Emergency UAV Delivery Framework: A Hybrid Approach to GPS Navigation and Visual Landmark Detection

Cao-Ky-Long U [1,2] and Nguyen Khac Toan [3,*]

[1] Faculty of Information Technology, University of Science, VNU-HCM, Ho Chi Minh City, Vietnam
[2] Vietnam National University, Ho Chi Minh City, Vietnam
[3] Institute of Intelligent and Interactive Technologies, University of Economics Ho Chi Minh City—UEH, Vietnam
Email: 24C11014@student.hcmus.edu.vn (C.K.L.U.); toannk@ueh.edu.vn (T.N.K.)
*Corresponding author

Abstract—**This study develops an innovative Unmanned Aerial Vehicles (UAV) framework incorporating GPS and Landmark Detection to refine relief delivery processes in emergency scenarios, particularly in expansive regions with intricate conditions. The system leverages extensive positioning capability of GPS to navigate UAV toward target zones, subsequently employing Landmark Detection to ascertain precise drop-off locations. Notably, an Archimedean spiral trajectory algorithm is deployed under unstable GPS conditions, enabling UAV to expand search coverage and enhance landmark detection capabilities, even in obscured areas. Experimental results in Vietnam reveal a precision drop rate of 98% at an altitude of 5 m, ensuring accurate delivery. Additionally, the landmark detection success rate under unstable GPS conditions achieved 100% when the overlap ratio of camera frames reached half the width of frame (0.5 W), demonstrating high efficacy in mitigating target omission risks. This approach not only minimizes delivery time but also enhances operational flexibility, facilitating rapid and precise UAV access to critical relief zones. The proposed system exhibits significant potential to deliver effective solutions for emergency relief missions, meeting stringent demands for speed, accuracy, and stability in time-sensitive delivery operations.**

Keywords—**Unmanned Aerial Vehicles (UAV)-based delivery, GPS-independent navigation, spiral search algorithm, UAV emergency delivery**

I. INTRODUCTION

With the continuous progress of technology, the field of robotics has seen remarkable advancements [1, 2]. Among these, Unmanned Aerial Vehicles (UAV) stand out for their diverse and impactful applications. UAV, represent autonomous aerial systems equipped with many technologies such as sensors, cameras, and navigation frameworks. Their versatility has enabled widespread application across diverse domains, including geospatial mapping, disaster response, and autonomous delivery within e-commerce [3]. In logistics, the exponential growth of e-commerce has imposed significant demands on optimizing delivery processes, necessitating accelerated speeds, cost reduction, and enhanced operational efficacy. UAV have emerged as a promising solution for "last-mile delivery", a critical phase in logistics often hindered by traffic congestion or challenging terrains [4]. The natural flexibility and adaptability of UAV effectively address these obstacles, driving substantial advancements in the logistics field. Their deployment marks a transformative milestone, offering unparalleled potential to streamline delivery systems while mitigating traditional logistical constraints.

The deployment of UAV in delivery operations has rapidly expanded due to their unparalleled advantages, including flexibility, superior accessibility, and reduced operational costs. In e-commerce and food industries, UAV facilitate swift and efficient delivery, particularly for small-scale orders within densely populated urban environments [5]. In healthcare, UAV have proven instrumental in transporting medications and medical supplies to remote or disaster-stricken regions where conventional road access is infeasible [6]. Notably, during rescue operations, UAV plays a pivotal role in delivering relief items to inaccessible areas, mitigating risks, and expediting aid in critical emergencies. However, operator-dependent UAV systems face limitations in handling large-scale payloads within constrained timeframes, especially during disaster scenarios. This underscores the urgent need for advanced autonomous systems to reduce human workload and enhance operational reliability in high-pressure environments.

To address challenges in automated UAV delivery, current methodologies emphasize enhancing drop-site identification to improve operational efficiency [7]. Each approach presents distinct advantages and limitations, tailored to specific scenarios. Semantic Segmentation, a prevalent technique, leverages deep learning models to segment images and identify critical areas such as rooftops, lawns, or other safe surfaces. This approach is extensively utilized in urban settings, where UAV must navigate narrow spaces and numerous obstacles to locate

Manuscript received February 10, 2025; revised March 5, 2025; accepted March 26, 2025; published July 4, 2025.

doi: 10.18178/ijmerr.14.4.374-383

secure drop sites. However, its limitation includes high computational demands and low performance under poor lighting conditions, such as at night or during adverse weather. Another research direction involves terrain analysis based on aerial imagery. UAV equipped with cameras detect flat surfaces and assess surrounding environments to select suitable delivery locations. This method is effective in simple terrains with minimal obstructions but shows significant limitations in complex landscapes or areas dense with obstacles. The GPS-based approach remains widely adopted due to its global positioning capabilities, enabling UAV to target specific coordinates for delivery. Nonetheless, its accuracy decreases in obstructed environments, such as indoor settings, under tree canopies, or urban areas with high-rise buildings, rendering drop-site reliability less dependable. To overcome GPS constraints, Simultaneous Localization and Mapping (SLAM) [8] has emerged as an advanced alternative. SLAM utilizes sensors like cameras or Light Detection and Ranging (LiDAR) [9] to create real-time maps and localize UAV within space. This technique proves particularly valuable in GPS-denied environments or scenarios requiring detailed mapping. However, its drawbacks include reliance on expensive sensors and substantial processing time, limiting its feasibility for low-cost or real-time applications. Each approach, while innovative, addresses specific operational contexts, highlighting the trade-offs between computational resources, environmental adaptability, and implementation cost. These comparative insights contribute to the development of more robust UAV delivery systems in diverse settings.

Current research mainly focuses on precise delivery, often requiring UAV to land prior to releasing packages. While this approach is suitable for pre-planned scenarios under normal conditions, it demonstrates significant limitations in disaster contexts such as floods or earthquakes. In such cases, the vast expanse of affected areas and the time-consuming nature of landing procedures substantially reduce the efficiency of relief efforts. Currently, local rescue teams frequently deploy helicopters to expedite the transport and drop supplies, enabling swift access to critical zones. Building upon these practical insights, this study proposes an automated UAV-based delivery methodology that integrates high-altitude package dropping to enhance time efficacy in relief missions. At the same time, this method ensures accuracy in determining drop-off locations, making it particularly suitable for relief packages, as they are carefully packed to withstand being dropped from significant heights.

