Evaluation of SSD Architecture for Small Size Object Detection: A Case Study on UAV Oil Pipeline Monitoring

Annisa Istiqomah Arrahmah *, Rissa Rahmania, and Dany Eka Saputra

Computer Science Department, School of Computer Science, Bina Nusantara University Bandung Campus, Jakarta, Indonesia; Email: rissa.rahmania@binus.ac.id (R.R.), dany.eka@binus.ac.id (D.E.S.)
*Correspondence: annisa.arrahmah@binus.ac.id (A.I.A.)

***Abstract*—Oil pipeline monitoring using Unmanned Airborne Vehicles (UAV) can be done by utilizing Deep Learning. Deep Learning can be used to automatically detect harmed or unauthorized objects near the pipeline for further action by the authority. Input video in the pipeline area taken from the UAV has unique characteristics. It has low resolution with dense composition object in the image. The detected object also has a small scale as the objects are far away from the UAV. Thus, the selection of the Deep Learning algorithm is important to get a desirable result with the following conditions. Single Shot Multi-Box (SSD) is one of the popular Deep Learning algorithms with fast calculation compared to others and suitable for real-time object detection. Previous works on this topic using low to medium altitude dataset (20–200 m). This paper provides an evaluation of SSD implementation to detect vehicles on high-altitude dataset (300 m). As much as 2482 dataset is fed into SSD architecture and trained to detect 3 class of vehicles. The result shows the mAP and mAR are 0.026360 and 0.067377, respectively. However, the low lost function value shows that the model is able to classify the object correctly. In conclusion, the SSD cannot process low density information to correctly locate the object.**

***Keywords*—oil pipeline monitoring, Unmanned Airborne Vehicles (UAV), deep learning, object detection, Single Shot Multi-Box (SSD) architecture**

I. Introduction

The transportation of midstream products in oil and gas company through land mostly uses pipelines because it is more convenient compared to shipping by railroad and highway. These big pipelines normally pass through public space [1], and consequently are highly vulnerable to risks that lead to equipment failure or loss, yielding to transmission disruption. Equipment failure or loss can be caused by various reasons including natural disaster, over-age structure or even third-party interference [2]. Loss by third party interference can be accidental or even worse, intentional, causing huge loss of oil products. These risks can potentially happen not only in several points but can be spread along the entire length of the pipeline which is thousands of kilometers extends from reachable and unreachable areas. Therefore, it is necessary to monitor and undertake surveillance of the pipeline and surrounding condition regularly for any possible security incident such as oil-theft.

Conventional pipeline inspection is commonly performed by the pipeline operator through Line Walk (on foot), survey, or by air (using helicopter). Security patrols and camera installation on several points of the pipeline are also added for the video surveillance monitoring process. Hence, the surveys and surveillance methods can be exhaustive, slow, costly and time consuming [3]. There are several studies that propose another method for pipeline monitoring to overcome the lack of conventional method by utilizing Internet of Things [4–7] or the usage of Unmanned Airborne Vehicles (UAV) [8–12]. Monitoring pipeline by IoT (Internet of Things) mostly utilizes sensor and camera that installed around the pipeline, thus, do not present the complete picture of the pipeline while monitoring by UAV can cover all condition with several advantages [3].

Oil pipeline monitoring by using UAV can be done by utilizing Deep Learning. Deep Learning can be used to automatically detect harmed or unauthorized objects near the pipeline for further action by the authority. Input video in the pipeline area taken from the UAV also has unique characteristics. It has low resolution image with dense object composition. The object also has a small scale because the video is taken from the UAV camera, while the UAV usually flies quite far from the ground. The selection of the Deep Learning algorithm is important to get a good result. Deep learning approach for object detection in aerial scenes captured by drones has been widely studied with specific challenge based on its application [13, 14].

However, all previous research on this subject is conducted on low to medium height (20–200 m). The works of Meng *et al.* [11] is conducted using dataset collected at the height of 100–200 m. While the actual implementation of pipeline surveillance could be conducted at high-altitude (around 300 m and above). The

Manuscript received May 30, 2023; revised July 21, 2023; accepted August 31, 2023.

difference in altitude and its effect on the quality of object detection needs to be investigated further.

