

This work was written as part of one of the author's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law.

Public Domain Mark 1.0

<https://creativecommons.org/publicdomain/mark/1.0/>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

**Please provide feedback**

Please support the ScholarWorks@UMBC repository by emailing [scholarworks-group@umbc.edu](mailto:scholarworks-group@umbc.edu) and telling us what having access to this work means to you and why it's important to you. Thank you.

# GADAN: Generative Adversarial Domain Adaptation Network For Debris Detection Using Drone

Masud Ahmed\*, Naima Khan\*, Pretom Roy Ovi\*,  
Nirmalya Roy\*, Sanjay Purushotham\*,  
Aryya Gangopadhyay\*, Suya You<sup>§</sup>

<sup>\*</sup>University of Maryland Baltimore County, USA

<sup>§</sup>DEVCOM Army Research Laboratory, USA

**Abstract**—In maritime, coastal, and riverine environments debris has become an abundant pollutant, posing a significant threat to aquatic life. Existing debris detection systems are mostly designed to detect debris for specific type of environment. Therefore, we propose GADAN, an autonomous debris detection model fusing RGB and thermal images, and experiment on both over land and aquatic environment. We collected data using the Parrot Anafi Thermal drone from different water streams, and construction sites near the aquatic environment. We employ a generative domain adaptation based architecture to generate the thermal image compatible with the RGB image. We postulate a two-stream network on these thermal and RGB pairs of images separately, and pass the concatenated features to object detecting YOLO model. We also demonstrated the performance of debris detection using only RGB and only thermal images. Our proposed GADAN network incorporating RGB and thermal images outperforms (with a mean average precision of 0.97) RCNN and single modality based models.

**Index Terms**—Object detection, generative adversarial network, domain adaptation

## I. INTRODUCTION

The industrial production of plastics has been increased dramatically over the last six decades. Plastic has been the leading commodity on the commercial market, with an approximate production rate of more than 380 million tons per year [1]. Approximately 60 – 79% of the waste produced from these plastic products has been deposited in landfills and the natural environment in recent years. If this trend continues, then it is estimated that there will around 12 billion tons of plastic waste in the nature environment [2]. Recent studies emphasized plastic pollution as an unending risk to the environment like water pollution and mentioned it as a threat to marine life [3], [4]. Therefore, plastic pollution has been recognized by many international organizations as a ubiquitous pollutant for nature and declared it as a serious global environmental issue.

Several previous works have been performed to monitor debris by visual assessments of the aquatic ecosystem. However, it is important to define the debris in considered environment. According to Coe and Rogers, any fabricated man-made product or waste that enters the marine environment is considered as marine debris [5]. Marine debris can be categorized into several types: plastics debris, smoking waste,

metal, glass, rubber, cloth, etc. [6]–[9]. This debris has a diverse range of affects on the marine environment, such as ghost fishing and entanglements, water poisoning, physical damage and blockage creation for aquatic lives, accelerating growth of aquatic insect vectors, etc. [10]. Similarly, road-side and building-side debris can have various damaging affects on the surroundings aquatic environment. However, cleaning these pollutants requires available manufacturing infrastructures and a lot of trained people involved in time-consuming manual debris detection processes [11].

Unmanned aerial system (i.e. drone) based technology can be used as an alternative way to monitor and detect debris. Fast developments in technology for sensing, battery, and low-cost digital cameras made the drones more efficient, secure, and flexible device to be used in various aspects of different domains [12]. Such drone-based systems are aimed at achieving more efficient services in inspection and maintenance with minimal human intervention. However, previous automated debris detection systems were experimented on specific environments and relied one image modality (i.e., sonar images [13], optical images, etc.). As thermal images provide useful information about the emissivity of different materials, we are motivated to incorporate thermal images for debris detection. But one of the main challenges of incorporating thermal cameras is increment of cost. Installation of a thermal camera on a drone is costly. Even if there are both RGB and thermal cameras on drone, there exists spatial mismatch between RGB and thermal images to some extent. In our proposed framework, we address these issues by using generative domain adaptation technique which can transform both RGB and thermal images into same dimension consisting with pixels from same spatial region and also generate the thermal images from the RGB and then incorporate them along with the RGB images. We develop a model for object tracking by fusing RGB and generated thermal data with an end-to-end two stream CNN network which can detect debris such as plastic debris, toxic metals, etc., autonomously.

