CNN-BASED ESTIMATION OF WATER DEPTH FROM MULTISPECTRAL DRONE IMAGERY FOR MOSQUITO CONTROL

Qianyao Shen[1], K.T.Y. Mahima[2], Kasun De Zoysa[2], Luca Mottola[3,4], Thiemo Voigt[3,4], Markus Flierl[1]

[1]KTH Royal Institute of Technology, Electrical Engineering and Computer Science, Sweden
[2]University of Colombo, School of Computing, Sri Lanka
[3]RISE Research Institutes of Sweden, Sweden
[4]Uppsala University, Department of Electrical Engineering, Sweden

qianyao@kth.se, yasasm@scorelab.org, kasun@ucsc.cmb.ac.lk, luca.mottola@ri.se, thiemo.voigt@angstrom.uu.se, mflierl@kth.se

ABSTRACT

We present a machine learning approach that uses a custom Convolutional Neural Network (CNN) for estimating the depth of water pools from multispectral drone imagery. Using drones to obtain this information offers a cheaper, timely, and more accurate solution compared to alternative methods, such as manual inspection. This information, in turn, represents an asset to identify potential breeding sites of mosquito larvae, which grow only in shallow water pools. As a significant part of the world's population is affected by mosquito-borne viral infections, including Dengue and Zika, identifying mosquito breeding sites is key to control their spread. Experiments with 5-band drone imagery show that our CNN-based approach is able to measure shallow water depths accurately up to a root mean square error of less than 0.5 cm, outperforming state-of-the-art Random Forest methods and empirical approaches.

***Index Terms*—** Bathymetry Retrieval, Multispectral Imagery, Convolutional Neural Networks, Drones.

1. INTRODUCTION

Dengue fever and Zika are two arboviruses that affect a significant proportion of the world's population [1]. The main vector species of both infections are *Aedes aegypti* and *Aedes albopictus* mosquitos, which breed in shallow and slow-moving water pools. Reducing and controlling potential breeding sites is key to limit their spread.

Identifying potential breeding sites requires rapid and accurate estimations of water depths over potentially large areas [2, 3, 4]. Satellite imagery only provides depth measurements at meter level, not to count the relatively high cost, the complex image processing required, and the high latency to deliver the estimates. Therefore, we explore the use of drones equipped with high-resolution multispectral cameras in order to estimate the depth of slow-moving water pools. Compared to satellites, drones can provide near real-time estimates with centimeter accuracy at low operational costs. These estimates can then be used to classify aquatic environments likely to harbor breeding larvae.

The outline of the paper is as follows: After discussing the limitations of the current state of the art in Sec. 2, we illustrate in Sec. 3 the design and training of a custom Convolutional Neural Network (CNN) for estimating shallow water depth using multispectral images acquired from drones. Our design is novel in that the network is robust to noisy multispectral images from drones, that the network adapts to various water bottom types, and that a very good accuracy is achieved in the centimeter range. Sec. 4 presents our obtained results for multispectral drone imagery in five datasets. Using 5-band drone imagery, we can estimate shallow water depths with a root mean square error (RMSE) of less than 0.5 cm and R^2 values of up to 0.99, outperforming state-of-the-art Random Forest methods and empirical approaches. Moreover, we study the impact of larvae in the water when estimating water depth. We observe, that the larvae in the water make the depth measurements slightly less accurate. Finally, we analyze the impact of selected spectral bands toward the model's depth estimation performance. In particular, the additional information in the Red Edge band is valuable.

This work has been partly funded by Digital Futures and the Swedish Research Council (Grant 2018-05024)

2. RELATED WORK

A large body of literature exists on bathymetry retrieval models, especially using remote imagery. Available methods are of analytical or empirical nature. Analytical methods examine the relationship between the spectrum values of images and water depth [5, 6], allowing bathymetry inversions without ground truth data. However, the complexity of models increases as a result of the numerous unidentified physical parameters that influence the spectrum measurements [7].