This work evaluates the accuracy of existing object detection model on high-altitude. This works uses case study on detecting the physical presence and tampering possibility at pipeline as one of the critical challenges in the distribution sector of the oil and gas company [15]. A video stream from a high-altitude UAV equipped with camera is obtained as the input with a low-resolution image and small yet dense objects are being captured. This works uses Single Shot Multi-Box (SSD) algorithm, since it is a lightweight object detection algorithm that suited for real-time detection in the operation site.

Based on that purpose, this work provides an evaluation of high-altitude object detection as its novelty, different on the previous work's medium-altitude object detection. Accuracy and loss metric is used to evaluate the result of object detection using the selected algorithm.

This rest of paper is divided into the following sections: Section II provides the related works for this research. Section III discusses the method used in this research. Section IV shows the evaluation result and discussion. The conclusion is given in Section V.

II. Literature Review

A specific application of object detection for pipeline monitoring using UAV also has its own challenge caused by the native of the environment, the object that wants to be detected and the type of UAV used to capture the video stream and the resolution of the image itself [16]. There are several studies that have been conducted for UAV based object detection, especially for pipeline monitoring. Ammour *et al.* [17] uses Deep Convolutional Neural Network (CNN) and Support Vector Machine (SVM) to detect cars from the medium-altitude UAV with a wide range application. Ammour *et al.* uses images taken from high resolution cameras mounted on the UAV. Jiao *et al.* [10] focus on detecting the oil spill while monitoring the pipeline using UAV. Deep CNN and Otsu algorithm are being used for the detection algorithm. Ukaegbu *et al.* [18] uses CNN to detect pipeline leakage and non-authorized employer or people without PPE. The paper focuses on the UAV system while the architecture of the Deep Learning (DL) is not deeply explained. Gleason *et al.* [12] use classic Machine Learning (ML) classification algorithms to detect vehicles near the pipeline to replace the current pipeline patrol. Four different algorithms are being compared to get the best result. The area coverage includes vegetation, rural roads, buildings, lakes, and rivers. The high-altitude UAV are being used to take the surveillance and a wide range of vehicles being detected from excavators to a private car. The accuracy within 85% vehicle detection rate. Meng *et al.* [11] already use Deep learning one stage object detector YOLOv3 to detect threat vehicles near the pipeline. A medium altitude UAV with 100–200 m height range is being used and the threat vehicle specified to excavator. The model is tested on the China's oil pipeline with maximum 99.4% recall rate.

In this paper, a video around pipeline in rural area, Indonesia is gathered using a high altitude UAV with a constant height of 300 m is used as an input. The video has a low resolution to detect three main unauthorized vehicles: car, truck, and motorcycle. Because the height of the UAV is constant, the size of the bounding box for each class tends to remain unchanged while each class has a different size of the bounding box. Therefore, Single Shot Multi-Box (SSD) neural network is a suitable solution for the condition. SSD is one of the latest deep learning methods and is known to be the simultaneous detection of objects with various sizes [19]. Compared to other algorithms, SSD has claimed with better performance while using lower resolution images [20]. SSD is also designed for object detection in real-time. SSD speeds up the process by eliminating the need for the region proposal network, thus it has fast processing compared to others.

III. Materials and Methods

Deep learning-based monitoring for the oil pipeline tampering is done by using deep learning method. Thus, the research is conducted based on deep learning process, which consists of data preparation, training/modeling, and testing shown in Fig. 1. The data collected from the UAV is preprocessed during the data preparation phase for the next step. In the training process, the training data is used to get the predictive model based on the selected deep learning algorithm. During the testing phase, the result from the training phase is used on the testing data to determine the algorithm's performance. The details of each phase are explained in the next step.

