

Active Contact Engagement for Aerial Navigation in Unknown Environments with Glass

Xinyi Chen, Yichen Zhang, Hetai Zou, Junzhe Wang and Shaojie Shen

Abstract—Autonomous aerial robots are increasingly being deployed in real-world scenarios, where transparent glass obstacles present significant challenges to reliable navigation. Researchers have investigated the use of non-contact sensors and passive contact-resilient aerial vehicle designs to detect glass surfaces, which are often limited in terms of robustness and efficiency. In this work, we propose a novel approach for robust autonomous aerial navigation in unknown environments with transparent glass obstacles, combining the strengths of both sensor-based and contact-based glass detection. The proposed system begins with the incremental detection and information maintenance about potential glass surfaces using visual sensor measurements. The vehicle then actively engages in touch actions with the visually detected potential glass surfaces using a pair of lightweight contact-sensing modules to confirm or invalidate their presence. Following this, the volumetric map is efficiently updated with the glass surface information and safe trajectories are replanned on the fly to circumvent the glass obstacles. We validate the proposed system through real-world experiments in various scenarios, demonstrating its effectiveness in enabling efficient and robust autonomous aerial navigation in complex real-world environments with glass obstacles.

Index Terms—Aerial Systems; Perception and Autonomy; Vision-Based Navigation; Software-Hardware Integration for Robot Systems.

I. INTRODUCTION

AUTONOMOUS aerial robots are increasingly expanding their roles beyond laboratory settings, undertaking real-world applications such as autonomous exploration [1], industrial inspection [2], and search and rescue operations [3]. A practical scene often features glass walls and windows, but exteroceptive sensors like cameras and LiDARs can suffer from them. These transparent surfaces can cause unreliable sensor measurements, posing significant challenges to autonomous robot navigation in real-world scenarios.

Researchers have explored various methods to address the challenge of detecting transparent glass obstacles in the environment. On the one hand, some studies have focused on detecting glass using a variety of non-contact sensors, including RGB cameras [4], LiDARs [5], etc. Leveraging advancements in machine learning, these methods can achieve high accuracy

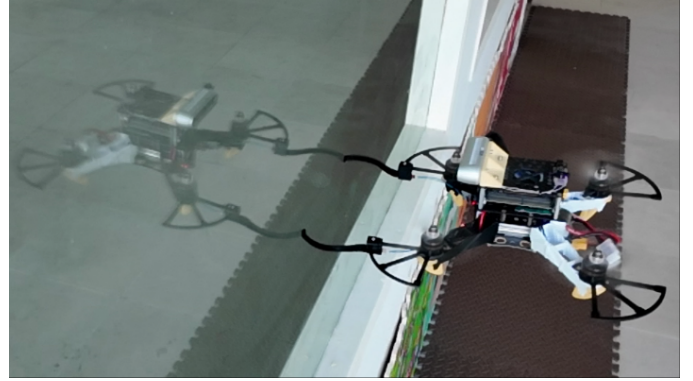


Fig. 1. A real-world experiment photo captured during a touch action, when the aerial vehicle actively contacts with a visually detected glass surface to confirm its presence using a pair of front-mounted contact-sensing modules.

in detecting transparent obstacles without physical contact. However, their performances are often influenced by factors such as lighting conditions [6] and angular uncertainty [7]. While achieving perfect accuracy in all circumstances remains challenging, a single detection failure can block a practically viable path in the case of false positive or even cause a crash for false negative. On the other hand, contact-resilient aerial vehicles have been designed to withstand collisions with unexpected obstacles [8, 9]. While these mechanical designs help prevent crashes, they often come with increased weight and complexity, which may compromise flight duration and narrow space maneuverability. Although contact information has been interpreted as discrete points and utilized for mapping, numerous collisions are often required to effectively model the transparent surfaces [10], which is inefficient and impractical. Furthermore, during autonomous navigations, these methods primarily rely on passive collision response [11], limiting the utility of contact information as a source of sensing about the environment. These concerns hinder the widespread adoption of autonomous aerial navigation in practical scenarios.

Imagine standing in front of a pristine glass door: you may first notice and suspect its presence, and then instinctively reach out to confirm it through touch. Motivated by this natural behavior, we propose a novel approach for robust autonomous aerial navigation in complex real-world environments with transparent glass obstacles. The proposed system combines the strengths of both sensor-based and contact-based glass detection methods, where active contact engagement provides reliable complementary information to a sensitive visual glass detector by offering true positive confirmation and false positive invalidation. When receiving new camera images, potential glass surfaces are detected from the color image, and their 3D point clouds are also computed. The informa-

Manuscript received: January 24, 2025; Revised April 2, 2025; Accepted April 28, 2025. This paper was recommended for publication by Editor Pascal Vasseur upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by HKUST Postgraduate Studentship, Research Grants Council Research Matching Grant Scheme Project RMGS20EG20, and HKUST-DJI Joint Innovation Lab. (Xinyi Chen, Yichen Zhang and Hetai Zou contributed equally to this work.) (Corresponding author: Xinyi Chen.)

All authors are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, China (e-mail: {xhencq,yzhangec,hzouah,jwanggj}@connect.ust.hk, eeshaojie@ust.hk).

Digital Object Identifier (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

tion of each potential glass surface is incrementally updated during the navigation. We integrate a lightweight contact-sensing module employing flex sensors, which provides touch sensing confirmation of the detected glass surfaces without hard collisions that risk unnecessary damage to both the agent and the glass. For robust autonomous navigation, the proposed system is able to generate navigation trajectories that actively engage touch actions (Fig. 1) with the visually detected glass surfaces consciously. If a touch event is detected, the presence of the glass surface is confirmed and precisely localized using its information and the vehicle pose at the moment of contact. The volumetric mapping is also updated on the fly, requiring only a single touch per glass surface. Otherwise, the potential glass surface is invalidated. We incorporate the proposed system with point-to-point aerial navigation tasks in unknown environments, aiming for more robust autonomous navigations in real-world scenes.