The rest of the paper is organized as follows: few relevant research works are mentioned in section II. Section III provides a brief overview of our work including the major

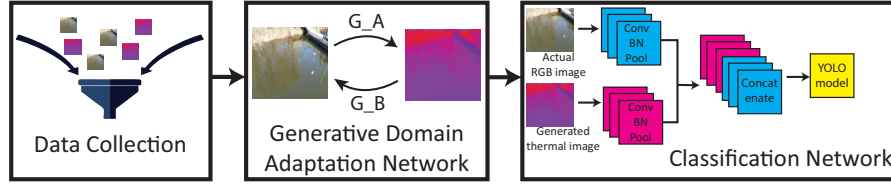


Fig. 1. Overview of the GADAN

contributions of the work. In section IV, we discuss debris data collection processes along with the definition of debris in our work. We describe the proposed methodology in section V in detail and present the experimental results with ablation studies in section VI. Finally, we conclude the article, outlining some potential difficulties in developing this process in section VII.

## II. RELATED WORK

In modern days of technology, high resolution drone images are being used for many creative purposes, such as, surveys on marine erosion [14], [15], wildlife studies [16], [17], post flood debris assessment [18]–[21], beach waste observation [22], for monitoring diverse river environments [23]–[25]. Image processing techniques are embedded with machine learning algorithms in several works to improve the accuracy of detection [26]–[28]. However, deep learning based debris detection systems are introduced in several recent studies. Ramalingam et al. proposed a cascaded model consisting of CNN and SVM to detect solid and liquid spill for autonomous floor cleaning robots [29]. Harmful objects on airfield pavements were detected by incorporating region proposal network and spatial transformer network with CNN classifier [30]. Another model for counting and segmenting woody debris including tree stumps in forest area using UAV images was developed using faster R-CNN [31]. In the majority of these works, debris detection systems were developed for a specific environment and mostly used only RGB images. In several recent works, it has been shown that fusing thermal image RGB improves the accuracy of detection [32]–[35]. However, incorporating a thermal camera into a drone will be an additional cost. Therefore, we foresee opportunities to generate thermal images from the RGB images and the incorporate that generated thermal images alongside. In several recent works, researchers are transferring knowledge from one domain to other domain [36], [37]. Inspired from these works we have developed the generative domain adaptation network that will generate thermal image from RGB image.

## III. OVERVIEW

We discuss the overall framework into three parts: data collection, generative domain adaptation, and predictive learning network. We collected RGB and thermal data using a drone, annotated the images with different types of debris. In the second phase, we address the challenge of incorporating both RGB and thermal images. We discuss the generated domain

adaptation network based model to generate RGB to thermal images and vice versa in the second step. In our experimentation, we generated the thermal image using generative domain adaptation technique from the RGB images collected from drone. Finally, we designed a generalized predictive network to detect the debris irrespective of any environment. Our two-stream predictive network consists of the CNN and YOLO model. The overall framework takes RGB images as input and generated thermal images, and used both the RGB and thermal images to predict the types of debris along with the bounding boxes. Our overall framework is depicted in figure 1. The major contributions of this work are as following.

- 1) We designed a generalized framework based on generative domain adaptation technique to detect debris irrespective of any environment.
- 2) We collected a novel dataset consisting of debris from land and aquatic environment using unmanned aerial system.
- 3) We demonstrated the benefits of incorporating both the thermal and RGB images for debris detection by performing ablation studies on RGB and thermal images separately.

## IV. DATA ACQUISITION

We captured drone videos from two different environments i.e., land and aquatic. RGB and thermal image pairs from the two different environments are shown in figure 2. Aquatic environment is mentioned by *Environment A* while *Environment B* denotes debris on building roofs and construction sites near water sources. We included this *Environment B* because there might be construction work beside rivers, and the pollutants generated that work can pollute the water. For data collection, we used Parrot Anafi Thermal drone which is equipped with two high precision cameras i.e., one FLIR radiometric thermal camera and one HDR camera mounted with 180° tilt 3-axis gimbal. The HDR camera has a 1/2.300 CMOS sensor with a lens with 20 mm focal length. The drone is also equipped with GPS/ GLONASS positioning systems, barometer and inertial measurement unit. The RGB and thermal videos are captured by the drone from 0.3 m – 3m distance from the objects. Under normal conditions, the smart battery provides approximately 23 min of flight time. We collected data mostly during daytime (i.e., between 9am to 6pm) for several days during Spring. The wind velocity was in between 1 mph to 2 mph, and temperature of the surrounding was in between 40°F to 65°F. All data were collected from

a designated area at the University of Maryland Baltimore County campus.