This study focuses on developing a method for accurate package drop-off location determination by integrating GPS and Landmark Detection. Specifically, the UAV begins its journey by navigating to the target area using coordinates provided by GPS. In cases where GPS signals are unavailable, a No-GPS mode is employed to temporarily guide the UAV until GPS functionality is restored. Upon reaching the designated area, UAV utilizes an onboard camera combined with a YOLO detection model to identify and accurately locate predefined landmarks. These landmarks serve as predefined indicators. Based on the positional information from landmarks, the UAV adjusts its trajectory and positioning to ensure precise delivery at the intended drop-off point. If the camera fails to detect landmarks within its field of view, a spiral search algorithm is deployed to locate the landmarks. This method combines wide-range positioning capability of GPS with the high accuracy of Landmark Detection in confined areas, along with advanced search techniques and aerial delivery mechanisms, to enhance both precision and delivery speed. The study aims to improve the efficiency of autonomous delivery systems by ensuring time-efficient and highly accurate package drops under real-world conditions. The key contributions of this research can be summarized as follows:

- Enhancing relief delivery efficiency through high-altitude payload deployment.
- Developing an integrated methodology combining GPS and landmark detection to improve delivery precision.

II. Related Work

The deployment of Unmanned Aerial Vehicles for last-mile delivery has gained increasing attention as a solution to optimize logistics, particularly in challenging environments such as dense urban areas and remote regions. Several methodologies have been proposed to address the key challenges of precision delivery, computational efficiency, and adaptability to real-world scenarios. This section reviews existing work on UAV delivery systems, focusing on GPS-based navigation, SLAM, visual landmark detection, semantic segmentation, and hybrid approaches.

GPS has long been a primary method for UAV navigation due to its ability to provide wide-area localization with global coverage [10]. In UAV delivery, GPS is often used for coarse navigation to guide drones to target locations. However, its limitations become evident in complex urban environments where GPS signals suffer from interference or obstruction by tall buildings and dense foliage. Eskandaripour and Boldsaikhan [11] highlighted these challenges in urban logistics, particularly for achieving precise positioning during the final phase of delivery. Despite its utility, GPS alone cannot ensure reliable drop-off accuracy in environments requiring sub-meter precision. To overcome these limitations, recent studies have combined GPS with visual methods. For example, Brunner et al. proposed a system where UAV relies on GPS for approximate navigation but use visual markers for the last few meters of localization [12].

SLAM has emerged as a robust alternative to GPS for UAV navigation, particularly in GPS-denied environments [13, 14]. SLAM combines real-time mapping with localization, leveraging sensors such as cameras, LiDAR, and Inertial Measurement Unit (IMU) to enable UAV to navigate autonomously. Steenbeek and Nex [15] developed a monocular visual SLAM system using inexpensive RGB cameras for UAV exploration in emergency conditions, highlighting the potential of method for real-time mapping in confined spaces. SLAM systems are particularly advantageous in environments where GPS is unreliable or unavailable, such as indoors, under dense canopies, or during natural disasters. However, SLAM systems present trade-offs between accuracy, sensor cost, and computational overhead. The use of dense SLAM methods, while effective, often demands significant computational resources, limiting their application in cost-sensitive scenarios. The challenge lies in balancing real-time processing with mapping precision, especially in dynamic environments.

While GPS and SLAM provide foundational navigation capabilities, the integration of visual algorithms has become increasingly necessary to address their inherent limitations. Visual algorithms, particularly those based on deep learning, enable UAV to interpret their surroundings more effectively by detecting landmarks, recognizing safe zones, and navigating around obstacles. One of them is Semantic segmentation that allows UAV to classify image pixels, identifying safe drop-off zones [16]. Meanwhile, landmark detection leverages distinctive markers or natural features for precise localization [17]. Together, these techniques address the limitations of GPS and SLAM, significantly improving delivery accuracy and adaptability in complex environments.

Semantic segmentation, a deep learning-based technique, has been widely adopted to identify safe drop-off zones in complex environments [18, 19]. By classifying pixels in an image, semantic segmentation enables UAV to differentiate between surfaces such as rooftops, lawns, roads, and sidewalks. Kannan and Min proposed a semantic segmentation-based approach for autonomous drone delivery, allowing UAV to identify and navigate to safe drop-off zones around houses [20]. This method ensures safe deliveries even in unstructured environments. However, semantic segmentation methods face challenges such as sensitivity to lighting conditions, computational demands, and the need for high-quality datasets [21]. The reliance on deep learning models necessitates significant onboard processing power, which can be prohibitive for lightweight UAV.

Landmark detection has been proposed as an effective method to enhance UAV delivery accuracy [22, 23]. Unlike GPS, landmark-based approaches rely on visual cues, such as fiducial markers, predefined features, or object detection algorithms, to localize UAV at short ranges. Visual markers, such as QR codes or ArUco tags, have been widely used to enable precise drop-offs. Innocenti et al. tested multiple fiducial marker systems for medicine delivery in smart cities, demonstrating that UAV equipped with high-resolution cameras can accurately locate drop-off points [24]. Brunner *et al.* [12] similarly employed visual navigation for balcony deliveries, where UAV identify visual markers placed on target locations. In more advanced systems, deep learning models are applied to detect landmarks and improve delivery precision. Xia *et al.* [25] proposed a computer-vision-based system leveraging semantic segmentation and house-aware structures to guide UAV toward specific drop-off points, such as front doors or garages. These methods significantly reduce the reliance on GPS and improve localization accuracy in cluttered urban environments.

Previous studies have focused on UAV delivery under normal conditions, achieving success in precise deliveries to doorsteps or balconies. These methods, however, require UAV to land, limiting their efficiency in disaster scenarios where quick delivery across large areas is critical. According to our survey, no prior research has specifically addressed emergency delivery through high-altitude package drops, which highlights the novelty of this approach. Landing processes in such contexts can significantly delay the distribution of relief supplies. This research proposes an approach combining GPS, landmark detection, and high-altitude drops to overcome these limitations. By allowing UAV to release packages accurately without landing, the method enhances speed and adaptability for emergency relief operations.

