Article

A Visual Aquaculture System Using a Cloud-Based Autonomous Drones

Naomi A. Ubina [1], Shyi-Chyi Cheng [1,*], Hung-Yuan Chen [1], Chin-Chun Chang [1] and Hsun-Yu Lan [2]

[1] Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung City 202, Taiwan; 20857002@mail.ntou.edu.tw (N.A.U.); 10757023@mail.ntou.edu.tw (H.-Y.C.); cvml@mail.ntou.edu.tw (C.-C.C.)

[2] Department of Aquaculture, National Taiwan Ocean University, Keelung City 202, Taiwan; 10833002@mail.ntou.edu.tw

* Correspondence: csc@mail.ntou.edu.tw

Abstract: This paper presents a low-cost and cloud-based autonomous drone system to survey and monitor aquaculture sites. We incorporated artificial intelligence (AI) services using computer vision and combined various deep learning recognition models to achieve scalability and added functionality, in order to perform aquaculture surveillance tasks. The recognition model is embedded in the aquaculture cloud, to analyze images and videos captured by the autonomous drone. The recognition models detect people, cages, and ship vessels at the aquaculture site. The inclusion of AI functions for face recognition, fish counting, fish length estimation and fish feeding intensity provides intelligent decision making. For the fish feeding intensity assessment, the large amount of data in the aquaculture cloud can be an input for analysis using the AI feeding system to optimize farmer production and income. The autonomous drone and aquaculture cloud services are cost-effective and an alternative to expensive surveillance systems and multiple fixed-camera installations. The aquaculture cloud enables the drone to execute its surveillance task more efficiently with an increased navigation time. The mobile drone navigation app is capable of sending surveillance alerts and reports to users. Our multifeatured surveillance system, with the integration of deep-learning models, yielded high-accuracy results.

Keywords: autonomous drone; cloud computing; visual aquaculture surveillance; object detection and activity recognition; remote sensing

1. Introduction

Food security and affordability are some of the major concerns of the world's growing population. It is a challenge to increase food production and make food more affordable and accessible to the general public. So far, many depend on fisheries, particularly in fish farming or aquaculture, as a food source. Previously, fish farming was a small-scale production that only addressed the need for a family or a small community's livelihood source. The increasing demand for aquaculture industrialization pushed small-scale farmers to improve their farming skills to increase production in their operations. One of the challenges to aquaculture property security is theft and human intrusions that can affect profit and farm operations, particularly to larger aquaculture sites with more financial investments and resources. Additionally, many of the aquaculture site locations are in open water and have higher risks or vulnerability to possible intrusions. Unauthorized vessels could access fish-cage areas for illegal fishing that causes profits losses. Most of the time, discovering a security breach occurs during harvest time. Others rely on on-site security, using barriers and employing guards to monitor and survey, but this entails additional expense from the owners. Aside from security threats, fish welfare and fish feed-

ing behaviors are significant factors for successful fish farming. Fish behaviors have practical and economic significance in fishery production. It has also become an essential theoretical basis to improve or guide fishing and production techniques by applying new technologies [1].

In the past, fish fed from natural food which came from natural water resources. Due to the development of aquaculture methods, farmers are now adopting artificial fish feeding. Since artificial feeding was adopted, the amount of feeding has become a great concern, so that neither inadequate feeding nor overfeeding occurs. If underfeeding affects the quality and quantity of fish yield, overfeeding decreases the profit of fish farmers due to food waste. Fish overfeeding also deteriorates the environment, particularly in water quality, and can further harm the fishing industry in general. To address this concern, farmers use direct feeding observation and manual human recording to monitor feeding, which is labor-intensive, time-consuming and energy-consuming. Additionally, human observation is prone to subjectivity and errors, and is not suitable as a continuous, accurate, and consistent source of information [2].

The aquaculture industry is still one of the fastest growing agricultural sectors despite many challenges, and is responsible for providing more available and affordable fish products on the market. The Department of Fisheries and Aquaculture under the Food and Agricultural Organization of the United Nations [3] reported that aquaculture is the leading source of worldwide fish production. Aquaculture's contribution to the world's total production remains consistent and rapid, from 25.7% in 2000 to 46.8% in 2016. This significant progress can be attributed to the farmer's utilization of aquaculture management software to reduce costs and optimize production [4]. Many technological innovations have been proposed and utilized to improve aquaculture production and management. However, many existing aquaculture management systems have problems with cost, mobility, efficiency, and functionality. Many of these current systems use expensive sensors for data collection. Additionally, site-wide monitoring of the entire aquaculture area requires the installation of multiple fixed cameras. Others use high-priced drones equipped with aerial cameras, sensors, and computational capability on board for site surveillance. Drones are reliable in monitoring, but they have a limited power source. Embedding the computing strength in the drone affects its real-time operations and limits its efficiency, mobility, and decreases navigation time [5]. Full system features are likewise expensive to acquire and require higher technical skills. The combination of radars, identification systems, drones, hovercrafts, thermal cameras, night vision, sonar, echo sounders, and countermeasures is costly. Our aquaculture surveillance provides physical site monitoring for aquaculture farms in order to monitor unauthorized individuals or ships for possible theft. The surveillance of fish behavior and growth will allow aquaculture farmers to monitor fish growth. Both surveillance methods can help ensure that no profit is lost due to security breaches, and that aquaculture production is optimized using fish feeding intensity, fish count, and fish-length estimation.

With the limitations of traditional surveillance systems, the higher cost, and the personnel requirement of commercial systems, our paper proposes a surveillance system that addresses the issues of cost, mobility, and efficiency using drones. Our method integrates computer vision to monitor aquaculture sites, such as fish feeding activities [6], inspecting nets, moorings, cages, and detecting suspicious objects (people, ships). These tasks require varying inspection facilities, adding difficulty to implementing a vision-based aquaculture surveillance system. To address this, we propose a low-cost and cloud-based autonomous drone equipped with a single camera to perform surveillance. The autonomous drone can capture the multiview video of the scene for inspection. The drone becomes an intelligent flying robot that captures distant objects and valuable data. These data will be processed using decision-making algorithms to provide information on how to optimize fish production and add security to the aquaculture site. The system has an aquaculture cloud, which is a private storage cloud that stores surveillance data for big data analysis.

The cloud receives the inspection data from the drone to perform object-recognition activities. This process is faster since the cloud has a higher and more powerful computing ability as opposed to using drones with onboard computational power. This ability enables the drone to have faster speed and longer navigation time for its inspection activities. The cloud is also capable of understanding the scene, detecting suspicious activity, and detection. Artificial intelligence using computer vision provides a real-time monitoring capability for an aquaculture site. Computer vision offers an ideal automatic, noninvasive, economical, and efficient method to monitor facility-specific activities at an aquaculture site, in order to address the problem of expensive sensors [7]. There are different types of object-detection and recognition models for the Artificial Intelligence (AI) services of our cloud system. However, conventional activity-recognition models cannot perform well on captured videos that are from a textureless water environment, especially for the evaluation of fish feeding intensity. To deal with this difficulty, we integrated our motion-estimation neural network to compute the optical flows from above-water images, since fish movements during feeding time produce optical flow. The use of optical flow enables the system to evaluate the intensity of fish feeding, and helps to detect excess feeding to help regulate the feeding process. Our approach provides an objective, more accurate and reliable method for monitoring fish feeding compared to human investigations, in order to increase agricultural productivity. Fish count and fish length are also important nonintrusive or noncontact methods for monitoring fish growth, fish welfare and production. The assessment of fish stocks helps determine the amount of food given to the fish. Estimation of fish length will help optimize feeding and determines the appropriateness of the fish to be sold on the market. Estimating the length is also an indicator of whether the food given is sufficient or not. Automating such methods will reduce contact and stress on the fish.

Lastly, the proposed system is a suitable alternative to expensive sensors and multiple cameras needed to cover the requirement for aquaculture site surveillance. Instead of using different recognition systems to monitor the inspection works individually, our proposed method combined various recognition models to form one framework that performs and executes various AI surveillance techniques and multiple inspection tasks. Additionally, to deal with usability problems, the system integrated a simple and easy-to-use mobile-based drone navigation application. Users can easily navigate the application and it will not require high technical skills in its operations. The drone is provided with a given altitude and camera position parameters to perform further surveillance at an intended location. Upon confirmation of suspicious activity on the site, the system triggers an alarm module that implements a vision-based Artificial Intelligence-Based Internet of Things (AIoT) system and sends the surveillance report to the user using the drone navigation application. As part of the notification, for the fish feeding intensity, the system alerts the user when the fish motion is no longer significant, which implies that feeding is already sufficient. Our proposed approach allows the deployment of a low-maintenance and cost-efficient system that utilizes a wireless network as its primary communication channel. The main contributions of our proposed system include:

1. The autonomous drone solves the complexities of installing fixed cameras to survey aquaculture sites and cages and collect the necessary information, which often requires electric power for the cameras to function. Drone autonomy can perform the inspection task automatically without full supervision from the user.
2. The integration of deep-learning models increases the reliability of surveillance functions.
3. The successful integration of various AI capabilities to detect various target objects (e.g., ships, fish tanks/cages/ponds) for site security, fish feeding evaluation, fish count and fish-length estimation in the cloud provides low-cost and scalable aquaculture surveillance. Cloud services can be extended to accommodate additional surveillance functions.
4. Drones provide an easier and faster way to collect big data from aquaculture sites, such as data on fish feeding, and act as a WIFI gateway for underwater cameras to

directly communicate with the cloud for other services such as fish count and fish-length estimation for improved aquaculture farm management and optimized production.

