Intelligent Laser Guided Swarm of Drones for Vessel Identification and Cargo Secure Retrieval

Munachiso Nwadike [1] **Steven Hoang**[1] **Khaled Dhawoud**[1] **Kirill Vishniakov**[1] **Zangir Iklassov**[1] **Dmitrii Medvedev**[1] **Yu Kang Wong**[1] **Karima Kadaoui**[1] **Mugariya Farooq**[1] **Sara Pieri**[1] **Asma Ahmed Hashmi**[1] **Roberto Guillen**[1] **Ruben Solozabal**[1] **Moayad Aloqaily**[1] **Mohsen Guizani**[1] **Martin Takáč**[1] **Wail Gueaieb**[1] **Ilir Capuni**[4] **Shahzad Khan**[5] **Rami Abielmona**[2,3]

[1]Mohamed bin Zayed University of Artificial Intelligence, [2]Larus Technologies Corporation, [3]University of Ottawa, [4]University of Montenegro, [5]Levadata

Team Description

A. Points of contact

- **Main:** Munachiso Nwadike munachisnwadike@gmail.com munachiso.nwadike@mbzuai.ac.ae
- **Alternative:** Steven Hoang steven.hoang22@gmail.com steven.hoang@mbzuai.ac.ae

B. UAE Based Supervisors

1. Dr. Moayad Aloqaily is the Managing Director of xAnalytics Inc., and a faculty member at MBZUAI. He is the symposium co-chair of IEEE GC22-IoT and Sensor Networks, and serves as the co-editor-in-chief of IEEE CommSoft TC eLetter since 2020. He is an Associate Editor of Ad Hoc Networks, Simulation Modelling Practice and Theory, Journal of Network and Systems Management, Cluster Computing, etc. He is a Senior IEEE Member, ACM Member, and a Professional Engineer Ontario (P.Eng.). He established the icNet Lab [link], where many projects on Autonomous UAV systems were developed.

2. Dr. Ruben Solozabal is a postdoctoral fellow at MBZUAI. He holds a PhD in applied Reinforcement Learning from the University of the Basque Country and he has previous experience in the industry as a R&D software developer. Some of the project he has worked for include the development of the industrial controllers and target integration on Simulink Coder, the development of the Software-in-the-loop designs as well as the integration of numerous protocols for industrial communication as EtherNet/IP and CANopen.

3. Prof. Mohsen Guizani is the Associate Provost for Faculty Affairs and Institutional Advancement at MBZUAI. He has received more than 30M US$ of research grants and holds over 10 US Patents. Dr. Guizani served as the Editor-in-Chief of IEEE Network and is currently serving on the Editorial Boards of IEEE Network and IEEE Internet of Things Journal. He served as chair of the IEEE Communications Society Wireless Technical Committee and the Chair of the TAOS Technical Committee. He is a Clarivate AI Highly Cited Researcher.

4. Prof. Martin Takáč is the Deputy Machine Learning Department Chair at MBZUAI. His research interests include the design and analysis of algorithms for machine learning, applications of ML in various scientific and engineering fields, optimization, and HPC. He serves as an Area Chair at several machine learning top-tier conferences including ICML, NeurIPS, ICLR, and AISTATS, in optimisation and reinforcement learning. Recently, he and his collaborators explored how Reinforcement Learning can be used by multiple robots (agents) for a classification task [1, 2, 3]. The agents (drones) have to learn how to achieve a given task and efficiently communicate and plan their movements to explore the space synergically. In this video [link] we show the done flying over the map to classify the the campus. The drone has a limited visual view as shown here [link] together with the classification prediction. A simulation in Gazebo for a network or drones for a MNIST classification can be seen here [link].

5. Prof. Wail Gueaieb is the founder and director of the Machine Intelligence, Robotics, and Mechatronics (MIRaM) Laboratory at University of Ottawa, Canada. Dr. Gueaieb holds more than 100 patents and articles in highly reputed venues for UAVs and Robotics. He serves in Editorial and Co-Chair roles in venues such as the ASME Journal of Dynamic

Systems, Measurement, and Control, International Journal of Robotics and Automation, IEEE Transactions on Instrumentation and Measurement, IEEE/ASME Transactions on Mechatronics and the IEEE Conference on Decision and Control.

C. International Supervisors

1. Dr. Rami Abielmona: is the Vice-President of R&D at Larus Technologies Corporation, responsible for research in sensor networking, multi-sensor data fusion and computationally intelligent computing architectures. He oversaw the development of Larus Total Insight [link], patented Decision Support System. He is an Adjunct Professor from the EECS at the University of Ottawa, Canada. He received the NSERC Industrial Research and Development Fellowship, IEEE MGA Achievement Award, and Ottawa Business Journal (OBJ) Top 40 Under 40 Award in 2011. He has a long experience in working on Drone systems and Simulations where distinguished papers were published on the same topic.

2. Dr. Ilir Çapuni: Doctor of Computer Science degree from Boston University. He has established Excellence Laboratories in Tirana from which several products and spin companies have emerged. Mr. Çapuni is one of the founders of the scientific conference "Balkan Communications Conference". He is a researcher in computer science and has co-founded the decentralized social network "FBRK.io".

3. Dr. Shahzad Khan: has a PhD from Cambridge University, UK. He is currently with VP Data Science at Levadata. He has a decade of experience in information retrieval (search technologies), machine learning, data mining, and analytics. 22 years of info-tech industry experience across the entire software development life-cycle. Experienced AI researcher and product manager with polished client-facing skills, who can lead an engineering team to deliver projects within the triple constraints (cost, time, and scope). He was an advisor on *DroneEntry*, a social portfolio management platform for drone operators [link].