We evaluate the effectiveness of the incremental glass surface detection module in a glass-rich corridor, and the advantage of active contact planning compared with baselines in simulation benchmarks. We also conduct a series of autonomous point-to-point navigation experiments in three distinct real-world scenarios to validate the proposed system. The experimental results demonstrate the system's capability of safe and robust navigation in real-world environments with transparent glass obstacles. In summary, the contributions are as follows:

- 1) An incremental visual transparent glass surface detection method, that extracts glass regions from camera images and maintains essential information about potential glass surfaces in the environment.
- 2) A lightweight contact-sensing module that provides touch confirmation of potential glass surfaces and invalidate false detections for non-glass surfaces, minimizing the risks associated with hard collisions while ensuring navigation efficiency.
- 3) An autonomous aerial navigation system that actively engages contacts with visually detected glass surfaces for safe and robust navigation in real-world scenarios. To the best of our knowledge, this is the first work to address the challenge posed by transparent glass surfaces in aerial navigation through an active approach.

II. RELATED WORK

A. Glass Detection with Non-Contact Sensors

Extensive research has been conducted on detecting transparent glass surfaces using a wide range of non-contact sensors. Color cameras are among the most commonly used sensors for this purpose. Studies in computer vision typically focus on glass identification and segmentation in a single color image, leveraging cues such as contextual information [4, 12], glass blurriness [6], and boundary information [13]. Ultrasonic sensors offer an alternative approach for detecting glass obstacles, but they are often limited by angular uncertainty [7] and require fusion with RGBD cameras [14] or laser scan data [15] for reliable detection. Laser rangefinders utilize the specular reflection of laser beams from the glass surface, achieving an

accuracy rate of 95% [16]. Photoelectric sensors are widely used in industrial settings to inspect transparent objects, but their limited field of view and fixed installation requirements make them unsuitable for mobile robots [17]. Polarized light cameras have also been studied for identifying transparent objects [18] and measuring depth [19]. Nevertheless, they are highly susceptible to ambient lights in real-world environments and remain expensive due to limited availability. Given the inherent limitations of non-contact sensors in achieving perfect accuracy, our work adopts a hybrid approach by combining non-contact sensor detections with physical contact to enhance the reliability of glass detection in practical applications.

B. Contact-Resilient Aerial Vehicle Design

Numerous studies have focused on developing aerial robots that can withstand collisions. For instance, an early work [8] designs a protective frame with a gimbal system that allows it to rotate independently from the inner body. The Tiercel [10] consists of a carbon fiber frame and collision bumper to sustain collision impacts. The ARQ [11] introduces a compliant arm design with Hall effect sensors for collision detection, which is later evolved to the s-ARQ [20] with a single compliant arm. The RMF-Owl [21] employs a carbon-form sandwich material for main rigid component fabrication and integrates passive flaps to count collisions and adjust flight aggressiveness. A survey [9] reviews various other designs for contact-resilient aerial vehicles. However, most of these contact-resilient aerial vehicle mechanical designs are overly complicated, heavy and cumbersome. Although the origami-inspired design Rotorigami [22] provides a lightweight protection system, it only supports passive collision sustenance since no sensor is equipped for collision detection and characterization. Different from these approaches, our system consciously engages gentle contact with the environment using a lightweight yet effective contact-sensing module, which is easy to integrate with most quadrotors without much modifications to the structure.

C. Autonomous Navigation with Contacts and Active Methods

With the development of contact-resilient mobile robots, researchers have investigated various roles of contacts in autonomous vehicles with different emphases. Some works leverage collisions to generate pseudo-measurements for localization to estimate robot states and reduce odometry uncertainty [23, 24]. Others focus on collision recovery control, enabling robots to recover from unexpected collisions and return to a normal operational state [25, 26]. Additionally, some studies aim to improve mapping by detecting passive collisions with unseen obstacles and estimating their positions in the map [27, 28]. However, only a few studies have considered utilizing contact information for autonomous navigation in unknown environments with transparent glass obstacles. Mulgaonkar *et al.* [10] employ multiple passive collisions and their corresponding position to estimate the dimensions of a rectangular room with glass walls, which is inefficient and impractical for real-world applications.

Active methods have been studied for a wide range of autonomous robotics tasks since the pioneering work on active

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE I
INFORMATION STORED IN A GLASS SURFACE DATA STRUCTURE G_i

Notation	Description
\mathbf{c}_i	Centroid of the glass surface
\mathbf{n}_i	Normal of the glass surface
P_i	Polygon vertices of the glass surface boundary
C_i	Point cloud of the glass surface

perception [29] in the late 1980s. For instance, active localization drives the robot toward informative locations to improve its state estimation [30, 31]. Active SLAM focuses on building the most accurate and complete map of the environment [32, 33]. Active tracking enables a robot to dynamically adjust its motion to maintain an optimal view of a target [34, 35]. To the best of the authors' knowledge, this work is the first to introduce active contact engagement for autonomous aerial navigation in glass-rich unknown environments.

III. INCREMENTAL GLASS SURFACE DETECTION

While the glass detection problem has been extensively studied in the computer vision community [4, 6, 12, 13], these methods typically concentrate on a single image frame. For the purpose of autonomous navigation, we integrate a simple yet effective visual glass detection module (Sec. III-A) that can be deployed on edge devices for real-time processing. Furthermore, we develop an incremental approach (Sec. III-B) that fuses potential glass surface detections from consecutive frames, maintaining the information of glass surfaces in the environment in a data structure detailed in Table I.

A. Visual Glass Detection

On transparent glass obstacles, exteroceptive sensors like cameras and LiDARs can produce noisy or invalid measurements, due to a lack of opaque and detectable features. For autonomous navigation, we first need to efficiently detect potential glass surfaces from individual frames of onboard RGBD camera images.