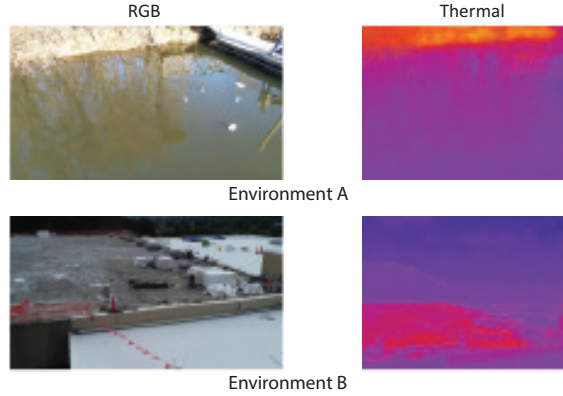


Fig. 2. Sample images of the dataset

The size of the RGB images is  $1920 \times 1080$  pixels, while the size of the thermal image is  $1440 \times 1080$  pixels. As shown in figure 2, the scene captured by RGB images are not same as the thermal images and vice versa. Due to slight differences in RGB and thermal camera placement in the drone, this spatial mismatch happens between RGB and thermal images. The frequency of RGB videos is 30 fps, whereas the frequency for the thermal videos is captured at 15 fps. We trained our proposed model GADAN with 500 pairs for training and 100 pairs for testing. In this pilot project, our data collection was limited to the designated area inside the campus due to permission oriented issues. We plan to increase the size of the dataset with diverse locations in future.

We annotated the dataset using the CVAT application. Any item that is not belong to the natural state of the environments is considered debris. For example, plastic bags, empty cans, or pieces of separate bricks lying on the floor, etc. are labeled as debris. We annotated four types of debris, such as, plastic debris, metal debris, glass debris, and other debris.

## V. METHODOLOGY

The processing pipeline of the GADAN network consists of two parts, thermal image processing by GAN and debris tracking by Yolo model.

### A. Thermal Image Processing

The core principle of our approach is to incorporate information from both the thermal images with RGB images. As there exists dimension and pixel mismatches between RGB and thermal images, we performed transformation on the RGB and thermal images to get the same sized images with common pixels from both type of images. Notable attempts have been made based on generative adversarial networks (GAN) to address these issues between unpaired images [38]–[40]. Following the literature, we implemented the GAN based unpaired image to image translation using cycle consistency loss presented in Eq.1 [41].

$$\mathcal{L}_{ccl} = \|Gen_{T \rightarrow RGB}(Gen_{RGB \rightarrow T}(X_{RGB})) - X_{RGB}\|_1 + \|Gen_{RGB \rightarrow T}(G_{T \rightarrow RGB}(X_T)) - X_T\|_1 \quad (1)$$

$X$  denotes the spatial information of the RGB and thermal images. GAN usually consists of a generator and a discriminator network. The generator consists of Encoder, transformer and decoder, that takes RGB images and transforms the RGB image into thermal image so that it looks like thermal target images. While the discriminator is basically a patchGAN that provides feedback to generator by classification the generated images as fake or real. From this feedback, generator improves its image generation quality. This loop continues until both generator and discriminator are trained well.

### B. Classification Network Design

The classification model consists of a two-stream network. In each stream, RGB and thermal images pass through several convolution, max pooling, and batch normalization layers. Convolution and batch normalization operations were performed using the Eq 2 and 3.

$$X_{conv}^{(i)} = \sum_{i=1}^n (x^{(n)} \times w^{(k)}) + b \quad (2)$$

$$X_{BN}^{(j)} = \frac{(X_{conv}^{(i)} - \mu)}{\sqrt{\sigma^2 + \xi}} \quad (3)$$

Here,  $\mu$  and  $\sigma$  are the mean and variance of the batch. There are five convolution layers and four max pooling layers before concatenating the features from RGB and thermal input stream. Next, we pass the concatenated output to the YOLO model which is able to predict the bounding boxes and the types of the debris.