The remaining sections of the paper are as follows: Section 2 describes the framework of the system, performance evaluation, and experimental procedures and materials. Section 3 presents key system features and experimental results, while Section 4 provides the discussion and future works; Section 5 summarizes the work and our conclusion.

2. Materials and Methods

2.1. System Components and Functions

This section provides details and components of our visual surveillance system. The framework in Figure 1 contains three main elements: drone, Drone Navigation App, and scene cloud.
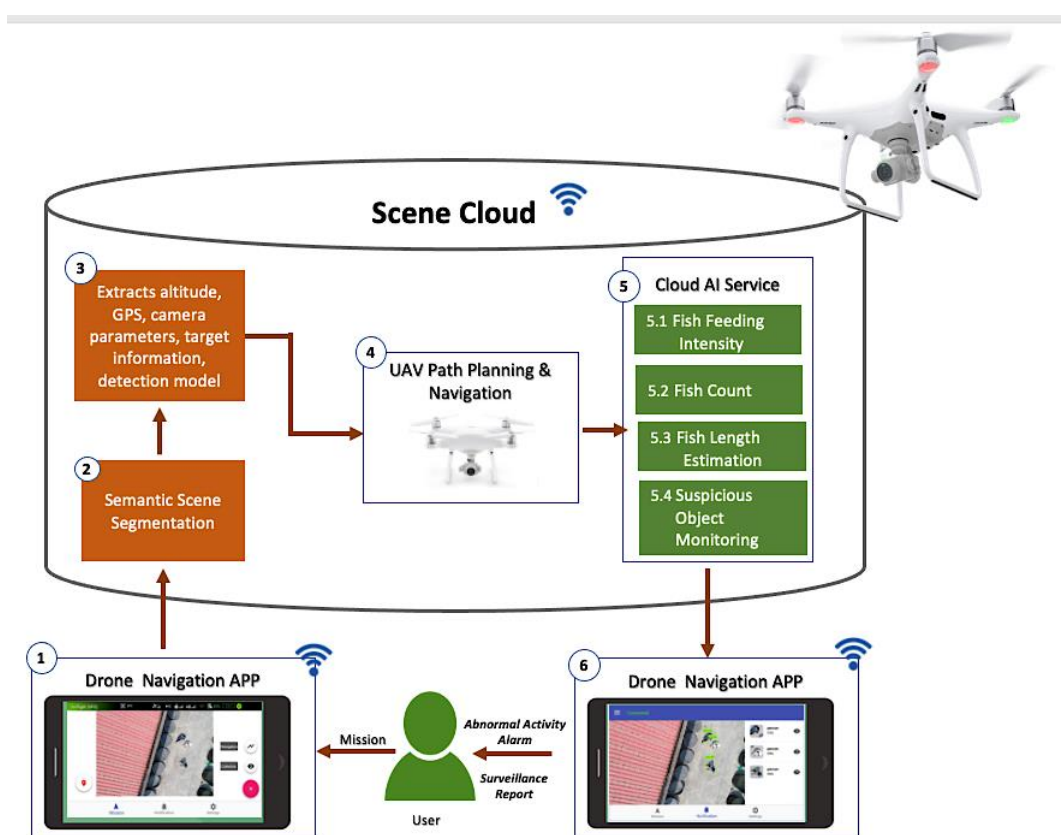

Figure 1. System framework.

The drone navigation app is a personalized mobile application installed on the user's device. It sends instructions and receives data from the drone, and at the same time sends data and receives alerts or notifications from the scene cloud. Using the Drone Navigation App, the user selects a mission that the drone must execute. This mobile application, developed using the Java language, can be installed on mobile devices with IOS or Android platforms and WIFI capability to connect to the drone and the cloud of the area.

After the user selects the mission, the drone navigation app will command the drone to fly at high altitude above the aquaculture farm to capture the aerial image of the aquaculture site. It will then send these images to the cloud for semantic scene segmentation to locate the target object in the field. Semantic scene segmentation labels or classifies spe-

cific regions of captured images for scene understanding. It analyzes the concept and nature of objects and recognizes them and their corresponding shapes in the scene. First, it detects the object, identifies its shape, and then performs object classification [8].

For semantic scene segmentation, the locations of the objects taken from the aerial images were already known, manually marked, and saved in the scene cloud database. These objects can be aquaculture tanks, box nets, ships, and personnel. Mask R-CNN [9] is a deep-learning convolutional neural network utilized for image segmentation and distinguishes the different objects in the images. Detection of various objects in the scene is needed to perform physical surveillance of the aquaculture site, which will allow the drone to identify its target object for its mission. Understanding the scene will allow the drone to navigate its target object correctly.

We integrated a two-dimensional (2D) semantic representation of the scene to locate the necessary objects in the aquaculture site. The result of the semantic scene segmentation is the visual and geometrical information of the semantic object that defines the checkpoints for the inspection work. Each checkpoint is associated with the corresponding global positioning system (GPS) signal. The altitude of the drone in Figure 2 locates the 2D objects such as aquaculture tanks and cages.

Figure 2. Semantic scene segmentation.

The scene cloud functions as the brain of the system, equipped with different services capable of decision making and analysis. The altitude, GPS information, camera parameters, target information, and detection model are the information stored in the cloud used for the unmanned aerial vehicle (UAV) path planning and navigation. The scene cloud utilizes the Google Firebase platform as its architecture and the scene dataset keeps the status of the scene updated. A database-management application was designed for its management so that users can easily search for information.

The drone path planning and navigation are the core of this paper. It contains the mechanisms, design, and procedures of its autonomous capability. Planning the path to navigate the target object or facility is relevant in performing surveillance for the drone to arrive at its target or destination safely. This plan will enable the drone to avoid obstacles during the flight which may cause damage, in order to safely navigate to its destination using the navigation path-planning service of the scene cloud. We integrated various algorithms and techniques for the drone's autonomous navigation capability. The details of the drone path planning and navigation are in Section 2.2.

Once the target object is established, the autonomous drone uses the calculated path from the path planning to navigate to the target object and perform individual monitoring tasks (e.g., fish feeding intensity and suspicious object monitoring). The drone once again takes the video of the monitoring tasks and sends it to the scene cloud to execute the corresponding AI service. The AI services process the data and sends an alert for abnormal activity, suspicious events, or the level of the fish feeding. Instead of utilizing the drone to include graphics processing, the AI services, as part of the cloud, save a large amount of power consumed by graphics processing, enabling the drone to use its power to perform surveillance. The convolutional neural networks and other technical details for the semantic scene segmentation and cloud AI services are presented in Section 2.3.

Additionally, the drone can serve as a gateway to connect the underwater cameras (e.g., stereo camera system, sonar camera) to the cloud AI services to perform additional fish surveillance such as fish count and fish-length estimation using the underwater camera.

Suspicious object monitoring can detect suspicious objects, which could be ships or humans, that can cause possible security threats to aquaculture farms. If a human is identified as a suspicious object, the corresponding AI service instructs the drone to continue the navigation. The computed altitude and camera pose parameters will be the basis for capturing the facial image of the suspicious human object. The captured image is sent to the cloud to perform further face recognition using FaceNet [10] deep-learning model. It will compare the newly captured images with the images of the authorized persons or staff in the scene database. This approach will distinguish workers from nonworkers or authorized from nonauthorized individuals and alerts the user for detected intruders. In executing the AI service for a fish-feeding-intensity mission, the cloud must first detect a fish cage or pond in the aquaculture site. Upon detection of the target object, the cloud will provide the altitude and camera position parameters for the drone to auto-navigate and proceed to the area of the fish cages. When the drone arrives at its target destination, it will then capture the feeding activity at the fish cage facility. The captured video is sent to the cloud for further processing and executes the fish-feeding-intensity evaluation and sends the feeding level to the users using the drone navigation APP.

2.2. Drone Path Planning and Navigation

For path planning, the scene cloud will use the GPS information of the detected object or the target goal (e.g., fish boxes for fish feeding intensity) to plan the path and sends this information to the user's device. This user's device has the drone navigation app installed and controls the drone using its WIFI channel as the communication device. The drone is also equipped with WIFI to communicate with the user's device and scene cloud. The D* Lite algorithm [11], works by reversing the A-star (A*) search framework [12] using an incremental heuristic method for its search functions. Its key feature involves using previous search results to identify the path-planning requirement instead of solving each search from the start. If connections between nodes are created, the data are modified and only affected nodes are recalculated. It also identifies the best path or the shortest path with the lowest navigation cost using the following equation:

$$f(n) = g(n) + h(n)$$

where $f(n)$ is the total score from the starting point through node n to the target point. Meanwhile, $g(n)$ is the actual distance from the starting point to node n and $h(n)$ is the estimated distance from node n to the target point. When $h(n) = 0$, no calculation is made. If $h(n)$ is the actual distance from node n to the target point, then node n is the node on the best path.

