

IMPROVING THE ACCURACY OF AN OIL SPILL DETECTION AND CLASSIFICATION MODEL WITH FAKE DATASETS

N. A. BUI ^{1*}, Y. G. OH ¹, I. P. LEE ¹

¹ Dept. of Geoinformatics, University of Seoul, Seoul, South Korea
anbn@office.uos.ac.kr, ohrgon@naver.com, iplee@uos.ac.kr

KEY WORDS: Oil spill detection, Generative adversarial networks-GANs, Data augmentation, Dual attention mechanism, Pix2Pix.

ABSTRACT:

Deep learning is a popular tool for object detection, including oil spill detection. However, acquiring sufficient data for training deep learning models can be challenging, particularly for offshore oil spill accidents. Data augmentation is an effective solution to this issue. This study proposes a data augmentation method using a conditional-GAN model, specifically Pix2Pix, to generate dummy datasets of oil spills. These datasets were used to train the DaNet model for oil detection and classification. Results show that using the dummy datasets improves the mIoU and f1-score to 2.56% and 1.69%, respectively, and enhances the accuracy of classifying of each oil in the model. This approach not only improves the accuracy of the deep learning model but also presents a direction for data enhancement in detection or segmentation tasks for formless objects, such as oil spills, cracks, water seepage, and mildew.

1. INTRODUCTION

In the past five years, oil spill detection has become more widespread thanks to advancements in artificial intelligence, particularly deep learning. Research in this field now goes beyond detecting oil spills to classifying the types of oils involved in an accident. This information can help environmental managers respond more quickly and accurately, with greater flexibility in treatment options.

With the increasing computing power of computer systems and supporting libraries like Keras and Caffe, implementing and applying deep learning models is theoretically simple. However, in practice, preparing a dataset for training a deep learning model poses challenges such as insufficient data, imbalanced datasets, and fragmented data (Maharana et al., 2022). Additionally, when the dataset is too small, deep learning algorithms perform poorly (Tyagi et al., 2020).

Data collection for oil spill detection is challenging for two main reasons. Firstly, oil spill accidents resulting from shipwrecks or oil pipeline failures are rare and often occur in offshore locations. Secondly, it is difficult to obtain photographic data during oil spill accidents due to transportation limitations and limited range of the flying vehicles (Benjamin et al., 2023). Although unmanned aerial vehicles-UAVs can assist with data collection, they are limited by battery capacity and communication distance with the remote controller. In addition, oil classification poses further difficulties due to the imbalanced proportion of oils in a given oil crash. As a result, utilizing deep learning for oil detection and classification requires addressing two main issues: lack of data and unbalanced datasets.

Since collection additional data is not feasible; therefore, data augmentation techniques offer a viable solution. Data augmentation can address data scarcity issues, unbalance data and prevent overfitting of models also. These data augmentation techniques can be categorized into three main groups: image level, object level, and pixel level, as discussed in previous studies (Liu et al., 2019; Zhang et al., 2021). The figure 1 illustrates an example of three types of data augmentation.

Image-level augmentation involves geometric transformations that alter the entire image, creating a new image that resembles the real world while being distinct from the original. Examples of commonly used geometric transformations include vertical or horizontal flips, cropping, rotation, affine transformation, so on and so forth.

Object-level data augmentation is a technique that applies geometric transformations to individual objects or a small set of objects within an image to generate a new image (Zhang et al., 2021).

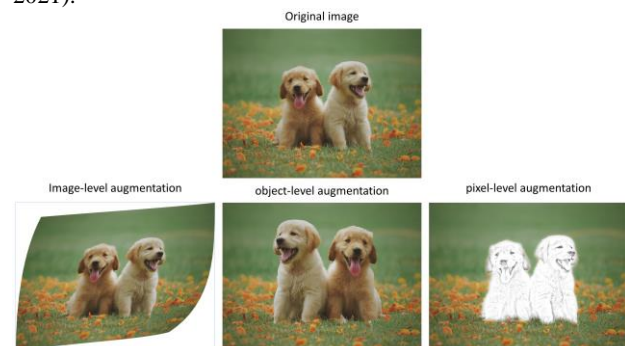


Figure 1. Three types of data augmentation.

Pixel-level data augmentation involves altering the pixel values of an entire image or a group of pixels, often through techniques like data diffusion, image generation, adding noise, and adjusting contrast.