III. Literature Review

A. Overview

The UAV system is designed with an automated operation process to optimize delivery in various environmental conditions, as illustrated in Fig. 1. The process begins by identifying starting and destination points. GPS signals are periodically checked to ensure stability. If the GPS signal is stable, UAV operates in GPS mode, navigating to the target location using global coordinates to adjust its flight path. Conversely, if the GPS signal is unavailable or unstable, UAV switches to No-GPS mode, utilizing sensors such as the Inertial Measurement Unit (IMU), pressure sensors, and cameras to estimate its position and navigate using control algorithms.

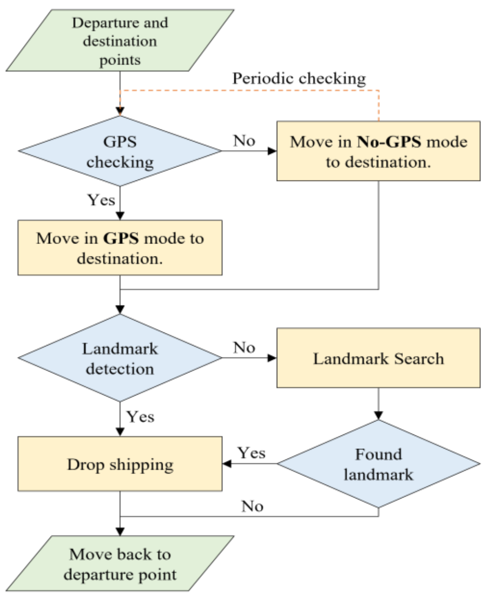

Fig. 1. Structure of proposed UAV delivery method.

Upon reaching target area, the UAV activates a landmark detection function based on deep learning models to accurately identify drop-off location. If the landmark is detected, system automatically aligns and performs the delivery. If the landmark is not immediately identified, the UAV deploys a search algorithm based on an Archimedean spiral trajectory to expand the scanning area. The search process continues until landmark is located or a predefined limit is exceeded. After completing delivery mission, UAV returns to starting point to conclude the operation cycle.

B. Autonomous Navigation

This is a critical capability of UAV systems, enabling them to operate independently in diverse environments. The proposed UAV system supports two distinct modes of navigation: GPS navigation and No-GPS navigation, ensuring flexibility and adaptability in many different environmental conditions. These two modes are designed to complement each other, addressing limitations of each approach and enhancing the overall reliability and precision of UAV operations. Below, we provide a detailed explanation of each mode and their integration into the UAV system.

1) GPS navigation

GPS navigation relies on the Global Positioning System to determine the position of UAV and guide it to predefined points or target locations. The proposed method does not focus on improving navigation; instead, existing algorithms from the ArduPilot library are utilized for guidance.

Initialization: The UAV connects to the ground control system (GCS) using MAVLink protocol and switches to GPS navigation mode. The system automatically checks to ensure that the UAV is armable and ready for operation.

Takeoff: The takeoff process is executed by gradually increasing the thrust of motors. To ensure a stable and safe ascent, the thrust is adjusted in two stages. For the first stage, initial thrust is gradually increased to 140% of the total weight to quickly lift the UAV off ground. However, it is not excessively strong to avoid sudden acceleration, which could destabilize the UAV. In the second stage, as the UAV reaches approximately 60% of target altitude, the thrust is slightly reduced to ensure a smooth and precise approach to target altitude. This prevents the UAV from overshooting the target altitude or oscillating unstably.

Navigation: The UAV navigates to target locations using GPS coordinates. To optimize travel time, the navigation system enables the UAV to fly in a straight line when there are no obstacles. However, this approach becomes less effective in areas with dense obstacles, such as tall forests. An obstacle avoidance algorithm [26] is integrated with the straight-line navigation method to help the UAV bypass obstacles along its flight path while maintaining time efficiency.

Landing: After completing the mission, the UAV returns to its starting position and lands.

2) No-GPS navigation

No-GPS navigation is designed for environments where GPS signals are unavailable or unreliable. This method uses accelerometers, gyroscopes, and barometers to estimate the UAV position and altitude. The UAV operates in No-GPS mode, which allows it to perform autonomous tasks without GPS data.

Initialization: The UAV connects to ground control system and switches to No-GPS navigation mode. The system checks the UAV in a similar manner to the initialization process in GPS mode.

Takeoff: The takeoff method is similar to that in GPS navigation. However, to control the UAV position during takeoff, a Quaternion-based approach is used. Detail of approach was shown in section a) Quaternion-based approach.

Navigation: Propose method used the Inertial Navigation System for navigation. Detail of approach was shown in section b) Inertial Navigation System.

Landing: The landing process is the same as in GPS navigation.

a) Quaternion-based approach

Specifically, Quaternions are a mathematical representation used to describe the orientation of an object in 3D space. They are superior to Euler angles as they avoid limitations such as gimbal lock and provide a compact and efficient way to represent rotations. In UAV navigation, particularly during the takeoff phase without GPS, the attitude control system based on quaternions ensures the UAV maintains stable attitudes (roll, pitch, yaw) to perform vertical ascent and counteract external disturbances. The rotation angles are denoted as Roll (ϕ), Pitch (θ), Yaw (ψ). The mathematical connection between the No-GPS takeoff algorithm and quaternion-based attitude control can be summarized as follows:

Thrust Control: The thrust (T) is calculated based on current altitude of UAV (*h-current*) and target altitude (*h-target*):

$$T = \begin{cases} DEFAULT\_THRUST, & if\ h_{current} < 0.6 \cdot h_{target} \\ SMOOTH\_THRUST, & if\ h_{current} \geq 0.6 \cdot h_{target} \end{cases} \tag{1}$$