Since drone navigation is a dynamic planning problem due to its actual changing environments or characteristics, possible obstacles are present during the drone navigation process. The drone should be able to collect information and send it to the scene cloud to plan the path of an unfamiliar map in real-time. The D* Lite [11] algorithm for path

planning works for a dynamic environment and is the most suitable for drone navigation. It uses the current position as the starting point for the calculation and searches backward from the target point to the starting point. In drone navigation using the D* Lite algorithm to navigate a three-dimensional space, for the space coordinates of the starting point s_{start}, and target point s_{goal}, the x axis and y axis correspond to the GPS coordinates of longitude and latitude positions, respectively; the z axis represents the drone's navigational height. Figure 3 is a schematic diagram of drone navigation, and the blue grid is the drone flight map. This map is a plane composed of four points $(x_{start}, y_{start}, z_{start})$, $(x_{start}, y_{start}, z_{goal})$, $(x_{goal}, y_{goal}, z_{start})$, $(x_{goal}, y_{goal}, z_{goal})$. When encountering obstacles, the edge is updated, and the path is re-panned. Hence, if an obstacle (white box) suddenly appears during the navigation at the top, a new plan or path is created from the scene cloud and sends this information for the drone to avoid the obstacle until it reaches its target destination or path.
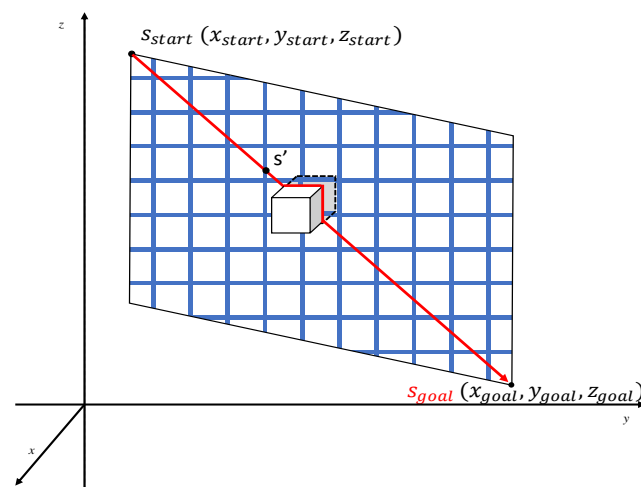

Figure 3. The navigation schematic diagram.

Figure 4 represents the flow chart of our drone navigation. As an initialization and to create the navigation map, the start-point and target point are needed. After building the plan map, the estimation of the shortest path (Figure 3) follows next. The shortest path ensures that the drone has the minimum navigation cost from its current to target location [13]. The drone now uses the computed shortest path in its navigation and checks if the starting point is the same as the target point. Once it reaches an equivalent value for the starting point and target point, the navigation task or mission is complete.

Otherwise, if the starting point is not the same as the target point, the navigation continues. It will then compute and finds the minimum vertex position to change the current starting point and moves to this new position. During the navigation, it will check for new obstacles in the area. If obstructions are detected, the path is recomputed using the shortest path function. If not, it will first update the current path plan, and then it will continue to calculate the fastest path. The path planning computation iterates until the drone's starting point is the same as its target point. The drone has reached its destination and completed its task once the start point is equal to the target point.

The sequence interaction diagram shows the order of interaction between the app, drone, cloud database, and the cloud services and how they work together to execute and complete visual surveillance. Figure 5 is the sequence diagram for the fish feeding intensity. Initially, the user selects the mission using the Drone Navigation App, which serves as the communication channel for the drone, cloud database, and cloud service for the drone take-off, capturing images and setting the route or path for the drone to navigate. The drone obtains the necessary parameters for navigation from the cloud database, the

internal and external camera parameters, field knowledge for field analysis, path planning, and feeding monitoring.
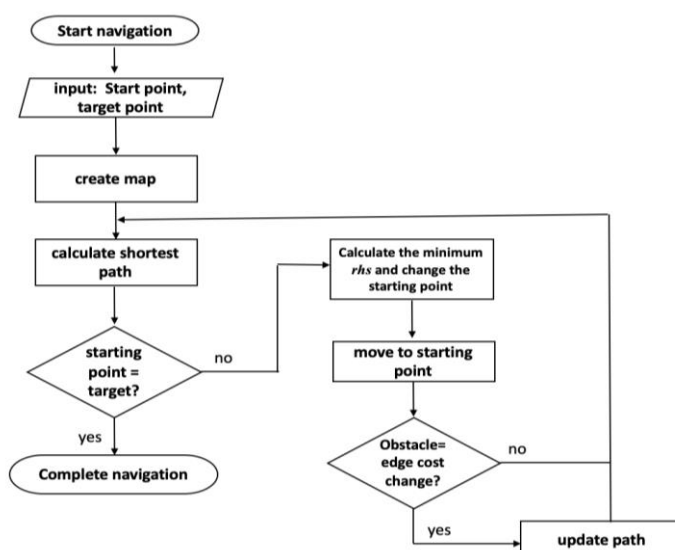

Figure 4. Flowchart diagram of the drone navigation.

When the drone takes off for the first time, it will perform scene analysis. The drone is first navigated at an altitude of 100 m to capture the image of the scene and sends it to the cloud database. After receiving the captured images, the cloud semantic understanding function searches for the location of the monitored objects from the scene. Once the captured scene is analyzed, it will send this information to the drone navigation app to perform drone navigation. The cloud provides the GPS coordinates of the drone, the location of the target object, the knowledge about the scene, route navigation to guide the drone towards the correct destination for monitoring. The drone continues to navigate until it reaches its target destination. The algorithm previously mentioned recalculates or modifies the navigation path in case there are obstacles along the way to avoid damage to the drone. Once the drone reaches the destination of the assigned mission or task, it will capture the fish feeding activity and sends data to the cloud AI service to evaluate the feeding level. This feeding level is forwarded to the user using the navigation app.

Figure 6 is the sequence diagram for monitoring the personnel in the aquaculture site. The same steps are integrated as for drone navigation in Figure 5. The only difference is that the personnel monitoring task is continuous even after obtaining the checkpoint. The drone monitors all individuals found on the site, one at a time. It captures the facial image of each person and sends it to the cloud AI service to perform face recognition. Once the cloud detects a suspicious person based on the facial images sent by the drone, the cloud alerts the user using the navigation app. If there is no suspicious person from the people present in the aquaculture site, the mission ends.
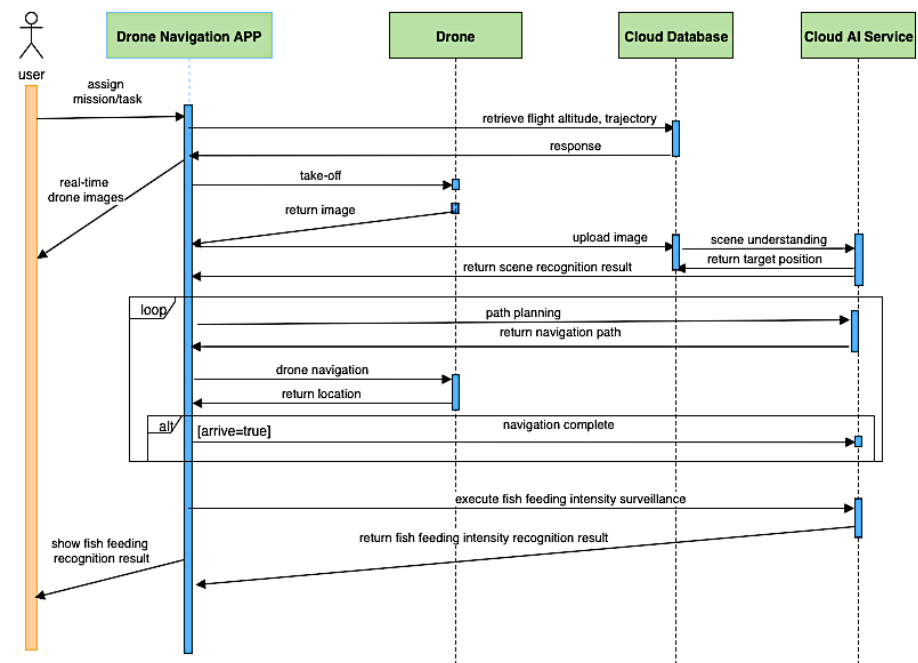
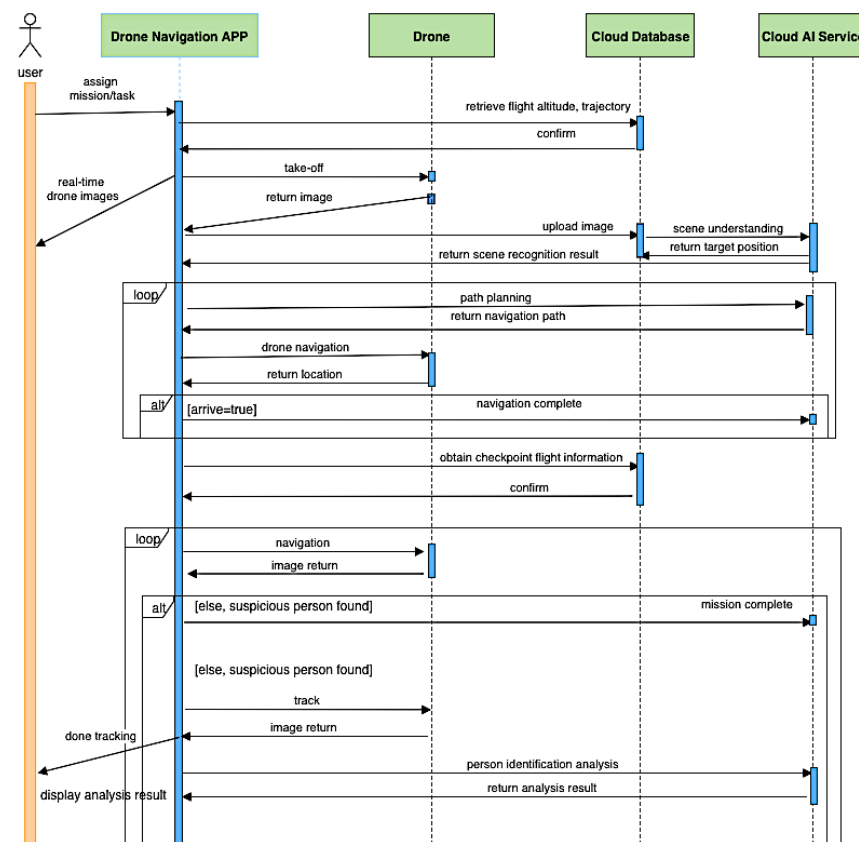**Figure 5.** Sequence interaction for fish feeding intensity.

