Improved Robotic Platform to perform Maintenance and Upgrading Roadworks: The HERON Approach

Grant Agreement Number: 955356

D5.2: Drones Implementation

Work package	WP5: Development of the Robotic Platform, Improved Navigation and integration with the sensing devices
Activity	Task 5.2: Deployment of the Aerial Drones and integration with the UGVs
Deliverable	D5.2: Drones Implementation
Authors	Nikos Bakalos, Iason Katsamenis, Anastasios Doulamis, Nikolaos Doulamis, Dimitris Kalogeras, Manthos Bimpas
Status	Final (F)
Version	1.0
Dissemination Level	Public (PU)
Document date	30/04/2024
Delivery due date	30/09/2023
Actual delivery date	30/04/2024
Internal Reviewers	Christos Polykretis (STWS), Miquel Cantero (ROB)
	This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 955356.

Document Control Sheet

Version history table			
Version	**Date**	**Modification reason**	**Modifier**
0.1	21/08/2023	Initial Table of Contents and basic structure of the deliverable	Iason Katsamenis
0.2	12/09/2023	Initial draft	Nikos Bakalos
0.3	15/09/2023	Data-capturing sessions	Iason Katsamenis
0.4	27/10/2023	Updated Deliverable structure	Nikos Bakalos
0.5	01/03/2024	Updated data-capturing sessions and object detection techniques	Iason Katsamenis
0.6	01/04/2024	Localization framework and integration of the UAV information with the UGV	Nikos Bakalos
0.7	17/04/2024	Initial deliverable version ready for internal review	Nikos Bakalos, Iason Katsamenis
1.0	30/04/2024	Final version ready for submission	Nikos Bakalos, Iason Katsamenis

Legal Disclaimer

Table of Contents

List of Tables

List of Figures

Abbreviation Lists

Table 1: Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
UGV	Unmanned ground vehicle
UAV	Unmanned Aerial Vehicle
CV	Computer Vision
WP	Work Package
RTK	Real Time Kinematics
LiDAR	LIght Detection And Ranging
CMOS	Complementary Metal Oxide Semiconductor
RI	Road Infrastructure
MP	Megapixel
PoI	Point of Interest
GPS	Global Positioning System
CDMA	Code-Division Multiple Access
RUP	Removable Urban Pavement
FCC	Federal Communications Commission
GNSS	Global Navigation Satellite System
GLONASS	GLObalnaya NAvigatsionnaya Sputnikovaya Sistema
CSP	Cross Stage Partial
YOLO	You Only Look Once
mAP	Mean Average Precision
SAM	Segment Anything Model
FoV	Field Of View
EXIF	Exchangeable Image File Format

Table 2: Abbreviations of the Partners' names

Short name	Participant organization name
ICCS	Institute of Communications and Computer Systems
ACCI	Acciona Construcción S.A.
OLO	Olympia Odos Operation S.A.
UGE	Université Gustave Eiffel
ETHZ	Eidgenössische Technische Hochschule Zürich
ROB	Robotnik Automation
CORTE	Confederation of Organisations in Road Transport Enforcement
STWS	SATWAYS - Olokliromenes Lyseis Asfaleias Kai Amynas-idiotiki Epicheirisi Parochis Ypiresion Asfaleias (Iepya)-etaireia Periorismenis Efthynis
RISA	RisaSicherheitsanalysen Gmbh
INAC	InnovActs
IKH	Ainoouchaou Pliroforiki SA -IKnowHow-
RG	Resilience Guard Gmbh

Executive Summary

This deliverable is written in the framework of WP5 – Development of the Robotic Platform, Improved Navigation and integration with the sensing devices of the HERON project under Grant Agreement No. 955356. Deliverable 5.2, namely "Drones Implementation" provides a detailed description of the utilized inspection drones within the HERON framework that are coordinated to operate in a flexible manner to facilitate maintenance works and the pre-/post-intervention phase. This report illustrates the outcomes of Task 5.2, titled: "Deployment of the Aerial Drones and integration with the UGVs" corresponding to M15-M28 of the HERON project's period.

To this end, the document presents an overview of the implementation process of the drone platform with a highlight on the hardware, the localization framework and the utilized computer vision perception mechanisms using aerial imaging data. Thus, the UAVs can gather pertinent data from targeted scenes, serving dual purposes: aiding in preoperative planning and supporting ongoing operations. Lastly, D5.2 details the coordination and communication between the different robotic platforms and in particular the integration of aerial information with the Unmanned Ground Vehicle and the broader HERON system.

1 Introduction

1.1 Purpose of the Document

The document contains deliverable D5.2 “Drone Implementations”. This deliverable is a compilation of the work that was completed in the framework of Task 5.2 “Deployment of the Aerial Drones and Integration with the UGVs”. D5.2 provides a comprehensive overview of the technological enhancements achieved in the HERON project, focusing specifically on aerial drone capabilities. This document delineates the deployment of drones, a robust localization framework, and state-of-the-art computer vision algorithms, including object detection and image segmentation, which are founded on the deep learning models developed in WP3. Furthermore, D5.2 elaborates on the seamless integration of these aerial capabilities with the Unmanned Ground Vehicle (UGV) and the broader HERON system, facilitating the synergistic benefits and operational efficiency gained through this multimodal communication.

The remainder of this document is organized as follows: Initially, Section 2 provides an overview of the implementation of the HERON drone system, detailing its design and deployment strategies. Section 3 delves into the data-capturing sessions, outlining the procedures and methodologies employed during inspection missions within the HERON project. Subsequently, Section 4 elaborates on the localization framework utilized by the system, discussing the various techniques used to precisely determine the drone's position and orientation. Following this, Section 5 explores perception using aerial imaging data, examining how visual information captured by drones is processed and analyzed by state-of-the-art computer vision algorithms to detect and classify road infrastructure defects and PoIs. In Section 6, the integration of aerial information with the UGV is discussed, highlighting how data collected by drones informs and enhances the operations of the HERON system. Lastly, Section 7 concludes this report.