In contrast to analytical techniques, empirical methods rely on the availability of in-situ bathymetry data for model validation. Several papers are based on the work by Lyzenga [8, 9] and consider various applications [10, 11, 12, 13, 14]. Empirical approaches do not always need absolute radiometric or atmospheric adjustments [11, 15]. Moreover, they can handle different datasets as long as the bottoms of the water pools are similar [16].

For example, Yu et al. [17] use three different ratios of reflectance band pairs from multispectral images to reduce the impact of the bottom material and the type of water body. This way, an empirical model using the ratio of band pairs can be used to obtain the water depth. However, Yu's method considers only a depth-dependent linear relation between reflectance and depth. Unfortunately, this linear relation is not sufficiently accurate due to the diversity of bottom types and the noise in drone images.

Mahima et al. [4, 18] experiment with the bathymetric log-ratio algorithm [9] for estimating shallow water depths in the centimeter range using multispectral drone images. Also this empirical method

Table 1: Bands and corresponding sensor range data.

Band	Range
Blue	475 nm center, 32 nm bandwidth
Green	560 nm center, 27 nm bandwidth
Red	668 nm center, 16 nm bandwidth
Red Edge	717 nm center, 12 nm bandwidth
NIR	840 nm center, 57 nm bandwidth

(a) DJI Drone with MicaSense RedEdge-MX sensor.

(b) Set of water buckets used for the experiments.

Fig. 1: Experimental setup.

provides depth estimates of limited accuracy. It is not sufficiently robust to the noise in drone images.

On the other hand, machine learning techniques exist that are capable of mapping the complex relationship between multispectral data and depth values. For example, Manessa et al. [19] propose a Random Forest (RF) algorithm to estimate water depth in shallow coral reef water bodies based on single beam echo sounder samples and Worldview-2 satellite images. In this work, they use the surface reflectance of six visible bands.

Ai et al. [20] explore CNNs to retrieve water depth in marine water areas from remote sensing images. The authors use the local connectivity property of CNNs to exploit the local spatial correlation of image pixels for bathymetry inversion. The accuracy of various topologies employing various window widths and band combinations is further investigated in [21].

However, the majority of existing research in the area is based on satellite imagery, resulting in estimates in the range of meters. This is not sufficient for identifying potential mosquito breeding sites.

3. CNN-BASED DEPTH ESTIMATION

We illustrate the datasets we use, the corresponding preprocessing, and the network architecture.

3.1. Datasets

We use a DJI Phantom 4 drone equipped with a MicaSense RedEdge-MX sensor [22] to collect multispectral image datasets. RedEdge-MX provides RGB and multispectral outputs with a resolution of 2 cm per pixel while flying at a height of 60 meters. It records five multispectral bands at 1.2 MP with the wavelength and bandwidth ranges as depicted in Table 1.

Each RedEdge-MX camera kit contains a light sensor for irradiance and sun angle measurements as well as a reflectance panel for on-the-ground calibration. These radiometric calibration tools account for different weather and light conditions and are useful to obtain more accurate data.

Table 2: Dataset features.

Dataset	Height	Larvae
Dataset 1	varying altitudes 5-15 m	with and without
Dataset 2	about 8 m	without
Dataset 3	about 8 m	without
Dataset 4	about 5 m	with and without
Dataset 5	about 10 m	with and without

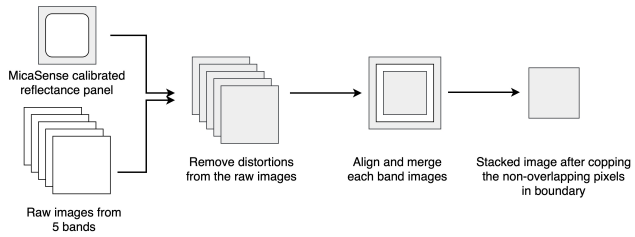

Fig. 2: Steps to generate stacked 5-band images from the MicaSense RedEdge-MX sensor.