D. Graduate Students

Areas of Expertise: *CV, ML, ECE, Mechanical & Control, Management and Safety*

1. Munachiso Nwadike is an MSc in Machine Learning at MBZUAI. He obtained his BSc in Computer Science at New York University. His work on SautiDB, a machine learning based accent translation projects was funded by USD$8000 grant from Deep Learning Indaba. His works have led to workshop publications in AAAI and NeurIPS. **[CV and ML]**

2. Kirill Vishniakov is an MSc student in Computer Vision at MBZUAI, doing research on self-supervised learning. He obtained BSc degree in Mathematics and Computer Science at St. Petersburg Polytechnic University. He worked as a data scientist at Soter Analytics developing AI solutions and computer vision systems. Kirill holds the rank of Kaggle Competition Expert. **[CV and ML]**

3. Roberto Guillen is a MSc in Machine Learning at MBZUAI, doing research in optimization methods. He previously worked at Google as software apprentice and a full time software engineer with JP Morgan. Roberto holds a bachelor degree in Computer Science from the Monterrey Institute of Technology and Higher Education. **[ML and Management and Safety]**

4. Dmitrii Medvedev is currently obtaining the MSc In Machine Learning in MBZUAI. Holds MSc in Petroleum engineering with 12 years of experience providing technical solutions for offshore and onshore drilling operations for ExxonMobil, Total, Gazpromneft, ADNOC, Lukoil, Rosneft, Eni, and NNPC. His expertise includes seamless and welded pipe manufacturing (steel making, rolling, heat treatment, finishing), corrosive environments, material selection for complicated wells (horizontal and extended-reach with high pressure / high temperature / CO2 / H2S) and well design. **[Mechanical and Control, Management and Safety]**

5. Karima Kadaoui is an MSc student in Machine Learning at MBZUAI. Her work on conversion of impaired human speech has been recognised in Wired Middle East magazine. Karima worked as software developer for CommonShare, New York. She obtained a BSc in Computer Science from Al Akhawayn University, Morocco. **[ML and Mechanical and Control]**

6. Zangir Iklassov is a PhD student in Machine Learning at MBZUAI. Zangir holds BSc and MSc degrees in applied Mathematics and Economics from Nazarbayev University. He was the senior team lead at ReLive Intelligence working as a computer vision engineer on object detection and tracking algorithms. His MSc research focuses on deep reinforcement learning. **[ML and Mechanical and Control]**

7. Steven Phong Hoang is a student at MBZUAI pursuing a Master's in Machine Learning. He currently holds a bachelors degree in Computational and Data Science from George Mason University. **[ML and Robotics]**

8. Yu Kang Wong is an MSc student in Machine Learning at MBZUAI and has a Bachelors in Computer Science(AI) from the University of Malaya. He worked as a full-stack developer for Ifast, Malaysia. His current research area encompasses sub-manufacturing and reinforcement learning. **[ML and Robotics]**

9. Mugariya Farooq is an MSc student in Machine Learning at MBZUAI and has a BSc in Electronics and Communication Engineering from National Institute of Technology. Her work in plant disease classification and Yield prediction in the hackathon organised by the Government of Abu Dhabi was crowned with the second place. **[ML and Management and Safety]**

10. Khaled Dawoud is an MSc student in computer vision at MBZUAI. Khaled holds BSc in electronics and telecommunication from King Abdulaziz University, Saudi Arabia. He has 4 years of experience working in leading GCC telecom companies such as Saudi Telecom, STC Solutions, and Al-Taknia for telecommunication. **[ML, ECE and Mechanical and Control]**

11. Sara Pieri is pursuing an MSc degree in computer vision at MBZUAI. Sara holds a degree in computer engineering from Sapienza Università di Roma. She collaborated on research in object detection and instance segmentation applications in outdoor environments at ALCOR Lab, Rome. Her current research thesis focuses on human-object interaction. **[CV and ML]**

12. Asma Ahmed Hashmi is an MSc student in Machine Learning at MBZUAI. Her research involves developing probabilistic graphical models to estimate the ground truth from noisy data to enhance image segmentation. She has a BSc in Mathematics from University of Massachusetts, Boston. She worked as a data analyst for 3 years in Persivia Inc. **[ML and Robotics]**

Technical and Societal Impact Considerations

UAVs are emerging as a promising technology that can offer a lot to the Abu Dhabi community in applications such as connected vehicles, surveillance and monitoring, border security, etc. There are many technical and societal impacts of the proposed system, including, civil safety, security of critical infrastructures, inspection and early intervention, and military application. The methods investigated in this paper are designed to be performed fully autonomously, which completely eliminates the risk of human harm to first responders, operators, etc. Moreover, the system helps to mitigate the illegal maritime activities such as piracy, illegal fishing, and espionage.

Technical Approach

1. Introduction

The ultimate goal of this white paper is to locate target vessels and transfer payloads of weights 1kg and 10kg onto the Unmanned Surface Vehicle (USV) from the vessel, with the aid of Unmanned aerial vehicles (UAV) and a robotic arm. To the best of our knowledge, there are no publicly available solutions that are capable of preforming surveillance, detection of illegal activities, intervention, and safe retrieval of important objects. In addition to the technical requirements of the white paper, and contents of the Frequently Asked Questions, we provide some technical assumptions to show that the team has investigated the aforementioned aspects. The target vessels are at least the size of the USV, such that they match the payload carrying capacity of the USV. We use this assumption in several places, including the specifications for the visual detection task. We also assume that the 10kg object volume is bigger than the 1.0kg object.

2. UAV/USV Swarm Management

2.1 Laser Guided Target Resolution

Specific security applications, such as laser guided missiles have sparked intensive research efforts into laser guided unmanned device technology [4]. We outline a procedure for identification of target vessels at long range using laser guided UAV technology.

Definition 2.1 (Object of Interest)**.** This includes vessels, payloads, target vessel platform boundaries, center, or any form needed for carrying, pushing, or placing operations.

Definition 2.2 (Observation)**.** A task whereby UAV's attempt to locate or determine the properties of an object of interest. May be executed with the aid of sensors and cameras.

Definition 2.3 (Working Task)**.** This includes carrying pushing, or placing an object of interest.

Definition 2.4 (Workers)**.** Worker UAV's $\{W_1, \cdots, W_n\}$ will be responsible for executing the working task. In this work, we take $n \geq 2 + 3$, (see section 2.2 on introduction to UAVs). n is a hyperparameter adjustable to the number of payloads.

Definition 2.5 (Global Observers)**.** B_d is the global observer, whose task is to locate all the vessels in a target-agnostic manner. It just search vessels, not for markers, and communicates to the other observers and workers where to go using laser.

Definition 2.6 (Observers)**.** $\{O_1, \cdots, O_m\}$ are close-range observers which coordinated to examine objects of interest. Since we set aside the worker UAV's and the global observer, we have $m = 20 - n - 1$.

The first contribution of this paper is to devise an algorithm to use lasers mounted on the global observer to guide the other UAV's towards target vessels. The key question in this algorithm relates to how we will beam the laser and how the other UAVs target the identified vessel. To achieve this, we investigated two methods: beam-riding guidance [5, 6], and semi-active laser homing (SALH) [7, 8]. The second contribution of the paper, is the use of secondary safety backups to avoid any single point of failure to the system. This method involves a mechanism called chain of command where a secondary observer drone can take the role of the global observer.

Note that for the laser beam emanating from global observer B_d to be used to guide the worker and observer UAV's, they must be in the field of view of the laser. This condition is achieved by virtue of the proposed Algorithm 1, since all other UAV's follow B_d closely in a tight formation from the start gate.