Figure 6. Sequence interaction for personnel monitoring.

2.3. Experimental Procedures and Materials

The semantic object detection used the mean average precision, suspicious person detection integrated intersection over union, and person identification utilized the equation TP/(TP+FP) x 100. We used the following formula for fish feeding intensity:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \; x \; 100\%$$

where TP (true positives) is a result that a positive class was predicted correctly; TN (true negative) is a result that a negative class was predicted correctly; FN (false negative) is when a positive class gets a negative class result; and FP (false positive) is when a negative class gets a positive result. Aside from accuracy, we also measured the precision, recall, and F-score value for the fish feeding-intensity evaluation. The definition of F-score is as follows:

$$F - score = \frac{TP}{TP + \frac{1}{2}(FN + FP)} \; x \; 100\%$$

The experimental aquaculture site is at Aqua Center, Gong Liao, New Taipei City, Taiwan. The experimental data used were RGB videos captured using the Phantom 4 Pro V2.0. The drone navigation and DB management apps used the Java programming language for its implementation. Similarly, we used an Intel Core i7 3.4 GHz personal computer and NVIDIA GeForce 1080ti GPU for neural network model training. Two GoPro cameras was used as the stereo camera lens for the fish-length estimation.

The details of the different convolutional neural networks used, their corresponding parameters, training and evaluation methods, and datasets are presented in the following sections:

2.3.1. Convolutional Neural Networks (CNNs)

Deep learning has become famous for optimizing the results of various tasks for classification, recognition, or detection problems. CNNs are inspired by the concept of natural visual perception of living creatures [14]. CNN is a deep-learning method, taking images as input and performing processing using the idea of convolution to extract features or characteristics of the images. It generates the required output in the last layer, a fully connected layer where each node is linked to all other nodes in the previous layer [15]. Each convolution has a convolutional layer for feature extraction. Convolution preserves the relationship between pixels as it uses various filters to detect edges, blur, and sharpen images.

One component of convolution is activation functions, which define the weighted sum of its input nodes and then transform it as an output of the layer. They are nonlinearities that take a single number and perform mathematical operations on it. Furthermore, the addition of pooling layers reduces the computational requirement and reduces the number of connections between layers [14]. All AI services in this article used CNN as their methodology for learning features.

2.3.2. Semantic Segmentation

For site monitoring, there are different objects, such as box nets, ponds, and houses, for consideration. The Mask R-CNN deep-learning model [9] recognizes these objects present in the scene. The Mask RCNN first proposes regions of the possible objects from the input image. It will then predict those objects, refine the corresponding bounding box, and generate the object's mask on a pixel level using the proposed regions as its basis. We trained the Mask R-CNN using the image sequences captured from the aquaculture site at different times and perspectives. We manually marked the locations of aquaculture tanks, box nets, work platforms, and personnel in the captured images.

Figure 7 shows the images captured using the drone with an altitude of 150 m since it can cover most of the objects in the aquaculture site. The objects marked are fish ponds, box nets, and houses. There were two types of fish ponds: box nets and the ordinary pond. Fish ponds were further classified into rounded or square fish ponds. The box nets' location is in the open sea with 12 m diameter. Box nets were used for fish feeding-behavior analysis. Each object in the aquaculture site covers at least 30% of the scope of the screen before marking. A total of 114 images were manually labeled using LabelMe for the Mask R-CNN training and 200 images were used for testing. The parameters used for training the Mask R-CNN are learning _rate = 0.001, weight_decay = 0.001, and ephocs = 1000.

Figure 7. Aerial image of Gongliao Aquatic Experiment Center; the upper row is the original data; the bottom row is the labeled training data.

2.3.3. Fish Feeding Intensity

The fish feeding intensity is a part of the AI service which can evaluate the level of satiety of fish. As shown in Figure 8, we combined two deep neural network models: CNN-based optical flow estimation for motion estimation and CNN-based fish feeding-intensity evaluation to generate higher-accuracy results. For the optical flow estimation, we used video interpolation-based CNN [16] to estimate and generate the optical flow from our RGB input image sequences. The generated optical flow is used as input to the I3D model to determine the level of feeding intensity as none, weak, medium, or strong. The I3D model, as a two-stream approach, takes both the RGB video and the sequences of our preprocessed optical flow images [17] as inputs.

In preparing the dataset for the I3D model, we generated 216,000 frames or 6,750 video segments from 24 collected fish feeding activity videos. Each image frame is manually labeled using one of the four categories: none, weak, medium, or strong. These image frames were resized into smaller sizes (256 × 256 pixels) and cropped into 224 × 224 patches. The leave-one-out cross-validation method was the basis for the distribution of the training and testing dataset. In training the I3D model, we used 4,000 epochs with a batch size of 30, a learning rate of 0.0001, and a decay rate of 0.1. The complete details for the fish feeding-intensity evaluation are in a paper by Ubina et al. [18].
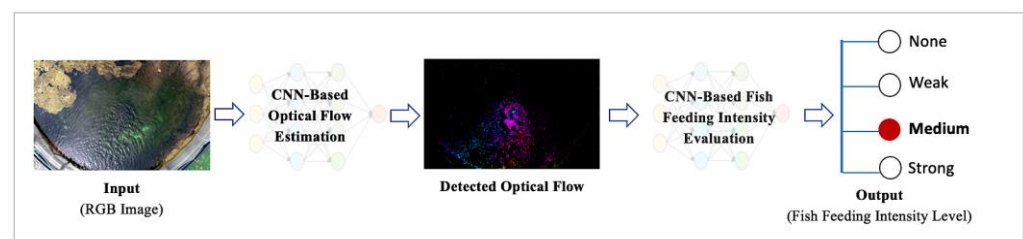
Figure 8. Fish feeding-intensity evaluation procedure.

2.3.4. Fish Count

To achieve real-time monitoring for fish, we applied an architecture for fish detection using YOLOV4 [19], which is very popular in terms of speed and high efficiency in object detection. In retraining the YOLOV4, we used 200 images, and each image is labeled with

around 110 fish. The batch_size is equal to 8, and the number of iterations is 2000. To determine the number of fish for each frame, we calculated the total number of fish as 30,000, with an actual volume of 11,996.57 cubic meters and a collected image frame of 45,000. We compared the ratio of the average number of fish for each image frame with the total number of fish. To calculate this, we used the following formula:

$$\sum_{k=1}^{15} X_k : 30,000 = \sum_{k=t}^{t+14} X_k : \text{Estimated total number of fish}$$

To calculate the density, we divided the estimated value of the total fish by the actual volume, where the unit used is the number of fish per cubic meter.

2.3.5. Fish-Length Estimation

We used a stereo-image camera system to estimate the fish body length to calculate the fish density. To determine the fish body length, we used a deep stereo-matching neural network and used the left image to reconstruct the right image. We also derived the optical flow or movement of the fish for each frame using its forward and backward motions to generate an intermediate frame.

We generated the optical flow of each image using a pretrained, unsupervised deep-learning optical-flow model based on the video interpolation method [16]. The optical flow is the basis for calculating the disparity value of the two images by summing up the pixelwise residual displacement for each pixel in the target fish object. In computing the 3D coordinates of the pixel value, the disparity is combined with the camera parameters to reconstruct the 3D point cloud of the target object to estimate the fish body length.

Instead of using a single image frame to estimate the body length, we considered the different postures of the fish for each image frame. By tracking the fish across frames, we calculated the body length for each frame and obtained the average, which serves as the final body length. We integrated the Principal Component Analysis (PCA) method in analyzing the length, width, and height of the 3D fish using the front and side positions.