1.2 Intended Audience

This specific deliverable report is public and therefore can be accessed by any interested stakeholder. Envisioned stakeholders involve, amongst others, the HERON end users. In particular, these are road operators, who are agents in the monitoring procedure as well as monitoring information consumers. Other envisioned stakeholders could be those interested in consuming aerial information to develop autonomous inspection drones.

1.3 Relation with other Deliverables

The guidelines set forth in HERON deliverables D2.1 and D2.2, titled "End-user needs and KPIs report" and "Architecture specification" respectively, lay the groundwork for this document. D2.1 outlines the users' requirements, encompassing an analysis of current practices, needs, and expectations from infrastructure stakeholders. Meanwhile, D2.2 provides specifications for the HERON platform architecture, along with guidelines and toolsets for development activities. These deliverables directly address the challenges and limitations faced by the UAV inspection framework of the HERON system, aiming to facilitate maintenance works and the pre-/post-intervention phase and streamline RI maintenance procedures effectively.

Deliverable 5.2 outlines the deployment and integration of aerial drones with the HERON UGV, aimed at facilitating maintenance operations and the pre-/post-intervention phase and thus its outcomes will be utilized in Task 5.3 “Adaptation of the HERON Robotic Platform”. Moreover, this integration interfaces with all the technical and development WPs within the HERON project, including the deep CV perception systems (WP3), the motion and high-level planner (WP4), and the integrated incident management system designed to support decisions made by road operators and managers (WP6). Lastly, the perception and localization methodologies outlined in this deliverable require rigorous testing, evaluation, and refinement during the pilots conducted as part of WP7, "Field integration, demonstration, and validation activities".

2 Overview of the Drone Implementation

2.1 Deployed equipment

For the purposes of this deliverable two drones were deployed, for data-capturing sessions, the DJI Air 2S and the DJI Mavic 2 pro. The data-capturing sessions (described in section 3) were executed to (i) assess the WP3 computer vision models performance when extensively using UAV imaging data, (ii) gather training data for extending the models with additional samples, (iii) assess the localization framework that will be used.

For the upcoming HERON trial activities and future data-capturing sessions, ICCS is advancing its technological capabilities by acquiring a more sophisticated UAV—the DJI Matrice 350 RTK. This state-of-the-art drone is equipped with advanced depth perception sensors utilizing LiDAR technology, enhanced imaging and localization sensors, and expanded operational capacities. The procurement process for this UAV is currently, significantly augmenting ICCS's research and operational capabilities.

2.1.1 DJI Air 2S

Figure 1: The DJI Air 2S UAV

The DJI Air 2S drone is equipped with numerous advanced features optimized for high-performance flying and media capture. Overall, the DJI Air 2S is designed for both amateur and professional users who require high-end performance, excellent imaging capabilities, and reliable drone behavior in diverse environmental conditions. Key specifications include [1]:

- **Aircraft Specifications:** It has a takeoff weight of 595 g and offers dimensions that vary when folded (180×97×77 mm) and unfolded (183×253×77 mm). The drone can ascend and descend at speeds of up to 6 m/s and can reach a maximum service ceiling of 5000 m above sea level. Its maximum flight time is approximately 31 minutes without wind, and it can cover a distance of 18.5 km.
- **Camera Capabilities:** The Air 2S features a 1" CMOS sensor capable of capturing 20 MP still images and supports a variety of formats and sizes. It offers impressive video resolutions up to 5.4K at 30 fps. The camera system includes multiple photography modes, such as Burst shooting and Auto Exposure Bracketing.
- **Battery and Power:** The intelligent flight battery has a capacity of around 3500 to 3750 mAh, providing a substantial energy store of about 40.42 to 41.4 Wh, and supports a max charging power of 38 W.

- **Flight Performance:** The drone can fly at speeds up to 19 m/s in its fastest mode and offers robust wind resistance of up to 10.7 m/s. It includes advanced flight modes facilitated by its GPS, GLONASS, and GALILEO systems.
- **Gimbal and Sensing Systems:** The 3-axis gimbal provides stability and supports a wide range of motion for detailed angle control. The sensing system is comprehensive, including forward, backward, downward, and upward sensors, enhancing its obstacle avoidance capabilities.
- **Transmission and Control**: The drone uses DJI's O3 transmission system for robust, high-definition video transmission over distances up to 12 km under FCC regulations. It supports various mobile device connectors and offers an extended remote controller battery life.

2.1.2 DJI Mavic 2 Pro

Figure 2: The DJI Mavic 2 Pro drone

The DJI Mavic 2 series, which includes the Mavic 2 Pro, is designed to cater to diverse photography and videography needs with sophisticated features tailored for both models. The Mavic 2 series combines high-quality imaging capabilities with robust flight performance and safety features, making it a leading choice for advanced aerial photography and videography. Key specifications include [2]:

- **Camera Capabilities:** The Mavic 2 Pro boasts a 1” CMOS sensor with 20 million effective pixels and a lens with an adjustable aperture range of f/2.8 to f/11. The Mavic 2 Zoom features a smaller 1/2.3" CMOS sensor with 12 million effective pixels and a unique 24-48 mm zoom lens. Both models support a range of video resolutions up to 4K, various photography modes, and have advanced color profiles like Dlog-M and D-Cinelike for enhanced post-production flexibility.
- **Aircraft Specifications:** Both drones have similar physical dimensions, with a takeoff weight of approximately 907 g for the Pro and 905 g for the Zoom. They can ascend at speeds up to 5 m/s in S-mode, feature a max flight time of 31 minutes, and can operate over a maximum distance of 18 km without wind.
- **Performance and Sensing:** The series is equipped with OcuSync 2.0 technology for video transmission, offering a maximum distance of up to 10 km under FCC regulations. The drones include an omnidirectional obstacle-sensing system, which enhances their safety by providing detailed environmental awareness, though it has specific limitations in certain modes and directions.
- **Gimbal and Stability:** Both models are equipped with a 3-axis gimbal that ensures stable footage. The control ranges and vibration resistance of the gimbals differ slightly between the two models, reflecting their specialized use cases.