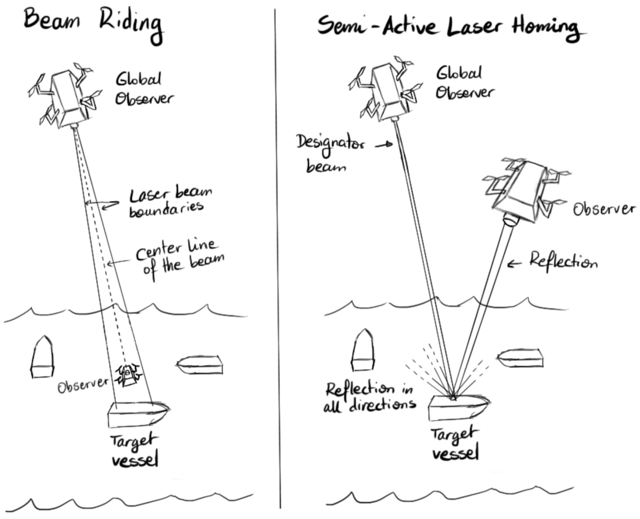

Figure 1. Laser guidance techniques proposed for the identification of target vessels. In Beam Riding, a global observer casts a laser beam on the target. An observer then uses a laser seeker to align its trajectory with the center line of the beam. In Semi-Active Laser Homing, an observer uses the reflection of a global observer's laser beam on the target vessel.

Algorithm 1 Algorithm on Target Resolution via Laser Guidance

```
Require: $B_d$ positioned at the start gate (middle of one region edge).
   $\{O_1, \cdots, O_m\}$ and $\{W_1, \cdots, W_n\}$ positioned in starting formation around
   $B_d$, and capable of maintaining this formation after takeoff.
Ensure: Dimensions of search region, depth $D$ and width $W$ in Figure 3,
   can be input to $B_d$ at the start gate.
   → $B_d$ moves forward $D/2$ units into the depth of the boundary region.
   → $B_d$ rises to altitude H
   → $B_d$ detects the set of all vessels $S_n = \{V_i\}$, and begins tracking them,
   pointing lasers at them as they move.
   for $V_i \in S$ do
      → $B_d$ sends a single observer $O_i$ to decide if $V_i$ is target vessel
      if $V_i$ is a target vessel then
         → $O_i$ finds all payloads in a $V_i$
         → video streaming from $O_i$ to operator is initiated
         → Operator selects the payload
         → $O_i$ builds a 3D model of the selected payload
         → $O_i$ sends the 3D model to the $B_d$
         → $B_d$ identifies the correct UAV gripper type from 3D model
         if all payloads are found or command to return then
            break
         end if
      end if
   end for
   → volumes of payloads are compared on $B_d$ to detect which is 1kg
   → $B_d$ sends the correct type of worker UAV to each payload
   → The payloads are placed on the USV
```

2.2 Swarm of Drones: An Overview

Nature and biological science frequently provide inventive inspiration. Research directions in Artificial Neural Networks (ANNs) have often been inspired by the brain, right from the early work of Dr. Frank Rosenblatt on the perceptron [9]. Aeronautical engineering, since the time of Leonardo da Vinci, has also taken inspiration from the mechanisms of bird flight [10, 11]. It is only natural that our proposed path for coordination between UAV's would take directly after a phenomenon we observe in nature. We propose that since there are multiple types of tasks mandated by our problem statement, it is also only natural that we have multiple kinds of UAV's. Each type of UAV will have different adaptations to its kind of task, similarly to the distinct adaptations of and roles played by bees in a colony.

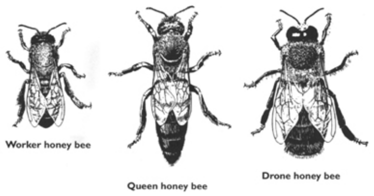

Figure 2. Different members of a bee colony have modifications to suit their swarm role swarm. Source: NASA Goddard Spaceflight Center.

We propose 2 types of UAV's:

- **Type A (Worker):** the "worker", or "carrier" UAV's are responsible for object lifting. To accommodate heavy lifting capacity, they will be large and heavy. They will also be equipped with robotic grippers. We can cover the 1.0kg object with 3 types of UAV grippers. We require 2 additional worker UAV's when dragging the 10.0kg object with a net or a rope. They will also have RGB-D cameras.
- **Type B (Observer):** the observer UAV's are specialised for localising payloads, and targeted vessels. They are equipped with an array of powerful visual cameras and sensors to manage this task. All UAVs not engaged as workers, will serve as observers.

2.2.1 Advantages

1. Specialising UAV's to specific roles reduces the burden on UAV's to quickly switch between inspection and intervention mode. Instead of having a UAV stops inspection in order to perform intervention, while trying to remember the locations of where it last saw the moving target vessels, we can make sure it only needs to worry about one kind of task at a time. Otherwise, we would need to introduce sophisticated path planning algorithms to decide on how to move through the search region.
2. If we focus on UAV's on both inspection and intervention, we will require to build them specially for the tasks to be performed. For example, the observer UAV's can afford to be smaller, since they are thought carry less load, and this would lead to a longer battery life, allowing them to hover in the air for longer periods.
3. If we do not have a large collective camera footprint from the various UAV cameras, the leader UAV or USV can partition the rectangular search region into parallel sub-rectangle regions, and lead the observer UAV's through 1 sub-region at a time, as shown in Figure 4.

2.3 Technologies for Maintaining Formation

The task of UAV localisation is non-trivial. Simultaneous Localisation and Mapping (SLAM) is known to be infeasible for large areas, since landmarks cannot be uniquely identified [12, 13] . This problem is exacerbated in marine context, where we have little to no meaningful landmarks. We consider Dead Reckoning as a main localization approach to be used in conjunction with laser guidance. This approach uses an initial reference combined with the estimation of the trajectory computed using estimates of angular velocity measurements from the gyroscope, and acceleration vectors from the accelero-meters. The challenge with dead reckoning is that it suffers from drifts off the projected path, due to accumulation of errors and noise in sensor measurements. Many works have attempted to solve this problem, for instance, one research project investigated the use of periodic motion instead of straight line trajectory in order to allow peak-to-peak distance estimation with Weinberg's approach [14]. Another work makes use of a Multi-Target Gaussian Conditional Random Field (MT-GCRF), to correct navigational errors, as well as for avoiding UAV collisions [15]. This paper leverages the state-of-the-art techniques for dead reckoning error correction.