This dynamic adjustment ensures that the UAV ascends smoothly and avoids sudden changes in altitude. In Quaternion Conversion, the desired roll, pitch, and yaw angles are converted into a quaternion using the following equations

$$\begin{aligned} w &= \cos\left(\frac{\psi}{2}\right)\cos\left(\frac{\phi}{2}\right)\cos\left(\frac{\theta}{2}\right) + \sin\left(\frac{\psi}{2}\right)\sin\left(\frac{\phi}{2}\right)\sin\left(\frac{\theta}{2}\right) \\ x &= \cos\left(\frac{\psi}{2}\right)\sin\left(\frac{\phi}{2}\right)\cos\left(\frac{\theta}{2}\right) - \sin\left(\frac{\psi}{2}\right)\cos\left(\frac{\phi}{2}\right)\sin\left(\frac{\theta}{2}\right) \\ y &= \cos\left(\frac{\psi}{2}\right)\cos\left(\frac{\phi}{2}\right)\sin\left(\frac{\theta}{2}\right) + \sin\left(\frac{\psi}{2}\right)\sin\left(\frac{\phi}{2}\right)\cos\left(\frac{\theta}{2}\right) \\ z &= \sin\left(\frac{\psi}{2}\right)\cos\left(\frac{\phi}{2}\right)\cos\left(\frac{\theta}{2}\right) - \cos\left(\frac{\psi}{2}\right)\sin\left(\frac{\phi}{2}\right)\sin\left(\frac{\theta}{2}\right) \end{aligned} \tag{2}$$

where:
w, x, y, z are the quaternion components.

ϕ (roll), θ (pitch), and ψ (yaw) are the desired angles in radians.

b) Inertial navigation system

Basic INS algorithm consists of three main steps: Initialization, Data Collection, and Position Calculation. First, during the Initialization step, the system is set up with initial parameters, including the position (r_0), velocity (v_0) and orientation (C_0). The initial position $r_0 = [x_0, y_0, z_0]^T$ defines the geographic coordinates, the initial velocity $v_0 = [v_{x0}, v_{y0}, v_{z0}]^T$ describes the speed at initialization moment, and the orientation matrix C_0 or quaternion q_0 represents the initial orientation of object. INS continuously collects data from inertial sensors. The accelerometer measures acceleration in the body frame $a_b = [a_x, a_y, a_z]^T$ and the gyroscope measures angular velocity $\omega = [\omega_x, \omega_y, \omega_z]^T$. These data serve as the basis for calculating subsequent positioning parameters.

The Position Calculation step includes orientation updates, acceleration transformation, and velocity and position computation. First, the orientation of object is updated based on the measured angular velocity. If quaternions are used, the orientation update at step k+1 is performed using the following formula.

$$q_{k+1} = q_k + \frac{1}{2}\Delta t \cdot \Omega(\omega_k) \otimes q_k \tag{3}$$

In this, matrix $\Omega(\omega_k)$ and quaternion multiplication $\otimes$ are defined as:

$$\Omega(\omega_k) = \begin{bmatrix} 0 & -\omega_x & -\omega_y & -\omega_z \\ \omega_x & 0 & \omega_z & -\omega_y \\ \omega_y & -\omega_z & 0 & \omega_x \\ \omega_z & \omega_y & -\omega_x & 0 \end{bmatrix} \tag{4}$$

After updating the orientation, the measured acceleration in the body frame is transformed into the world frame using orientation matrix C_k. With $g = [0, 0, -9.81]^T$ m/s^2 represents the gravitational acceleration. The formulas for acceleration, velocity, and position are presented below.

$$\begin{aligned} a_\omega &= C_k \cdot a_b + g \\ w_{k+1} &= v_k + a_\omega \cdot \Delta t \\ r_{k+1} &= r_k + v_{k+1} \cdot \Delta t \end{aligned} \tag{5}$$

3) Object detection and intelligent position adjustment algorithm

The You Only Look Once (YOLO) model is employed for real-time object detection [27], enabling the UAV to identify the target drop-off location by processing images captured from its onboard camera. In this study, the YOLOv11 [28] architecture is utilized to take advantage of its lightweight design, which reduces hardware requirements on the UAV while still ensuring real-time performance. YOLOv11 is the latest improved version in the YOLO model series, designed to enhance object detection performance and expand its applicability to various computer vision tasks. The structure of YOLOv11 is built upon key components, including Spatial Pyramid Pooling-Fast (SPPF), C2PSA (Convolutional Block with Parallel Spatial Attention), and C3K2 (Cross Stage Partial with Kernel Size 2), each of these components plays a critical role in optimizing the performance of model. Key architectural modules of YOLOV11 were shown in Fig. 2.

SPPF is an optimized spatial pooling block designed to enhance the ability to extract spatial information from images of various sizes and scales with SPP [29] background. This block aggregates features from multiple spatial regions in the image into a single feature, enabling the model to detect objects regardless of their size or position. Compared to previous versions, SPPF has been refined to increase processing speed and reduce computational costs while maintaining high accuracy, even in real-time tasks. It is a crucial component for improving feature extraction efficiency in the backbone of YOLOv11.

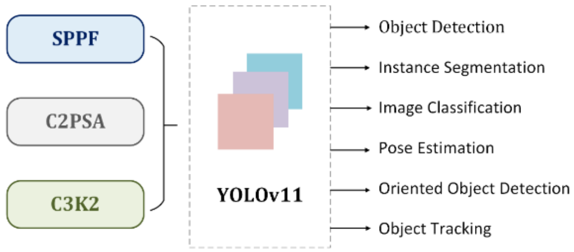

Fig. 2. Key architectural modules in YOLO11 [27].

C2PSA is a convolutional block combined with a parallel spatial attention mechanism, designed to enhance the ability of model to focus on important regions in an image. This attention mechanism allows the model to recognize fine details or occluded objects, improving accuracy in object detection and segmentation. By integrating spatial attention, C2PSA enables the model to concentrate on critical information while minimizing the impact of noise in the image. This is a significant improvement over previous versions, where the ability to focus on small regions was often limited.

C3K2 is a new enhancement of the Cross Stage Partial (CSP) architecture, utilizing a smaller kernel size (kernel size = 2). This modification not only reduces the number of parameters and increases processing speed but also improves feature extraction capabilities. C3K2 enables the model to maintain high performance in object detection tasks, even for objects with complex shapes or those appearing at various scales. C3K2 is integrated throughout the components of YOLOv11, from the backbone to the neck and head, ensuring comprehensive optimization.

Thanks to these improvements, YOLOv11 supports a wide range of important computer vision tasks, including object detection, object segmentation, image classification, pose estimation, oriented object detection, and object tracking. The model is also optimized for applications across various scales, from edge devices with limited resources to high-performance computing systems, due to its balance between speed and accuracy. These advancements make YOLOv11 one of the leading models in the field of real-time computer vision.