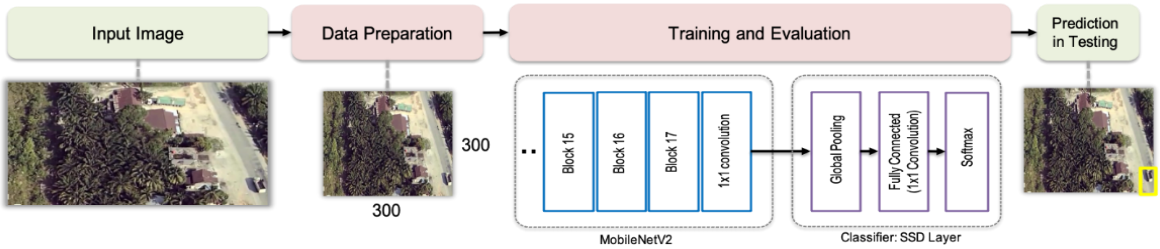


Figure 1. The research method.

A. Data Preparation Phase

Several videos from an UAV (quadcopter type UAV) with 300 m altitude (high altitude) are taken along the oil pipeline on the rural area in Indonesia as the raw data for the training and testing phase. The environment of the area near pipeline are buildings, common road, rural road and

vegetations. The dataset consists of 2482 images acquired by taking the video frames from the raw videos. This custom dataset is configured by using COCO configuration. Example of the dataset can be seen in Fig. 2. Object labelling is conducted for each image in the dataset. The labels are "mobil" (car), "motor" (motorcycle) and "truk" (truck). These three objects are identified as potential unauthorized objects near the pipeline that have the possibility to tamper the pipeline with the high impact. The Region of Interest (ROI) is also determined to each image as the detected object only valid as the potential unauthorized vehicle if the existence of the object is in the area near the pipeline. The dataset is then divided into 65% for training, 15% for validation and 20% for testing. The ratio between each class label of the dataset can be seen in Fig. 3. The Imbalance Ratio (IR) is 3 and categorized as moderate data imbalance [21].

Figure 2. The example of the dataset.

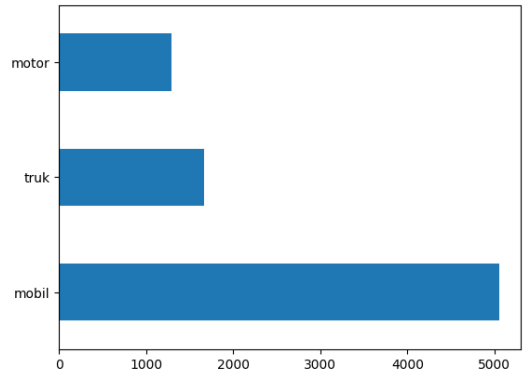


Figure 3. Class label composition.

B. Training and Evaluation Phase

The training dataset has a unique characteristic because the desired detected object comes from aerial images which make the object become small and dense. Single Shot Multi-Box Detector (SSD) is chosen as the deep learning algorithm for vehicle detection with the consideration mentioned in Section II. The SSD for real time object detection has two components, namely backbone and head SSD [20]. The backbone model is the pre-trained image classification network as a feature extractor. In this research, MobileNet V2 [22] is used for the feature extractor while the original SSD using VGG-16 feature extractor [23]. The backbone result is 320×320 feature maps. The SSD head is the convolutional layer added at the end of the backbone for object detection. The architecture can be seen in Fig. 4 [24].

For the training stage, the batch size and epoch used in this research are 24 and 50,000 respectively as the validation result shows the following configuration gives the highest accuracy value. The result of the training phase is the predictive model for the testing phase. During the evaluation phase, loss and precision information are gained for analysis purposes.

C. Testing Phase

During the testing phase, another video from the high altitude UAV near the pipeline in rural area is taken and an inference process is done to the tested video. A predictive model gained from the training phase is used to identify the unauthorized vehicles near the pipeline. If the presence of the potential object is detected, a bounding box is formed around the object along the pipeline for further checking by the authority. In this phase, the prediction result of the undesired vehicles (car, motorcycle, or truck) near the pipeline is also obtained.