2.3.6. Suspicious Object Monitoring

To detect suspicious objects in the site, we adopted a streamlined architecture using depthwise separable convolutions with MobileNet [20]. This is a lightweight and efficient model that reduces the number of parameters, so it meets the real-time calculation requirements. The depthwise separable convolution decomposes a standard convolution into deep convolution (depthwise convolution and 1x1 pointwise convolution). Figure 9 shows the framework of the depthwise separable convolution using the input (D_F, D_F, M). The convolution kernel (D_K, D_K, M, N) can produce (D_G, D_G, N) output after convolution. The depth separable convolution together with the input size of (D_F, D_F, M) and the convolution kernel $(D_K, D_K, 1, M)$ are convolved to produce (D_G, D_G, M). This output is further convolved with the point-by-point convolution kernel $(1, 1, M, N)$ and adds the result of the M channel to produce the result (D_G, D_G, N), which is an output of the same channel as the standard convolution kernel. This method has reduced the amount of calculation by eight to nine times. We used LabelMe to manually provide the label of the collected aerial images for the suspicious person dataset. The suspicious objects may appear both in inland aquaculture farms (people) or in the open sea (ship vessels or boats). The data sets were captured involving suspicious persons with various postures in the field, such as walking, fish stealing, and sitting.

In training the MobileNets [20], we loaded the pretrained model ssd_inception_v2_coco and set the batch_size to 24 to increase the speed and accuracy. The COCO dataset was utilized to fine-tune the final parameters of the neural network. The dataset contains 3,328 images at 20 different checkpoints, with at most five persons in each image. We utilized 80% of those data for training and 20% for testing.
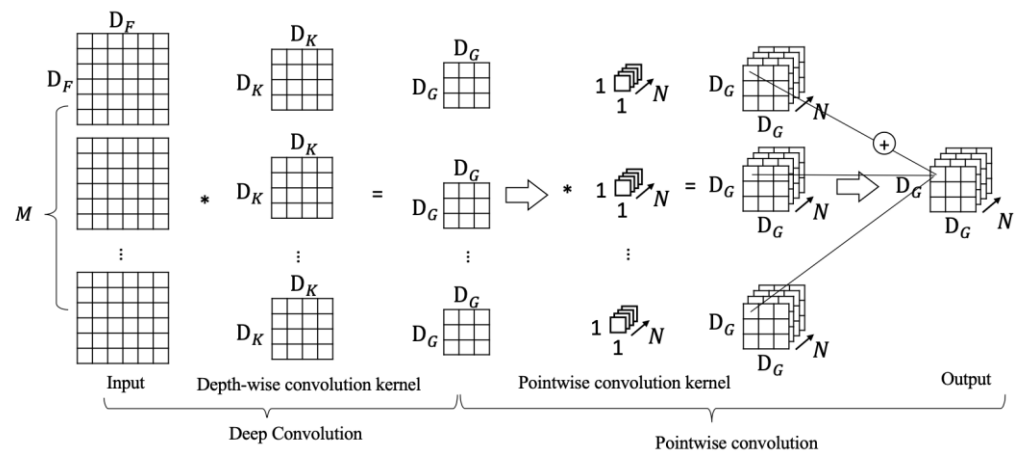
Figure 9. Depthwise separable convolution.

3. Results

3.1. Autonomous Drone for Navigation and Surveillance

Figure 10 shows the drone's initialization process with the drone navigation app installed on the mobile device. The user can select a mission such as suspicious person detection or fish feeding evaluation; the user can execute only one mission at a time. The drone navigation app initializes the drone's take-off. The drone was designed to be mobile and autonomous. Once initialized by the user, the drone can execute its mission and reach the target facility without any user intervention. Drone surveillance or site monitoring is the main function of the drone. This task is achieved by the drone using its installed RGB camera to capture real-time images and videos as a form of surveillance and sends this directly to the cloud for processing through its AI services. The surveillance report generated by the AI services sends this information to the user through its drone navigation app, as shown in Figure 11. Once the drone has completed its task or cannot detect the target facility, it will go back to its original starting position.
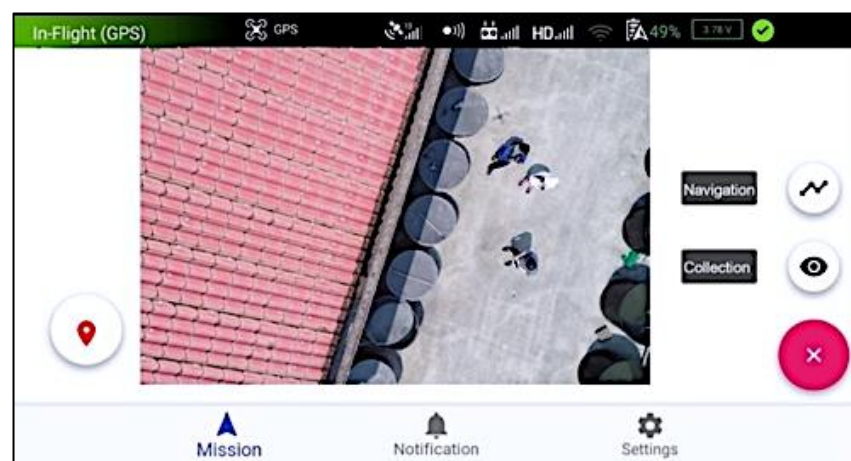

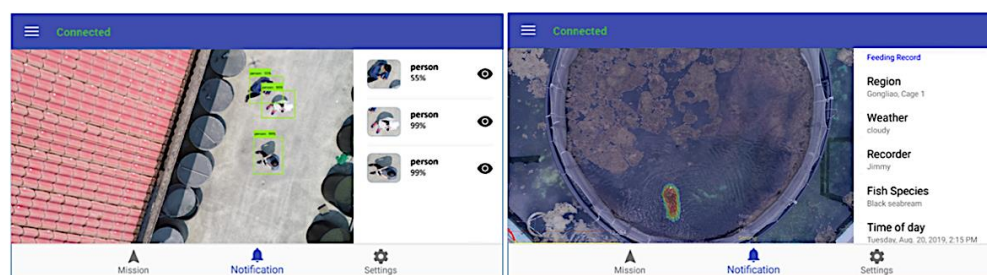**Figure 10.** Mission selection by the user.

Figure 11. Surveillance report using the navigation application.

3.2. Semantic Scene Segmentation

The drone is equipped with a camera and takes RGB images with different views and perspectives. These images captured by the drone are sent to the cloud to perform semantic scene segmentation. This function enables the system to understand the various facilities or objects present in the scene. For surveillance, we integrated a 2D semantic representation of the scene to locate the necessary objects in the aquaculture site. The semantic model represents the visual and geometric information of the semantic object that defines the checkpoints for the inspection work. Each checkpoint is associated with the corresponding GPS signal and altitude of the drone. Figure 12 shows the results for the detection of semantic objects in which objects were highlighted or colored (fish cages, houses, and fishponds) in the surveyed area. This detection capability using Mask R-CNN has 95% accuracy.

Figure 12. The results of semantic object detection.

3.3. Cloud AI Services

The cloud AI services are where the deep-learning models execute their classification and identification functions. The first AI service is the fish feeding-intensity evaluation, which is responsible for evaluating the satiety level of the fishes in the detected aquaculture cages or tanks. The drone needs to capture the feeding activity and send it to the cloud for evaluation. The feeding intensity has four levels: none, weak, average, and strong. After its evaluation, it sends the result as a form of surveillance report to the user. The information provided by the cloud will help the user decide whether to continue or stop the feeding process. Figure 13 shows the different fish feeding intensities based on videos captured by the drone during its surveillance, where (a) is the raw image and (b) is the processed image, which is the basis for classifying the level of fish feeding intensity.

A nonwave image texture shows a "none" feeding intensity and indicates that no significant movement from the fish manifests with zero feeding activity. However, images with larger wave textures show a "strong" feeding intensity that manifests active fish movements during the feeding process.

Altitude and camera angle are important considerations when capturing videos during the fish feeding activity. Higher altitude denotes a smaller image size, and a change in camera angle would also make a target object bigger or smaller. To address this, we first collected different images of fish feeding activity using a drone with different heights and viewing angles (Figure 14). Then, we identified the combination in terms of altitude and perspective that yields the highest accuracy rate. The combination of the lowest height (4 m) with the top view (top of the cage) produced the highest accuracy rate for the fish feeding-intensity evaluation.
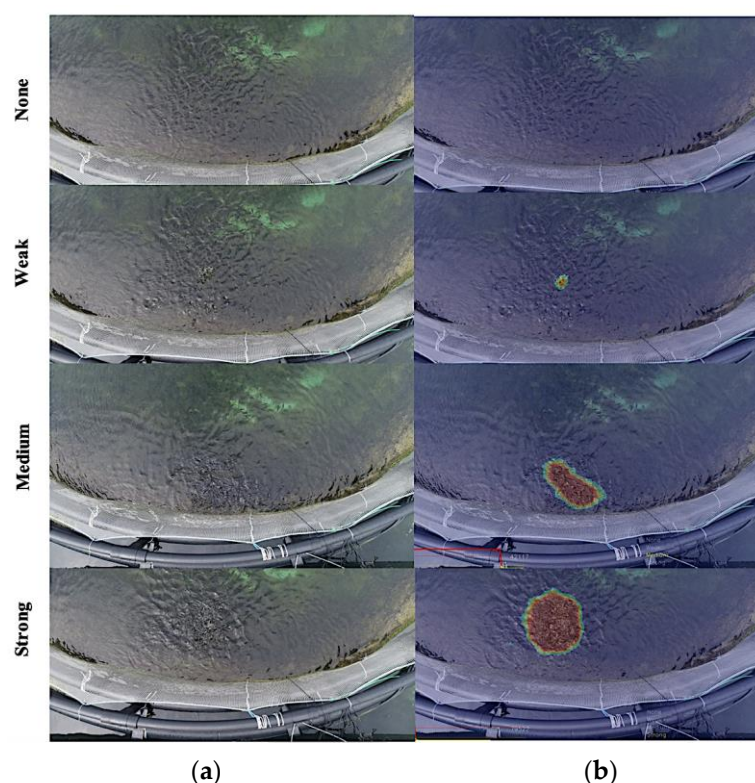