The YOLO model is pre-trained on a custom dataset tailored to the delivery environment (e.g., rooftops, lawns, or predefined markers). This model is loaded into the onboard system of UAV to perform real-time inference during flight. The UAV captures frames from its onboard camera and resizes each frame to match the input dimensions required by the YOLO model (640×640 pixels). This ensures compatibility with the model and optimizes detection performance. The YOLO model processes resized frame to detect objects and generate bounding boxes. Bounding box coordinates: (x_1, y_1, x_2, y_2), where (x_1, y_1) and (x_2, y_2) represent the top-left and bottom-right corners of the bounding box. Euclidean Distance was denoted as *distance*. Based on the position of the selected target relative to image center, the algorithm outputs directional indicators (a, b) to guide the UAV movement.

Horizontal Direction:

$$a = \begin{cases} -1 & if\ distance_x < -Threshold_x \\ 1 & if\ distance_x > Threshold_x \\ 0 & if\ \left|distance_x\right| \le Threshold_x \end{cases} \tag{6}$$

Vertical Direction:

$$b = \begin{cases} -1 & if\ distance_y < -Threshold_y \\ 1 & if\ distance_y > Threshold_y \\ 0 & if\ \left|distance_y\right| \le Threshold_y \end{cases} \tag{7}$$

Once the target is detected, the UAV adjusts its position to align with the target drop-off location. The Intelligent Position Adjustment Algorithm ensures precise movement by dynamically controlling the velocity of UAV in response to the target position. The directional indicators (a, b) from the object detection module. The UAV continuously adjusts its position until the target is centered in the camera view. The adjustment logic is as follows:

- If $a = -1$: Move left by sending a velocity command for moving left.
- If $a = 1$: Move right by sending a velocity command for moving right.
- If $b = -1$: Move forward by sending a velocity command moving forward.
- If $b = 1$: Move backward by sending a velocity command for moving backward.
- If $a = 0$: The target is centered, and no further adjustment is needed.

The integration of object detection and position adjustment ensures seamless operation during the UAV delivery process. First, the UAV captures a frame using its onboard camera and detects the target drop-off location using the YOLO model. Based on the detected target position relative to the image center, directional indicators (a, b) are generated to guide the movement of UAV. Using these indicators, the UAV dynamically adjusts its position by sending velocity commands to align itself with the target. This adjustment process is repeated iteratively until target is centered in the view of camera. Once the target is accurately centered, the UAV confirms the drop-off location and releases the package, completing the delivery process with precision.

4) Landmark search

In practice, GPS can sometimes experience significant inaccuracies. This can result in the landmark being out of the view of UAV camera after the UAV navigates to the delivery location based on GPS coordinates. To address this issue, an efficient search method is proposed, where the UAV automatically flies along an Archimedean spiral trajectory to expand the search area. The equation for the Archimedean spiral trajectory is described by Eq. (8), where r is the spiral radius, φ is the rotation angle, and a, b are parameters that control the spacing between spiral loops. UAV follows this trajectory, gradually expanding the search range while ensuring uniform scanning of the surrounding area, increasing the likelihood of detecting the landmark and reducing search time. This method is particularly suitable for scenarios requiring precise localization in open spaces or complex environments. Here, a represents the starting radius, and b determines the distance between two consecutive spiral loops. In this algorithm, parameter a was set by 0. In the Cartesian coordinate system, (x, y) was defined by Eq. (9).

$$R = a + b \cdot \varphi \tag{8}$$

$$\begin{aligned} x &= r \cdot \cos(\varphi) \\ y &= r \cdot \sin(\varphi) \end{aligned} \tag{9}$$

Currently, the three common image formats returned by cameras are rectangular, square, and circular (fisheye). Among these, fisheye images can be converted into a rectangular format using transformation algorithms. Square images can also be considered as a special case of rectangular images. Therefore, to ensure compatibility across a wide range of devices, this algorithm is designed for cases where the view of camera is rectangular with dimensions (h, w).

In Fig. 3, the UAV executes the search algorithm by flying along a spiral trajectory from start point to end point. The blue and orange areas represent the view of camera at different positions. Due to the oblique angle from the camera to the edges of field of view, objects may partially or completely obscure the landmark. To minimize this effect, the algorithm establishes an overlapping area between two fields of view with dimensions (h, m). The parameter m is configured based on the terrain of each specific area.

$$b = \frac{(w - m)}{2\pi} \tag{10}$$

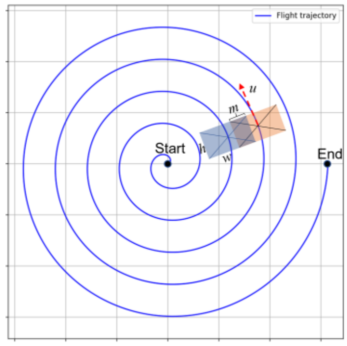

Fig. 3. Illustration of the flight trajectory of the landmark search algorithm.

The stopping condition of this search algorithm is also a hyperparameter determined by the implementer, denoted as φ. In Fig. 3, it is set $\varphi = 10\pi$. During the flight along the trajectory, the UAV direction is tangential to the curve at any given moment. When the view of UAV is represented by the orange area, its flight direction is illustrated by the red vector u.

$$\begin{aligned} u_x &= -(a + b\cdot\varphi)\sin(\varphi) + b\cos(\varphi) \\ u_y &= (a + b\cdot\varphi)\cos(\varphi) + b\sin(\varphi) \end{aligned} \tag{11}$$

IV. Experiment

A. Experiment Setup

1) Enviroment

The experiment was conducted in a rural area in Vietnam with open space, minimal obstacles, and suitable conditions for simulating the operation of the UAV system. The UAV was equipped with an ICM-42688-P sensor, which supports accurate measurements and ensures stability throughout the flight. A 1080p RGBD camera was integrated to provide high-quality images, enabling landmark detection in various scenarios. UAV has a maximum payload capacity of 4 kg, and for the experiment, a 3 kg package was used to ensure stability during operation. The package dimensions were 30×30×30cm, and it was designed with a surface that provides good grip to eliminate the possibility of bouncing to another position when dropped from above. Landmark was designed as a circular shape with a radius of $R_{landmark}$ = 1 m, shown in Fig. 4. It featured a red outer ring, a white circle inside, and a red circle in the center, forming a simple yet easily recognizable symbol from above. In practice, landmarks were selected and placed in diverse locations, including areas with partial obstructions, to test the detection capability of system under real-world conditions.