- **Battery and Power:** The intelligent flight battery for both models has a capacity of 3850 mAh, which supports extended flight times and efficient power management. The drones can be operated in temperatures ranging from -10°C to 40°C, making them versatile for various climatic conditions.
- **Additional Features:** The drones support microSD cards for expanded storage and come with a range of automated flight modes, enhancing their usability for both professional cinematographers and hobbyists.

3 Data-Capturing Sessions

As mentioned before, during the last period, the project has performed a number of data-capturing sessions, to gather additional data for the drone. These data were used in assessing the performance of the already established models, developed in WP3, and for gathering additional training samples to improve the operational capabilities of such models. These data-capturing activities took place in two different places, the university campus of the National Kapodistrian University in Athens Greece, and on the UGE premises in Bouguenais, near Nantes .

3.1 National Kapodistrian University Campus – Athens, Greece

The National Kapodistrian University, Athens' largest educational institution, is situated in the Zografou borough. In the summer of 2023, a portion of the campus roads sustained significant damage due to extreme weather events. Located in close proximity to ICCS and within the same urban area of Athens, Greece, the university campus was an ideal choice for research-related activities as it allowed drone operations without the need for special permits. Consequently, it was selected as a site for data-capturing activities. Over the course of Autumn and Winter 2023-2024, three data-capturing sessions were conducted. These sessions involved deploying drones at various altitudes and under different weather and lighting conditions, enabling comprehensive data collection. A view of the overall area, as captured by the UAVs can be viewed below.

Figure 3: View from UAV of the road surface at National Kapodistrian University of Athens

3.2 UGE Premises – Bouguenais, France

To proceed with the Removeable Urban Pavement (RUP) use case, a data-capturing specifically targeting the RUP scenario took place in France in February 2024. The session took place in the Nantes region.

In 2020, as part of a mechanical and hydraulic design study on RUP, the Gustave Eiffel University (Nantes, France) constructed a test section measuring 8.51 meters in length and 2.31 meters in width.

(a) (b)

Figure 4: (a) Construction of the test structure at Gustave Eiffel University, (b) Fatigue Carousel (Gustave Eiffel University – Nantes, France).

This section comprises slabs with a 46 cm edge and 23 cm thickness, including a surface layer of 4 cm porous concrete for the 22 double-layer hexagonal slabs with a weight of around 280 kg, 4 cm hydraulic concrete for the 15 half-edge slabs, and 2 single-layer quarter-slabs. These slabs have undergone an accelerated test with a fatigue carousel. Since 1984, Gustave Eiffel University has had a system to test pavements under heavy vehicle loads at speeds up to 100 km/h (see Figure 4(b)). This helps validate new structures by testing for fatigue before moving to larger road tests. So, from November 2020 to February 2021, the RUPs structure was subjected to 200,000 cycles of loading by a half-axle with dual wheels loaded at 65 kN (which corresponds on average to traffic between 45 and 91 buses/day/direction for 20 years). One of the current objectives is to identify potential structural anomalies of RUPs through multi-technique methods on the surface and subsurface [3]. The overall section can be viewed in the figure below.

Figure 5: RUP Section in Nantes UGE Campus.

4 Localization Framework

The deployed equipment has access to three different positioning systems namely GPS, GALILEO, and GLONASS. In this section, we briefly present those positioning systems.

4.1 GALILEO

GALILEO represents the European Union and European Space Agency's venture into satellite navigation technology. Conceived as an independent global navigation satellite system, GALILEO aims to reduce dependence on foreign systems like GPS and provide Europe with strategic autonomy in navigation services. The project traces its origins to the early 2000s, with the European Commission spearheading its development in collaboration with various European and international partners. After years of planning, research, and technological development, GALILEO became operational in 2016, marking a significant milestone in Europe's space exploration and positioning capabilities.

4.1.1 Technical Specifications

GALILEO operates through a constellation of 24 active satellites, comprising three orbital planes with eight satellites each. These satellites emit signals across three frequency bands: E1 (around 1.575 GHz), E5a (around 1.176 GHz), and E5b (around 1.207 GHz). Unlike GPS and GLONASS, GALILEO's signal architecture incorporates binary offset carrier modulation, which enhances signal robustness and interoperability. The signals broadcast by GALILEO satellites enable precise positioning, navigation, and timing measurements, utilizing advanced techniques such as multiplexed binary offset carrier modulation and triple-frequency signal reception.

4.1.2 Localization Accuracy

GALILEO promises high-precision positioning capabilities, with expected accuracies within one meter for most users and even greater accuracy achievable through advanced techniques such as real-time kinematic (RTK) positioning and precise point positioning (PPP). The availability of signals across multiple frequency bands and advanced modulation schemes enhances GALILEO's resistance to signal interference and improves positioning accuracy, especially in challenging environments. Moreover, GALILEO incorporates authentication and integrity mechanisms to ensure the trustworthiness and reliability of its signals, bolstering confidence in its positioning solutions.