When augmenting data, two factors must be ensured: firstly, the creation of a new image that is both distinct from the original and realistically representative of the real world; secondly, must create of a corresponding annotation image, also known as a mask image. Without a corresponding mask image, the time and cost required for data labelling increases significantly, rendering the augmentation method less practical. For example, while data diffusion is effective in creating realistic new images, but it does not generate corresponding annotation images. Thus, data diffusion-based approach is not very effective. Liu et al. (2019) demonstrated the efficacy of augmenting the Cityscapes dataset

by randomly selecting each object in each class and combining them into a new image, resulting in a 2.1% increase in accuracy. This research employs pixel-level data augmentation to enhance the precision of a deep learning model used for identifying and categorizing oil type in a spill. This technique generates fake oil spill photographic images using a Pix2Pix model based on fake oil spill annotated images. The generation of these annotated images is randomized but selective to ensure a balanced dataset and to improve accuracy evenly. The main contributions of this paper are as follows:

- 1) Proposed a pipeline to train a model for detecting and classifying oil spills, even with an imbalanced dataset;
- 2) Proposed three algorithms to generate a synthetic dataset for objects that lack a fixed shape and scale;
- 3) Improved the accuracy of both mIoU and macro average f1-score for overall performance.

2. LITERATURE REVIEW

This research builds upon the work of Bui et al. (2023) who trained the DaNet model to detect and classify oil spills into four types based on its color such as Black oil, Brown, Rainbow and Silver oils. The objective of this research is to enhance the accuracy of oil type classification.

As mentioned above, data augmentation is considered a promising method for improving the accuracy of deep learning models, and it has been extensively explored for the past six years. Yang et al. (2022) provide an overview of the effectiveness of data augmentation in semantic segmentation. They point out that data augmentation can improve mIoU from 0.53% to 2.61% with ISANet model when experimented on the PASCAL VOC dataset.

In object-level data augmentation, Zhang et al. (2021) applied geometric transformation methods such as rotation, scaling, flipping, and shifting to individual objects in the PASCAL VOC 2012 dataset. The results show that using object-level data augmentation improves the mIoU accuracy by an average of 1.5%, with the highest improvement being 2.4%. Liu et al. (2019) used conditional GANs to create a new dataset based on the Cityscapes dataset. The results were very promising with a 2.1% mIoU increase.

After analyzing the above studies and considering the peculiar nature of oil, which has no fixed shape, this study proposes three algorithms to create annotation images of oil spills that are random and realistic and that also ensure the two basic requirements of data augmentation mentioned above. These algorithms aim to produce annotation images in which the proportion of the each oil shown is relatively equal. Once these annotation images are created, a conditional GAN, specifically pix2pix, will be employed to generate corresponding images that is similar to photographic images.

3. METHODOLOGY

In this part, our proposed pipeline will be demonstrated. This pipeline involves training the DaNet and Pix2Pix models simultaneously using a small dataset. After testing the dataset with DaNet, the results will be used to analyze then generate three dummy datasets using Pix2Pix, each with a different algorithm. These datasets will then be combined with the original data and used to retrain the DaNet model. This process will be repeated until accuracy reaches saturation, as demonstrated in Figure 2.

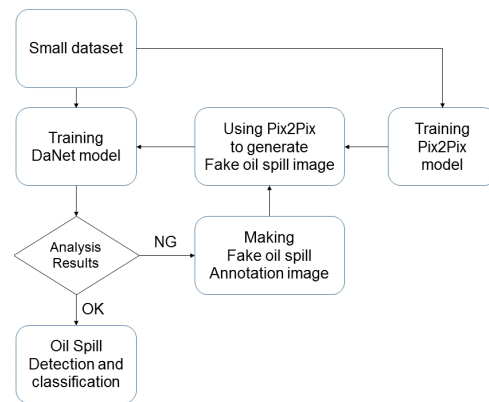


Figure 2. The proposed pipeline for improving the accuracy of the oil classification and detection model.

3.1 Image Annotation

An annotation image is a same-sized image as the original one, used for annotating objects by assigning them the same pixel value and classifying them accordingly. Typically, objects to be detected and classified are referred to as the foreground, while the remaining ones are referred to as the background. Figure 3 illustrates an example of an original image and its corresponding annotation image. Annotate an image of an oil spill is a challenging task. Distinguishing between different types of oils can be difficult, and it becomes even more challenging to separate their boundaries. Additionally, noise elements like sunlight, fog, seaweed, etc. further complicate the task. As aforementioned, annotation a photo is a time-consuming task. To address this issue, this study will take a opposite process by generating fabricated annotation images, which will then be used to create dummy photographic images.