(**a**)   (**b**)

Figure 13. Fish feeding intensity results using four levels of feeding intensity. (**a**) original image; (**b**) processed image.
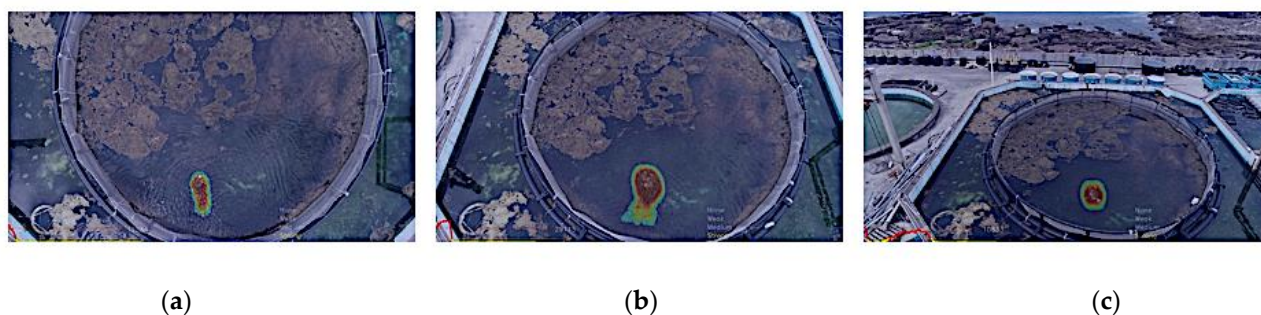

(**a**)   (**b**)   (**c**)

Figure 14. Different angular perspective of the drone camera capture. (**a**) View one (top) (**b**) View two (30 degrees adjustment (**c**) View three (50 degrees adjustment).

Figure 15 shows the best viewing angle (top view) using 4 m altitude results for the I3D training process for the accuracy evaluation and training loss of the fish feeding-intensity evaluation. The orange line represents the training curve, while the blue line is the validation curve where the training curve converged at 3000 steps. Using the best altitude and camera angle result, we tested the accuracy of our deep-learning model to evaluate fish feeding intensity. In training our deep-learning model, we used the following manually labeled datasets for each fish feeding-intensity level: 90 image sequences for none, 398 image sequences for weak, 373 image sequences for medium, and 401 sequences for strong. Each image sequence contains 80 frames, and 70% of the data set was used for the training data, while the remaining 30% is for the test data.
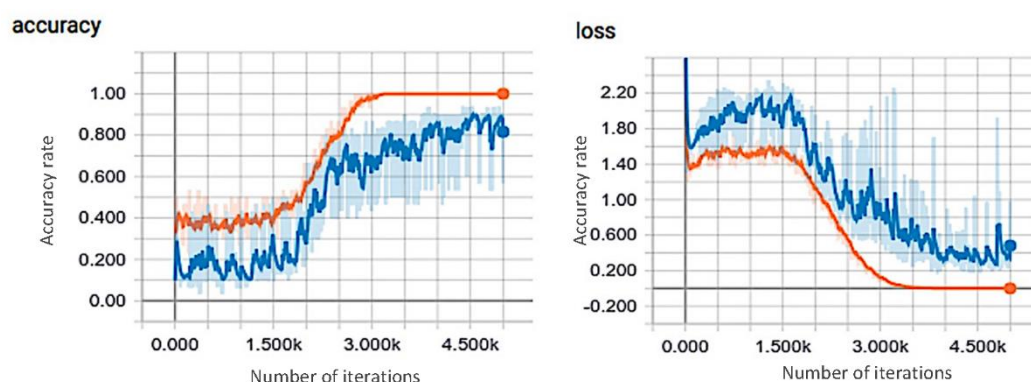

Figure 15. Training history for top view using 4 m altitude.

Our approach uses I3D [17], a convolutional neural network for video classification, and our video interpolated optical flow. We also implemented the approach of Zhou et al. [2] for the fish feeding-intensity evaluation using LeNet [21]. We used the same dataset and hardware to provide a realistic environment. The CNN architecture using Lenet [21] has three convolutional layers, two down-sampling layers, and two full-convolutional layers. We compared our accuracy results with that of Zhou et al. [2], shown in Table 1. Our proposed approach using I3D and video interpolated optical flow generated the highest accuracy rate of 95%, while the method of Zhou et al. [2] integrated with our video interpolated optical flow has 80% accuracy. Using I3D and RGB images, our approach has only 80% accuracy, while Zhou et al. [2] has a lower accuracy of 75%. The results of Zhou et al. [2] with video interpolated optical flow contributed to a 5% increase in accuracy compared with the RGB results.

Table 1. Comparison of our proposed fish feeding-intensity evaluation with other methods.

Method	Accuracy
I3D and video interpolated optical flow (our proposed)	95%
I3D and RGB	80%
Zhou et al. [2] and video interpolated optical flow	80%
Zhou et al. [2] and (RGB)	75%

Table 2 contains the results for the accuracy, precision, recall and F-score values for each class for the fish feeding intensity, where accuracy has the highest rate at 0.967 and precision the lowest rate at 0.912. The overall performance for the F-score is at 0.923 which is highly acceptable.

Table 2. Accuracy, precision, recall and F-score results for fish feeding-intensity evaluation.

Class	Accuracy	Precision	Recall	F-Score
None	0.994	0.980	0.996	0.948
Weak	0.975	0.809	0.895	0.850
Medium	0.941	0.8755	0.905	0.891
Strong	0.956	0.982	0.946	0.964
Average	0.9667	0.912	0.936	0.923

One of the activity recognition features of the aquaculture surveillance system is person detection (Figure 16), which renders up to 99%. We also provide a label for the detected persons for easier recognition. Detecting ship vessels is also added to the detection of suspicious objects, as shown in Figure 17. We used our ship vessel dataset to train the MobileNet [20] deep-learning model for the detection and rendered a 90% accuracy. Person and ship detection is important for aquaculture site monitoring to make sure that only authorized individuals are allowed to enter the site and intruders are monitored for better site safety. It will also keep aquaculture facilities, especially fish cages and ponds, safe from possible sabotage or theft from intruders.

Figure 16. The results of the person detection and recognition.

To further confirm whether the identity of a detected person is suspicious or not, the drone will track/follow the person and capture the facial image of that person (Figure 18). The drone sends the captured image to the scene cloud for face recognition using the FaceNet [10] deep-learning model; this will determine if the person is authorized or not by comparing the captured image with the information of authorized personnel stored in the cloud database. Once an intruder or unauthorized person is detected, the system sends an alarm to the user through the drone navigation app. We have used three different altitudes (3, 4, and 5 m) to determine the best altitude parameter for drone image capture. Using 4 m as the best altitude, where we have obtained an 87% accuracy detection rate.
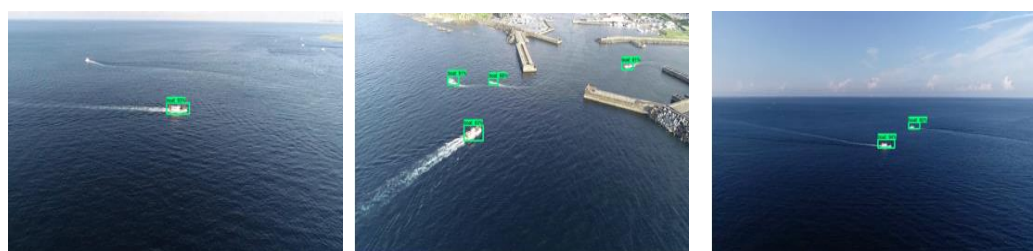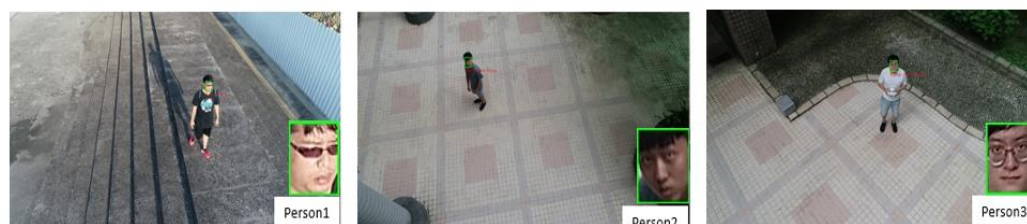
Figure 17. Fishing boat/ship detection results.

Figure 18. Face recognition detection results.