The proposed visual glass surface detection module is detailed in Line 1-15 of Alg. 1. We observe that though glass surfaces are transparent, they often have visible edges and frames that can produce reliable measurements. Hence we attempt to utilize these boundaries to detect and extract glass surfaces from RGBD images. Firstly, the regions of glass segmentations \mathcal{S} are extracted from the color image \mathcal{I} with an object detection and instance segmentation framework YOLOv8 [36]. A segmentation with a confidence score exceeding a threshold τ_s is valid and considered as a potential glass surface (Fig. 2(a)). The boundary of each segmentation is dilated to obtain a set of boundary points H_i with stable depth measurements. We query the pixels of the boundary points in H_i in the synchronized and aligned depth image \mathcal{D} (Fig. 2(b)) and project them to the 3D world coordinates to obtain point set Q_i (Fig. 2(c)). Due to the noisy depth measurements, Q_i is further fitted into a plane π_i using the least square method RANSAC [37]. The inlier points are projected to the fitted plane π_i to obtain \hat{Q}_i . The convex hull of \hat{Q}_i is then computed, yielding the glass surface boundary as a polygon defined by

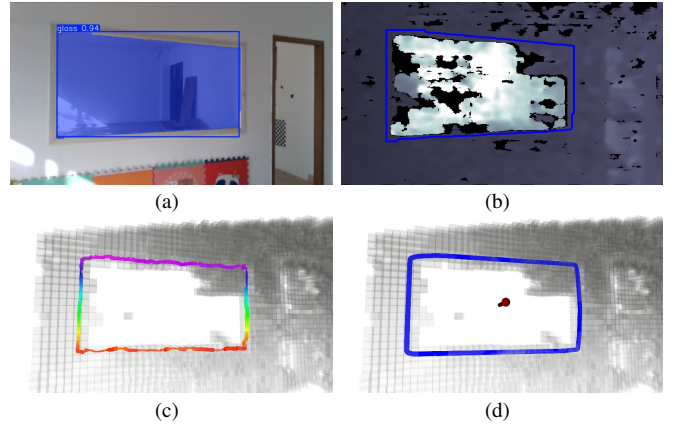


Fig. 2. An illustration of the proposed visual glass surface detection module. (a) Glass segmentation from the color image. (b) Boundary points in the depth image. (c) The 3D point set of the glass surface boundary in the world frame. (d) The polygon of the glass surface on the fitted plane. In (c)(d), the grey voxels are occupancy grids constructed from the depth image.

its vertices P_i (Fig. 2(d)). The centroid \mathbf{c}_i of the glass surface is computed as the average of the points in P_i , and the normal \mathbf{n}_i of the glass surface is estimated as the normal of the fitted plane π_i . Lastly, a 3D point cloud C_i is sampled within the polygon formed by P_i to represent the glass surface obstacle for volumetric mapping refinement in later stages. At this point, we have obtained the essential information of each potential glass surface G_i contained in the current single image frame.

B. Incremental Glass Surface Update Across Frames

As the glass surfaces may not be fully observed in a single frame, fusing the detection results across different frames becomes crucial for maintaining information of the glass surfaces in the environment. Whenever new sensor measurements are received during the navigation, results G_i from the visual glass detection in Sec. III-A should be aggregated with the existing glass surface list \mathcal{G} incrementally. The overall incremental update process is described in Line 16-22 of Alg. 1.

To fuse the detection results across different frames, data association is required to recognize whether the incoming glass surface G_i is an observation of one of the existing surfaces in \mathcal{G} or a new glass surface. For each newly detected glass surface G_i and each existing $G_j \in \mathcal{G}$, we first compute the angle between their normal directions $\theta(\mathbf{n}_i, \mathbf{n}_j)$ and the Euclidean distance between their centroids $d(\mathbf{c}_i, \mathbf{c}_j)$. An angle exceeding a threshold τ_n or a distance exceeding τ_c indicates that they are different surfaces. For surfaces with similar normal directions and centroids, we further compute their Intersection over Union (IoU) score $IoU(i, j)$. We project points in P_i to the plane holding G_j to obtain jP_i . Then we compute the 2D IoU score $IoU(i, j)$ as $IoU(i, j) = \text{area}(\mathcal{P}_i \cap {}^jP_j) / \text{area}(\mathcal{P}_i \cup {}^jP_j)$, where the polygons jP_i and \mathcal{P}_j are defined by vertices jP_i and P_j . For each incoming glass surface G_i , the highest IoU score $\hat{IoU}(i)$ and its index k can be computed by

$$\hat{IoU}(i) = \max_j IoU(i, j), \quad k = \arg \max_j IoU(i, j). \quad (1)$$

In the case that the highest IoU score $\hat{IoU}(i)$ is lower than a threshold τ_{iou} , we treat G_i as a new potential glass surface and

Algorithm 1: Incremental Glass Surface Detection

Input: RGB image \mathcal{I} , depth image \mathcal{D}
Output: Updated glass surfaces list \mathcal{G}

```

1  $\mathcal{G}_{\text{new}} \leftarrow \emptyset$ 
2  $\mathcal{S} \leftarrow \text{Segmentation}(\mathcal{I})$ 
3 foreach glass sementation mask  $s_i \in \mathcal{S}$  do
4    $H_i \leftarrow \text{MaskBoundary}(s_i)$ 
5   foreach boundary point  $\mathbf{h}_j \in H_i$  do
6      $d_j \leftarrow \mathcal{D}(\mathbf{h}_j.x, \mathbf{h}_j.y)$ 
7      $\mathbf{q}_j \leftarrow \text{ProjectToWorld}(\mathbf{h}_j, d_j)$ 
8      $Q_i.\text{add}(\mathbf{q}_j)$ 
9    $\pi_i \leftarrow \text{FitPlaneRANSAC}(Q_i)$ 
10   $\tilde{Q}_i \leftarrow \text{ProjectToPlane}(Q_i, \pi_i)$ 
11   $P_i \leftarrow \text{ConvexHull}(\tilde{Q}_i)$ 
12   $\mathbf{c}_i, \mathbf{n}_i \leftarrow P_i.\text{centroid}, \pi_i.\text{normal}$ 
13   $C_i \leftarrow \text{samplePointCloud}(P_i)$ 
14   $G_i \leftarrow \text{new GlassSurface}(\mathbf{c}_i, \mathbf{n}_i, P_i, C_i)$ 
15   $\mathcal{G}_{\text{new}}.\text{add}(G_i)$ 
16 foreach  $G_i \in \mathcal{G}_{\text{new}}$  and  $G_j \in \mathcal{G}$  do
17    ${}^jP_i \leftarrow \text{ProjectToPlane}(P_i, G_j.\text{plane})$ 
18    $I\hat{o}U(i), k \leftarrow \text{ComputeMaxIoU}(P_j, {}^jP_i)$ 
19   if  $I\hat{o}U(i) < \tau_{iou}$  then
20      $\mathcal{G}.\text{add}(G_i)$ 
21   else
22      $G_k.\text{update}(G_i)$ 