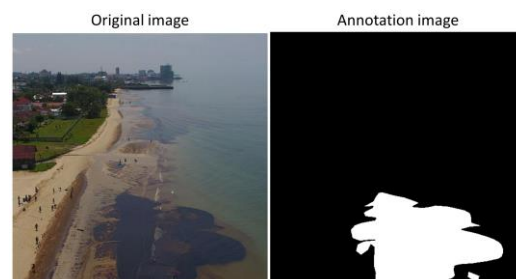


Figure 3. Original image and its annotation image.

3.2 Dataset and Segmentation model

The research dataset comprises aeronautical photographs of oil spill incidents, each with a corresponding annotated image at 512x512 resolution. Figure 4 shows the number of images and the distribution proportion of the oils. The dataset is unbalanced, with an uneven contribution proportion of the oils. Notably, silver oil represents the highest percentage, while brown and black oils account for the lowest percentage.

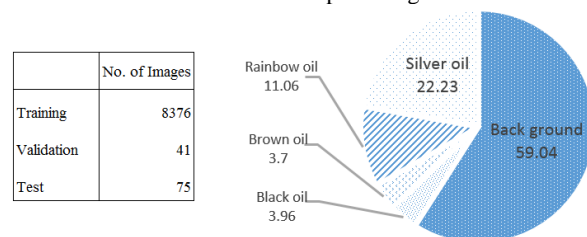


Figure 4. No. of images and proportion in the original dataset.

In their research, Bui et al. (2023) trained the DaNet model to detect and classify oil. Figure 5 shows the correlation between the contribution proportion (the percentage of a class in a dataset) and accuracy for each oil. The model's accuracy generally corresponds to the percentage of each oil. However, the accuracy of silver oil, despite having a high contribution proportion, is the lowest.

3.3 Making fake oil spill annotation images

This research proposes three algorithms for creating dummy oil spill annotation images, which are as follows:

- 1) Increasing the contribution proportion of one or more classes;
- 2) Increasing the contribution proportion of all classes;
- 3) Enhancing the contribution proportion of all classes by generating completely dummy data.

Algorithm 1:
Increasing the contribution proportion of one or more classes.

Require:
All original annotation images (A)

Ensure:
Annotation image size = 512x512

BEGIN

- 1 **for** each image in A **do**
- 2 **extract** desired classes' shapes and positions
- 3 **return** classes dictionary (CD)
- 4 **for** each image in A **do**
- 5 **pick** randomly desired classes (r) in CD
- 6 **paste** r into image
- 7 **return** new annotation image

END

Algorithm 2:
Increasing the contribution proportion of all classes.

Require:
All original annotation images (A)

Ensure:
Annotation image size = 512x512

BEGIN

- 1 **for** each image in A **do**
- 2 **extract** all classes' shapes and positions
- 3 **return** classes dictionary (CD)
- 4 **call** n = the number of images in A
- 5 **for** i in n **do**
- 6 **make** empty image (E_i)
- 7 **pick** randomly desired classes (r) in CD
- 8 **paste** r into E_i
- 9 **return** new annotation image

END

Algorithm 3:
Enhancing the contribution proportion of all classes by generating completely dummy data.

BEGIN

- 1 **call** n = number of needed annotation images
- 2 **for** i in n **do**
- 3 **make** empty image (E_i)
- 4 **make** randomly shape of all classes (S_i)
- 5 **paste** S_i into E_i
- 6 **return** new annotation image

END

Algorithm #1 extracts all shapes and positions of the layers to be enhanced from the original dataset. This results in a dictionary where the classes are the keys, and their corresponding shape and position are the values. Then, a random element of enhanced-classes is selected from the dictionary and pasted into each original annotated image to ensure that all images contain the desired classes, while preventing overfitting when training through random selection. Figure 7 shows some annotation images created from three algorithms and their corresponding dummy images.

Algorithm #2 is similar to algorithm #1, except that it enhances all classes at once, ensuring that each image has all classes needed detection and classification. When pasting classes' element into an image, it is important to pay attention to the order. In this study, the order must be silver, rainbow, brown, and black oil due to the characteristics of the oil. Failure to follow this order may result in silver oil covering the entire area of the remaining oil layers in the image.