YOLO [42] is a powerful algorithm for object detection with real-time efficiency on a GPU device. Even though YOLOv3-spp is much more detailed, it is very slow, even on GPU, for real-time processing. We therefore chose to use YOLOv3-tiny, which guarantees to optimize both precision and running time. The YOLO-Tiny [43] edition predicts bounding boxes using anchor boxes and multi-label classification, so we can detect separate bounding boxes for the same object, or we can classify the same object into two distinct classes (with different scores). If an object is observed twice in the current picture, we filter the corresponding bounding boxes to acquire a single specified object.

The network architecture for our proposed model is showed in Fig 3. The whole model has been trained for 100 epochs, where in each batch we have provided 16 images. Learning rate, momentum and decay of the model are 0.001, 0.9, and 0.0005, respectively.

## VI. RESULTS AND ANALYSIS

The generated thermal images and corresponding RGB images are shown in figure 4. We generated thermal images from the RGB images based on the provided actual thermal

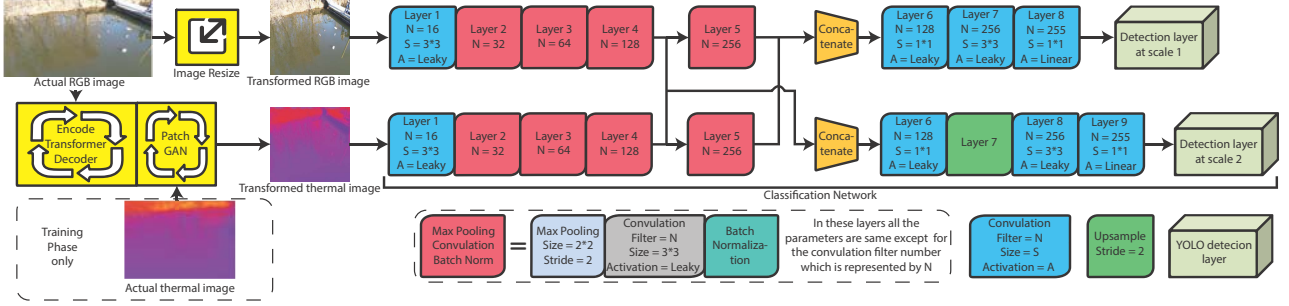


Fig. 3. Flow diagram of the proposed GADAN network

TABLE I  
MODEL PARAMETERS

Layers	Input RGB	Input Thermal	Parameter	Activation
1	Conv	Conv	Filter 16, Size 3*3	Leaky
2	Pool	Pool	Size 2*2, Stride 2	
3	Conv	Conv	Filter 32, Size 3*3	Leaky
4	Pool	Pool	Size 2*2, Stride 2	
5	Conv	Conv	Filter 64, Size 3*3	Leaky
6	Pool	Pool	Size 2*2, Stride 2	
7	Conv	Conv	Filter 128, Size 3*3	Leaky
8	Concatenate layer 7			
9	Route to layer 7			
10	Pool	Pool	Size 2*2, Stride 2	
11	Conv	Conv	Filter 256, Size 3*3	Leaky
12	Concatenate layer 11			
13	Conv		Filter 128, Size 1*1	Leaky
14	Conv		Filter 256, Size 3*3	Leaky
15	Conv		Filter 255, Size 1*1	Linear
16	Yolo Layer			
17	Route to layer 12			
18	Conv		Filter 128, Size 1*1	Leaky
19	Upsample		Stride 2	
20	Concatenate layer 19 and 8			
21	Conv		Filter 256, Size 3*3	Leaky
22	Conv		Filter 255, Size 1*1	Linear
23	Yolo Layer			

images. The generated thermal images look similar to the original thermal images. The resolution of generated thermal image is  $256 \times 256$ . Therefore, we have also converted the RGB into  $256 \times 256$ .