D. Metric Performance

In measuring the object detection model performance, mean Average Precision (mAP) and mean Average Recall (mAR) can be used. The mAP and mAR are calculated by finding the Average Precision for each class and then averaging it for all class in the dataset, following Eqs. (1) and (2), respectively [25].

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \quad (1)$$

$$mAR = \frac{\sum_{i=1}^{k} AR_i}{k} \quad (2)$$

The Average Recall is calculated by averaging recall value over all Intersection of Union $IoU \in [0.5, 1.0]$ and can be computed as two times the area under the recall-IoU curve Eq. (3) [1].

$$AR = 2 \int_{0.5}^{1} recall(o)\, do \quad (3)$$

The Average Precision calculation is similar to Average Recall. It calculates the average precision value for all Intersection of Union Eq. (4) [1].

$$AP = \int_{0}^{1} precision(r)\, dr \quad (3)$$

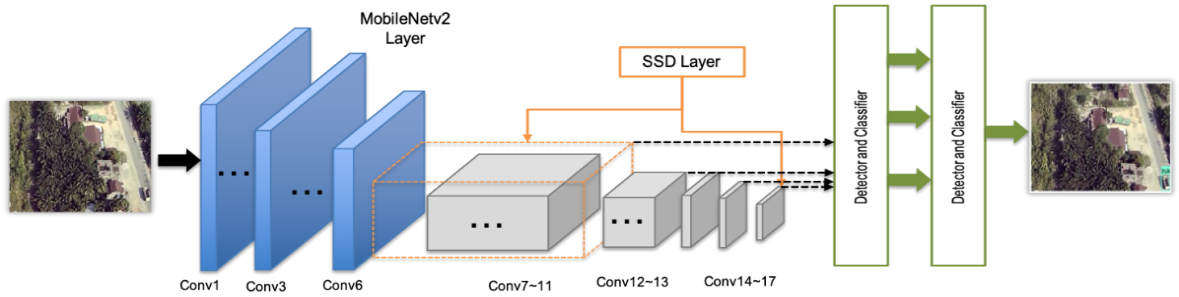


Figure 4. SSD-MobileNetV2 Architecture with 300×300 resolution input image.

IV. Result and Discussion

During the inference process, the example output of the video frames from high altitude UAV can be seen in Fig. 5. The pipeline is located on the right side of the road. An object identified as truck and motorcycle on the road near the pipeline is captured with low probability, while the truck outside the ROI is not identified.

Evaluation of the model performance is conducted using accuracy and loss metrices. Recall and precision are calculated for accuracy. Localization, classification, generalization, and total loss are presented as the loss metrices.

Figure 5. The inference result on the testing video frames.

A. Object Detection Performance

The evaluation metric at step 50,000 can be seen in Table I. The figure contains the whole performance for each IoU (Intersection of Union) threshold. Based on the result, the mAP is 0.026360 and the mAR is 0.067377.

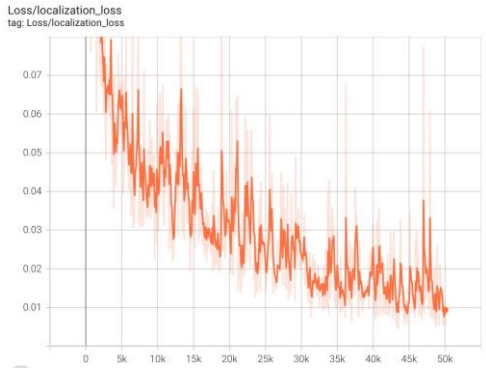
TABLE I. Performance at 5000 Epoch

Average Recall		Average Precision	
$AR^{max=1}$	0.067	AP^{IoU}	0.026
$AR^{max=10}$	0.106	$AP^{IoU=.50}$	0.065
$AR^{max=100}$	0.118	$AP^{IoU=.75}$	0.018
AR^{small}	0.103	AP^{small}	0.019
AR^{medium}	0.101	AP^{medium}	0.028
AR^{large}	−1.000	AP^{large}	−1.000