Our setup is shown in Fig. 1. Since mosquito larvae breed in shallow water pools, we use buckets of water to emulate the environment. To obtain depth-accurate image data, we progressively add water to the bucket and fly the drone over the bucket to perform sensor measurements using the RedEdge-MX sensor.

We create a total of five datasets, each at varying altitudes, ranging from 5 to 15 m above the ground. We place small stones at the bottom of the buckets to emulate an urban scenario and to have the sensor's spectral reflectance behaviour be consistent across all buckets. Dataset 1, 4 and 5 have buckets with and without larvae. If there are larvae, they are near the water surface. Table 2 gives an overview.

3.2. Data Preprocessing

Since the MicaSense camera lenses are not aligned mechanically [23], each spectral image must be aligned with the other bands to create meaningful data. Image alignment and stacked image generation is a three-step process. First, we use the MicaSense Calibrated Reflectance Panel (CRP) to remove distortions and get the reflectance image. Specifically, the reflectance image can be obtained by extracting the area of the image containing the Lambert panel of the CRP, determining its radiance to reflectance scale factor, and then scaling the entire image by that factor. Then, we align each band to a common band by a two-dimensional (affine) transform between images. Finally, we merge and crop the aligned images. The process of aligning raw images from the five spectral bands and producing a stacked multispectral image is shown in Fig. 2.

Our goal is to estimate the water depth for a given measurement window in a stacked 5-band image. A measurement window of 1×1 pixels is very small and noisy. Large measurement windows will capture spatially varying effects, such as reflections and shadows. Such extreme sizes of measurement windows are not beneficial. Therefore, we choose windows smaller than 10×10 pixels. To find the best trade-off between noise and spatial effects, we will explore the three window sizes 5×5, 7×7, and 9×9 pixels. For simplicity, we extract multiple measurement windows inside the buckets and label them with the corresponding water depth to generate the data for our bathymetric CNN model.

Table 3: CNN architectures.

Model	Layer
Model 1	2 Conv2D, kernel size = 2×2
Model 2	2 Conv2D, kernel size = 3×3
Model 3	3 Conv2D, kernel size = 2×2
Model 4	3 Conv2D, kernel size = 3×3

Table 4: Different CNN architectures and measurement windows.

Model	5×5		7×7		9×9	
	RMSE	R^2	**RMSE**	R^2	**RMSE**	R^2
Model 1	1.09	0.94	0.82	0.97	0.71	0.98
Model 2	1.21	0.93	0.86	0.96	0.76	0.97
Model 3	1.02	0.95	0.79	0.97	0.75	0.97
Model 4	NA	NA	0.88	0.96	0.69	0.98

3.3. Network Architecture

In order to exploit the local spatial correlation among neighboring image pixels, we design a CNN regression network to estimate the water depth. At least one convolutional layer is needed to remove the noise. Our experiments show that multiple layers are appropriate to obtain accurate depth estimates. To build an efficient architecture for accurate depth estimates, we consider different numbers of convolutional layers and varying kernel sizes.

As discussed in the previous section, our maximum window size is 9×9. Therefore, we explore architectures with two or three convolutional layers and kernel sizes less than 3×3. In particular, we evaluate four different architectures. Larger measurement windows will provide more information to the network, however, entail more noisy data. More layers seem to be necessary to learn the noise patterns and to remove them. On the other hand, smaller windows will provide less information and the network has to deal with less noisy data. Less layers seem to be necessary to learn the noise patterns and to remove them. Table 3 lists several models and their parameters.

In the network, the first convolutional layer should attenuate the noise and extract depth features from the input. We use a rectified linear activation function. Further, we utilize batch normalization layers to prevent overfitting. More convolutional layers can be added to improve the quality of the extracted depth features. Then, the three-dimensional data is flattened and connected to a fully-connected layer. For the output layer, we use a linear activation function to accomplish the depth estimation task. An example of the CNN network is shown in Fig. 3.