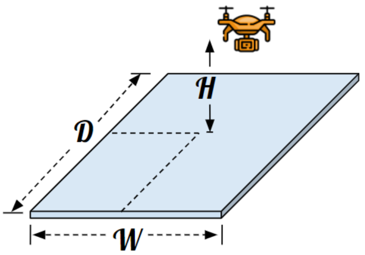

Figure 3. For a fixed UAV altitude H, a UAV would have a fixed camera footprint, or field of view region. H may be adjusted to suit the camera viewing resolution. The camera footprint may not cover a desired region, and so the UAV may need adjust H. The area D×W = 10km^2.

2.4 Communication Mechanisms

Since most of the commercial UAVs do not include built-in features to communicate with other UAVs. We propose a communication system based on a mounted micro-controller that is connected to the UAV ports and has nRF24 chip. nRF24 is a single 2.4GHz transceiver designed for wireless applications that need a few external components to construct. 2.4GHz is a common frequency band for WiFi in UAE [16]. The data rate supported by the nRF24L01 is 2Mbps and it has a range of 1.1 Km approximately. Depending on which subset of UAV's we need to communicate with, we can alternate between the unicast and/or multicast topologies. This will be vital to coordinating the swarm. For example, a broadcast signal from a leading UAV A Type can be used to tell other observers UAV's Type B to move to a new search region within the boundary. For example, a multicast signal sent by a leading UAV to specific worker UAV's to tell them to start the intervention (i.e a unicast signal from 1 observer UAV to 1 worker UAV would apply for the 1.0kg object, while a multicast signal to multiple worker UAV's would apply in case of the 10.0kg).

2.5 Return-To-Home Functionality

The requirements request a way for the UAV's to come back to the starting point on demand. As our baseline, one of our UAVs will be mounted with a beacon which serves as a means to signal to other UAV's. This UAV may remain at the start gate, while the other UAVs and the USV will continue their task. Alternatively, we can use a robotic arm mounted to the USV to detach a beacon and leave it at the start gate in the beginning. If we assume a rectangular search region, then the UAV's can simply move directly from their original location in a straight line toward the return to home beacon. Indeed, this approach can work for any convex shaped search region. For example, in the Figure 4, the convex nature of a rectangular boundary region is such that any UAV at point **x** can return in a straight line to point **y** without exiting the boundary region.

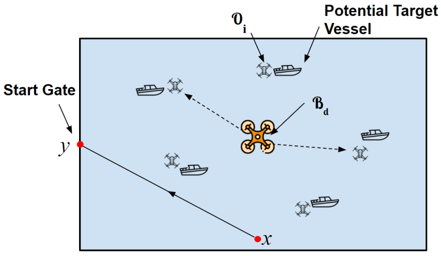

Figure 4. We may leave beacon at the start gate to serve as guide for the return-to-home functionality. Laser guidance by global observer conducts the close-range observers.

3. Computer Vision

Computer Vision is a necessary component for any automated system including UAV's and other types of UAVs. The ability to accurately identify target objects in varied weather, lighting conditions, viewpoints and scale is crucial for tasks which involve physical interaction. We will have different types of computer vision systems for close-range observers, global observer and workers.

3.1 Computer Vision for Global Observer

The task for global observer B_d is to locate all the vessels inside the boundary region without specifying whether it's target or not. After the localization of vessels it will point out the infrared laser to the detected vessels and use object tracking [17] to update the position of the laser as the vessels move. Requirements for B_d are:

- Very high resolution camera, either 4K or 8K.
- Heavy-weight deep learning object detection and tracking model with a very high accuracy. Probably a two-stage detector based on Faster-RCNN [18].
- Laser that will be pointed to the detected vessels.

3.2 Computer Vision for Close-range Observers

Small observers will perform an observation task in a close proximity, these include:

- Target vessels identification using several stage approach with 2D and 3D object detection models.
- 3D model generation of the payloads.
- Object pickup and placement related tasks: find the center of an object, decide on the gripper type, ensure dragging of the object is correct.

Requirements for close-range observers:

- Lidar and RGB-D camera which will be used in generation of the 3D models of a vessel and payloads, as well as capturing images for marker identification.
- Lightweight 2D object detection model based on EfficientDet[19] or Yolo [20] to detect target

vessels from 2D RGB image.

- 3D object detection model or template matching algorithm to compare the 3D models of target vessels, if needed, to the reference model.
- 3D object detection model for classification of payloads to decide the optimal gripper type.
- Edge detection algorithm [21] to ensure that the worker UAV's do not drag the 10kg object overboard.

3.2.1 Camera positioning

In [22] authors provided an interesting insight that detection accuracy depends on the UAV camera angle and flying height. For example, they showed that two CenterNet-based models with ResNet-101 and ResNet-18 encoders, which hugely differ in terms of performance speed, have comparable detection accuracy when averaged over angle domain. If averaged across model domain, the best detection results are produced when camera was in Medium-Right angle of 73-90°. We think that camera angle in range of 70-90° will allow to capture more surface of the object and, thus, more distinctive features, compared to a top-down approach in which only the top part of the object's surface will be captured. Therefore, we will try to maintain a camera angle close to a 70-90° degree range. Flying height should not be too high, as the quality of the detection will suffer because markers and payloads can be of a very small size. We plan to fly UAV's on the altitude range of 2-50 meters, if needed the UAV will fly closer to the object of interest.

3.2.2 Detection of Target Vessels and Objects

Deep Learning Model: There is a trade off between accuracy and speed of detection. It is mentioned that the target vessels will have some distinctive features (markers). It has been also stated that the markers will be provided in the form of blurry images to mimic the environment of the person making the photo from their phone. Therefore, we need to be able to do the fine-tuning of the model with the very limited low quality data, thus we need to pretrain our DL model on a dataset relevant to maritime applications such as [22]. Moreover, as the finetuning will be performed on a small amount of data we might require a 2-stage detector based on Faster-RCNN [18] in order to get a good accuracy. However, 2-stage detectors cannot do inference in real-time on edge devices. Therefore, we may use optimization techniques such as knowledge distillation [23, 24], pruning [25, 26], and quantization [27, 28]. Moreover, it is possible to accelerate the inference with hardware optimizations [29]. We plan to train our model with different color-based data augmentations [30] to mitigate the problem of sunlight reflection from water, extreme brightness and inconsistent water color. Additionally, we can use thermal cameras to enhance the object detection accuracy. In our experiments we will try one-stage models: CenterNet [31], EfficientDet [19] and YOLO[20]; as well as two-stage models based on Faster R-CNN. We will pretrain these models on the SeeUAV'sea dataset and fine-tune using the markers which will be provided later by the organizers. If needed we will label additional data ourselves and apply optimization techniques listed above to speed up the model performance.