4.1.3 Coverage

GALILEO provides global coverage, ensuring that at least four satellites are visible from any point on Earth at any given time. The constellation's distribution across multiple orbital planes enhances coverage and ensures robust positioning capabilities worldwide. GALILEO satellites are strategically positioned to complement existing navigation systems like GPS and GLONASS, offering interoperability and redundancy. Additionally, GALILEO's emphasis on open access and compatibility with other systems promotes international cooperation and collaboration in the field of satellite navigation.

4.2 GPS

The Global Positioning System (GPS) represents a landmark achievement in satellite-based navigation technology, originating from the United States Department of Defense's efforts in the late 20th century. Conceived as a military initiative, GPS aimed to provide precise positioning, navigation, and timing services for military operations, enhancing situational awareness and strategic capabilities. The system's development commenced in the 1970s, with the first satellite launch in 1978, and achieved full operational capability in 1995, heralding a new era of global navigation. Since then, GPS has undergone continual upgrades and enhancements, expanding its utility to civilian and commercial applications worldwide.

4.2.1 Technical Specifications

GPS operates through a constellation of 24 satellites orbiting the Earth, arranged in six orbital planes. Each satellite emits signals in the L-band frequency range, specifically on two frequencies: L1 (around 1.575 GHz) and L2 (around 1.227 GHz). These signals facilitate precise positioning and timing measurements, utilizing a technique known as trilateration. By receiving signals from at least four satellites, a GPS receiver can calculate its three-dimensional position, as well as its precise time. GPS employs a code-division multiple access (CDMA) scheme for signal transmission, enabling multiple signals to coexist within the same frequency band.

4.2.2 Localization Accuracy

GPS offers high localization accuracy, typically ranging from a few meters to centimeters, depending on the quality of received signals and environmental conditions. Modern GPS receivers equipped with advanced processing capabilities and multi-frequency support can achieve sub-meter-level accuracy under optimal circumstances. Factors such as satellite geometry, signal strength, and atmospheric conditions influence the accuracy of GPS positioning. Additionally, differential GPS (DGPS) techniques, which involve correcting GPS measurements using reference stations, further enhance accuracy for critical applications.

4.2.3 Coverage

GPS provides global coverage, ensuring that at least four satellites are visible from any point on Earth at any given time. The distribution of satellites across multiple orbital planes enhances coverage and ensures robust positioning capabilities worldwide. Moreover, GPS satellites are continuously monitored and maintained to ensure their operational integrity and availability. GPS signals penetrate through most atmospheric and environmental conditions, making them suitable for a wide range of applications, from outdoor navigation to precision agriculture and aviation.

4.3 GLONASS

GLONASS, an acronym for Global'naya Navigatsionnaya Sputnikovaya Sistema, emerged from the Soviet Union's ambitious space program in the 1970s. Conceived as a counterpart to the American GPS (Global Positioning System), GLONASS aimed to establish a global navigation satellite system that could provide precise positioning and timing services for both

military and civilian purposes. The system's development culminated in its operational deployment in 1993, marking a significant milestone in Russia's space exploration endeavors. Over the years, GLONASS has undergone extensive modernization efforts, overseen initially by the Russian Space Forces and later by the Russian Space Agency, ensuring its relevance and effectiveness in the rapidly evolving field of satellite navigation.

4.3.1 Technical Specifications

GLONASS operates through a constellation of satellites orbiting the Earth, currently comprising 24 active satellites distributed across three orbital planes. These satellites continuously emit signals in the L-band frequency range, specifically on two frequencies: L1 (around 1.602 GHz) and L2 (around 1.246 GHz). Unlike GPS, GLONASS employs frequency-division multiple access (FDMA) for signal transmission, distinguishing its signal architecture. The signals broadcast by GLONASS satellites enable precise positioning and timing measurements, relying on trilateration principles similar to those used by GPS. By receiving signals from at least four satellites, a GLONASS receiver can calculate its three-dimensional position, as well as its precise time.

4.3.2 Localization Accuracy

The localization accuracy of GLONASS depends on several factors, including the number of satellites in view, the quality of received signals, and environmental conditions. Under optimal circumstances, modern GLONASS receivers can achieve positioning accuracies comparable to those of GPS, typically ranging from a few meters to sub-meter levels. Notably, GLONASS offers dual-frequency signals on its satellites, transmitting both the L1 and L2 frequencies. This feature enhances the accuracy of positioning calculations and improves resistance to signal interference and multipath effects, particularly in urban environments or areas with dense foliage.

4.3.3 Coverage

GLONASS provides global coverage, ensuring that at least four satellites are visible from any point on Earth at any given time. The distribution of satellites across multiple orbital planes enhances coverage and ensures robust positioning capabilities worldwide. Moreover, GLONASS satellites are strategically positioned to complement the coverage provided by GPS, offering redundancy and improved positioning accuracy, especially in regions where GPS signals may be obstructed or degraded.

4.4 Combining Positioning systems

Combining GPS, GLONASS, and GALILEO systems significantly enhances localization accuracy through a technique known as multi-constellation or multi-GNSS positioning. By utilizing signals from multiple satellite constellations simultaneously, users can mitigate the limitations and exploit the strengths of each system, resulting in improved accuracy, availability, and robustness. By tracking satellites from multiple systems (multi-constellation tracking), the receiver can compute position solutions based on a larger number of satellites, improving accuracy and reliability, especially in challenging environments with obstructed or

degraded signals. Moreover, the combination of multiple GNSS systems provides resilience and redundancy in case of signal blockage, outage or interference.

Moreover, as analyzed in D3.4, by employing Real Time kinematics we can significantly increase the accuracy of the GNSS positioning solutions to achieve centimeter-level accuracy. RTK sensors were not available during the data-capturing sessions. However, they will be available during the trial implementations.

5 Perception using aerial imaging data

5.1 Object Detection

During this implementation, we retrained the models described in Section 4.1.2 of D3.1, and in particular the state-of-the-art object detection framework You Only Look Once (YOLO), using data from the missions described in Section 3.