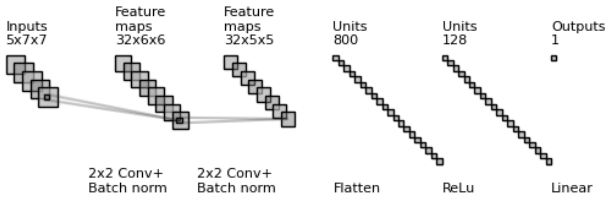

Fig. 3: Example CNN architecture (model 1 and window size=7).

4. EXPERIMENTAL RESULTS

Before we assess various CNN architectures, we discuss the metrics we use for the evaluation.

4.1. Metrics

Our primary objective is to obtain accurate estimates of water depth. Thus, we focus on comparing predicted and actual depth values. For the assessment, we use the Root Mean Square Error (RMSE) of the depth values and the R-squared (R^2) value to capture the unexplained variation of the depth values.

The root mean square error captures the magnitude of the average prediction error and is a measure of accuracy. On average, low RMSE values indicate accurate depth prediction values, whereas high RMSE values entail inaccurate depth prediction values.

The R-square value measures the proportion of the variance in the dependent variable that can be explained by the independent variable. It captures the quality of the regression model. The closer it is to one, the more effective is the regression model. When the prediction model makes no errors, the R-square value is 1. On the other hand, if the R-square value is less than 0, the prediction model fails.

4.2. Assessment of CNN Architectures

In the following, we assess our CNN architectures, including the choice of the data measurement window. As mentioned above, we evaluate three measurement window sizes and four distinct CNN architectures. For the assessment, we use three datasets, i.e. *Dataset 1* (water depth ranges from 1-10 cm), *Dataset 2* (water depth ranges from 1-18 cm) and *Dataset 3* (water depth ranges from 1-16 cm). For these three datasets, we use all five bands of the sensor, where 80% of the data is used for training and 20% for testing. The hyperparameters are the same for all configurations. We train with 500 epochs at a learning rate of 0.0001 and a batch size of 500.

Table 4 depicts the experimental results for each CNN architecture and measurement window size. We observe that the performance of the CNN model improves with increasing window size. That is, a larger number of samples within a measurement window will improve the accuracy of the prediction. Therefore, we choose a 9×9 window to generate the network input data for the following experiments. Further, we see that a deeper network and a larger kernel size result in better regression performance. Based on these observations, we choose a CNN model with three convolutional layers and 3×3 kernels (Model 4). Note, three convolutional layers

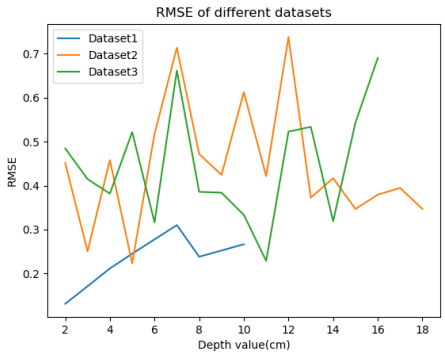

Fig. 4: RMSE vs. depth values for Dataset 1, 2, and 3.

Table 5: Depth estimation with or without larvae.

Model	Dataset 4		Dataset 5	
	with larvae	without larvae	with larvae	without larvae
RMSE	1.35	0.73	1.04	0.70
R^2	0.89	0.96	0.90	0.98

Table 6: Performance of our CNN model when using subsets of available bands.

Band	Dataset 1		Dataset 2		Dataset 3	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
5 bands	0.23	0.99	0.46	0.99	0.45	0.98
RGB	0.40	0.96	0.74	0.97	0.68	0.97
RGB+R	0.25	0.98	0.53	0.98	0.53	0.98
RGB+N	0.48	0.95	0.71	0.97	0.82	0.95

with 3×3 kernels require input windows larger than 5×5. Therefore, the corresponding results cannot be reported.