Confidence Tresholding for target vessel identification: To mitigate the risks of false positives and false negatives in target vessel detection we will follow the protocol with several confidence levels depicted in Figure 5:

1. Predict the confidence C of a vessel being a target using only information from 2D RGB camera.
2. There will be three confidence levels, based on two thresholds t_{small} and t_{large}:
   1. If $C < t_{\text{small}}$, then reject the target vessel from being a target.
   2. If $C > t_{\text{large}}$, then accept the target vessel as a being target.
   3. $t_{\text{small}} \leq C \leq t_{\text{large}}$, then additionally generate the 3D model of a candidate vessel and match it against the provided reference.

3.2.3 3D model generation

Two different options are considered for 3D modeling:

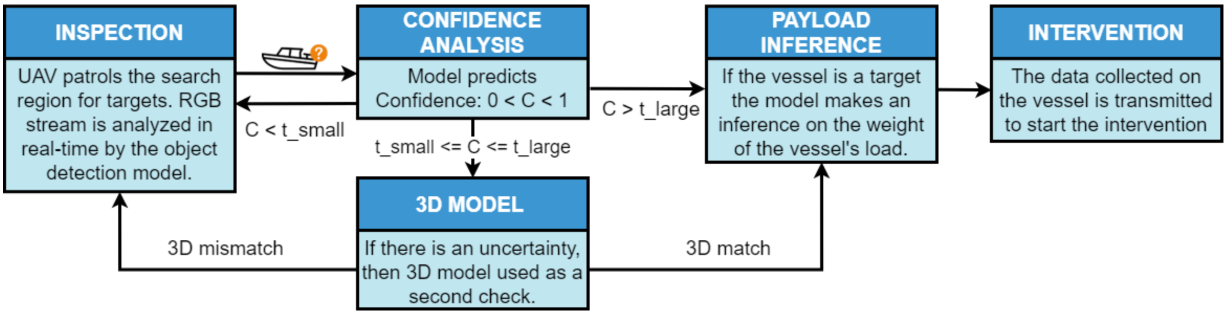
INSPECTION
UAV patrols the search region for targets. RGB stream is analyzed in real-time by the object detection model.
CONFIDENCE ANALYSIS
Model predicts Confidence: 0 < C < 1
C < t_small
t_small <= C <= t_large
C > t_large
PAYLOAD INFERENCE
If the vessel is a target the model makes an inference on the weight of the vessel's load.
INTERVENTION
The data collected on the vessel is transmitted to start the intervention
3D MODEL
If there is an uncertainty, then 3D model used as a second check.
3D mismatch
3D match

Figure 5. Process of target vessel detection and payload inference by close-range observers. To ensure the redundancy of our system and avoid failure cases, we propose a 2-tier target vessel identification approach based on confidence thresholding for detected vessels. RGB cameras make an initial estimate of target probability, while a more computationally intensive 3D model is used when the confidence is lower that a threshold. Only when the UAV/USV swarm confirm a match with the target vessel the process of assessing and intervening upon a payload begins. The confidence thresholds t_{small} and t_{large} would be hardcoded into the system as a hyperparameter during model pretraining.

1. **Lidar**: We can use Lidar to generate a 3D model of the target vessel which will be matched against a reference using confidence intervals as discussed in section 3.1.4. Furthermore, 3D models for the 1kg and 10kg objects will be created in the form of point clouds. First, the Lidar will produce a point cloud and then a 3D object detection algorithm will segment the object from the point cloud [32, 33, 34, 35, 36]. This model will be to be used to calibrate the arms that will be used grasp the object.

   The error of Lidar measurements is proportional to the distance from the target object. For instance, certain models of Velodyne Lidar, have a measurement range of 100m up to an accuracy of ±3cm, for a horizontal field of view is 360° [37]. As a result, once the operator localises a target, an observer UAV will increase its proximity to the target to ensure we are within the appropriate Lidar range. The UAV will scan the target such an accurate dimensions of the object can be obtained.
2. **3D Camera (RGB-D)**: We use depth cameras to produce a 4-dimensional tensor of certain resolution consisting of R, G, and B color components and D distance information for every pixel. It can work in real-time producing up to 90 fps. Accuracy depends on the distance showing an average result of 0.5mm. There are 2 main technologies used in depth cameras: structured light and stereo depth. We will rely on the latter, since it mimics vision of human-beings.

Whereas Lidar typically provides a more accurate depth information in the form of point cloud, it doesn't recognize color. Moreover, 3D cameras are typically much cheaper than Lidar, which also requires post-processing and additional computational resources to work with high-dimensional data of point cloud. In contrast, RGB-D camera yields a simpler representation of a 3D image. Another alternative is estimating the Structure-from-Motion.

3.3 Computer Vision for Workers and USV

Workers will be responsible for object pick up and placement procedure, which is based on [38]. We discuss the details of object grasping from a perspective of deep reinforcement learning further in section 4.2. However, we note that worker UAV's and USV will be equipped with RGB-D cameras as in [38] in order to build perform depth measurements needed for grasping heatmaps. Deep learning models would be deployed to identify the gripper type [39].

4. Robotic Intervention

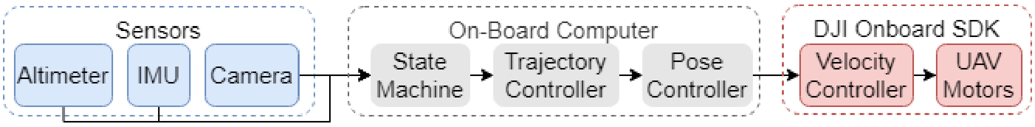

Figure 6. The overall control system consists on a multi-level schema. On a higher level, a state-machine handles the different phases on the challenge. The state machine will run on the on-board computer on the UAV and will set control points to the high level controllers as the trajectory or pose controllers. These modules will rely on the DJI SKD low-level controllers as the integrated velocity controller to maneuver the UAV.

4.1 UAV Specifications

The DJI's commercial UAV platform will be the basis of our solution. The solid integration this manufacturer offers between the hardware and software makes DJI's aerial platform the best choice for the present challenge. Additionally, we plan to increase the UAV's control capabilities complementing it with an external edge device running ROS.

4.1.1 Robotic Operating System

The core component of robot software is the Robotic Operating System (ROS) [40] that serves to integrate all the robotic components and algorithms into one interoperable machine. ROS is modular and there exists a wide variety of drivers as well as state-of-the-art algorithms that allow to control a wide variety of ROS compatible components.