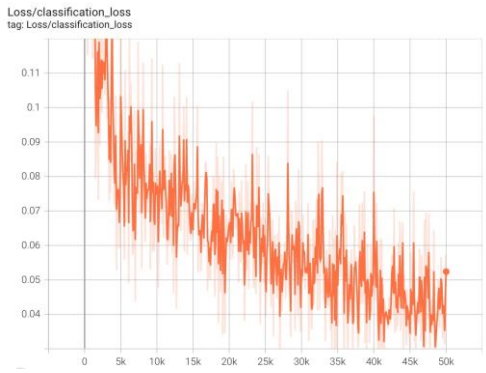
It is stated that the value of mAP and mAR are small, around 2% and 6%. It is also similar with other IoU threshold, the value of mAP and mAR is less than 10%. This also means that the precision and recall for all IoU thresholds are also low. The low precision means that the True Positive (TP) over all positive result is low, caused by a high result of False Positive. The low recall means that the True Positive (TP) over all data that should be predicted as positive is low, caused by a high result of False Negative. The result also means that the algorithm cannot detect the vehicle that exists on the video and sometimes detect a vehicle where in the video the vehicle is not present.

All the loss result can be seen in Fig. 6. Localization loss is the mismatch between the ground truth box with the predicted boundary box. Based on the Fig. 6(a), the loss value on the 5000 step already 0.07 and its value decreases as the number of steps increases. At step 50,000, the value hits 0.01. In this paper, MobileNetV2 is used for the classification process. Based on the Fig. 6(b), the classification loss is also low with 0.11 at 5000 steps and become 0.04 at 50,000 steps. The total loss of the algorithm can be seen in Fig. 6(c). At 5000 steps, the total loss is 0.22 and it decrease and hit 0.1 at 50,000 steps. Based on the results, overall loss value is near zero. While SSD only penalizes prediction from positive matches, the loss result is low. Based on the low loss during training and low accuracy during testing and evaluation, the model is overfitting. This is because there is class imbalance in the dataset obtained from the UAV. If the composition of the class in the dataset is considered, the class imbalance could affect the overfitting. The IR ratio of the dataset is above 1.5 and considered moderate. When the model is trained with imbalanced datasets, it can overfit the training samples from the minority class but not generalize well during the testing [26]. Class imbalance ratio can also be defined as the ratio between the number of pixels of the background and the number of pixels of the ROI in one

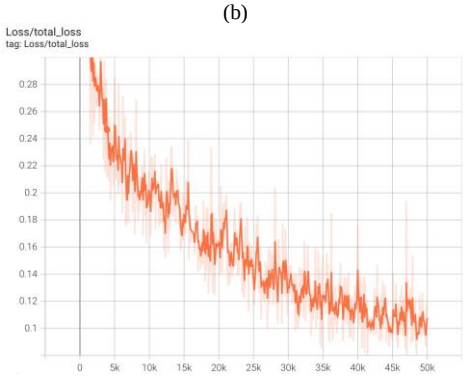
image [26]. Based on Table II, the majority class has 27.8×27.31 pixel, about less than 10% from the total pixel in one image, w.r.t the data set.



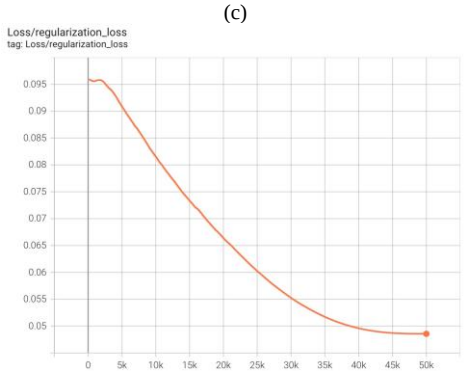
(a)



(b)



(c)



(d)

Figure 6. Loss result: (a) Localization loss (b) Classification loss (c) Total loss (d) Regularization loss.

B. SSD for Small Object and Low Resolution Image

In this paper, dataset is obtained from UAV with average flight altitude 300 m and categorized as high-altitude UAV. All the dataset is used for the training and testing. The paper focuses on evaluating the object detection performance based on SSD-MobileNetV2 architecture while the input image has a low resolution, and the detected objects have a small scale relative to the image size with a complex background which impacts the dataset for training and testing. The resolution of the image generated from the high altitude UAV is 640×360 pixels and resized into 300×300 pixels. For each detection class, the average bounding box for the ground truth can be seen in Table II. This input is also used for testing and evaluation.