Finally, we use the chosen CNN (Model 4) and assess the accuracy of depth estimates for given depth values. Fig. 4 shows the RMSE for given depth values, separately for Dataset 1, 2, and 3.

4.3. Impact of Larvae in the Water

If larvae breed in shallow water, they are usually located near the water surface. These larvae may actually affect the accurate measurement of the water depth. In recent work, Gabriel et al. [24] observe that there is a variation in spectral reflectance when comparing water with and without mosquito larvae.

To study this further, we additionally collect *Dataset 4* and *Dataset 5*. Note, the former places the drone at an altitude of 5 m, the later at an altitude of 10 m. Both include water buckets with water depths from 1 to 15 cm and water buckets with and without mosquito larvae. Then we train a Model 4 network using both samples with and without larvae.

Table 5 gives the obtained results. For Dataset 4 and 5, the RMSE is slightly larger for water buckets with larvae when compared to those without larvae. From this data, it appears that the accuracy of the depth measurements is slightly lower for water buckets with larvae. A similar observation can be made when comparing the R-squared values. Hence, the observation in [24] may have also a minor impact on the accuracy of depth measurements. On the other hand, we have also a slightly different camera angle for the water buckets with larvae. We will investigate this in future work.

The main goal of our research is to accurately detect potential mosquito larvae habitats. The current results above suggest that our CNN estimates the depth of the water with sufficient accuracy to determine which water pools may contain mosquito larvae.

4.4. Impact of Selected Spectral Bands

The previous experiments use all five available bands of our multispectral sensor. In the following, we study the impact of using subsets of the available bands. This may allow us to use cheaper sensors with less bands. Therefore, we choose selected combinations of bands and analyze Dataset 1, 2, and 3.

The results in Table 6 show that using all five bands produces the best regression results. This is expected as the availability of richer information may result in more accurate estimates. The performance of RGB + Red Edge bands (RGB+R) is almost as good as that of 5 bands. On the other hand, RGB + Near Infrared (NIR) bands (RGB+N) perform similar to RGB bands only, or even worse. This indicates that adding the NIR band to RGB bands will not help much to improve the accuracy of the depth estimates. Finally, using only RGB bands comes in around third place in our experiments.

Therefore, 5-band ground data is preferred when accuracy is the only factor. But if costs have to be considered too, RGB + Red Edge bands are a good compromise.

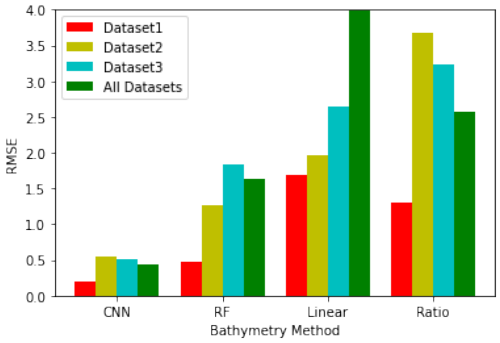

Fig. 5: Comparison with existing methods.

4.5. Comparison with Existing Methods

Finally, we benchmark our CNN model and compare with machine learning and bathymetry regression methods, in particular, Random Forests as well as linear and ratio bathymetry regression.

Fig. 5 depicts the RMSE for Dataset 1, 2, 3, and their union. We observe that our CNN model achieves the lowest RMSE. The Random Forest (RF) outperforms both linear and ratio bathymetry regression. Linear and ratio bathymetry regression assume a linear relationship between log-reflectance and depth, whereas RF and CNN are machine learning algorithms that can also consider other dependencies. In addition, CNNs can exploit the local spatial correlation of image pixels for more accurate water depth estimation. The adaptive attenuation of noise in the multispectral images appears to be beneficial too. Both observations may give our CNN an advantage over the Random Forest.