Moreover, we updated our architecture to the v8 version of the YOLO model, the overall architecture of which can be viewed in Figure 6. YOLO architectures are designed for real-time object detection where detection is performed in a single pass of the image through the neural network. This approach divides the image into a grid and predicts bounding boxes and class probabilities for each grid cell. The general progression in YOLO architectures focuses on increasing accuracy, speed, and efficiency. The network is comprised of the following building blocks:

1. **Backbone:** The backbone is a variation of the CSP (Cross Stage Partial) networks, optimized for feature extraction at multiple scales. YOLOv8 uses a CSP-based backbone that includes multiple CSP layers.
2. **Neck:** The neck of YOLO models usually features additional convolutional layers and pathways to enhance feature integration across different scales. YOLOv8 employs a PANet (Path Aggregation Network) for efficient multi-scale feature aggregation.
3. **Head:** The detection head applies convolutional layers to the feature maps to predict bounding boxes, objectness scores, and class probabilities. YOLOv8 utilizes anchor boxes predefined to match the aspect ratios of common objects in the training datasets.

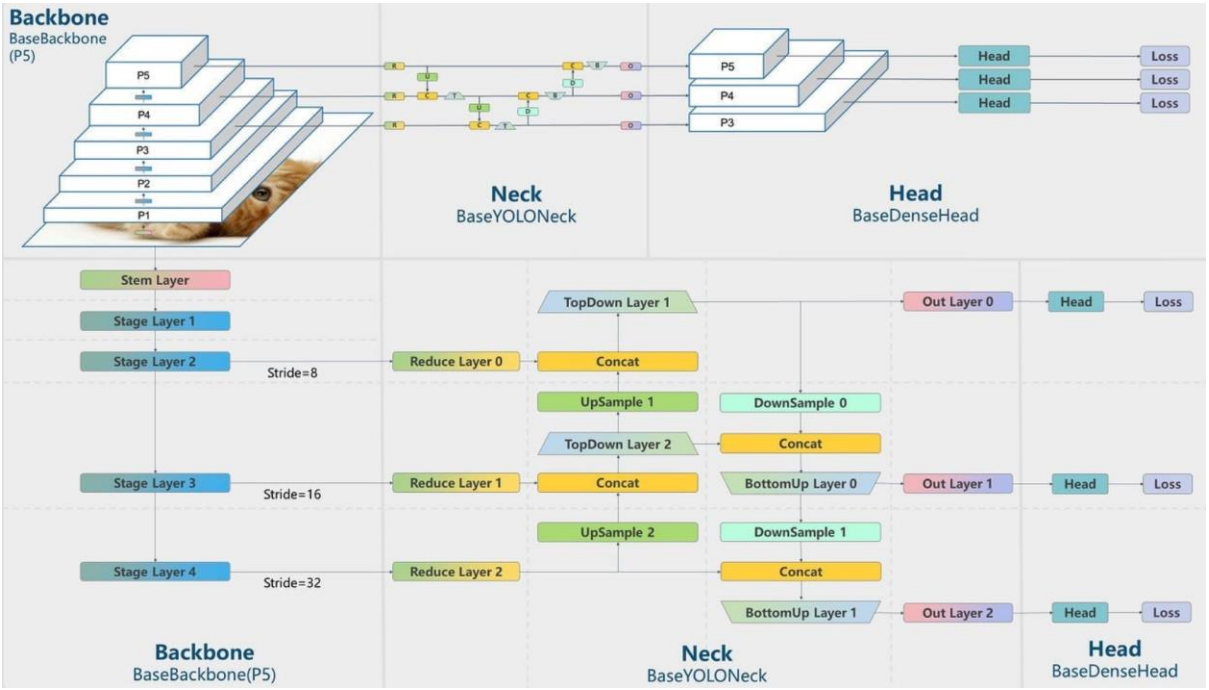
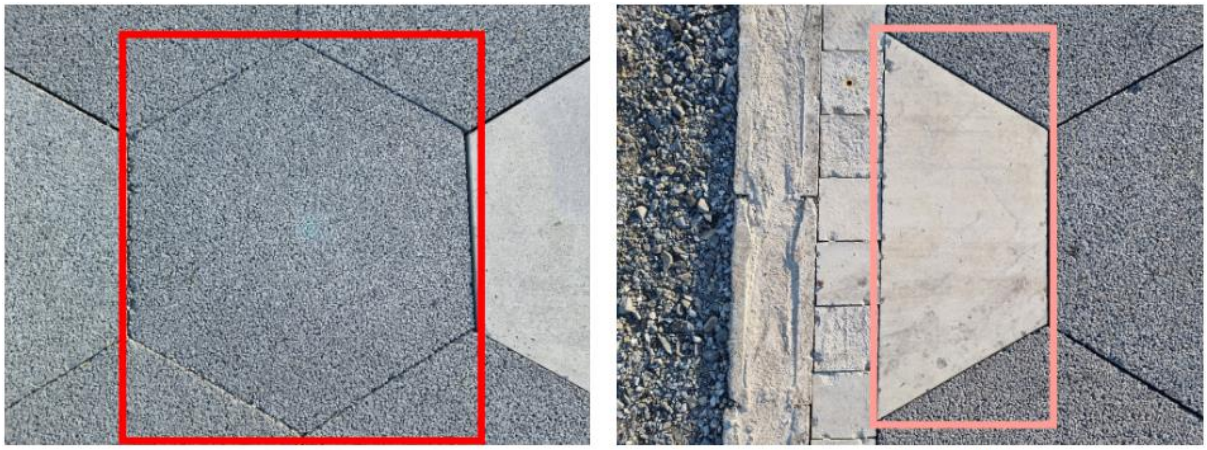


Figure 6: YOLO v8 network architecture.