TABLE II. OBJECT CLASS RESOLUTION

Size (pixels)	Object Class Average Ground Truth Resolution		
	Truck	**Car**	**Motorcycles**
Width	41.64	27.80	16.80
Height	40.27	27.31	17.28

For comparison, similar paper by Meng *et al.* [11] in Section II uses UAV to detect excavators along the pipeline in China with average range of flight latitude between 100 and 200 meters with the image size of 1920× 1080 pixels and 350 images is used for training. The detailed comparison can be seen in Table III. The dataset is obtained from the internet and from the UAV with composition of 70% and 30% respectively. For the testing and evaluation, the input image is resized into 320×320 and one-stage detector YOLOv3 is used to detect the excavator. However, there is no information about the ground truth scale of the object. The result of this experiment has a high recall with the value of 96.8%.

TABLE III. MODEL COMPARISON

	Meng *et al.* [11]	**Proposed method**
UAV latitude	100–200 m	300 m
Image resolution	1920×1080 pixels	640×360 pixels
Object class	Excavator	Car, Truck, Motorcycle
Object Detector	YOLOv3	SSD-MobileNetV2

Based on the result in Section IV.A, the basic SSD-MobileNetV2 architecture for low resolution and dense image with small scale object has an exceedingly small value of mAP and mAR. If we compare the performance with another one stage detector such as YOLOv3 in paper [11] and only consider the input size that enters the first stage in the detector, the result shows that YOLO is superior to SSD. This result is in line with the statement that the SSD accuracy is inferior to other architecture if the size of the object is small [19]. If the resolution of the detected object is considered, based on Table II, each of the objects occupies much fewer pixels in the image. This is because the vehicles are far enough from the UAV and the video taken from the UAV is also small. Thus, the process during the feature map in SSD makes the object

not detected because SSD uses bigger feature map than other architecture to speed up the process [20].

Based on the result and discussion above, there are two possible reasons why SSD shows limitations while detecting objects with a small scale caused by the altitude of the UAV that captured the image. First, the desired object is difficult to detect, even using the largest output feature map. Based on SSD-MobileNetV2 architecture, the largest output feature maps are scaled at around one-eighth of the input image. For example, the truck class with the largest pixel among all the classes has an average number of only 41.64×40.27 pixels. By using the largest output feature map, resulting only 5.2×5 pixels for the desired object. In addition, the motorcycle class as the smallest class has a result of 2.1×2.16 pixels. Thus, there is insufficient information to classify and locate the desired objects in the image. Second, based on the architecture, SSD model extracts multiple layers to detect and classify the object. During the backbone, subsampling and pooling operation, detailed information of the desired object could be lost and disturbed by the dense type of image [27]. Thus, the feature used to locate the small objects might lack detailed information to accurately localize the objects.

Upon the challenge of small object detection in a low-resolution image and the SSD limitation, future research could be conducted to improve the detection performance. It can be done by exploiting powerful contextual information by enlarging the small region that occupied the small object or modifying the feature extractor process [28]. Yet, the detection speed should also be considered while this option applied in the UAV object detection.

V. Conclusion

The purpose of this study is to evaluate the performance of SSD architecture in detecting objects around pipeline site at high-altitude. The experiment resulted in mean Average Precision and mean Average Recall of 0.026360 and 0.067377, respectively. The results show that the mAP and mAR are low, meaning that the True Positive value over False Positive and False Negative value is low. Thus, the accuracy of SSD architecture is low for low resolution image with small scale object. It is because the SSD architecture uses lower resolution layers and smaller feature maps to detect objects, to speed up the process. Therefore, SSD architecture is not suitable for small-scale object detection in a low-resolution image. Even though the accuracy of SSD architecture is not desirable, the classification, localization, generalization, and total loss have a desirable value, that is close to zero. SSD only penalizes prediction from positive matches; thus, SSD is good enough for localization and classification for the true positive result.