5. CONCLUSION

We have proposed a water depth retrieval model based on CNNs. We have conducted an experimental evaluation to explore the extent to which CNNs can be used for bathymetry in shallow water areas when using multispectral 5-band drone imagery. Our CNN-based approach is able to measure shallow water depths accurately up to a root mean square error of less than 0.5 cm, outperforming state-of-the-art Random Forest methods. Further, we have analyzed the impact of using subsets of available multispectral bands. Here, we have noticed the significance of the Red Edge band. Finally, we have explored the impact of larvae in the water when estimating water depth. Here, we have found that the impact is small enough to still allow us to estimate the water depth with sufficient accuracy. According to our observations, CNNs are advantageous for our problem as they exploit the local spatial correlation in images as well as attenuate adaptively the noise in the measurement windows.

6. REFERENCES

[1] Mosquito Reviews, "Statistics for mosquito-borne diseases and deaths," Online, Accessed on 2023-01-31, Available at https://mosquitoreviews.com/learn/disease-death-statistics.

[2] Michelle C. Stanton, Patrick Kalonde, Kennedy Zembere, Remy Hoek Spaans, and Christopher M. Jones, "The application of drones for mosquito larval habitat identification in rural environments: a practical approach for malaria control?," *Malaria Journal*, vol. 20, no. 1, pp. 1–17, 2021.

[3] Andrew Sutherland, "Mosquito management for ponds, fountains, and water gardens," *UC IPM Retail Nursery & Garden Center IPM News*, vol. 3, no. 2, Jun 2013.

[4] KTY Mahima, Malith Weerasekara, Kasun De Zoysa, Chamath Keppitiyagama, Markus Flierl, Luca Mottola, and Thiemo Voigt, "MM4Drone: A multi-spectral image and mmwave radar approach for identifying mosquito breeding grounds via aerial drones," in *International Conference on Pervasive Computing Technologies for Healthcare*. Springer, 2022, pp. 412–426.

[5] Zhongping Lee, Kendall L. Carder, Curtis D. Mobley, Robert G. Steward, and Jennifer S. Patch, "Hyperspectral remote sensing for shallow waters: 2. Deriving bottom depths and water properties by optimization," *Appl. Opt.*, vol. 38, no. 18, pp. 3831–3843, Jun 1999.

[6] David Lyzenga, "Remote sensing of bottom reflectance and water attenuation parameters in shallow water using aircraft and Landsat data," *International Journal of Remote Sensing*, vol. 2, pp. 71–82, 01 1981.

[7] Jakob J Assmann, Jeffrey T Kerby, Andrew M Cunliffe, and Isla H Myers-Smith, "Vegetation monitoring using multispectral sensors—best practices and lessons learned from high latitudes," *Journal of Unmanned Vehicle Systems*, vol. 7, no. 1, pp. 54–75, 2018.

[8] David R. Lyzenga, "Passive remote sensing techniques for mapping water depth and bottom features," *Applied Optics*, vol. 17, no. 3, pp. 379–383, 1978.

[9] Richard P. Stumpf, Kristine Holderied, and Mark Sinclair, "Determination of water depth with high-resolution satellite imagery over variable bottom types," *Limnology and Oceanography*, vol. 48, no. 1part2, pp. 547–556, 2003.

[10] Emily C. Geyman and Adam C. Maloof, "A simple method for extracting water depth from multispectral satellite imagery in regions of variable bottom type," *Earth and Space Science*, vol. 6, no. 3, pp. 527–537, 2019.

[11] Mehdi Gholamalifard, Tiit Kutser, Abbas Esmaili-Sari, Ali A. Abkar, and Babak Naimi, "Remotely sensed empirical modeling of bathymetry in the southeastern Caspian Sea," *Remote Sensing*, vol. 5, no. 6, pp. 2746–2762, 2013.