```

add it to the list \mathcal{G} . Otherwise, G_i and G_k are considered to be observations for the same glass surface. Hence, we update the information for the existing glass surface G_k using the new detection G_i . A new fitted plane π'_k for the glass surface G_k is computed using points from P_i and P_k , following a process similar to Sec. III-A. The normal of this plane is considered the updated normal \mathbf{n}'_k of the glass surface G_k . The updated polygon vertices of G_k are then computed as

$$P'_k = {}^{k'}P_i \cup {}^{k'}P_k, \quad (2)$$

where ${}^{k'}P_i$ and ${}^{k'}P_k$ are the projection of P_i and P_k onto the plane defined by the updated normal \mathbf{n}'_k and centroid \mathbf{c}'_k . The union function in (2) calculates the spatial set theoretic union that combines two polygons into one rather than a simple concatenation of the vertices. Finally, the updated centroid \mathbf{c}'_k is computed as the average of the points in P'_k .

IV. ACTIVE CONTACT ENGAGEMENT IN AUTONOMOUS AERIAL NAVIGATION

Building on the extracted glass surface information from camera images, the proposed system can next confirm their presence by purposefully contacting the detected glass using a lightweight contact-sensing module (Sec. IV-A). Utilizing touch actions (Sec. IV-B), we integrate active contact engagement into a fundamental autonomous point-to-point aerial navigation task (Sec. IV-C).

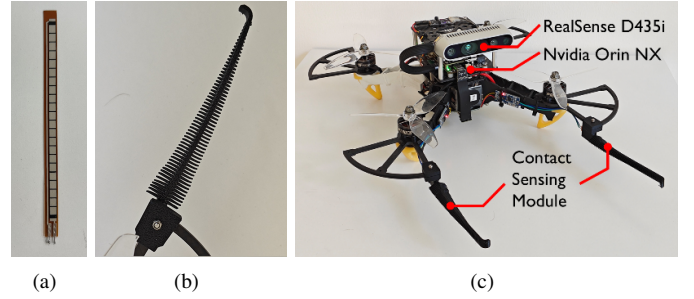


Fig. 3. The figure illustrates (a) the flex sensor used in our design, (b) the proposed contact-sensing module, (c) the aerial vehicle for flight experiments.

A. Contact-Sensing Module

We design a lightweight contact-sensing module mounted in the front of the aerial vehicle to detect physical contact with environmental obstacles, as shown in Fig. 3. The module consists of a 4.5-inch flex sensor that measures deflection and is encased in a 3D-printed TPU shell. Upon contact with obstacles, the contact-sensing module bends physically, causing the resistance of the flex sensor to change. This resistance is converted into a voltage signal, digitized by a calibrated analog-to-digital converter, and transmitted to the high-level onboard computer. The soft and deformable TPU shell serves two purposes: it protects the flex sensor from fractures due to excessive bending and acts as a suspension system to mitigate the effects of aerial vehicle body vibrations and air turbulence on the flex sensor. This design ensures the sensor remains durable and functional during physical contact with obstacles. Weighing only 3.3 g each, the lightweight contact-sensing module is portable and easy to install on various aerial platforms. Unlike hard collisions, the proposed contact-sensing module introduces negligible impact on the aerial vehicle during interactions, reducing the risk of damage to the vehicle. Additionally, a previous study [10] has revealed the correlation between collision impact and state estimation failure. By enabling gentle touch sensing, the module enhances the survivability of onboard state estimation, further improving the robustness of autonomous navigation.

B. Touch Action for Active Contact Engagement

Before discussing autonomous navigation with active contact engagement, we first introduce how a *touch action* is performed when the aerial vehicle intentionally approaches and attempts to make contact with a target potential glass surface. Assume the target surface has a centroid \mathbf{c} and a unit normal vector \mathbf{n} pointing towards the aerial vehicle. The touch action begins at a ready pose, defined by the position $\mathbf{c} + \delta_s \mathbf{n}$, where δ_s is a predefined distance, and an orientation that aligns the contact-sensing module with the target surface. The planner then generates a smooth trajectory from the ready pose to the ending position $\mathbf{c} - \delta_e \mathbf{n}$, maintaining a constant yaw and a conservative touch speed v_{touch} to ensure gentle contact. During the flight, if the contact-sensing module detects a touch event, the aerial vehicle confirms the presence of the glass surface. The contact position can be computed using the vehicle's pose and the physical parameters of the contact-sensing module. This contact position provides precise

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

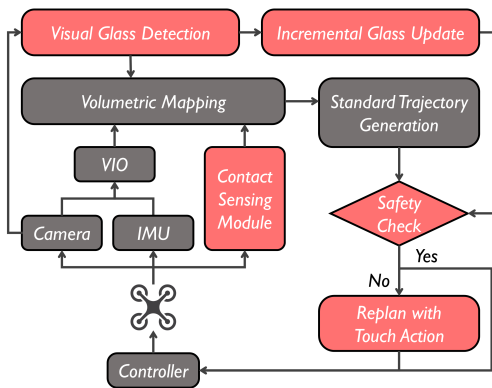


Fig. 4. The figure illustrates the pipeline of the proposed autonomous navigation system, with the pink blocks highlighting components related to active contact engagement.

information about the exact location of the glass surface. Subsequently, the volumetric map is updated using the glass surface information, where its point cloud C_i is offset along the surface normal \mathbf{n}_i to ensure the contact position lies on the surface. The aerial vehicle then returns to the ready pose for safety and subsequent replanning. If the aerial vehicle reaches the ending position without triggering the contact-sensing module, it records the potential glass surface as invalid and proceeds with the navigation. A photo capturing the moment of a touch action is shown in Fig. 1.

C. Autonomous Navigation with Active Contact Engagement

Active contact engagement enhances the safety of point-to-point navigation in unknown environments, particularly when visual sensor measurements of transparent glass surfaces are unreliable. In such cases, the occupancy map may incompletely represent glass surfaces, increasing the risk of crashes if the aerial vehicle attempts to traverse these areas. Equipped with the visual detection module and contact-sensing capability, the aerial vehicle obtains awareness of transparent glass obstacles in the environment during autonomous navigation.