Regarding the RGB images gathered during the first data-capturing session (see Section 3.1), we conducted an evaluation of the aforementioned YOLO model specifically for the pothole detection process. On the other hand, regarding the drone images gathered during the second data-capturing session (see Section 3.2) two classes were considered, based on the shape, material, and color of the slab, as well as its location on the RUP structure. In particular, (i) Slab_H represents the central slabs characterized by a hexagonal shape and dark gray color (see red bounding box in Figure 7a), while (ii) Slab_T refers to the slabs positioned on the sides of the RUP structure, featuring a trapezoidal shape and light gray color (see pink bounding box in Figure 7b).



(a) Hexagonal slab (Slab_H). (b) Trapezoidal slab (Slab_T).

Figure 7: The two different slab classes.

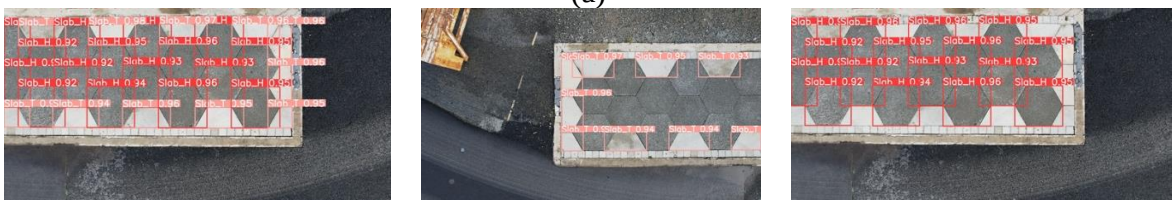
Our analysis included examining key metrics that have been already presented and analyzed in Section 4.5.4 of D3.1 and Section 3.7 of D3.3, such as precision, recall, and mean average precision (mAP), providing a comprehensive understanding of the model's performance and its potential real-world applicability. The model performance demonstrated in the respective test sets is presented in the following Table below.

Table 3: YOLO v8 performance evaluation.

Metric	Pothole	Hexagonal slabs	Trapezoidal slabs
Precision	94.5%	99.9%	99.7%
Recall	91.8%	100%	98.2%
mAP50	94.1%	99.5%	99.5%
mAP50-95	56.5%	98.0%	98.1%

Moreover, the model outputs from all the aforementioned data-capturing activities (see Section 3) are presented in the figure below.

(a)



(b)

Figure 8: Indicative outputs of the YOLOv8 models.

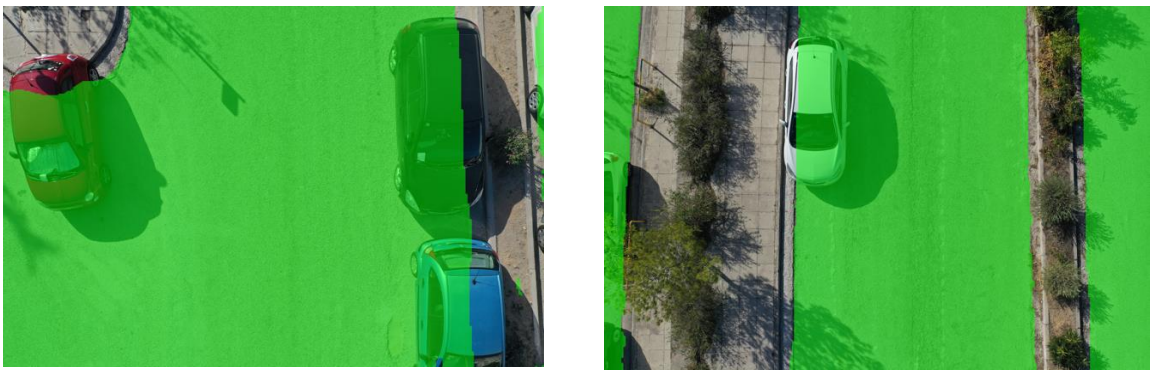
5.2 Image Semantic Segmentation

The segmentation of the captured issues has been one of the main topics of research for HERON, the results of which were presented in WP3 deliverables. However, the increased distance of imaging used for the UAVs creates significant problems in accurately calculating the segmentation mask of small road defects. To this end, we used the data gathered from the data-capturing sessions of section 3 to retrain our algorithms with additional training samples and elaborate on their performance. In this section, we present the results of these efforts.

It is worth mentioning here, that while the segmentation of the defect is critical for HERON's efforts, this information is critical only during the repairing phase, to accurately guide the robot. For UAV imaging, it is more critical for segmentation algorithms to capture more general area information such as segmenting the road surface, to ensure that the defects are correctly calculated.

5.2.1 U-Net Architecture

Even after retraining, the segmentation of small defects such as potholes or cracks did not achieve suitable performance. Thus, for the UAV detection, we will be using only the object detection methods for defect detection. However, the U-Net architecture presented in Figure 10 and analyzed in Section 4.2 of D3.2 will be used for segmenting the road surface as an additional check in the defect identification [4]. For road surface segmentation our approach achieved very nice results as presented in the figure below. The architecture of the selected segmentation model is presented in Figure 10.



