from perturbations. Yet challenges remain: communication latency causes 0.12m tracking errors, and sensor range below 50m degrades performance in dense urban canyons. Future research will focus on adaptive delay compensation, context-aware sensor fusion, and energy-aware cost functions, while integrating chance-constrained optimization with factor graphs holds promise for safer urban air mobility.

## 5. Conclusion

Window-level UAV delivery represents a transformative frontier in urban logistics, offering a promising solution to the growing demand for fast, efficient, and straightway last-mile delivery in dense metropolitan environments. This paper has systematically explored the core technological problems that enable safe, precise, and scalable autonomous drone delivery to building windows, including high-precision positioning, robust window recognition and intelligent path planning. By analyzing related studies in visual localization, deep learning-enhanced SLAM, LiDAR-based relative navigation, hybrid detection methods, and reinforcement learning-driven planning, this paper highlighted how these individual technologies address the unique challenges of GPS-denied, dynamic, and structurally complex urban environments.

The fusion of absolute and relative localization—leveraging deep learning-based visual georeferencing, semantic SLAM, and LiDAR-relative navigation—ensures centimeter-level accuracy even under signal blockage and environmental variability. These multi-sensor integration strategies effectively suppress drift and enhance robustness, forming a reliable spatial foundation for subsequent mission-critical tasks. In window recognition, it is demonstrated that hybrid approaches combining geometric efficiency with deep semantic understanding offer a balanced solution: point cloud slicing enables rapid facade analysis, while instance segmentation and keypoint-heatmap fusion ensure high accuracy and grouping robustness, especially in repetitive or occluded high-rise layouts. Furthermore, the integration of reinforcement learning with dynamic window fusion and formation control has shown significant promise in enabling real-time obstacle avoidance and multi-UAV coordination, achieving smoother trajectories, faster convergence, and improved safety in cluttered airspace.

While progress has been made, several challenges continue to exist. Sensor degradation under adverse weather, limited generalization across diverse architectural styles, and energy constraints during vertical operation have not been thoroughly investigated. To address current limitations, future work should focus on adaptive sensor fusion for robustness in dynamic urban environments, AI-based lifelong learning to improve window detection across diverse architectures, and energy-aware motion planning for efficient vertical navigation. Additionally, the development of scalable coordination strategies represents a critical requirement for multi-UAV operations in high-density airspace. As urban air mobility systems and regulations continue to advance, integrating reliable perception, precise localization, and intelligent decision-making will be key to enabling safe, autonomous window-level drone delivery within smart cities.

## References

[1] Zhang, X., Tian, Y., Lin, F., Liu, Y., Ma, J., Szatmáry, K. S., & Wang, F. Y. (2025). LogisticsVLN: Vision-Language Navigation For Low-Altitude Terminal Delivery Based on Agentic UAVs. arXiv preprint arXiv:2505.03460.

[2] Tomasz Dudek, Karolina Kaśkosz, Optimizing drone logistics in complex urban industrial infrastructure, Transportation Research Part D: Transport and Environment, Volume 140, 2025, 104610, ISSN 1361-9209, https://doi.org/10.1016/j.trd.2025.104610.

[3] Rinaldi, M., Primatesta, S., Bugaj, M., Rostáš, J., & Guglieri, G. (2024). Urban Air Logistics with Unmanned Aerial Vehicles (UAVs): Double-Chromosome Genetic Task Scheduling with Safe Route Planning. Smart Cities, 7(5), 2842-2860.

[4] O. Y. Al-Jarrah, A. S. Shatnawi, M. M. Shurman, O. A. Ramadan and S. Muhaidat, "Exploring Deep Learning-Based Visual Localization Techniques for UAVs in GPS-Denied Environments," in IEEE Access, vol. 12, pp. 113049-113071, 2024.

[5] Norbelt, M., Luo, X., Sun, J., & Claude, U. (2025). UAV Localization in Urban Area Mobility Environment Based on Monocular VSLAM with Deep Learning. Drones, 9(3), 171.

[6] Bodi MA, Zhenbao LIU, Feihong JIANG, Wen ZHAO, Qingqing DANG, Xiao WANG, Junhong ZHANG, Lina WANG, Reinforcement learning based UAV formation control in GPS-denied environment, Chinese Journal of Aeronautics, Volume 36, Issue 11, 2023, Pages 281-296, ISSN 1000-9361.

[7] S.M. Iman Zolanvari, Debra F. Laefer, Slicing Method for curved façade and window extraction from point clouds, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 119, 2016, Pages 334-346, ISSN 0924-2716.

[8] Shengming Li, Linsong Xue, Lin Feng, Cuili Yao, Dong Wang, Hybrid Convolutional-Transformer framework for drone-based few-shot weakly supervised object detection, Computers and Electrical Engineering, Volume 102, 2022, 108154, ISSN 0045-7906.

[9] Nordmark, N., & Ayenew, M. (2021). Window Detection In Facade Imagery: A Deep Learning Approach Using Mask R-CNN. ArXiv, abs/2107.10006.

[10] Li, CK., Zhang, HX., Liu, JX. et al. Window Detection in Facades Using Heatmap Fusion. J. Comput. Sci. Technol. 35, 900–912 (2020).

[11] Chuanbo Wu, Wangneng Yu, Guangze Li, Weiqiang Liao, Deep reinforcement learning with dynamic window approach based collision avoidance path planning for maritime autonomous surface ships, Ocean Engineering, Volume 284, 2023, 115208, ISSN 0029-8018.

[12] R. Chai, A. Tsourdos, A. Savvaris, S. Wang, Y. Xia and S. Chai, "Fast Generation of Chance-Constrained Flight Trajectory for Unmanned Vehicles," in IEEE Transactions on Aerospace and Electronic Systems, vol. 57, no. 2, pp. 1028-1045, April 2021.

[13] P. Yang and W. Wen, "Tightly Joining Positioning and Control for Trustworthy Unmanned Aerial Vehicles Based on Factor Graph Optimization in Urban Transportation," 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), Bilbao, Spain, 2023, pp. 3589-3596.