2) Model configuration for detection

YOLOv11 model was used in the experiment, offering outstanding advantages in terms of fast processing and high accuracy for real-time detection. The training dataset consisted of 1,651 images, each containing a single landmark, with 10% of the images adjusted to simulate cases where the landmark was partially occluded by up to 30%. The training parameters were configured as follows: 100 epochs, batch size 16, and learning rate 0.01.

Fig. 4. The designed landmark on the left and its real-world counterpart on the right.

3) Experiment process

The experiment was divided into two scenarios to evaluate the effectiveness of the UAV system under different operating conditions. In the first scenario, the GPS functioned normally, and the UAV used GPS coordinates to determine the drop-off location. In the second scenario, GPS errors were simulated by artificially altering the GPS coordinates to test the ability of system to search for and detect the landmark under adverse conditions. Specifically, a random nearby destination was assigned to the UAV. Each scenario included a total of 150 drop-offs, divided into three groups with different altitudes: 3 m, 5 m, and 10 m. The landmarks were placed in diverse locations, including areas with obstacles, to assess the detection capability in complex situations. The accuracy of drop-offs was evaluated at three levels: If the center of package landed within a circle with a radius of 1m, it was considered accurate. If the center of package landed outside this radius, it was considered inaccurate.

B. Detection Model Evaluation

In this study, YOLOv11 model was trained and evaluated on a dataset consisting of 1,651 images, including real-world scenarios and simulated cases where landmarks were partially occluded. The detection results were assessed using metrics Precision (P), Recall (R), mAP50, and mAP50-95 as follows.

Table I shows that the YOLOv11 model achieved high accuracy in detecting landmarks across various scenarios. A Precision of 99.4% indicates the accuracy of model in correctly identifying landmarks when proposed. A Recall of 99.6% confirms that the model almost never missed any landmarks in the dataset. Notably, the mAP50 value of 99.5% demonstrates the consistent object detection performance of model under standardized conditions. For mAP50-95, the value of 97.6% highlights the ability of model to maintain high effectiveness across multiple IoU thresholds.

TABLE I. Evaluation Metrics of Detection Model

Metrics	P	R	mAP50	mAP50-95
Value	0.994	0.996	0.995	0.976

The confusion matrix in Fig. 5 illustrates the performance of model at a confidence threshold of 0.5 in classifying two classes: class 0 (landmark) and class 1 (background). The model correctly predicted 327 instances of landmarks (True Positive—TP) and misclassified only 3 background instances as landmarks (False Positive—FP). However, there was 1 actual landmark that was misclassified as background (False Negative—FN), and there were no misclassifications from background to landmark (True Negative—TN is 0).

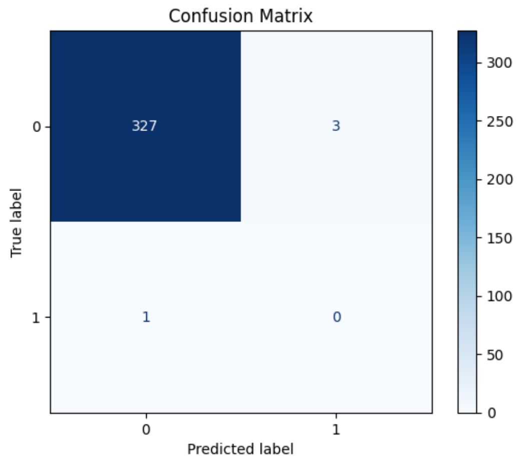

Fig. 5. Confusion matrix of landmark detection model.

Fig. 6 illustrates the relationship between Precision and the confidence threshold of the landmark detection model. The curve shows that model achieves high precision even at low confidence levels. At a confidence threshold of 0.769, the model reaches maximum precision of 1.00, demonstrating its strong classification capability when identifying objects in the landmark class. Additionally, the model exhibits stable performance across the entire confidence range, with precision only slightly decreasing at confidence thresholds between 0.2 and 0.769.

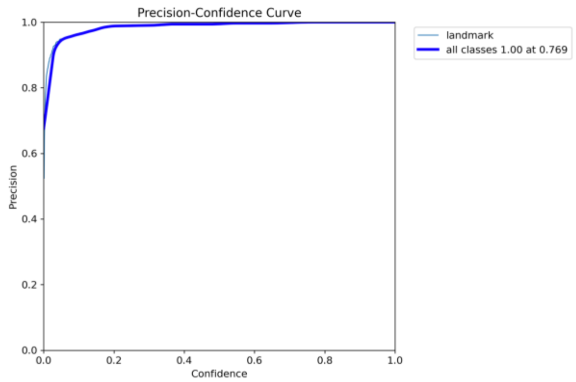

Fig. 6. Precision-confidence curve of landmark detection model.

Fig. 7 above is the Recall-Confidence curve, illustrating the relationship between detection capability, also called recall, and confidence threshold of the model for the landmark class. The curve shows that the model achieves a high Recall value (close to 1.0) at confidence levels below 0.7. This indicates that the model can detect nearly all landmarks in the dataset when the confidence threshold is set as low as 0.5.

C. *Drop-Shipping Evaluation*

1) GPS-accessed case

In the stable GPS scenario, the UAV operated with high accuracy when performing package drops at different altitudes. The results are presented in Table II.

TABLE II. EVALUATION RESULTS OF PACKAGE DROPPING FROM DIFFERENT ALTITUDES IN GPS-ACCESSED CASE

Case	3 m	5 m	10 m
Number of Accurate Drops	50	49	46
Number of Inaccurate Drops	0	1	4
Average Distance to Landmark Center (m)	0.32	0.41	0.73

Table II shows that at a low altitude (3 m), the UAV achieved the highest drop accuracy (100%) with the lowest average error from the landmark center (0.32 m). As altitude increased, the accuracy slightly decreased, and the average error gradually increased. The results from the stable GPS scenario demonstrate that the model and the algorithm, which combine GPS with Landmark Detection, have strong localization capabilities, especially under robust GPS conditions. The packages were accurately dropped into the target area with minimal error. This highlights the high applicability of the system in ideal conditions or areas with strong GPS signals.