While the first and second algorithm utilize the shapes and positions of enhanced-classes present in the data set, the third principle generates the shapes and positions of the enhanced-classes randomly, resulting in a more natural and realistic data set while avoiding overfitting.

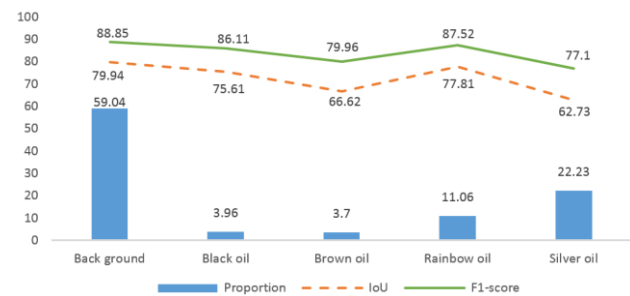


Figure 5. Correlation between accuracy and proportion.

3.4 Training Pix2Pix model and generate fake image

Pix2Pix is a conditional GAN model designed for image-to-image translation, which was first introduced in 2018 by Isola et al. The model consists of a Generator and a Discriminator, each with a distinct but related training process. Specifically in this research, the Generator is responsible for translating an annotation image into a photographic image, while the Discriminator is trained to distinguish between real images and those generated by the Generator. To train the GAN model, the Discriminator is trained for one or more epochs, followed by the training of the Generator for one or more epochs. These steps are repeated to continue to train both the Generator and Discriminator networks. Figure 6 depicts the structure of the GAN model and its training process.

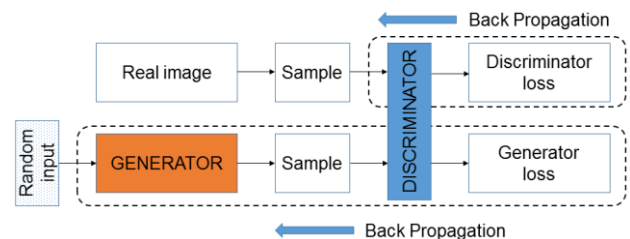


Figure 6. GANs' structure and its training process.

Source: <https://developers.google.com/machine-learning/gan/discriminator>

This study utilized the original dataset, which consisted of oil spill images and their corresponding annotation images, to train the Pix2Pix model over 800,000 iterations. Two examples of images generated from the Pix2Pix model are displayed in figure 7.

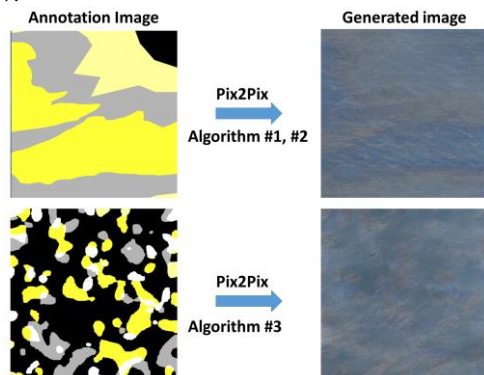


Figure 7. Fake annotation images and its fake generated photographic images.

4. EXPERIMENTAL RESULTS

In Figure 5, a correlation between contribution proportion and accuracy is observed, indicating that improvements are needed for the silver and brown oils. To address this, algorithm #1 will be applied in order to create a dataset named "BRSV". This dataset will add two classes of silver and brown in each image. Additionally, algorithm #2 will be used to create an "ALL"-dataset, where all four oils are present in each image. Finally, Algorithm #3 will be applied to create a "FAKE"-dataset, which features all four oils with completely random shape, size, and position in each image. The original dataset will be combined with these three datasets which generated using the Pix2Pix model and used to train the DaNet model for oil detection and classification.

All training processes were implemented on a Linux system with a 16Gb RAM and a Geforce RTX3090 D6X 24Gb GPU, and each dataset will be trained with 200 epochs.

Table 1 displays the training results of three datasets. It is evident that the average accuracy, as well as the accuracy of the classes, improve significantly when dummy datasets are utilized. The BRSV dataset enhances the Silver and Brown oil grades by an average of 5.05%. Furthermore, the IoU of the classes is uniformly improved. To provide a more objective assessment, f1-score is also calculated, which demonstrates the ratio between false positives and false negatives, reflecting the model's classification accuracy. Table 2 indicates that the f1-score is also significantly enhanced when dummy datasets are used, with an accuracy increase of 1.69%.