Particularly, we envision that our solution will rely on a muti-level control approach with tight integration in ROS. The overall solution is depicted in Fig. 6. A high level state-machine will be in charge of controlling the different phases in the challenge. While for the low-level control tasks, as the velocity control, we plan to rely on the control the DJI Software Development Kit (SDK) [41] offers. We also envision the coexistance of several additional controllers as the trajectory or pose controllers that has to be specifically designed for the challenge.

The overall control system work as follows. The state-machine would be the central unit on the UAV's operational system. The state-machine receives commands from other UAV's as well from the internal task-specific sensors and outputs high level operational signals. These signals are to be processed by specific controllers as the trajectory or pose controllers. These components have direct communication with the DJI SDK and will actuate on a close control loop over it to control the monouvers of the UAV.

In addition, due to the complexity the solution requires (control of multiple systems on complex operations as grasping) we also explore the possibility of developing our own adaptive controller for tighter interaction between the robotic arm and the UAV. Notice that the solution requires the coordination of control modules of different nature, e.g. deep neural networks with traditional numerical control. Therefore, we forecast that a thigh integration between the different control modules will be key to the success of the challenge.

4.1.2 Simulation Package - Gazebo

Simulation is a key component on the process of designing and testing control algorithms as well as training Machine Learning models without the safety risk and time consume that these processes involve in the real-world. Particularly, the simulation toolbox we plan to use is Gazebo [42]. Gazebo includes a robust physics engine, a high quality graphical interface and a convenient programmatic language that enables compatibility with a large quantity of the current robotic solutions. Gazebo will be used in conjunction with ROS to perform Software-in-the-loop simulations.

4.1.3 Computational Hardware

The UAV will be modified to carry an external computer such as an Intel NUC or Jetson Xavier NX [43]. A USB accelerator containing a TPU could be included to perform CV tasks. These options not only are low power but also have a small weight footprint. Additionally, the

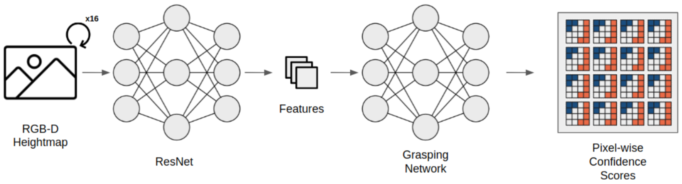

Figure 7. A neural network that computes features and 3D positions from RGB-D orientations. Features are used for predicting probabilty of grasping success at each pixel location with grasping position and angle.

computing units will be shielded to avoid disruptions on the on-board sensors.

4.2 Robotic Arm Manipulation

Traditionally, manipulation tasks utilize physics-based methods that require optimizing for control parameters and approximating physical dynamics. Given that the target object will not be stationary and subject to changing dynamics, these methods require knowledge of properties that are challenging to estimate [44], potentially leading to poor generalization. To address this, we propose a more resilient method using a combination of reinforcement learning and inverse kinematics. We leverage Zeng et. al's [38] prior work to train the arm to grab each target object. The gripping mechanism will consist of a neural network $f(I, p)$, where I represents the visual observation of the area containing the target object and p represents the position of the area the arm is to places the object. The goal of the neural network is to predict motion key parameters such that results in the robot grasping the target object and placing it in the target location. In summary, the proposed gripping mechanism consists of 2 components: 1) a convolutional residual network [45] that outputs the spatial feature representation and 2) a grasping mechanism that predicts the parameters.

The visual input I is captured as a RGB-D image and projected onto a 3-dimensional point cloud. We do this to account for the tight integrations required to measure the grasping distances in a dynamic environment. The point cloud is then converted into a heightmap with color (RGB) and height-from-bottom channels (D), with each pixel corresponding to a unique 3D location in the surface of the target vessel where the target object rests. I is then inputted into the CNN, for which we use ResNet, which outputs a spatial feature representation to be inputted into the grasping module.The overall architecture can be viewed in Figure 7.

4.2.1 Grasping Network

The visual feature representation is fed into a ResNet, which outputs a probability map with the same dimensions as I, which each pixel representing the probability of grasping success. Different grasping angles are learned by rotating the input heightmap by different orientations before passing the it into ResNet [46]. The key motion parameters are determined by the pixel with the highest probability. Once the object is grasped, the robot initiates a reverse sequence of actions to place the object at position p.

4.2.2 Grasping Reflex

The grasping reflex is an open-loop, using a collision-free inverse kinematics solver for motion planning [47], with no prior assumptions being made on the shape of the target object. The grasping reflex carries out a top-down grasping action centered at the 3D location. When the action is initiated, the arm approaches the target object with an open end-effector (if the chosen end-effector is a suction grip, then there is no starting position as the form of the end-effector stays constant) until the 3-dimensional center of the the end-effector meets the position. The end-effector then activates its gripping mechanism and lifts upwards.

4.3 USV and UAV Robotic Arm Gripper

This gripper would essentially serve as the "palm" for the arm, to be used in grasping the 10kg object securely. Since we do not have prior knowledge of object shape, we provide 2 alternatives for the UAV to grasp the 1kg object from the target vessels and transport it to the USV:

1. Quick Change Module: A quick change module allows us to swap a gripper easily between 3 different types of grippers in the gripper station which will be on a platform of the

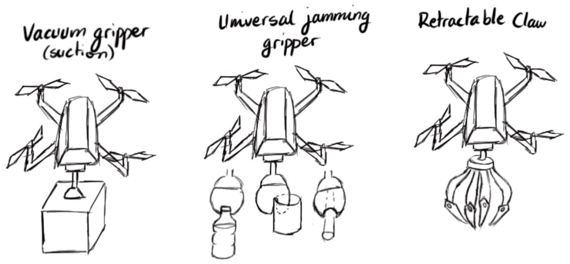

Figure 8. In contrast to settling on a single gripper type for the UAV, we will 3 different kinds of UAVs equipped with different kinds of grippers. RGB-D camera installed slightly above the gripper to help with grasping the object.

USV, and stabilised against wave influenced motion similarly to the robotic arm (see section 4.5.1). A sample video of this technology can be at this reference [48]. The 3 gripper types are: vacuum gripper, retractable claw gripper (three-fingers) and universal jamming gripper [49]. Each of the grippers, as we see in Table 1, can hold onto various shapes and types of objects. The shape and the type of the object, which dictates which grippers to use, will be identified by our computer vision component by the shape of the object.

2. **Combining Gripper**: Using a combination of a vacuum component and fingers, we yield reliable gripping consistency. The respective vacuum and finger components of the gripper will complement each other.

Types of Grippers	Shape of Object
Vacuum	box, object with flat surface, glass
Retractable Claw(three-fingers)	medium size, irregular shape, long, cylindrical objects
Universal Jamming	small size, irregular shape

Table 1. The types of objects that can be covered by each gripper.