[12] Sheng Ma, Zui Tao, Xiaofeng Yang, Yang Yu, Xuan Zhou, and Ziwei Li, "Bathymetry retrieval from hyperspectral remote sensing data in optical-shallow water," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 2, pp. 1205–1212, 2013.

[13] Dimosthenis Traganos, Dimitris Poursanidis, Bharat Aggarwal, Nektarios Chrysoulakis, and Peter Reinartz, "Estimating satellite-derived bathymetry (SDB) with the Google Earth Engine and Sentinel-2," *Remote Sensing*, vol. 10, no. 6, pp. 859, 2018.

[14] Chunzhu Wei, Qianying Zhao, Yang Lu, and Dongjie Fu, "Assessment of empirical algorithms for shallow water bathymetry using multi-spectral imagery of Pearl River delta coast, China," *Remote Sensing*, vol. 13, no. 16, pp. 3123, 2021.

[15] Jared Kibele and Nick T. Shears, "Nonparametric empirical depth regression for bathymetric mapping in coastal waters," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 11, pp. 5130–5138, 2016.

[16] Isabel Caballero and Richard P. Stumpf, "Retrieval of nearshore bathymetry from Sentinel-2A and 2B satellites in South Florida coastal waters," *Estuarine, Coastal and Shelf Science*, vol. 226, pp. 106277, 2019.

[17] Yu Rui-hong, Xu You-peng, Liu Ting-xi, and Li Chang-you, "Reversing water depth in shallow lake of arid area using multi-spectral remote sensing information," *Advances in Water Science*, vol. 20, no. 1, pp. 111–117, 2009.

[18] K.T.Y. Mahima, Malith Weerasekara, Kasun De Zoysa, Chamath Keppitiyagama, Luca Mottola, Thiemo Voigt, and Markus Flierl, "Poster: Fighting dengue fever with aerial drones," in *Proceedings of the 2022 International Conference on Embedded Wireless Systems And Networks*, 2022, pp. 206–207.

[19] Masita Dwi Mandini Manessa, Ariyo Kanno, Masahiko Sekine, Muhammad Haidar, Koichi Yamamoto, Tsuyoshi Imai, and Takaya Higuchi, "Satellite-derived bathymetry using random forest algorithm and worldview-2 imagery," *Geoplanning J Geomatics Plan*, vol. 3, no. 117, pp. 117–126, 2016.

[20] Bo Ai, Zhen Wen, Zhenhua Wang, Ruifu Wang, Dianpeng Su, Chengming Li, and Fanlin Yang, "Convolutional neural network to retrieve water depth in marine shallow water area from remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2888–2898, 2020.

[21] Yustisi Ardhitasari Lumban Gaol, "Convolutional neural networks for satellite-derived bathymetry," Tech. Rep., TU Delft Architecture and the Built Environment, July 2021, Available at https://repository.tudelft.nl/islandora/object/uuid:662ac71f-6373-4128-8eeb-163b8e727b72.

[22] MicaSense, "MicaSense RedEdge-MX integration guide," Online, Accessed on 2022-01-15, Available at https://support.micasense.com/hc/en-us/articles/360011389334-RedEdge-MX-Integration-Guide.

[23] MicaSense, "Why are the images from the camera not aligned to each other?," Online, Accessed on 2022-03-04, Available at https://support.micasense.com/hc/en-us/sections/4420292047895-General-FAQs.

[24] Gabriel Carrasco-Escobar, Edgar Manrique, Jorge Ruiz-Cabrejos, Marlon Saavedra, Freddy Alava, Sara Bickersmith, Catharine Prussing, Joseph M Vinetz, Jan E Conn, Marta Moreno, et al., "High-accuracy detection of malaria vector larval habitats using drone-based multispectral imagery," *PLoS Neglected Tropical Diseases*, vol. 13, no. 1, 2019, doi 10.1371/journal.pntd.0007105.