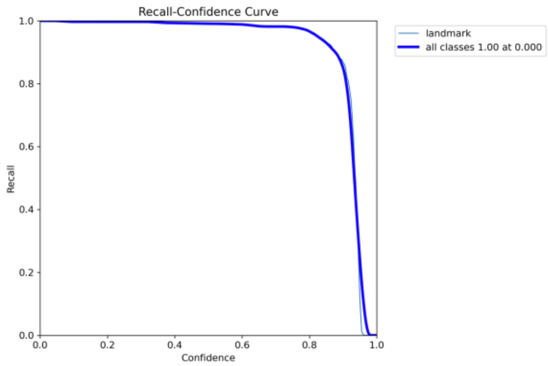

Fig. 7. Recall-confidence curve of landmark detection model.

2) GPS-denined case

In reality, GPS signals are not always stable and can often be affected by environmental factors, causing GPS coordinates to become inaccurate. This can result in the UAV moving to the wrong location and failing to locate the landmark. The proposed method allows UAV to fly along a spiral trajectory from the GPS-designated position to search for the landmark. Table III presents the experimental results in the case of unstable GPS with a parameter m = 0.25 W. The term W denotes for width of camera view. The results show that using search algorithms has almost no impact on the distance error during package drops compared to the stable GPS scenario.

Table IV provides a more detailed evaluation of the ability to locate landmarks as influenced by the parameter mmm. The results show that as the overlap level mmm increases from 0.25 W to 0.33 W and then 0.50 W, landmark detection rate improves significantly, reaching

94%, 98%, and 100%, respectively. This demonstrates that increasing the overlap level helps minimize the likelihood of missing landmarks, especially in areas with complex terrain or obstacles.

TABLE III. EVALUATION RESULTS OF PACKAGE DROPPING FROM DIFFERENT ALTITUDES IN GPS-DENIED CASE

Case	3m	5m	10m
Number of Accurate Drops	47	46	43
Number of Inaccurate Drops	3	4	7
Average Distance to Landmark Center (m)	0.35	0.43	0.69

TABLE IV. THE SUCCESS RATE OF LANDMARK DETECTION WITH DIFFERENT M VALUES

M = 0.25 W	M = 0.33 W	M= 0.50 W
94%	98%	100%

Additionally, experiments were conducted to evaluate the time efficiency of the proposed method. The experimental results, presented in Table V, show significant differences between the various delivery scenarios. Each result is averaged over 10 measurements using the UAV configuration described earlier. Compared to the scenario where the UAV lands on the ground before delivering the package (11.43 s), other scenarios demonstrate significantly shorter delivery times. Specifically, direct dropping from an altitude of 10m is the fastest option, taking only 0.27 s, as it does not require the UAV to descend. The 0.27 s represents the time required for release mechanism to operate. In the scenario where the UAV descends to an altitude of 5 m before dropping, the total time is reduced to 6.38 s, saving 44% compared to the landing scenario. Similarly, when the UAV descends to an altitude of 3 m before dropping, the total time is 8.02 s, a 30% reduction compared to the full landing scenario.

TABLE V. EVALUATION OF DELIVERY TIME AT DIFFERENT ALTITUDES

Case	Drop Shipping Time (s)
Direct Drop from 10 m	0.27
Descend to 5m before dropping	6.38
Descend to 3m before dropping	8.02
Descend to Ground before Delivering	11.43

V. DISCUSSION

By integrating GPS and Landmark Detection, the system enables the UAV to accurately determine the drop-off location without the need to land, thereby minimizing processing time and increasing operational efficiency. The Archimedean spiral trajectory algorithm allows the UAV to expand its search area when GPS is unstable, ensuring the ability to detect landmarks in complex environments. The method of dropping packages from above not only saves time but also enhances flexibility, allowing the UAV to easily access challenging areas without relying on terrain conditions. These advantages highlight the high applicability of the system in emergency situations, such as disaster relief operations.

In this study, the YOLO model was utilized to leverage its lightweight architecture while maintaining efficiency in detection. However, the application of deep learning models always comes with hardware requirements, which can lead to high costs in practical deployment. Another important factor in the system is the parameter mmm, which represents the overlap level between camera fields of view during the landmark search process. This parameter directly affects the ability to detect landmarks, and the time required to complete the task. In the study, m was tested with fixed values such as 0.25 W, 0.33 W, and 0.50 W, However, to optimize the system for diverse real-world conditions, additional field experiments are needed to adjust mmm appropriately for each specific scenario.

VI. CONCLUSION

This study developed a UAV system integrating GPS and Landmark Detection, providing an effective solution for optimizing delivery in emergency situations and complex environments. The method combines the global positioning capability of GPS with the precision of Landmark Detection, enabling the UAV to accurately determine the drop-off location without the need to land, thereby saving time and improving operational efficiency.

The system demonstrated high performance in real-world experiments, even under unstable GPS conditions. The search algorithm based on the Archimedean spiral trajectory enabled the UAV to effectively expand its scanning range and improve landmark detection capability. Notably, the system achieved a 98% package drop accuracy at an altitude of 5m, highlighting its ability to maintain high precision even when operating at medium altitudes. The results also showed an optimal landmark detection rate (100%) when the overlap level M = 0.50 W. With its high flexibility, this solution can make a significant contribution to improving efficiency in autonomous delivery and providing support in emergency or disaster situations.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Cao-Ky-Long U conceptualized the research framework and developed the UAV integration system of GPS and Landmark Detection. Nguyen Khac Toan designed the experiments, implemented the YOLO-based object detection model, and analyzed the experimental data. Cao-Ky-Long U conducted practical trials, optimized the UAV navigation algorithms, and drafted the manuscript. Both authors reviewed, edited, and approved the final manuscript.

FUNDING

This research is funded by University of Economics Ho Chi Minh City—UEH University, Vietnam.