This section describes the integration of active contact engagement into the point-to-point aerial navigation task in unknown environments. Based on a standard trajectory planner, the diagram for the proposed autonomous navigation system with active contact engagement is shown in Fig. 4. Given a start and a goal position, the planner first generates a standard trajectory using the volumetric map derived from conventional sensor measurements. A safety check is then performed to identify any intersection with potential glass surfaces by sampling along the trajectory (green trajectory in Fig. 5(a)). If the trajectory is safe without intersection with any potential glass surface, the aerial vehicle follows the standard trajectory to the goal position. Otherwise, the planner generates a new trajectory to the ready pose of the first intersected surface and decelerates the vehicle to v_{touch} to execute a touch action (blue trajectory in Fig. 5(a)). During the touch action (Sec. IV-B), the volumetric map is updated with the contact information (Fig. 5(b)). Once the touch action is completed and the glass surface is confirmed or invalidated, the planner replans another standard trajectory to the goal position using the updated volumetric map (Fig. 5(c)). This

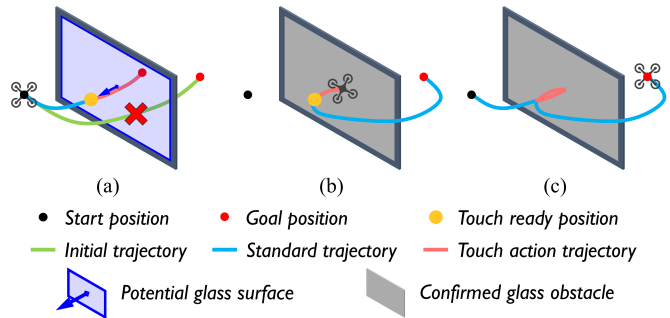


Fig. 5. An illustration of the proposed autonomous navigation system with active contact engagement in a point-to-point aerial navigation task.

process repeats iteratively until the goal position is reached. Note that the point-to-point navigation task represents only one of many potential applications for the proposed system. It can be extended to more comprehensive tasks by replacing the standard trajectory generation component in Fig. 4 with a corresponding planner tailored to the specific task.

V. EXPERIMENTS AND RESULTS

A. Implementation Details

For the onboard software, the object segmentation framework YOLOv8 [36] utilized in Sec. III-A is custom-trained using transparent surface data aggregated from open-source datasets [4, 38, 39]. The spatial set theoretic union in Sec. III-B leverages the Boost C++ Libraries¹. The standard trajectory for the point-to-point navigation task in Sec. IV-C is generated using the planner proposed in [40]. Table II summarizes the parameters setting for the proposed algorithm, which is used throughout the experiments.

In the real-world flight experiments (Sec. V-D to V-F), we customize the contact-sensing module based on an open-source lightweight quadrotor platform [41], as shown in Fig. 3(c). The quadrotor is equipped with an NVIDIA Jetson Orin NX² 16 GB, a NxtPx4 autopilot [42], a global-shutter stereo camera RealSense D435i³ and a pair of contact-sensing modules (Sec. IV-A), which has a take-off weight of 895 g. The linear and angular dynamics limitations of the quadrotor are set to $v_m = 1.0$ m/s, $a_m = 1.0$ m/s², $\xi_m = 1.05$ rad/s, $\xi_m = 1.05$ rad/s². The localization of the quadrotor is provided by a visual-inertial state estimator VINS-Fusion [43]. All the experiments are conducted fully onboard without other external infrastructure, such as ground stations or motion capture systems.

B. Incremental Glass Detection in a Glass-Rich Corridor

We first conduct a unit testing on the incremental glass surface detection capability of the proposed system. In this experiment, data is collected in a glass-rich corridor using a handheld device equipped with a RealSense D435i camera and a Livox Mid-360 LiDAR⁴. Initially, the agent can only

¹https://beta.boost.org/doc/libs/1_58_0/libs/geometry/doc/html/geometry/reference/algorithms/union_.html

²<https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-orin>

³<https://www.intelrealsense.com/depth-camera-d435i>

⁴<https://www.livoxtech.com/mid-360>

TABLE II
PARAMETERS SETTING

Parameter	Section	Notation	Value
Segmentation confidence	III-A	τ_s	0.75
Normal difference threshold	III-B	τ_n	0.65 rad
Centroid difference threshold	III-B	τ_c	1.0 m
IoU threshold	III-B	τ_i	0.1
Touch action start distance	IV-B	δ_s	1.0 m
Touch action end distance	IV-B	δ_e	1.0 m
Touch action velocity	IV-B	v_{touch}	0.2 m/s

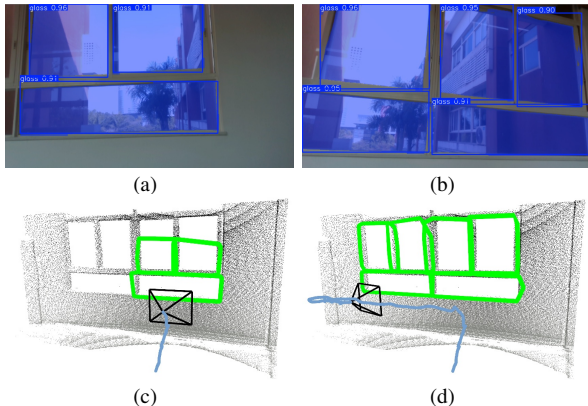


Fig. 6. Experiment results of incremental glass surface detection. The left and right columns showing the initial and final stages of the experiment respectively. The upper row shows the first-person view alongside the visual segmentation results. The lower row displays the trajectory and results, with the grey point cloud for visualization and ground truth comparison.

capture a partial view of the glass windows, as shown in Fig. 6(a)(c). With more measurements received, the visual detection module incrementally updates the detected glass surfaces in the scene, completing the glass surface detection of all glass windows, as shown in Fig. 6(b)(d). Please note that the LiDAR is used solely for reconstructing the scene for ground truth visualization and its data is not used in the proposed system. The experiment results demonstrate that the proposed glass detection module is capable of incrementally updating glass surface information in real time.