Type of Oil	Original mIoU (%)	BRSV mIoU (%)	ALL mIoU (%)	FAKE mIoU (%)
Background	79.90	81.80	83.17	82.40
Black	76.97	70.77	76.67	81.67
Brown	67.05	68.17	67.95	66.12
Rainbow	77.13	79.12	79.89	80.56
Silver	61.39	70.36	67.56	60.04
Mean	72.49	74.04 (+1.55)	75.05 (+2.56)	74.23 (+1.74)

Table 1. Comparison between the mIoUs of the original dataset and 3 fake datasets.

Type of Oil	Original F1-score (%)	BRSV F1-score (%)	ALL F1-score (%)	FAKE F1-score (%)
Background	88.85	89.99	90.81	90.35
Black	86.11	82.88	86.79	89.91
Brown	87.52	81.07	80.92	79.61
Rainbow	77.10	88.34	88.82	89.23
Silver	79.96	82.60	90.64	75.31
Mean	83.91	84.98 (+1.07)	85.60 (+1.69)	84.88 (+0.97)

Table 2. Comparison between the f1-scores of the original dataset and 3 fake datasets.

Figure 9 illustrates some oil classification and detection outcomes on the test set, providing a visualization of the results. Although the combination of the original dataset with the dummy datasets yields better results, there are still some inadequacies that need to be addressed. These remaining issues will be discussed in the following section.

5. DISCUSSIONS

This section will discuss the impact of using three dummy datasets during training.

5.1 Efficiency of the BRSV-dataset

Based on Tables 1 and 2, and the proportion of brown and silver oil contributions in Table 3 of the BRSV dataset, a positive correlation between accuracy and contribution proportion was observed. Specifically, as the contribution proportion increased, the mIoU and f1-score improved. Additionally, when combining the BRSV dataset, the training process showed faster convergence and improved ability to overcome local minimums, as evidenced in Figure 8. Notably, the model with the combined BRSV dataset outperformed the model trained on the original dataset at iterations 80,000, 150,000, and 180,000 for silver oil and at iteration 180,000 for brown oil.

Type of Oil	Old proportion (%)	New proportion (%)	Changed (%)
Background	59.04	56.79	↓ 2.25
Black	3.96	3.83	↓ 0.13
Brown	3.70	5.53	↑ 1.83
Rainbow	11.06	10.64	↓ 0.42
Silver	22.23	23.21	↑ 0.98

Table 3. Comparison between Original and BRSV proportion.

Figure 9 illustrates a few predicted results of the model. It is observed that enhancing the brown and silver classes results in the model "forgetting" some of the remaining classes. For instance, in case (a), the model failed to detect the entire black oil patch near the shore. However, the model performed better in detecting small oil patches and was less confused by the shadow of the tree in case (d). In case (c), the model still faced challenges in distinguishing between sun glint and silver oil regions, resulting in a failure to detect a portion of the silver oil region.

5.2 Efficiency of the ALL-dataset and FAKE-dataset

The purpose of creating the ALL and FAKE datasets was to balance the proportions of oils in the dataset, as easy to see in

Tables 4 and 5. The accuracy of the mIoU and f1-score determinations between oils is also relatively balanced. The graph presented in Figure 10 demonstrates that using these datasets helps to reduce local minimums and achieve faster convergence of the model.

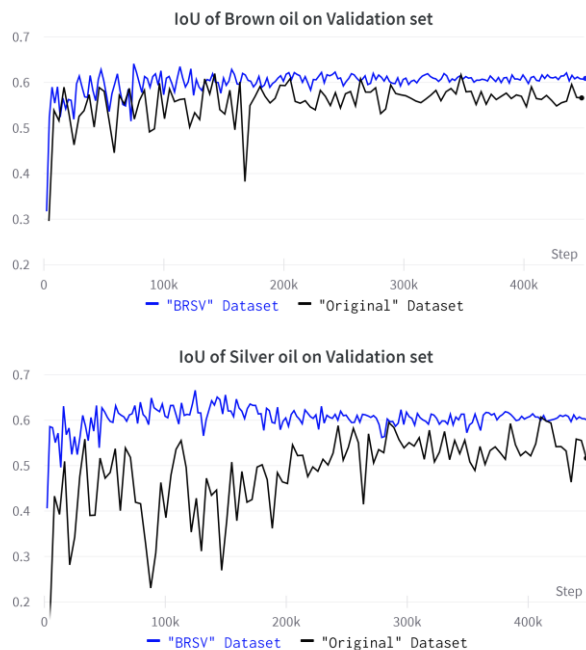


Figure 8. IoU on validation set of Silver and Brown oil using BRSV dataset.