4.4 Dragging Mechanisms using UAVs

In order to drag the big object a UAV will act as an observer and 2 UAV's as workers that hold onto each side with a net. First, the observer will determine the direction where the object should be moved to. Then, the workers will calculate the centre of the object from the shape and push the object slowly at that level. The observer will continuously measure the distance between the object and the edge of the vessels so that it won't fall into the sea. When the object is near the edge of the vessels, the observer will send signals to the worker to stop pushing and let the USV robotic arm pick up the object. We provide a visualisation of this technology in Figure 10.

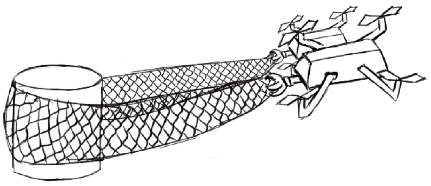

Figure 9. UAVs moving the 10kg object. Worker UAVs will determine the centre of the object, hold to each side with a rope, and slowly push to the vessel edge.

4.5 Stabilization of Wave-Influenced Motion

4.5.1 Robotic Arm Stabilisation

As the USV can have a rocking motion due to the waves, a stabilization system is needed for the robotic arm. One way of achieving that is by using a gyroscope based stabilisation platform. Specifically, the platform will utilise a 9-axis Inertial Measurement Unit (IMU) [50] that integrates a 3-axis accelerometer, a 3-axis gyroscope and a 3-axis magnetometer.

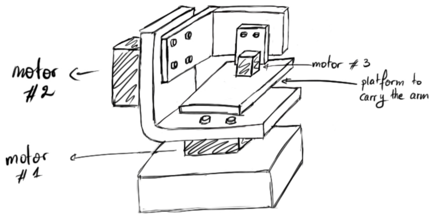

Figure 10. Gyroscope mechanism responsible for stabilisation. 3 servo-motors would correct the tilt on each of the 3 coordinate axes.

Similarly to a camera gimbal stabiliser, it would serve as a base upon which the robotic arm would sit, reactively counterbalancing the wave motion.

4.5.2 USV Positioning

When a target has been identified, the USV shall approach the location of the target vessel. One-to-two UAVs will play the role of a beacon by hovering in the air above the optimal point of contact with the board of the target vessel. During the loading operations, the USV shall be the subject to waves motion and wind exposure, as well as to the relative movements of the target vessel.

Leveraging our earlier assumption about volume of targeted vessels, we assume that the target vessel will be at least as large as the USV. The USV will therefore be able to exert a gentle pushing against a target vessel without damaging it. The thrust of USV engines can be directed against the body of the target vessel in order to maintain an inertial state. While the targeted vessel is in motion, appropriate adjustment of the longitudinal component to be considered.

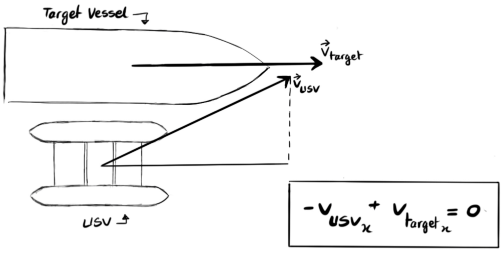

Figure 11. To counterbalance the effect of wave influence motion on the USV, we will contact the target using bumper such that it exerts a diagonal force vector on the target vessel. We ensure the horizontal components of this force vector, V_{usv} equals to the target vessel V_{target}. This way, it maintains an overall inertial state (10kg object intervention).

5. Risks and Mitigation

The team are aware of the different types of risks that might encounter the proposed solution. They also have the expertise to design an appropriate backup solutions in case of emergency situations. For example, the system is designed in away to account for fault-tolerance techniques. Moreover, the proposed system relies on duplicates of drones to avoid single point of failure. Below, we briefly discuss risks and the mitigation with respect to some individual components.

5.1 UAV Communication and Path Management

1. The laser will be infrared, so it is not visible to human on the target vessel.
2. The drones will orient themselves in a fixed formation around the global observer which will avoid the drones from crashing into each other. They will have proximity sensors to allow them to know when closely approaching other drones or objects.

5.2 Computer Vision

1. In case if the vessels/payloads are small in size, the altitude and speed of the UAVs will be adjusted accordingly to the size of the vessels. In addition, we will also generate a 3D model of the object to be picked up.
2. To tackle the possible failure of detection model due to the bad image quality, we will be training the DL models with data augmentations [30, 51, 52]. We will also make inference more robust by using the temporal property of video data [53].
3. Our multi-step approach with confidence levels allows us to reduce the rate of false positives and false negatives. False positive rate can be controlled by adjusting the t_{large}, whilst false negative rate can be controlled by adjusting t_{small}.

5.3 Robotics

1. The issue of connecting properly to the grippers due to wave motion will be handled by the stabilizer system mentioned in Section 3.5. The robotics arm or grippers will be sending frequent signals to notify when grasp onto objects.

References

[1] Guangyi L, Arash A, Martin T, and Nader M. Classification-aware path planning of network of robots. In *International Symposium Distributed Autonomous Robotic Systems*. Springer, 2021.

[2] Hossein M, Mohammadreza N, Martin T, and Nader M. Multi-agent image classification via reinforcement learning. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5020–5027. IEEE, 2019.

[3] Guangyi L, Arash A, Martin T, Héctor M, and Nader M. Distributed map classification using local observations. *arXiv preprint arXiv:2012.10480*, 2020.

[4] Vadim S, Vaclav K, and Alexandr S. Optical detection methods for laser guided unmanned devices. *Journal of Communications and Networks*, 20(5):464–472, 2018.

[5] Qi G, Yu J, Ai X, and Jiang J. Multi-missile coordination high precision guidance and control method for beam-riding guidance. In *Journal of Physics: Conference Series*, number 1. IOP Publishing, 2019.

[6] Qian P, Jianguo G, and Jun Z. Integrated guidance and control system design for laser beam riding missiles with relative position constraints. *Aerospace Science and Technology*, 98:105693, 2020.

[7] Siyuan G, Hui L, Hongwei Z, Xin Z, and Juan C. Improve the detection range of semi-active laser guidance system by temperature compensation of four-quadrant pin detector. *Sensors*, 19, 2019.

[8] WANG S, LI L, CHEN W, and SUN M. Improving seeking precision by utilizing ghost imaging in a semi-active quadrant detection seeker. *Chinese Journal of Aeronautics*, 2021.

[9] Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.

[10] IVOR B Hart. Leonardo da vinci's manuscript on the flight of birds.