C. Simulation Benchmark for Aerial Navigation

We conducted benchmark simulations to evaluate the proposed active contact navigation planner against two baseline methods: (1) *Non-contact*: A vision-only planner that treats all visually detected glass surfaces as obstacles without physical interaction, and (2) *Contact-based*: An adaptation of a state-of-the-art planner designed for glass-rich environments [10] that performs passive contact reactions and marks regions near contact points as occupied space.

Due to simulation constraints in rendering realistic camera measurements on transparent and reflective glass, the 2D visual glass detection and segmentation is directly provided by the simulator for all methods. As illustrated in Fig. 7(a), blue regions indicate correctly detected glass surfaces (true positives), and red regions show the traversable openings that are incorrectly detected as glass (false positives), which are both only for visualization purposes. The planners are required to perform real-time mapping of occupancy map and glass surfaces during navigation. The experiment consists of 10

TABLE III
SIMULATION BENCHMARK RESULT STATISTICS

Method	Duration (s)		Path Length (m)		#Contact [†]	
	Avg	Std	Avg	Std	Avg	Std
Non-contact	282.0	8.4	288.3	6.1	-	-
Contact-based	265.2	13.7	248.9	7.7	35.6	3.4
Proposed	223.2	2.7	159.7	0.5	5.2	0.4

[†] The number of physical contacts with glass surfaces.

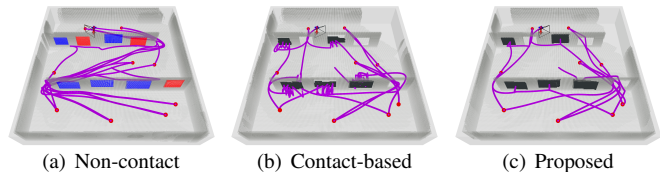


Fig. 7. The executed trajectories in simulation benchmark experiments. The red dots indicate the randomly sampled target positions.

point-to-point navigation tasks, with randomly sampled non-linear target positions. Each method was tested for 5 runs, with quantitative results summarized in Table III and representative trajectories shown in Fig. 7. The results demonstrate that the non-contact visual-only method suffered from conservative detours, while the contact-based method required numerous passive contacts to fully reconstruct the glass surfaces. In contrast, The proposed active contact method consciously confirmed or invalidated the detected glass surfaces in a single touch, yielding the shortest flight time and path length among the baselines.

D. Navigating a Space with a Glass Door

We conduct a series of field experiments to validate the proposed framework for autonomous navigation in environments with transparent glass obstacles. In the first scenario, a glass door frame is placed horizontally in a corridor between the start and goal positions of the aerial robot, as shown in Fig. 8(a). Using a standard perception and trajectory generation system, the aerial vehicle fails to perceive the glass surface, leaving a hole in the volumetric map. Consequently, it attempts to fly through the door frame along a minimum-time trajectory, as depicted in Fig. 8(b). This unsafe trajectory would inevitably lead to a crash when colliding with the glass surface, which is not executed due to safety concerns. In contrast, our proposed system with glass awareness actively engages with the detected glass surface by performing a touch action. After confirming the presence of the glass surface, the aerial vehicle dynamically replans its trajectory, circumventing the obstacle by flying above the glass door frame, as shown in Fig. 8(c). The entire trajectory including the touch action spans 6.58 m and is completed in 24.23 seconds. This experiment demonstrated that when a glass surface obstructs the navigation path, the proposed system can effectively detect and engage contact with the glass surface for safe navigation in real-world scenarios.

E. Navigating a Space with Two Glass Surfaces

The second scene involves a room with dimensions $8 \times 7 \times 2$ m³, which is separated by two fixed glass windows with

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

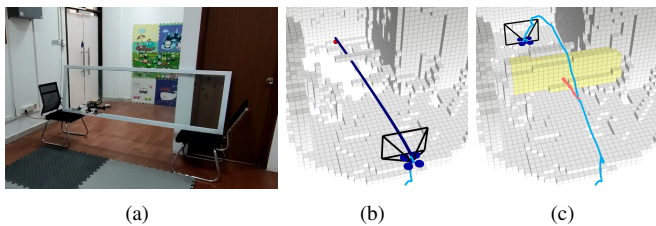


Fig. 8. Real-world navigation experiment traversing through a glass door. (a) A scene photo captured during the experiment. (b) The unsafe trajectory generated by a standard perception and planning system. (c) The final executed trajectory and volumetric map using the proposed method, with glass surface indicated in light yellow and the touch action marked in pink.

half of the room inaccessible, as shown in Fig. 9(1a). The aerial vehicle is initially placed on one side of the room, with the goal position set on the opposite side (the red dot in Fig. 9(1c)). During the experiment, the aerial vehicle attempts to reach the goal by flying through the right glass window, triggering the first touch action. After confirming and updating the volumetric map with the right glass window, the aerial vehicle replans a new trajectory through the left glass window. A second touch action is triggered there and the left window is also marked occupied. At this point, no feasible trajectory to the goal position can be resolved, and the aerial vehicle terminates the task and hovers in a safe location. The navigation task with a trajectory length of 9.48 m is completed in 27.56 seconds. The data from the contact-sensing modules during the experiment are plotted in Fig. 9(1b). This experiment highlights the system’s ability to detect multiple glass surfaces and autonomously terminate the task when no feasible path is available, ensuring safety in complex environments.