3.4. Drone as a WIFI Gateway Channel

Communication through wireless technology seems to be very popular for Internet of Things (IoT) devices due to increases in connectivity and coverage. The drone can be optimized for this capability serving as a WIFI gateway for underwater cameras installed in aquaculture ponds and cages, especially in the offshore open-sea area where one needs to take a boat/vessel to reach the site. In the study by Guillen-Perez et al. [22], they explored the capability of drones to provide WIFI networks. They presented a comprehensive characterization in deploying an aerial WIFI network by integrating a WIFI node that serves as an access point for the connection. We used the same concept of the drone as a WIFI gateway. This method provides better and more reliable communication and increases network connectivity [22] to monitor aquaculture fish cages and nets.

The drone is equipped with a camera to capture images of the water surfaces to evaluate the intensity of fish feeding. However, it is limited in conducting underwater monitoring due to its distance from the water. A typical monitoring setup for aquaculture farms uses underwater cameras, such as stereo cameras and sonar cameras. Cables connect cameras to a land-based computer system for control. For physical or wired connections, the hubs send the communication signals to the control system. With this setup, there is a problem when collecting underwater video and images for the fish count and fish-length estimation, for example. Since the size of underwater data is large, it will be difficult to send them to the cloud due to the distance of the cage and the on-site land-based control system.

Using the drone as a WIFI gateway eliminates the use of long cables and hubs for connection. Instead, it acts as a communication channel for the underwater cameras and the control system in order to have a reliable connection and improved communication. The architecture for the drone as a WIFI gateway can be seen in Figure 19, which shows the setup of how communication takes place. Equipping underwater cameras with an end-computing system enables wireless communication, which can help eliminate hubs and cables. The drone equipped with a WIFI app serves as an alternative way to communicate outside of the land-based control system, since the drone can directly send the data to the cloud AI service for processing. It is also possible to add sensors such as water-quality monitoring in which the drone can provide a connection to the cloud as a WIFI gateway.

Figure 19. Drone as WIFI communication channel from offshore cages to cloud AI services.

Figures 20 and 21 are the added AI services for aquaculture monitoring using the drone for the WIFI gateway channel. As an illustration of the data transmission using the drones as the WIFI gateway from underwater cameras, the images or videos captured are sent to the cloud using the end-computing system's WIFI capability. Cloud AI services will perform fish detection using the YOLO4 deep-learning model. The fish detection model can process 20 frames per second. The detected fish with bounding boxes are counted and generate the fish count results.

Figure 20. Fish detection and counting results.

In addition, the estimation of the length of the fish as part of the cloud AI services in Figure 21 shows the estimated size and weight of the fish. The average size in terms of width and height is used to estimate the final fish weight. Body-length estimation has an error rate of less than 5%.
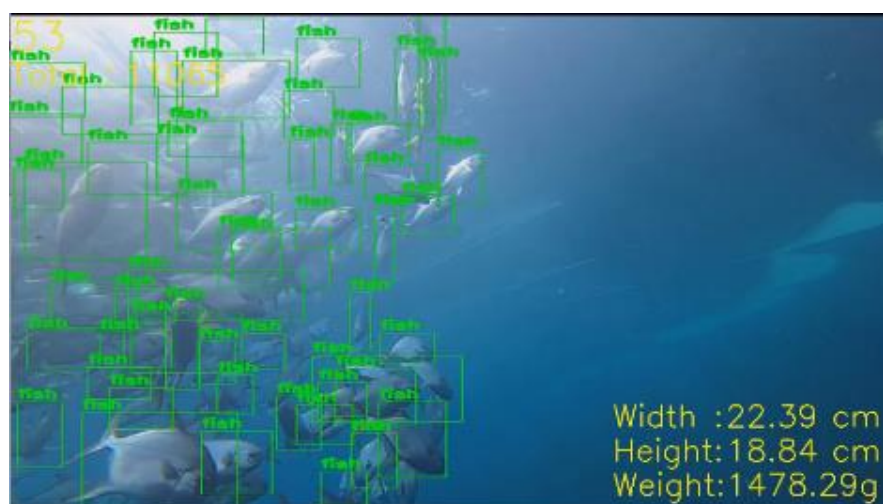
Figure 21. Fish-length and density estimation.

3.5. Execution Time

Since communication networks now employ higher bandwidth and faster speed, communication delays are no longer a problem to achieve real-time processing. The drone uses its camera to capture video from the aquaculture environment and takes only ten frames per second to deal with the time delay constraint. Fish feeding activity only captures ten frames per second to reduce the bandwidth for transmission. Thus, the data are sent more rapidly to the cloud services for processing and to achieve real-time analysis. Furthermore, the drone connects the end-computing system to extract the underwater images which are sent to the cloud for fish detection frame-by-frame, in order to estimate the fish lengths and counts. In this case, the drone is a WIFI gateway for establishing a reliable communication channel for offshore cages.

Table 3 shows the time needed to execute the different tasks for the drone to perform visual surveillance. It takes around a maximum of 7 min to finish one mission. Table 4 describes the type of AI service, the corresponding format of the image, and the time it takes to execute the function.

Table 3. Time requirement to execute the tasks.

Tasks	Time Requirement (in Seconds)		
	Start	End	Total
Drone take-off for scene capture	00	40	40
Semantic scene segmentation	40	42	2
Navigation (based on the flight distance)	42	1:42	60
Suspicious person detection & face recognition	1:42	4:12	150
Cage detection & fish feeding intensity	1:42	6:42	300
Underwater fish detection	1:42	6:42	300
Ship recognition (15 checkpoints)	1:42	6:42	300

Table 4. Time requirement for the AI services.

Services	Data	Time(s)
Scene understanding	Resolution: 1920 × 1080 Resize: 640 × 360 FPS: 30	0.33
Suspicious person or ship detection	Resolution: 3480 × 2160 Resize: 1920 × 1080	0.13

	FPS: 30	
Cage detection and fish feeding intensity	Resolution: 1920 × 1080	
	Resize 640 × 360	0.20
	FPS: 30	
	Resolution: 1920 × 1080	
Face recognition	Resize 1280 × 720	0.24
	FPS: 60	
	Resolution: 1920 × 1080	
Fish counting	Resize: Resize 640 × 360	0.67
	FPS: 30	
	Resolution: 1920 × 1080	
Fish-length estimation	Resize: 640 × 360	0.67
	FPS: 30	

4. Discussion

The key innovation of our proposed approach is the integration of an autonomous drone for combined visual surveillance. The power of drone navigation was designed based on the concept of an eagle eye, capable of flying above the aquaculture site to perform area surveillance. Its autonavigation capability can perform navigation based on the instructions or commands provided by the drone navigation app or the scene cloud. It does not need the intervention of users to control its direction once the mission is decided and confirmed. The autonomous drone can understand the position of the target objects through the information provided by the cloud, which makes it more intelligent than the usual drone navigation scheme. The drone can also execute a new path based on the path planning generated by the cloud, unlike the nonautonomous drone, which only follows a specific path.

There are many challenges in implementing drones for visual surveillance, including limited payload capacity and difficulty processing large data sets, such as a huge number of videos [23]. Drones have limited capabilities in terms of memory, processors, and energy. It is essential to optimize its ability to capture and collect numerous data from the site for monitoring and to allocate positions with minimum power consumption [24]. However, to address this limitation, we transferred the computationally demanding processing requirement to the cloud using artificial intelligence and deep-learning methods. Drones can communicate to the cloud using the Internet to perform the exchange of data and information and processing. The use of drones for surveillance can overcome the weakness of limited power, increasing its navigation time [25] and enabling it to be a round-the-clock surveillance system [26].

We integrate state-of-the-art technologies to perform surveillance to ensure that the best technologies and approaches have better potential and capabilities, which can be provided through our proposed method. We used different deep-learning models in our scene cloud to perform artificial intelligence services such as face recognition, fish feeding-intensity evaluation, fish count and fish-length estimation.

The various state-of-the-art applications for smart farming are presented in the work of [27], in which the UAV is one of the data-capture techniques. In our paper, we provide the details and functions of the drone as an autonomous vehicle to navigate and collect data from the aquaculture site. The drone is effective and affordable in its image-capture capabilities. It is also capable of wireless infrastructure and field connectivity [28] to transmit and receive data. One of the primary issues of drones in their data capturing that we dealt with is the quality of the captured images. Since we used a drone to capture the entire aquaculture site, the size of these images tends to be very small considering the total area required to cover. To address this, we have integrated a deep-learning model using Mask R-CNN for semantic scene segmentation, where even small objects can be detected and identified. In terms of data storage and data transfer, the images captured by the

drone are directly sent to the cloud-based platform for faster, safer processing and transferring of data.

We added system capabilities based on the requirements of accessibility and scalability. To address accessibility issues, we have designed a graphical user interface, the drone navigation app, to execute missions and monitor the drone surveillance activity. In terms of scalability and to enrich the capability of our system, it is possible to extend or add new AI services or functions such as plate-number detection for ship vessels. The user can send instructions to the drone by selecting a mission to execute a surveillance function by capturing images of the aquaculture site and sending them in real time to the cloud for further processing to provide users with warnings or alerts.

The drone is low in price and easy to deploy to collect information for the evaluation of fish feeding intensity. Additionally, it is not loud in noise emission, so it does not affect the feeding activity of the fish. The sound transmission of drones from the air to the water is poor compared to motorized ships and boats [29]. Therefore, the generated sounds would not affect the evaluation of fish feeding.