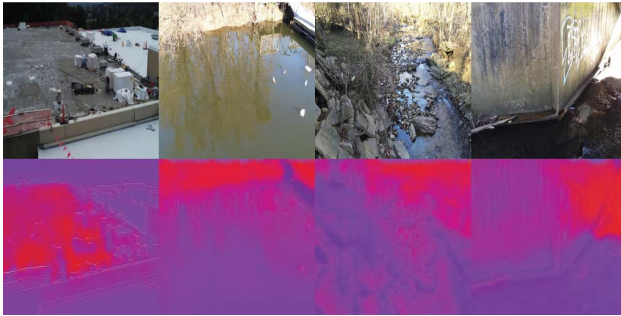


Fig. 4. Output generated thermal image

We evaluated the efficiency of the network on two environments both separately and jointly. For joint experiment with both environments, we have combined all data for training and testing. However, as evaluation metrics, we have calculated the average precision (AP) and mean average precision (MAP). Besides, in order to make an estimate of the precision, we have considered the IOU threshold at 0.5. Then we have determined the AP and MAP using the Eq 4 and 5

$$AP = \frac{1}{I} \sum_{r \in [0,1]} p_{inter_p}(r) \quad (4)$$

$$MAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (5)$$

Here,  $r$  is the recall value of precision-recall (PR) curve,  $p_{inter_p}$  is the interpolated precision value estimated in each point of  $r$ ,  $I$  is the total point considered in the PR curve and  $N$  is the number of the classes, where  $N = 4$  in Environment A and where  $N = 3$  in Environment B.

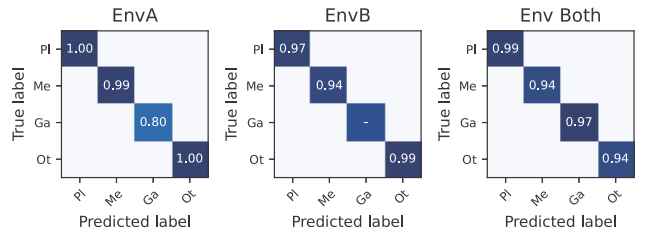


Fig. 5. Average precision for debris detection with GADAN

In figure 5, we presented AP of GADAN for predicting different types of debris namely, plastic(PI), metal (Me), glass (Ga), others (Ot). We used both RGB and thermal images for the two stream network and present the results for three different environmental settings (i.e., Env A, Env B, Both). For the Env B, AP for glass debris is empty, because of no presence of glass on that environment. However, on testing dataset, fusion of RGB and thermal images achieves better AP (i.e., approximately 0.96) on average when both environments are considered and provides approximately similar MAP for all three environments. Though in the marine environment, it is difficult to track the glass debris, the fusion model performed



TABLE II  
RESULT COMPARISON OF GADAN WITH OTHER NETWORKS

	EnvA	EnvB	Both
YOLO Model with generated thermal image only	0.86	0.90	0.86
YOLO Model with RGB only	0.95	0.93	0.93
Single stream YOLO model with RGB and generated thermal image	0.93	0.90	0.91
Two stream RCNN model with RGB and generated thermal image	0.95	0.96	0.95
Two stream YOLO model with RGB and generated thermal image (GADAN)	0.96	0.97	0.97

well compared to only RGB or thermal based network with a mean average precision of 0.97. Considering only RGB or thermal images from both environments provides lower AP and MAP for detecting glass and metal debris compared to plastic and other debris. We also performed ablation study by using different settings and the results are mentioned in table II. From the table it is visible that GADAN outperforms other networks.

Some samples of debris detection using our proposed model is demonstrated in figure 6.



Fig. 6. Visual output results of the model

## VII. CONCLUSIONS

In this work, we present a drone based system to track debris using RGB and thermal cameras. We generated the thermal image from the corresponding RGB image, and designed the debris detection model based on YOLO model both in land and water. Although our model with the fusion of RGB and thermal images outperforms the model developed using only RGB or thermal images, it has few limitations. In the marine

environment, the network finds it difficult to track the glass debris. Because of drone flying restriction rules, we were unable to collect data from lot of places and during night time. As future extension of this work, we plan to explore more types of debris from wide variety of land and aquatic environments and design deeper network.

## ACKNOWLEDGMENT

This research is partially supported by the NSF CAREER Award # 1750936, ONR under grant N00014-18-1-2462, and U.S. Army grant W911NF2120076.