F. Navigating a Space with a Double Sliding Window

We conduct a more challenging point-to-point navigation experiment involving a double sliding glass window, as shown in Fig. 9(2a). The glass window is fully open, with the left half unobstructed and right half blocked by transparent glass. As illustrated in Fig. 9(2b), the onboard visual detection module cannot distinguish between the two halves of the window, identifying both regions as potential glass surfaces. A standard trajectory is initially generated towards the goal position, which intersects with the right half of the window. The aerial vehicle actively performs a touch action to confirm the presence of the glass surface and the volumetric map is updated by filling in the space occupied by the glass surface, as depicted in Fig. 9(2c). The planner then generates a new trajectory to the goal position using the updated volumetric map. During this phase, an intersection is detected between the new trajectory and the left half of the window during the second safety check, prompting a second touch action. The contact-sensing module does not register any contact this time, thereby invalidating the left glass surface. Finally, the aerial vehicle replans one last trajectory and successfully reaches the goal position with a total trajectory length 8.86 m and execution time 33.04 s. In this navigation task, the aerial vehicle successfully traverses the double sliding glass window by actively engaging with potential glass surfaces, even when the visual detection module provides only partially accurate

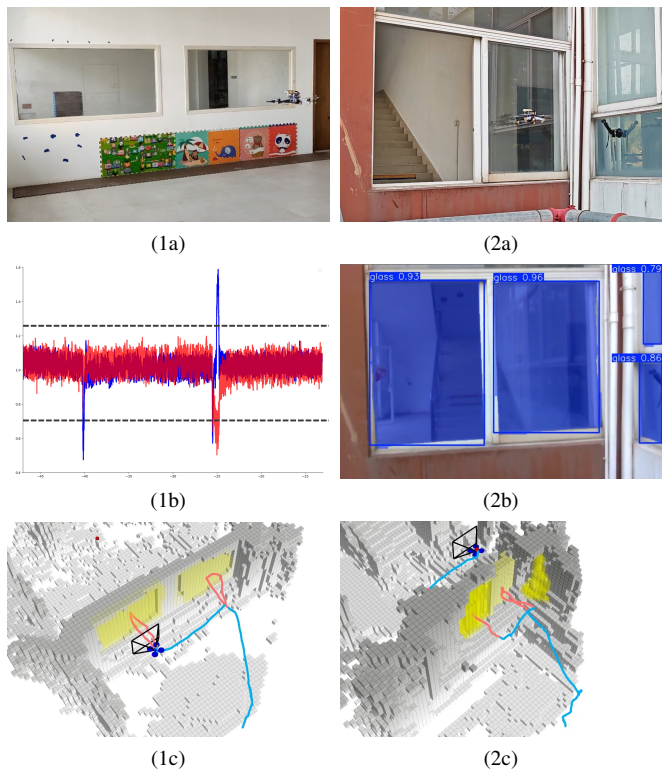


Fig. 9. Real-world navigation experiments (1) in a room with two fixed glass surfaces and (2) traversing through a double sliding window. (1a)(2a) Scene photos captured during the experiment. (1b) Data from the two contact-sensing modules, where the peaks exceeding the dashed black lines indicate touch action confirmations. (2b) A first-person view from the aerial vehicle, showing partially correct glass detection and segmentation. (1c)(2c) The final executed trajectory and volumetric mapping.

information. If the agent relied solely on visual glass detection without the capability of physical contact, the task would fail due to the high confidence scores of both halves of the window (Fig. 9(2b)). The three navigation experiments in Sec. V-D - V-F demonstrate the effectiveness of the proposed system in achieving safe and robust navigation in complex and real-world scenarios with transparent glass obstacles.

VI. CONCLUSIONS

In this work, we propose a novel approach for robust autonomous aerial navigation in unknown environments with transparent glass obstacles. Potential glass surfaces are detected and their information is incrementally maintained using visual sensor measurements. The aerial vehicle then actively engages in touch actions with these surfaces using a pair of lightweight contact-sensing modules to confirm or invalidate their presence. We integrate the proposed system into the fundamental task of autonomous point-to-point navigation in unknown environments. We validate the incremental glass detection module in a glass-rich corridor and assess the advantage of active contact planning in simulation benchmarks. A series of real-world experiments conducted in various scenarios demonstrate the system’s reliability and practicality for safe and robust autonomous aerial navigation in complex environments with glass obstacles.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