The results presented in Tables 1 and 2 suggest that the dataset created using algorithm #2 yields the most improvement in model performance. However, a closer examination of Figure 9 reveals that cases (a), (b), and (d) demonstrate that a dataset comprised of random elements in terms of shape, size, and position enables the model to detect smaller oil patches, but does not address the challenges posed by look-alike objects such as sunglint, fog, and cloud, etc., as case (c).

Type of Oil	Old proportion (%)	New proportion (%)	Changed (%)
Background	59.04	44.55	↓ 14.49
Black	3.96	19.82	↑ 15.86
Brown	3.70	6.49	↑ 2.79
Rainbow	11.06	13.5	↑ 2.44
Silver	22.23	15.6	↓ 6.63

Table 4. Comparison between Original and ALL proportion.

Type of Oil	Old proportion (%)	New proportion (%)	Changed (%)
Background	59.04	66.09	↑ 7.05
Black	3.96	6.64	↑ 2.68
Brown	3.70	4.75	↑ 1.05
Rainbow	11.06	8.46	↓ 2.60
Silver	22.23	14.24	↓ 7.99

Table 5. Comparison between Original and FAKE proportion.

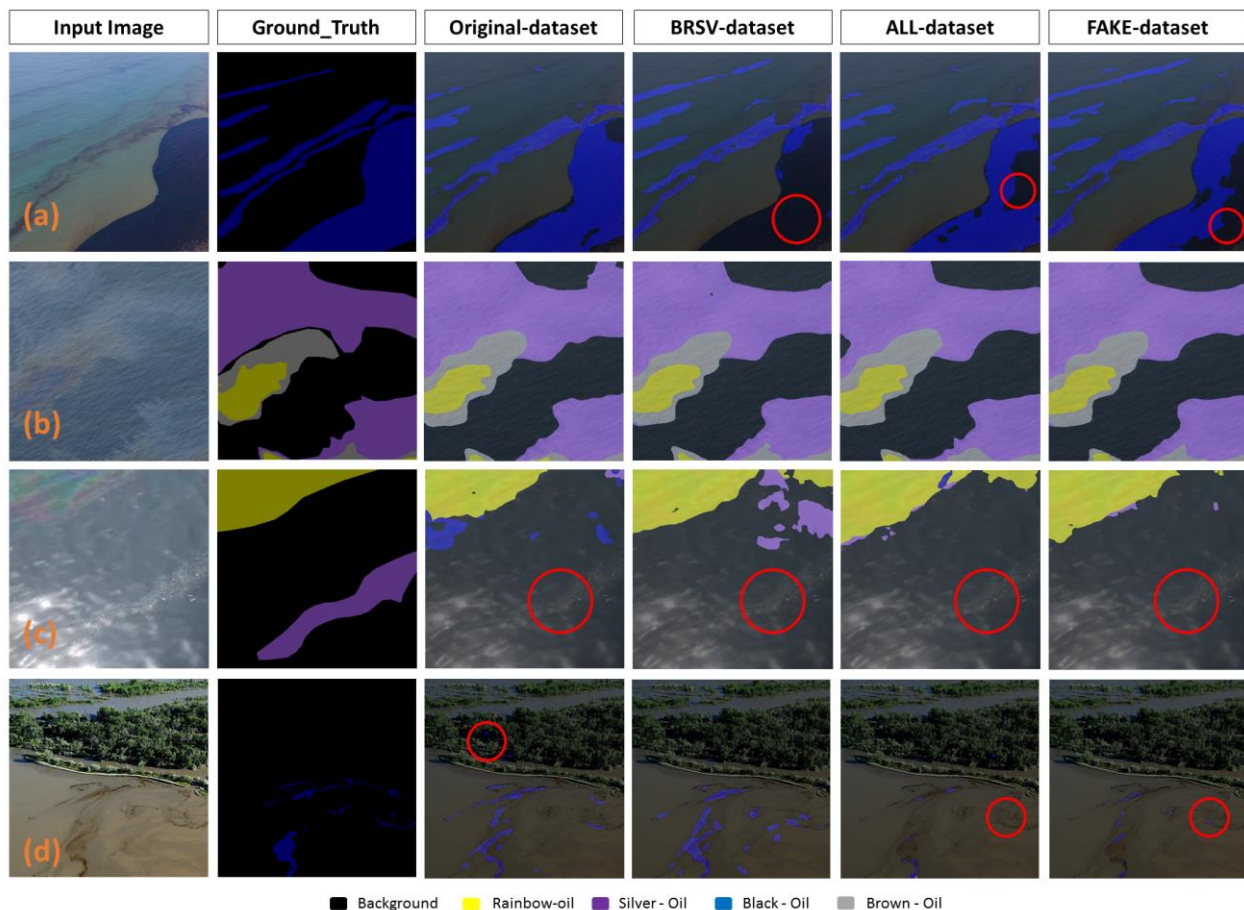


Figure 9. Oil detection and classification results on a few images of the test set.

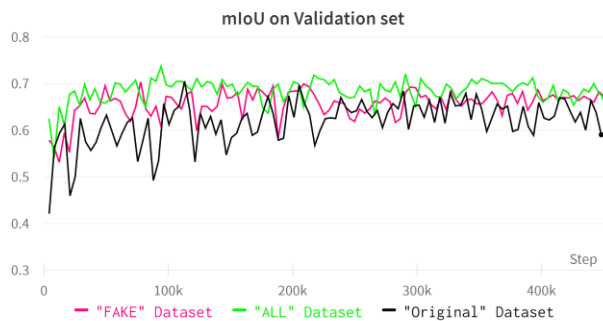


Figure 10. mIoU of three datasets on validation set.

6. CONCLUSION

This study proposes three algorithms to enhance data for detecting and classifying difficult objects like oil spills, cracks, and mould spots, which do not have a fixed shape or are hard to collect additional data for. Although the problem of distinguishing oil from look-alike objects remains unsolved, the results show promise with an increase in mIoU and f1-score by 2.56% and 1.69%, respectively. Moreover, these algorithms help improve the model's ability to detect small objects.

Based on the analysis, each algorithm of creating dummy data strengthens the model differently. In future studies, different models can be created, and the best results can be yielded using a module which combine all results from each different models by using highest-vote rule.

ACKNOWLEDGEMENTS