Based on this result, the works conclude that low accuracy result of SSD comes from the low-density information of the detected object. The low-density information reflected in poor performance of detecting the area in which the object resides. This problem might be solved by using image enhancing algorithms, to enrich the information of the image. The work of Truong *et al*. [29] might be used as inspiration to conduct this enhancement. However, further studies about the computational cost of such implementations should be conducted.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Annisa Istiqomah Arrahmah as first author wrote the paper and analyzed the data; Rissa Rahmania as second author conducted the research; Dany Eka Saputra as third author provided the data and became the advisor during the research; all authors had approved the final version.

Funding

This research is conducted under Kedaireka Grant from Indonesia Ministry of Education, Culture, Research and Technology.

Acknowledgment

The authors wish to thank Terra Drone Indonesia for their cooperation in this research to provide the data.

References

[1] A. Bahadori, *Oil and Gas Pipelines and Piping Systems: Design, Construction, Management, and Inspection*, 1st ed. 2017.

[2] I. P. A. Wiguna, N. Anwar, A. Widodo, and N. N. Rodhi, "Risk management effectiveness of oil and gas pipeline construction in Java Island—Indonesia," *IPTEK Journal of Proceedings Series*, vol. 3, no. 6, Dec. 2017. doi: 10.12962/j23546026.y2017i6.3308

[3] S. Marathe, "Leveraging drone based imaging technology for pipeline and RoU monitoring survey," presented at the SPE Symposium: Asia Pacific Health, Safety, Security, Environment and Social Responsibility, Kuala Lumpur, Malaysia, April 2019. doi: 10.2118/195427-MS

[4] S. Ahmed, F. Le Mouel, and N. Stouls, "Resilient IoT-based monitoring system for crude oil pipelines," in *Proc. 2020 7th International Conference on Internet of Things: Systems, Management and Security, IOTSMS 2020*, Dec. 2020. doi: 10.1109/IOTSMS52051.2020.9340197

[5] H. Yu and M. Guo, "An efficient oil and gas pipeline monitoring systems based on wireless sensor networks," in *Proc. 2012 International Conference on Information Security and Intelligent Control*, 2013.

[6] N. F. Henry and O. N. Henry, "Wireless Sensor networks based pipeline vandalization and oil spillage monitoring and detection: main benefits for Nigeria oil and gas sectors," *The SIJ Transactions on Computer Science Engineering & Its Applications*, vol. 3, no. 1, 2015.

[7] J. Sun, Z. Zhang, and X. Sun, "The intelligent crude oil anti-theft system based on iot under different scenarios," *Procedia Comput Sci*, vol. 96, pp. 1581–1588, 2016. doi: 10.1016/j.procs.2016.08.205

[8] F. E. Idachaba, "Monitoring of Oil and gas pipelines by use of VTOL-type unmanned aerial vehicles," *Oil and Gas Facilities*, vol. 5, no. 1, pp. 47–52, Feb. 2016. doi: 10.2118/172471-PA

[9] C. Gómez and D. R. Green, "Small unmanned airborne systems to support oil and gas pipeline monitoring and mapping," *Arabian Journal of Geosciences*, vol. 10, no. 9, p. 202, May 2017. doi: 10.1007/s12517-017-2989-x

[10] Z. Jiao, G. Jia, and Y. Cai, "A new approach to oil spill detection that combines deep learning with unmanned aerial vehicles," *Comput. Ind. Eng.*, vol. 135, pp. 1300–1311, Sep. 2019. doi: 10.1016/j.cie.2018.11.008

[11] L. Meng *et al.*, "Real-time detection of ground objects based on unmanned aerial vehicle remote sensing with deep learning:

Application in excavator detection for pipeline safety," *Remote Sens (Basel)*, vol. 12, no. 1, p. 182, Jan. 2020. doi: 10.3390/rs12010182
[12] J. Gleason, A. V. Nefian, X. Bouyssounousse, T. Fong, and G. Bebis, "Vehicle detection from aerial imagery," in *Proc. 2011 IEEE International Conference on Robotics and Automation*, May 2011, pp. 2065–2070. doi: 10.1109/ICRA.2011.5979853
[13] A. Jain *et al.*, "AI-enabled object detection in UAVs: Challenges, design choices, and research directions," *IEEE Netw*, vol. 35, no. 4, pp. 129–135, Jul. 2021. doi: 10.1109/MNET.011.2000643
[14] N. Dilshad, J. Hwang, J. Song, and N. Sung, "Applications and challenges in video surveillance via drone: A brief survey," in *Proc. 2020 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2020, pp. 728–732. doi: 10.1109/ICTC49870.2020.9289536
[15] W. Z. Khan, M. Y. Aalsalem, M. K. Khan, Md. S. Hossain, and M. Atiquzzaman, "A reliable internet of things based architecture for oil and gas industry," in *Proc. 2017 19th International Conference on Advanced Communication Technology (ICACT)*, 2017.
[16] F. Pasandideh, J. P. J. da Costa, R. Kunst, W. Hardjawana, and E. P. de Freitas, "A systematic literature review of flying ad hoc networks: State-of-the-art, challenges, and perspectives," *J Field Robot*, Jan. 2023. doi: 10.1002/rob.22157
[17] N. Ammour, H. Alhichri, Y. Bazi, B. Benjdira, N. Alajlan, and M. Zuair, "Deep learning approach for car detection in UAV imagery," *Remote Sens (Basel)*, vol. 9, no. 4, p. 312, Mar. 2017. doi: 10.3390/rs9040312
[18] U. F. Ukaegbu, L. K. Tartibu, M. O. Okwu, and I. O. Olayode, "Integrating unmanned aerial vehicle and deep learning algorithm for pipeline monitoring and inspection in the oil and gas sector," in *Proc. the International Conference on Artificial Intelligence and its Applications*, Dec. 2021, pp. 1–6. doi: 10.1145/3487923.3487924
[19] J.-H. Park *et al.*, "Automated identification of cephalometric landmarks: Part 1—Comparisons between the latest deep-learning methods YOLOV3 and SSD," *Angle Orthod*, vol. 89, no. 6, pp. 903–909, Nov. 2019. doi: 10.2319/022019-127.1
[20] W. Liu *et al.*, "SSD: Single shot multi-box detector," in *Computer Vision—ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. 2016, Springer, Cham., vol 9905, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2
[21] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J Big Data*, vol. 6, no. 1, p. 27, Dec. 2019. doi: 10.1186/s40537-019-0192-5
[22] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861, Apr. 2017.
[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, Sep. 2014.
[24] S. Arabi, A. Haghighat, and A. Sharma, "A deep learning based solution for construction equipment detection: from development to deployment," arXiv:1904.09021, Apr. 2019.
[25] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *Proc. 2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Jul. 2020, pp. 237–242. doi: 10.1109/IWSSIP48289.2020.9145130
[26] Z. Li, K. Kamnitsas, and B. Glocker, "Analyzing overfitting under class imbalance in neural networks for image segmentation," *IEEE Trans Med Imaging*, vol. 40, no. 3, pp. 1065–1077, Mar. 2021. doi: 10.1109/TMI.2020.3046692
[27] C. Sun, Y. Ai, S. Wang, and W. Zhang, "Mask-guided SSD for small-object detection," *Applied Intelligence*, vol. 51, no. 6, pp. 3311–3322, Jun. 2021. doi: 10.1007/s10489-020-01949-0
[28] A. M. Roy and J. Bhaduri, "DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism," *Advanced Engineering Informatics*, vol. 56, 102007, Apr. 2023. doi: 10.1016/j.aei.2023.102007
[29] N. Q. Truong, P. H. Nguyen, S. H. Nam, and K. R. Park, "Deep learning-based super-resolution reconstruction and marker detection for drone landing," *IEEE Access*, vol. 7, pp. 61639–61655, 2019. doi: 10.1109/ACCESS.2019.2915944