REFERENCES

- [1] Y. Zhang, X. Chen, C. Feng, B. Zhou, and S. Shen, "Falcon: Fast autonomous aerial exploration using coverage path guidance," *IEEE Transactions on Robotics*, 2024.
- [2] J. Nikolic, M. Burri, J. Rehder, S. Leutenegger, C. Huerzeler, and R. Siegwart, "A uav system for inspection of industrial facilities," in *2013 IEEE Aerospace Conference*. IEEE, 2013, pp. 1–8.
- [3] N. Michael, S. Shen, K. Mohta, V. Kumar, K. Nagatani, Y. Okada, S. Kiribayashi, K. Otake, K. Yoshida, K. Ohno *et al.*, "Collaborative mapping of an earthquake damaged building via ground and aerial robots," in *Field and service robotics: results of the 8th international conference*. Springer, 2014, pp. 33–47.
- [4] H. Mei, X. Yang, Y. Wang, Y. Liu, S. He, Q. Zhang, X. Wei, and R. W. Lau, "Don't hit me! glass detection in real-world scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3687–3696.
- [5] L. Zhou, X. Sun, C. Zhang, L. Cao, and Y. Li, "Lidar-based 3d glass detection and reconstruction in indoor environment," *IEEE Transactions on Instrumentation and Measurement*, 2024.
- [6] F. Qi, X. Tan, Z. Zhang, M. Chen, Y. Xie, and L. Ma, "Glass makes blurs: Learning the visual blurriness for glass surface detection," *IEEE Transactions on Industrial Informatics*, 2024.
- [7] M. Becker, R. C. B. Sampaio, S. Bouabdallah, V. d. Perrot, and R. Siegwart, "In-flight collision avoidance controller based only on os4 embedded sensors," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 34, pp. 294–307, 2012.
- [8] A. Briod, P. Kornatowski, J.-C. Zufferey, and D. Floreano, "A collision-resilient flying robot," *Journal of Field Robotics*, vol. 31, no. 4, pp. 496–509, 2014.
- [9] P. De Petris, S. J. Carlson, C. Papachristos, and K. Alexis, "Collision-tolerant aerial robots: A survey," *arXiv preprint arXiv:2212.03196*, 2022.
- [10] Y. Mulgaonkar, W. Liu, D. Thakur, K. Daniilidis, C. J. Taylor, and V. Kumar, "The tiercel: A novel autonomous micro aerial vehicle that can map the environment by flying into obstacles," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 7448–7454.
- [11] Z. Liu and K. Karydis, "Toward impact-resilient quadrotor design, collision characterization and recovery control to sustain flight after collisions," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 183–189.
- [12] E. Xie, W. Wang, W. Wang, M. Ding, C. Shen, and P. Luo, "Segmenting transparent objects in the wild," in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*. Springer, 2020, pp. 696–711.
- [13] H. He, X. Li, G. Cheng, J. Shi, Y. Tong, G. Meng, V. Prinet, and L. Weng, "Enhanced boundary learning for glass-like object segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15 859–15 868.
- [14] Z. Huang, K. Wang, K. Yang, R. Cheng, and J. Bai, "Glass detection and recognition based on the fusion of ultrasonic sensor and rgb-d sensor for the visually impaired," in *Target and Background Signatures IV*, vol. 10794. SPIE, 2018, pp. 118–125.
- [15] H. Wei, X.-e. Li, Y. Shi, B. You, and Y. Xu, "Multi-sensor fusion glass detection for robot navigation and mapping," in *2018 WRC Symposium on Advanced Robotics and Automation (WRC SARA)*. IEEE, 2018, pp. 184–188.
- [16] X. Wang and J. Wang, "Detecting glass in simultaneous localisation and mapping," *Robotics and Autonomous Systems*, vol. 88, pp. 97–103, 2017.
- [17] S. Judd, *Photoelectric sensors and controls: selection and application*. CRC Press, 1988, vol. 63.
- [18] H. Mei, B. Dong, W. Dong, J. Yang, S.-H. Baek, F. Heide, P. Peers, X. Wei, and X. Yang, "Glass segmentation using intensity and spectral polarization cues," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 12 622–12 631.
- [19] K. Ikemura, Y. Huang, F. Heide, Z. Zhang, Q. Chen, and C. Lei, "Robust depth enhancement via polarization prompt fusion tuning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 20 710–20 720.
- [20] Z. Liu, Z. Lu, A.-a. Agha-mohammadi, and K. Karydis, "Contact-prioritized planning of impact-resilient aerial robots with an integrated compliant arm," *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 4, pp. 2064–2072, 2023.
- [21] P. De Petris, H. Nguyen, M. Kulkarni, F. Mascarich, and K. Alexis, "Resilient collision-tolerant navigation in confined environments," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 2286–2292.
- [22] P. Sareh, P. Chermprayong, M. Emmanuelli, H. Nadeem, and M. Kovac, "Rotorigami: A rotary origami protective system for robotic rotorcraft," *Science Robotics*, vol. 3, no. 22, p. eaah5228, 2018.
- [23] T. Lew, T. Emmei, D. D. Fan, T. Bartlett, A. Santamaria-Navarro, R. Thakker, and A.-a. Agha-mohammadi, "Contact inertial odometry: Collisions are your friends," in *The International Symposium of Robotics Research*. Springer, 2019, pp. 938–958.
- [24] J. Li, J. Ji, Q. Wang, H. Yu, Y. Pan, and F. Gao, "Active collision-based navigation for wheeled robots," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024.
- [25] J. Zha, X. Wu, J. Kroeger, N. Perez, and M. W. Mueller, "A collision-resilient aerial vehicle with icosahedron tensegrity structure," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 1407–1412.
- [26] Z. Lu, Z. Liu, and K. Karydis, "Deformation recovery control and post-impact trajectory replanning for collision-resilient mobile robots," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 2030–2037.
- [27] Y. Mulgaonkar, A. Mäkinen, L. Guerrero-Bonilla, and V. Kumar, "Robust aerial robot swarms without collision avoidance," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 596–603, 2017.
- [28] T. Tomić, C. Ott, and S. Haddadin, "External wrench estimation, collision detection, and reflex reaction for flying robots," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1467–1482, 2017.
- [29] R. Bajcsy, "Active perception," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 966–1005, 1988.
- [30] V. Brunacci, A. Dionigi, A. De Angelis, and G. Costante, "Infrastructure-less uwb-based active relative localization," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 14 079–14 086.
- [31] X. Chen, Y. Zhang, B. Zhou, and S. Shen, "Apacc: Agile and perception-aware trajectory generation for quadrotor flights," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 17 858–17 864.
- [32] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [33] J. A. Placed, J. Strader, H. Carrillo, N. Atanasov, V. Indelman, L. Carlone, and J. A. Castellanos, "A survey on active simultaneous localization and mapping: State of the art and new frontiers," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 1686–1705, 2023.
- [34] A. Dionigi, S. Felicioni, M. Leomanni, and G. Costante, "D-vat: End-to-end visual active tracking for micro aerial vehicles," *IEEE Robotics and Automation Letters*, 2024.
- [35] Z. Han, R. Zhang, N. Pan, C. Xu, and F. Gao, "Fast-tracker: A robust aerial system for tracking agile target in cluttered environments," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 328–334.
- [36] G. Jocher, J. Qiu, and A. Chaurasia, "Ultralytics yolo," 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [37] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, pp. 381–395, 1981.
- [38] J. Lin, Y.-H. Yeung, and R. Lau, "Exploiting Semantic Relations for Glass Surface Detection," *Advances in Neural Information Processing Systems*, vol. 35, pp. 22 490–22 504.
- [39] J. Lin, Z. He, and R. W. Lau, "Rich Context Aggregation with Reflection Prior for Glass Surface Detection," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 13 410–13 419.
- [40] B. Zhou, F. Gao, L. Wang, C. Liu, and S. Shen, "Robust and efficient quadrotor trajectory generation for fast autonomous flight," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3529–3536, 2019.
- [41] Y. Zhang, X. Chen, P. Liu, J. Wang, H. Zou, N. Pan, F. Gao, and S. Shen, "Uniquad: A unified and versatile quadrotor platform series for uav research and application," *arXiv preprint arXiv:2407.00578*, 2024.
- [42] P. Liu, C. Feng, Y. Xu, Y. Ning, H. Xu, and S. Shen, "Omninx: A fully open-source and compact aerial robot with omnidirectional visual perception," *arXiv preprint arXiv:2403.20085*, 2024.
- [43] T. Qin, J. Pan, S. Cao, and S. Shen, "A general optimization-based framework for local odometry estimation with multiple sensors," *arXiv preprint arXiv:1901.03638*, 2019.