## REFERENCES

- [1] L. Milios, L. H. Christensen, D. McKinnon, C. Christensen, M. K. Rasch, and M. H. Eriksen, "Plastic recycling in the nordics: A value chain market analysis," *Waste Management*, vol. 76, pp. 180–189, 2018.
- [2] R. Geyer, J. R. Jambeck, and K. L. Law, "Production, use, and fate of all plastics ever made," *Science advances*, vol. 3, no. 7, p. e1700782, 2017.
- [3] T. S. Galloway, M. Cole, and C. Lewis, "Interactions of microplastic debris throughout the marine ecosystem," *Nature Ecology & Evolution*, vol. 1, no. 5, pp. 1–8, 2017.
- [4] J. Wang, Z. Tan, J. Peng, Q. Qiu, and M. Li, "The behaviors of microplastics in the marine environment," *Marine Environmental Research*, vol. 113, pp. 7–17, 2016.
- [5] J. M. Coe and D. Rogers, *Marine debris: sources, impacts, and solutions*. Springer Science & Business Media, 2012.
- [6] C. A. Ribic, T. R. Dixon, and I. Vining, "Marine debris survey manual," 1992.
- [7] I. Kiessling, "Finding solutions: derelict fishing gear and other marine debris in northern australia, report prepared for the national oceans office for the key centre for tropical wildlife management," *Charles Darwin University, Darwin*, 2003.
- [8] H. Otley and R. Ingham, "Marine debris surveys at volunteer beach, falkland islands, during the summer of 2001/02," *Marine Pollution Bulletin*, vol. 46, no. 12, pp. 1534–1539, 2003.
- [9] K. Edyvane, A. Dalgetty, P. Hone, J. Higham, and N. Wace, "Long-term marine litter monitoring in the remote great australian bight, south australia," *Marine Pollution Bulletin*, vol. 48, no. 11-12, pp. 1060–1075, 2004.
- [10] C. J. Moore, S. L. Moore, M. K. Leecaster, and S. B. Weisberg, "A comparison of plastic and plankton in the north pacific central gyre," *Marine pollution bulletin*, vol. 42, no. 12, pp. 1297–1300, 2001.
- [11] I. O. Commission *et al.*, "Unep/ioc guidelines on survey and monitoring of marine litter," 2009.
- [12] P. Liu, A. Y. Chen, Y.-N. Huang, J.-Y. Han, J.-S. Lai, S.-C. Kang, T.-H. Wu, M.-C. Wen, M.-H. Tsai *et al.*, "A review of rotorcraft unmanned aerial vehicle (uav) developments and applications in civil engineering," *Smart Struct. Syst.*, vol. 13, no. 6, pp. 1065–1094, 2014.
- [13] M. Valdenegro-Toro, "Deep neural networks for marine debris detection in sonar images," *arXiv preprint arXiv:1905.05241*, 2019.
- [14] B. Chen, Y. Yang, H. Wen, H. Ruan, Z. Zhou, K. Luo, and F. Zhong, "High-resolution monitoring of beach topography and its change using unmanned aerial vehicle imagery," *Ocean & Coastal Management*, vol. 160, pp. 103–116, 2018.