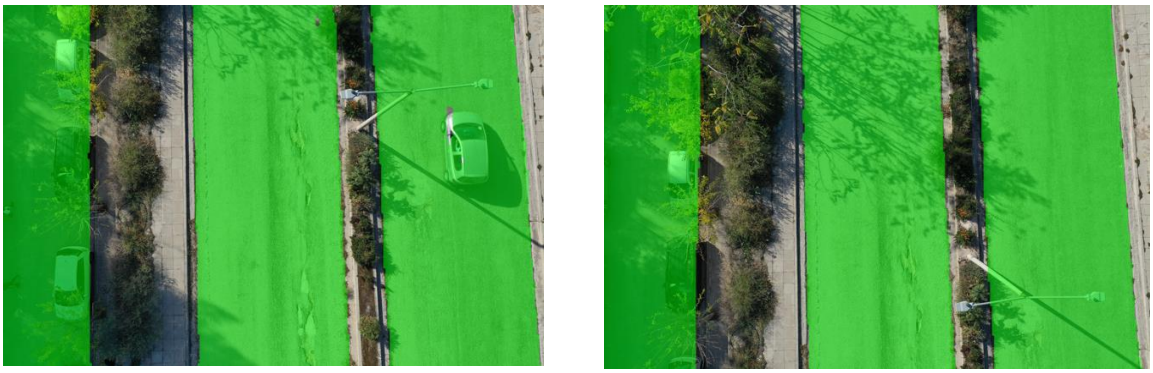


Figure 9: Outputs of the Deep Segmentation Model for road surface segmentation.

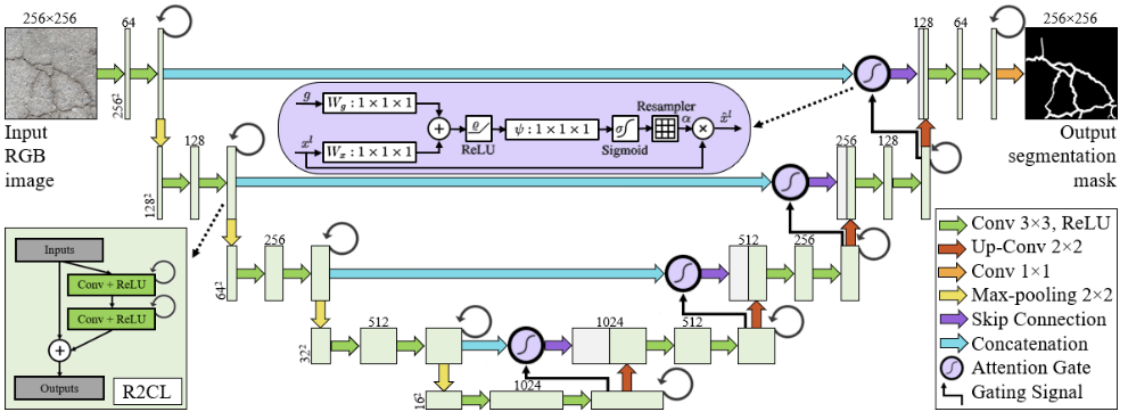


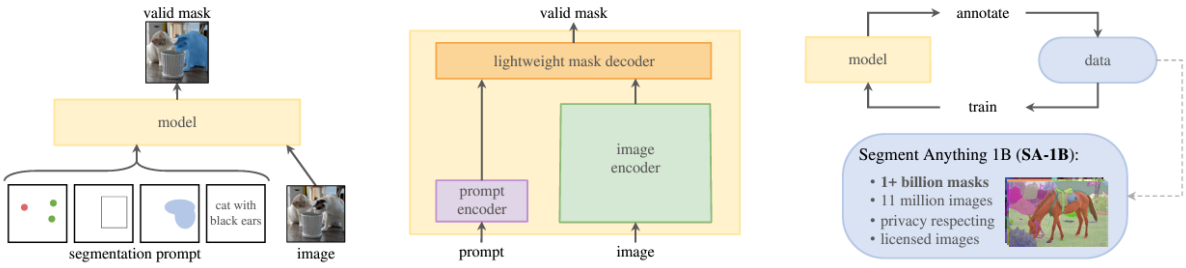
Figure 10: The selected deep U-Net architecture for image segmentation.

5.2.2 Segment Anything Model (SAM)

To increase the performance of the segmentation, we decided to also deploy the Segment Anything Model (SAM). SAM developed by Meta AI introduces a groundbreaking approach in the field of computer vision, particularly in the area of image segmentation. SAM is designed to address one of the most challenging problems in computer vision: segmenting a wide array of objects in images, even those for which it has not been explicitly trained. This capability marks a significant departure from traditional segmentation models, which typically require extensive training on labeled data for each specific object category they are expected to recognize and segment.

The core innovation behind SAM lies in its ability to generalize from known objects to unknown ones, leveraging a novel training methodology that combines supervised learning with self-supervised learning techniques. Supervised learning involves training the model on a dataset with labeled examples, where each image is annotated with precise segmentations of known objects. Self-supervised learning, on the other hand, allows the model to learn from unlabeled data, enabling it to infer structures and patterns without explicit annotations.

SAM's architecture is built upon advancements in deep learning and neural networks, incorporating elements such as convolutional neural networks (CNNs) for feature extraction and transformer models for understanding the relationships and context within images. This combination allows SAM to process and analyze images at a granular level, identifying and segmenting objects based on their shapes, textures, and contextual cues within the scene.



(a) **Task**: promptable segmentation (b) **Model**: Segment Anything Model (**SAM**) (c) **Data**: data engine (top) & dataset (bottom)

Figure 11: MetaAI's SAM architecture

One of the key benefits of SAM is its versatility across various domains and applications. In medical imaging, for example, SAM can assist in segmenting organs or anomalies in scans, aiding in diagnosis and treatment planning. In autonomous driving, it can help in identifying and segmenting road users and obstacles, contributing to safer navigation. Moreover, in environmental monitoring, SAM can segment different land uses and natural features in satellite images, supporting conservation and urban planning efforts.