[11] Erin M. How we lifted flight from bird evolution, Dec 2020.

[12] Josep A, Yvan P, Joaquim S, and Xavier L. The slam problem: a survey. *AI R&D*, 2008.

[13] Kevin Doherty, Dehann Fourie, and John Leonard. Multimodal semantic slam with probabilistic data association. In *2019 international conference on robotics and automation(ICRA)*, pages 2419–2425. IEEE, 2019.

[14] Artur Shurin and Itzik Klein. Qdr: A quadrotor dead reckoning framework. *IEEE Access*, 8:204433–204440, 2020.

[15] William P, Martin P, Daniel S, Ivan S, and Zoran O. Autonomous navigation for drone swarms in gps-denied environments using structured learning. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 219–231. Springer, 2020.

[16] TRDA, UAE Spectrum Outlook. `https://tdra.gov.ae/-/media/About/regulations-and-ruling/EN/UAE-Spectrum-Outlook-2020-2025-v1-0-pdf.ashx`.

[17] Yunhao Du, Junfeng Wan, Yanyun Zhao, Binyu Zhang, Zhihang Tong, and Junhao Dong. Giaotracker: A comprehensive framework for mcmot with global information and optimizing strategies in visdrone 2021. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 2809–2819, October 2021.

[18] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks, 2016.

[19] Mingxing T, Ruoming P, and Quoc L. Efficientdet: Scalable and efficient object detection, 2020.

[20] Xingkui Z, Shuchang L, Xu W, and Qi Z. Tph-yolov5: Improved yolov5 based on transformer prediction head for object detection on drone-captured scenarios, 2021.

[21] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[22] Leon V, Benjamin K, Martin M, and Andreas Z. Seadronessee: A maritime benchmark for detecting humans in open water, 2021.

[23] Geoffrey H, Oriol V, and Jeff D. Distilling the knowledge in a neural network, 2015.

[24] Zhiqiang S and Eric X. A fast knowledge distillation framework for visual recognition, 2021.

[25] Jonathan F and Michael C. The lottery ticket hypothesis: Finding sparse, trainable neural networks, 2019.

[26] Song H, Huizi M, and William D. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding, 2016.

[27] Vincent V, Andrew S, and Mark M. Improving the speed of neural networks on cpus. In *Deep Learning and Unsupervised Feature Learning Workshop, NIPS 2011*, 2011.

[28] Matthieu C, Yoshua B, and Jean-Pierre D. Training deep neural networks with low precision

multiplications, 2015.

[29] Alexandre B, Rami A, Miodrag B, and Emil P. Vessel identification using cnn-based hardware accelerators. In *2021 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 2021.

[30] Alexander B, Vladimir I, Eugene K, Alex P, Mikhail D, and Alexandr A. K. Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 2020.

[31] Xingyi Z, Dequan W, and Philipp K. Objects as points, 2019.

[32] Jongwon K and Jeongho C. Rgdinet: Efficient onboard object detection with faster r-cnn for air-to-ground surveillance. *Sensors*, 21(5):1677, 2021.

[33] Bin Y, Wenjie L, and Raquel U. Pixor: Real-time 3d object detection from point clouds, 2019.

[34] Liang P, Fei L, Zhengxu Y, Senbo Y, Dan D, Zheng Y, Haifeng Liu, and Deng Cai. Lidar point cloud guided monocular 3d object detection, 2021.

[35] Chenhang H, Hui Z, Jianqiang H, Xian-Sheng H, and Lei Z. Structure aware single-stage 3d object detection from point cloud. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[36] Shaoshuai S, Xiaogang W, and Hongsheng L. Pointrcnn: 3d object proposal generation and detection from point cloud. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[37] Kai-Wei C, Guang-Je T, Yu-Hua L, and Naser E. Development of lidar-based uav system for environment reconstruction. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 2017.

[38] Andy Z, Shuran S, Johnny L, Alberto R, and Thomas F. Tossingbot: Learning to throw arbitrary objects with residual physics, 2020.

[39] Masahiro F, Yukiyasu D, Ryosuke K, Gustavo R, Kenta K, Koji S, Rintaro H, Ryosuke A, Hironobu F, Shuichi A, et al. Bin-picking robot using a multi-gripper switching strategy based on object sparseness. In *2019 IEEE 15th International Conference on Automation Science and Engineering(CASE)*, pages 1540–1547. IEEE, 2019.

[40] Morgan Q, Ken C, Brian G, Josh D, Tully D, Jeremy L, Rob W, Andrew Y Ng, et al. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5. Kobe, Japan, 2009.

[41] Dji developer sdk. `https://developer.dji.com`, 2020.

[42] Gazebo - robot simulation made easy. `http://gazebosim.org`, 2022.

[43] Meet jetson. `https://developer.nvidia.com/embedded/jetson-benchmarks`, 2022.

[44] M. T. Mason and Kevin M. Lynch. Dynamic manipulation. *Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '93)*, 1:152–159 vol.1, 1993.

[45] Kaiming H, Xiangyu Z, Shaoqing R, and Jian S. Deep residual learning for image recognition. In *2016 IEEE Conference on CV and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[46] Andy Z, Shuran S, Kuan-Ting Y, Elliott D, Francois H, Maria B, Daolin M, Orion T, Melody L, Eudald R, Nima F, Ferran A, Nikhil C, Rachel H, Isabella M, Prem Q, Druck G, Ian T, Weber L, Thomas F, and Alberto R. Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching, 2020.

[47] Rosen D. *Automated Construction of Robotic Manipulation Programs*. PhD thesis, USA, 2010.

[48] *TripleA robotics Wingman automatic tool change on Universal Robots, Onrobot and Robotiq*. YouTube, Dec 2020.

[49] Eric B, Nicholas R, John S, Annan M, Erik S, Mitchell Z, Hod L, and Heinrich J. Universal robotic gripper based on the jamming of granular material. *Proceedings of the National Academy of Sciences*, 107(44):18809–18814, 2010.

[50] The importance of imu motion sensors, Sep 2019.

[51] Ola Tranum A, Frederik L, Håkon H, Stephanie K, and Tor A. J. Detection of objects on the ocean surface from a UAV with visual and thermal cameras: A ml approach. In *2021 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 81–90, 2021.

[52] Najda V, Alexandra L, Mohammad L, and Masoud D. Image synthesisation and data augmentation for safe object detection in aircraft auto-landing system. In *16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2021*, volume 5, pages 123–135. SciTePress, 2021.

[53] Andreas P and Klaus D. Robust semantic segmentation in adverse weather conditions by means of fast video-sequence segmentation, 2020.