- [15] I. L. Turner, M. D. Harley, and C. D. Drummond, "Uavs for coastal surveying," *Coastal Engineering*, vol. 114, pp. 19–24, 2016.
- [16] L. F. Gonzalez, G. A. Montes, E. Puig, S. Johnson, K. Mengersen, and K. J. Gaston, "Unmanned aerial vehicles (uavs) and artificial intelligence revolutionizing wildlife monitoring and conservation," *Sensors*, vol. 16, no. 1, p. 97, 2016.
- [17] G. P. J. IV, L. G. Pearlstine, and H. F. Percival, "An assessment of small unmanned aerial vehicles for wildlife research," *Wildlife society bulletin*, vol. 34, no. 3, pp. 750–758, 2006.
- [18] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. R. Murphy, "Floodnet: A high resolution aerial imagery dataset for post flood scene understanding," *IEEE Access*, vol. 9, pp. 89 644–89 654, 2021.
- [19] A. Sarkar and M. Rahnemoonfar, "Visual question answering: A deep interactive framework for post-disaster management and damage assessment," *UMBC Student Collection*, 2021.
- [20] —, "Uav-vqg: Visual question generation framework on uav images," in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2021, pp. 4211–4219.
- [21] —, "Vqa-aid: Visual question answering for post-disaster damage assessment and analysis," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 8660–8663.
- [22] C. Martin, S. Parkes, Q. Zhang, X. Zhang, M. F. McCabe, and C. M. Duarte, "Use of unmanned aerial vehicles for efficient beach litter monitoring," *Marine pollution bulletin*, vol. 131, pp. 662–673, 2018.
- [23] D. Bloom, P. A. Butcher, A. P. Colefax, E. J. Provost, B. R. Cullis, and B. P. Kelaher, "Drones detect illegal and derelict crab traps in a shallow water estuary," *Fisheries Management and Ecology*, vol. 26, no. 4, pp. 311–318, 2019.
- [24] D. Chabot, C. Dillon, O. Ahmed, and A. Shemrock, "Object-based analysis of uas imagery to map emergent and submerged invasive aquatic vegetation: a case study," *Journal of Unmanned Vehicle Systems*, vol. 5, no. 1, pp. 27–33, 2016.
- [25] M. A. Ezat, C. J. Fritsch, and C. T. Downs, "Use of an unmanned aerial vehicle (drone) to survey Nile crocodile populations: a case study at lake nyamithi, ndumo game reserve, south africa," *Biological Conservation*, vol. 223, pp. 76–81, 2018.
- [26] M. B. Lyons, K. J. Brandis, N. J. Murray, J. H. Wilshire, J. A. McCann, R. T. Kingsford, and C. T. Callaghan, "Monitoring large and complex wildlife aggregations with drones," *Methods in Ecology and Evolution*, vol. 10, no. 7, pp. 1024–1035, 2019.
- [27] B. Kellenberger, D. Marcos, S. Lobry, and D. Tuia, "Half a percent of labels is enough: Efficient animal detection in uav imagery using deep cnns and active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9524–9533, 2019.
- [28] M. E. Andrew and J. M. Shephard, "Semi-automated detection of eagle nests: an application of very high-resolution image data and advanced image analyses to wildlife surveys," *Remote Sensing in Ecology and Conservation*, vol. 3, no. 2, pp. 66–80, 2017.
- [29] B. Ramalingam, A. K. Lakshmanan, M. Ilyas, A. V. Le, and M. R. Elara, "Cascaded machine-learning technique for debris classification in floor-cleaning robot application," *Applied Sciences*, vol. 8, no. 12, p. 2649, 2018.
- [30] X. Cao, P. Wang, C. Meng, X. Bai, G. Gong, M. Liu, and J. Qi, "Region based cnn for foreign object debris detection on airfield pavement," *Sensors*, vol. 18, no. 3, p. 737, 2018.
- [31] L. Windrim, M. Bryson, M. McLean, J. Randle, and C. Stone, "Automated mapping of woody debris over harvested forest plantations using uavs, high-resolution imagery, and machine learning," *Remote Sensing*, vol. 11, no. 6, p. 733, 2019.
- [32] C. Li, D. Song, R. Tong, and M. Tang, "Illumination-aware faster r-cnn for robust multispectral pedestrian detection," *Pattern Recognition*, vol. 85, pp. 161–171, 2019.
- [33] H. Zhou, M. Sun, X. Ren, and X. Wang, "Visible-thermal image object detection via the combination of illumination conditions and temperature information," *Remote Sensing*, vol. 13, no. 18, p. 3656, 2021.
- [34] A. Banuls, A. Mandow, R. Vázquez-Martín, J. Morales, and A. García-Cerezo, "Object detection from thermal infrared and visible light cameras in search and rescue scenes," in *2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 2020, pp. 380–386.
- [35] P. Shaniya, G. Jati, M. R. Alhamidi, W. Caesarendra, and W. Jatmiko, "Yolov4 rgbt human detection on unmanned aerial vehicle perspective," in *2021 6th International Workshop on Big Data and Information Security (IWBIS)*. IEEE, 2021, pp. 41–46.
- [36] Y.-H. Kim, U. Shin, J. Park, and I. S. Kweon, "Ms-uda: Multi-spectral unsupervised domain adaptation for thermal image semantic segmentation," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6497–6504, 2021.
- [37] R. Zhang, J. Bin, Z. Liu, and E. Blasch, "Wggan: A wavelet-guided generative adversarial network for thermal image translation," in *Generative Adversarial Networks for Image-to-Image Translation*. Elsevier, 2021, pp. 313–327.
- [38] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2849–2857.
- [39] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789–8797.
- [40] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. Singh, and M.-H. Yang, "Diverse image-to-image translation via disentangled representations," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 35–51.
- [41] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [42] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [43] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.