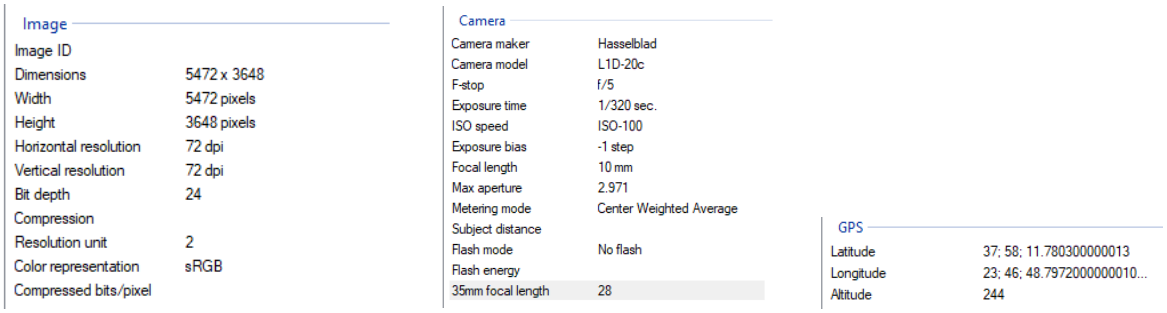
For our purposes, due to the fact that SAM has already been trained to receive YOLO style bounding boxes as additional inputs during the segmentation procedure, the deployment of the model allows the segmentation of defects even when captured from UAV's. Some example outputs of SAM are presented below.

Figure 12: Pothole segmentation from UAV data using YOLO v8 and SAM

6 Integration of Aerial information with the UGV

The HERON UGV should be able to handle the actual intervention procedures by itself, and the ingestion of the semantic information derived by the analysis of aerial footage from the computer vision and perception mechanisms should happen in a complimentary manner.

The UAV's output is timestamped and geolocalised images capturing areas of the road. The perception mechanisms of WP3 are used for analyzing these images and identifying defects using the semantic taxonomy described in section 4.6 of D3.1.



(a) (b) (c)

Figure 13: Details of the imaging data captured by the UAV (a) image resolution information, (b) camera information (c) localization information.

A critical component of this procedure is the translation of the issues identified in the camera coordinate system (pixel coordinate values) to a global coordinate system (GPS values).
As we can see in Figure 13(c), the GPS location of the image is available. These GPS locations are assigned to be the location of the middle pixel of the image. Translating pixel coordinates from an image taken by an Unmanned Aerial Vehicle (UAV) to GPS coordinates is a process that merges the dimensions of pixel space with real-world geography, taking into account the UAV's position, altitude, and the camera's intrinsic properties. To start, it's essential to grasp the metadata associated with the image, which includes the GPS location where the UAV captured the image, the UAV's altitude at the moment of capture, the camera's focal length, the sensor size, and the image's resolution.

The next step involves calculating the camera's field of view (FoV), which is crucial as it determines the visible area from a given distance. This calculation is based on the camera's focal length and the sensor's size. With the FoV known, the scale at which each pixel represents the real-world distance can be determined, a scale that varies with the UAV's altitude—the higher the UAV, the larger the area each pixel covers on the ground.

Mapping the pixel coordinates to GPS coordinates then involves identifying the pixel's position relative to the image's center and, by using the scale, converting this pixel offset into a real-world distance. With the UAV's GPS location as a reference, the GPS coordinates corresponding to the pixel can then be calculated. This process includes calculating the pixel's offsets from the center, converting these to real-world distances using the scale, adjusting for the UAV's orientation, and finally calculating the GPS coordinates through trigonometry and geographic coordinate system conversions. This transformation allows the exact measurement of spatial values on a pixel level. The final localisation accuracy has a spatial error identical to the error of the GPS sensor that the UAV is equipped with.

To support this process, the EXIF library was used for extracting metadata from images. The process's accuracy hinges on the precision of the UAV's GPS, the altitude's accuracy, and the camera's calibration.

7 Conclusions

In conclusion, this document has provided a comprehensive overview of the implementation process of the drone platform within the HERON framework. It has highlighted key aspects such as the hardware specifications, the localization framework, and the utilization of computer vision perception mechanisms based on aerial imaging data. Through these components, the drones can effectively gather relevant information from targeted scenes, aiding in preoperative planning and supporting ongoing operations. Additionally, the coordination and communication between the different HERON platforms, particularly the integration of aerial information with the UGV and the broader HERON system, have been detailed in D5.2.

Throughout this report, the implementation of the HERON drone system has been elucidated, covering its design and deployment strategies. The document has also outlined the procedures and methodologies employed during data-capturing sessions, presenting the inspection missions conducted within the HERON project. Furthermore, it has delved into the localization framework utilized by the system, discussing various techniques used to precisely determine the drone's position and orientation. Finally, the exploration of perception mechanisms using aerial imaging data has demonstrated how visual information captured by drones is processed and analyzed by state-of-the-art computer vision algorithms to detect, classify and segment various road infrastructure defects and PoIs.

References

[1] DJI Air 2S Specs. Available online: https://www.dji.com/gr/air-2s/specs (accessed on 15 April 2024).

[2] DJI MAVIC 2 Specs. Available online: https://www.dji.com/gr/mavic-2/info (accessed on 15 April 2024).

[3] Thierry Sedran, Eric Gennesseaux, Julien Waligora, Philippe Klein, Mai Lan Nguyen, Julien Cesbron, and Christophe Ropert. 2023. Development of a permeable removable urban pavement. In ISCR 2013-14th International Symposium on Concrete Roads.

[4] Katsamenis, I., Protopapadakis, E., Bakalos, N., Varvarigos, A., Doulamis, A., Doulamis, N., & Voulodimos, A. (2023, October). A Few-Shot Attention Recurrent Residual U-Net for Crack Segmentation. In International Symposium on Visual Computing (pp. 199-209). Cham: Springer Nature Switzerland.