References

[1] Z. Liu *et al.*, "Robot learning towards smart robotic manufacturing: A review," *Robotics and Computer-Integrated Manufacturing*, vol. 77, 102360, 2022. doi: 10.1016/j.rcim.2022.102360

[2] N. K. Toan *et al.*, "Development of humanoid robot head based on FACS," *International Journal of Mechanical Engineering and Robotics Research*, vol. 11, no. 5, pp. 365–372, 2022. doi: 10.18178/ijmerr.11.5.365-372

[3] N. Amarasingam *et al.*, "A review of UAV platforms, sensors, and applications for monitoring of sugarcane crops," *Remote Sensing Applications: Society and Environment*, vol. 26, 100712, 2022. doi: 10.1016/j.rsase.2022.100712

[4] V. Garg *et al.*, "Drones in last-mile delivery: A systematic review on efficiency, accessibility, and sustainability," *Transportation Research Part D: Transport and Environment*, vol. 123, 103831, 2023. doi: 10.1016/j.trd.2023.103831

[5] A. Li *et al.*, "Traffic management and resource allocation for UAV-based parcel delivery in low-altitude urban space," *Transportation Research Part C: Emerging Technologies*, vol. 143, 103808, 2022. doi: 10.1016/j.trc.2022.103808

[6] S. D. Eksioglu *et al.*, "Designing drone delivery networks for vaccine supply chain: A case study of Niger," *IISE Transactions on Healthcare Systems Engineering*, vol. 14, no. 3, pp. 193–213, 2024. doi: 10.48550/arXiv.2208.04357

[7] K. Messaoudi *et al.*, "A survey of UAV-based data collection: Challenges, solutions and future perspectives," *Journal of Network and Computer Applications*, vol. 216, 103670, 2023. doi: 10.1016/j.jnca.2023.103670

[8] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006. doi: 10.1109/MRA.2006.1638022

[9] S. E. Reutebuch *et al.*, "Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory," *Journal of Forestry*, vol. 103, no. 6, pp. 286–292, 2005. doi: 10.1093/jof/103.6.286

[10] J. Kwak and Y. Sung, "Autonomous UAV flight control for GPS-based navigation," *IEEE Access*, vol. 6, pp. 37947–37955, 2018. doi: 10.1109/ACCESS.2018.2854712

[11] H. Eskandaripour and E. Boldsaikhan, "Last-mile drone delivery: Past, present, and future," *Drones*, vol. 7, no. 2, 77, 2023. doi: 10.3390/drones7020077

[12] G. Brunner *et al.*, "The urban last mile problem: Autonomous drone delivery to your balcony," in *Proc. 2019 International Conference on Unmanned Aircraft Systems (ICUAS)*, IEEE, 2019, pp. 1005–1012. doi: 10.48550/arXiv.1809.08022

[13] G. Balamurugan *et al.*, "Survey on UAV navigation in GPS denied environments," in *Proc. 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, IEEE, 2016, pp. 198–204. doi: 10.1109/SCOPES.2016.7955787

[14] H. Wang *et al.*, "Intensity-slam: Intensity assisted localization and mapping for large scale environment," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1715–1721, 2021. doi: 10.1109/LRA.2021.3059567

[15] A. Steenbeek and F. Nex, "CNN-based dense monocular visual SLAM for real-time UAV exploration in emergency conditions," *Drones*, vol. 6, no. 3, 79, 2022. doi: 10.3390/drones6030079

[16] Y. Liu *et al.*, "A survey on semantic communications: Technologies, solutions, applications and challenges," *Digital Communications and Networks*, 2023. doi: 10.1016/j.dcan.2023.05.010

[17] K. Feng *et al.*, "Packages delivery based on marker detection for UAVs," in *Proc. 2020 Chinese Control And Decision Conference (CCDC)*, IEEE, 2020, pp. 2094–2099. doi: 10.1109/CCDC49329.2020.9164677

[18] S. Wang *et al.*, "Tolerant semantic segmentation for aerial logistics," in *Proc. DAGM German Conference on Pattern Recognition*, Springer, 2021, pp. 515–529. doi: 10.1007/978-3-030-92659-5_33

[19] H. Wang *et al.*, "Semantic segmentation algorithms for ground AGV and UAV medical transport scenes," in *Proc. 2022 16th ICME International Conference on Complex Medical Engineering (CME)*, IEEE, 2022, pp. 252–255. doi: 10.1109/CME55444.2022.10063288

[20] S. S. Kannan and B.-C. Min, "Autonomous drone delivery to your door and yard," in *Proc. 2022 International Conference on Unmanned Aircraft Systems (ICUAS)*, IEEE, 2022, pp. 452–461. doi: 10.48550/arXiv.2104.05503

[21] P. Mittal *et al.*, "Deep learning-based object detection in low-altitude UAV datasets: A survey," *Image and Vision Computing*, vol. 104, 104046, 2020. doi: 10.1016/j.imavis.2020.104046

[22] D. Dissanayaka *et al.*, "Review of navigation methods for uav-based parcel delivery," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 1, pp. 1068–1082, 2023. doi: 10.1109/TASE.2022.3232025

[23] A. Dawadee *et al.*, "A method for autonomous navigation of uavs using landmarks," in *Proc. the 16th Australian Aerospace Congress*, 2015, 022015.

[24] E. Innocenti *et al.*, "Uavs for medicine delivery in a smart city using fiducial markers," *Information*, vol. 13, no. 10, 501, 2022. doi: 10.3390/info13100501

[25] S. Xia *et al.*, "SSS: Towards autonomous drone delivery to your door over house-aware semantics," in *Proc. the 25th International Workshop on Mobile Computing Systems and Applications*, 2024, pp. 33–39.

[26] D. Wang *et al.*, "UAV environmental perception and autonomous obstacle avoidance: A deep learning and depth camera combined solution," *Computers and Electronics in Agriculture*, vol. 175, pp. 105523, 2020. doi: 10.1016/j.compag.2020.105523

[27] P. Jiang *et al.*, "A review of Yolo algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022. doi: 10.1016/j.procs.2022.01.135

[28] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," arXiv preprint, arXiv:2410.17725, 2024.

[29] K. He *et al.*, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015. doi: 10.1109/TPAMI.2015.2389824