Our visual surveillance system can detect suspicious objects such as humans or ships using the captured RGB image. The inclusion of surveillance reports or user alerts enables the users to be informed in case there is detected abnormal activity, such as suspicious objects or persons from the site.

In the next step of our research, we plan to integrate simultaneous inspection work by adding drones to perform multiple surveillance at the same time. We will also include ship plate-number detection to complete the inspection works for ship detection activity. We also want to extend our approach by integrating big data analytics to process the data collected from the site. It will provide information to farmers to help maximize the benefits of artificial intelligence for decision making. Lastly, we plan to integrate additional AI services for fish-behavior observation, water-quality assessments, broken and sea-grass net detection to provide complete underwater surveillance.

5. Conclusions

In this paper, we have presented our proposed visual surveillance system using a cloud-based autonomous drone for aquacultures sites. It successfully integrated different recognition models and artificial intelligence (AI) services with a high accuracy rate. Our proposed method can implement a low-cost and scalable aquaculture site surveillance system. The single camera of the drone provided an alternative solution to the electric power requirements of installing fixed cameras in aquaculture cages. The integration of deep-learning models into the various AI services increased the reliability of aquaculture surveillance offered by our system. Furthermore, the aquaculture cloud allowed the drone to provide precise inspection tasks despite its limited navigation time. The autonomous drone performed its surveillance functions automatically and did not require full supervision from its users. Fish farmers can take advantage of this autonomous capability. The scene cloud integrated a low-cost and scalable surveillance system with multiple-object detection capabilities. The system can accommodate additional surveillance services into its existing AI functions. The drone collects aquaculture data in a simpler and faster way, which could be analyzed to perform big data analytics that can help optimize farm production and management. Lastly, various inspection tasks obtained a very high accuracy along with fast operation speed, thus providing a reliable surveillance system for aquaculture sites.

Author Contributions: Conceptualization, N.A.U., S.-C.C., C.-C.C., H.-Y.L. and H.-Y.C.; methodology, N.A.U., S.-C.C. and C.-C.C.; software, H.-Y.C.; validation, S.-C.C. and H.-Y.C.; formal analysis, N.A.U., S.-C.C. and H.-Y.C.; investigation, S.-C.C. and H.-Y.C.; resources, S.-C.C. and H.-Y.L.; data curation, S.-C.C., H.-Y.L. and H.-Y.C.; writing—original draft preparation, N.A.U.; writing—review and editing, N.A.U. and S.-C.C.; visualization, N.A.U. and H.-Y.C.; supervision, S.-C.C. and C.-C.C.;

References

1. Niu, B.; Li, G.; Peng, F.; Wu, J.; Zhang, L.; Li, Z. Survey of Fish Behavior Analysis by Computer Vision. *J. Aquac. Res. Dev.* **2018,** *9*, doi:10.4172/2155-9546.1000534.

2. Zhou, C.; Xu, D.; Chen, L.; Zhang, S.; Sun, C.; Yang, X.; Wang, Y. Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision. *Aquaculture* **2019**, *507*, 457–465, doi:10.1016/j.aquaculture.2019.04.056.

3. Food and Agricultural Organization of the United Nations. 2018. The State of the World Fisheries and Aquaculture. Available online: http://www.fao.org/state-of-fisheries-aquaculture/2018/en (Accessed on 22 August 2021).

4. Zhou, C.; Lin, K.; Xu, D.; Chen, L.; Guo, Q.; Sun, C.; Yang, X. Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture. *Comput. Electron. Agric.* **2018**, *146*, 114–124, doi:10.1016/j.compag.2018.02.006.

5. Van Beeck, K.; Tuytelaars, T.; Scarramuza, D.; Goedemé, T. Real-Time Embedded Computer Vision on UAVs. In *Leal-Computer Vision—ECCV 2018 Workshops*; Taixé, L., Roth, S., Eds.; Lecture Notes in Computer Science; Springer: Cham, Germany, 2018.

6. Wishkerman, A.; Boglino, A.; Darias, M.J.; Andree, K.B.; Estévez, A.; Gisbert, E. Image analysis-based classification of pigmentation patterns in fish: A case study of pseudo-albinism in Senegalese sole. *Aquaculture* **2016**, *464*, 303–308, doi:10.1016/j.aquaculture.2016.06.040.

7. Zhou Y, Yu H, Wu J, Cui Z, Zhang F (2019b) Fish Behavior Analysis Based on Computer Vision: A Survey. In *Data Science. ICPCSEE 2019*; Mao, R.; Wang, H., Xie, X., Lu, Z., Eds.; Communications in Computer and Information Science; Springer: Singapore, 2019; Volume 1059, pp. 130–141, doi:10.1007/978-981-15-0121-0_10.

8. Saffar, M.H.; Fayyaz, M.; Sabokrou, M.; Fathy, M. Semantic Video Segmentation: A Review on Recent Approaches. *ArXiv* **2018**, doi:abs/1806.06172.

9. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988, doi:10.1109/ICCV.2017.322.

10. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823, doi:10.1109/CVPR.2015.7298682.

11. Koenig, S.; Likhachev, M. D* Lite. In Proceedings of the AAAI Conference of Artificial Intelligence (AAAI), Alberta, AB, Canada, 28 July-1 August 2002; pp. 476–483, ISBN: 978-0-262-51129-2.

12. Hart, P.; Nilsson, N.; Raphael, B. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Trans. Syst. Sci. Cybern.* **1968**, *4*, 100–107, doi:10.1109/TSSC.1968.300136.

13. Mackay, D.; DRDC Suffield,. *Path Planning with D*Lite, Implementation and Adaptation of D*Lite Algorithm, Technical Memorandum DRDC*; Suffield, Connecticut, Defense and Research Development Canada, 2005.

14. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377, doi:10.1016/j.patcog.2017.10.013.

15. Aloysius, N. Geetha, M. A review on deep convolutional neural networks. In Progress of the 2017 International Conference on Communication and Signal Processing (ICCSP), Tamilnadu, India, 6-8 April 2017; 0588-0592, doi:10.1109/ICCSP.2017.8286426.

16. Niklaus, S.; Mai, L.; Liu, F. Video Frame Interpolation via Adaptive Separable Convolution. In Progress of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–27 October 2017; pp. 261-270, doi:10.1109/ICCV.2017.37.

17. Carreira, J.; Zissermany, A. Quo Vadis, action recognition? A new model and the Kinetics dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4724-4733, doi:10.1109/CVPR.2017.502.

18. Ubina, N.; Cheng, S.C.; Chang, C.C.; Chen, H.Y. Evaluating fish feeding intensity in aquaculture with convolutional neural networks. *Aquac. Eng.* **2021**, *94*, doi:10.1016/j.aquaeng.2021.102178.

19. Bochkovskiy, A.; Wang, C.; Liao, H.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *ArXiv* **2020**, abs/2004.10934.

20. Howard, A.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *ArXiv* **2017**, doi:abs/1704.04861.
21. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. In Proceedings of the IEEE, Nov. 1998; Volume 86, pp. 2278-2324, doi:10.1109/5.726791.
22. Guillen-Perez, A.; R.; Sanchez-Iborra, Cano, M.; Sanchez-Aarnoutse, J.C.; Garcia-Haro, J. WiFi networks on drones. In Proceedings of the 2016 ITU Kaleidoscope: ICTs for a Sustainable World (ITU WT), Bangkok, Thailand, 14-16 November 2016; pp. 1-8, doi:10.1109/ITU-WT.2016.7805730.
23. Garcia, A., Ghose, K. Autonomous indoor navigation of a stock quadcopter with off-board control. 2017 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS), 2017, pp. 132-137, doi:10.1109RED-AUS.2017.8101656.
24. Mahmoud, S.; Mohamed, N.; Al-Jaroodi, J. Integrating UAVs into the Cloud Using the Concept of the Web of Things. *J. Robot.* **2015**, doi:10.1155/2015/631420.
25. Yu, Y.; Lee, S.; Lee, J.; Cho, K.; Park, S. Design and implementation of wired drone docking system for cost-effective security system in IoT environment. 2016 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 7-11 January 2016; pp. 369-370, doi:10.1109/ICCE.2016.7430651.
26. Chae, H.; Park, J.; Song, H.; Kim, Y.; Jeong, H. The IoT based automate landing system of a drone for the round-the-clock surveillance solution. In Proceedings of the 2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM), 7-11 July 2015, Busan, South Korea, 2015; pp. 1575-1580, doi:10.1109/AIM.2015.7222767.
27. Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.J. Big Data in Smart Farming—A review. *Agric. Syst.* **2017**, *153*, 69–80, doi:10.1016/j.agsy.2017.01.023.
28. Faulkner, A.; Cebul, K. *Agriculture Gets Smart: The Rise of Data and Robotics, Cleantech Agriculture Report 2014*; Cleantech Group: San Francisco, CA, USA, May 2014.
29. Erbe, C.; Parsons, M.; Duncan, A.J.; Osterrieder, S.; Allen, K. Aerial and underwater sound of unmanned aerial vehicles (UAV, drones). *J. Unmanned Veh. Syst.* **2017**, *5*, 92–101, doi:10.1139/juvs-2016-0018.