We express our appreciation for the assistance given by the project titled "Development of Technology for Automatic Generation and Presentation of Pollution Information Using the Field Survey Data of the Korea Coast Guard Agency" (1535000304), which provided support for this research.

REFERENCES

- Benjamin, K., et al., 2023: 1st Workshop on Maritime Computer Vision (MaCVi) 2023: Challenge Results. *WACV 2023 Workshop*, 265-302.
- Bui, N.A., Oh, Y.G., Lee, I.P., 2023: Oil spill detection and classification from airborne EOIR images using a deep learning model. *International Conference on Aquatic Science & Technology (i-CoAST2023)*.
- Goodfellow, I.J., Abadie, J.P., Mirza, M., Xu, B., Farley, D.W., Ozair, S., Courville, A., Benjo, Y., 2014: Generative Adversarial Networks. *Machine Learning*. <https://doi.org/10.48550/arXiv.1406.2661>.
- Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2018: Image-to-Image Translation with Conditional Adversarial Networks. *Computer Vision and Pattern Recognition*. <https://doi.org/10.48550/arXiv.1611.07004>.
- Liu, S., Zhang, J., Chen, Y., Liu, Y., Qin, Z., Wan, T., 2019: Pixel Level Data Augmentation for Semantic Image Segmentation Using Generative Adversarial Networks. *2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1902-1906.

Maharana, K., Mondal, S., Nemade, B., 2022: A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, Volume 3, Issue 1, 91-99. <https://doi.org/10.1016/j.gltpr.2022.04.020>.

Tyagi, A.K., Rekha, G., 2020: Challenges of Applying Deep Learning in Real-World Applications. *Challenges and Applications for Implementing Machine Learning in Computer Vision*. DOI: 10.4018/978-1-7998-0182-5.ch004.

Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., Shen, F., 2022: Image Data Augmentation for Deep Learning: A Survey. *Computer Vision and Pattern Recognition*. <https://doi.org/10.48550/arXiv.2204.08610>.

Zhang, Y., Zhang, J., Xu, X., 2021: Object Aug: Object-level Data Augmentation for Sematic Image Segmentation. *IEEE*. <https://doi.org/10.48550/arXiv.2102.0022>.