# Interpreting Abnormality of a Complex Static Scene using Generative Adversarial Network

Mahamat Moussa* and Chern Hong Lim†
* † MONASH University, Malaysia
E-mail: mahamat.moussaabbasali@monash.edu*
E-mail: lim.chernhong@monash.edu†

***Abstract*—Anomaly detection remains a difficult task in the computer vision and image processing field. Although several studies have been done to address this challenge, most of these studies focused on analyzing temporal features to determine abnormality. Examples of temporal features include behavioral changes and new object appearance in the target scene. In this paper, we are interpreting abnormality from a new perspective, which is static and complex image scene that focused on one object (airplane) using generative adversarial networks (GANs). Our interpretation of abnormality in such image intended to test two research hypotheses: 1) whether GANs can capture the cognitive features of abnormality from within a complex scene. 2) whether GANs can be used to generate more reliable datasets of abnormal scenes. In this work, we chose an airplane as the object of our experiment. We defined abnormal and normal scenes as follow: The scene is abnormal if the airplane involved in accidents (such as crash or fire), and normal otherwise (such as flying or landed airplane). A custom dataset is built for this experiment and it consists of two classes; normal and abnormal. We augmented each class to the double of its size using GANs, and then we created three different sets of datasets: (DS1, DS2, and DS3) to test our hypotheses. We applied four different supervised machine learning classifiers on each of these three sets, we repeated this step 3 times as follow: 1) pixel-based, 2) with applying Principal Component Analysis (PCA), and 3) with applying Local Binary Pattern (LBP). The overall results showed that GANs possess the capability of generating images that capture the abnormality features from the static complex scene.**

## I. Introduction

Deep learning (DL) models have been widely used to solve many complex problems recently. Variant models are being developed to address issues from different perspectives. Generative adversarial networks (GANs), which is one of the DL models that is capable to generate synthesis images, has gained lots of attention among researchers, and have become increasingly a popular model in the last few years [1]. With such popularity, many useful applications of GANs have been proposed to solve complex problems such as data synthesis [1], Attention prediction [2], Object generation, and security assurance [3].

One of the problems that GANs are used as a proposed solution, is anomaly detection [4]. Anomaly detection is a process of identifying an area or frame that is different from the usual ones, or those contain abnormality [4]. Generally, there are two main methods for anomaly detection, trajectory-based method, and patch-based method [5]. The former computes the difference between normal and abnormal trajectories, while the latter focus on feature studies to solve the problem. Although DL has solved many complex problems, anomaly detection remains a difficult and challenging task [5]. This is due to most of the prior researches focused on formulating anomaly detection as a temporal problem where temporal features are ought to be found to assure the abnormality of a scenario. For examples, new object appearance in a scene or sudden behavioral changes (from walking to running) as shown in figure 1.

(a) Abnormal scene when a car appears in the scene (UCSD anomaly dataset) [6]


(b) Abnormal scene on the right when the crowd behave abnormaly (transition from walking to running (UMN crowd dataset) [7]

Fig. 1: The figure shows how the abnormal scenes are interpreted by assuming them as a temporal problem. (a) A new object appears in the scene, (b) A sudden behavioral change.

This assumption aids to uncover the abnormality by identifying the differences in the image sequence using temporal analysis [9] but it has a limitation on interpreting abnormality from a static complex scene. For instance, to determine if an airplane in a static scene is normal (flying) or abnormal (in an accident), the above method is not capable to perform this. Unlike machines, human tends to understand abnormality by looking at the scene only once, such as the scenes shown in

Fig. 2: Samples of abnormal scenes [8]


Fig. 3: samples of normal and abnormal scenes

figure 2. Machines are yet to have this capability. The main reason is that human analyzes an image at first glance by understanding the logic of the scene. This could be based on intuitive physics, cognitive knowledge [8] or common sense. A lacking of interpretability in such a way hinder the anomaly detection in a static complex scene.

In this paper, we interpret abnormality from the perspective of a static complex scene using generative adversarial networks (GANs) and other Machine Learning (ML) algorithms. The abnormal scenes we study consist of unusual events that are not likely to happen in our daily life. Our main goal is to test two hypotheses. First, to understand whether GANs can capture the cognitive features of abnormality when it generates synthesis images from complex abnormal scenes. Second, whether these synthesis images can be used as reliable datasets for determining abnormality, therefore, can be a solution for the lack of abnormal datasets availability, which can be used for future study in cognitive intelligence [8]. As a preliminary study, we selected an airplane to be our target object in this research. We defined abnormal and normal scenes as follow: The scene is abnormal if the airplane involved in accidents (such as crash or fire), and normal otherwise (such as flying or landed airplane).

A custom dataset is built to carry out this experiment. It consists of two classes; “abnormal" and “normal" . Each class has a total of 1000 real-world images of airplane accidents (as abnormal), and flying or landed airplane (as normal). The size of each class is augmented using GANs to generate synthesis examples. We created three sets out of this datasets: DS1, DS2, and DS3, and applied four different ML classifiers on these three sets. Each classifier is tested three times (with different settings ) to interpret if GANs is capable to capture the cognitive features of abnormality in the classification tasks. Figure 3 shows some examples of the normal and abnormal images from our dataset.

The following sections of this paper are organized as follow: In section II we cover briefly some background studies related to anomaly detection. Section III explains our methodology of interpreting the abnormality of a complex static scene using GANs in details. The experiment results will be discussed in Section IV and finally a conclusion is made in Section V.

## II. BACKGROUND

Several studies have been done to solve abnormality detection using different methods [10]. Recent studies used GANs as well as convolution neural network (CNN) to examine, develop methodology and solve abnormality detection.

For example, [11] argues that abnormality detection consider a difficult task due to two reasons. First, lack of datasets that contain enough examples of abnormality. Second, there is no single definition for abnormality, which means every case study may have a different definition of abnormality. In [11], they used GANs to solve abnormality detection in crowded areas. They introduce a new method by which GANs is trained to understand the internal structure of only normal scenes instead of abnormal ones. As a result, GANs would not be able to generate abnormal data. During the testing phase, the real-time data is compared with the data generated by GANs, which help in detecting abnormal areas [11].

Similarly, [12] uses CNN to solve the same problem by implementing a self-supervised framework and abnormal event detection network (AED-Net) to detect abnormal events. Reference [12] uses a surveillance image as input data and

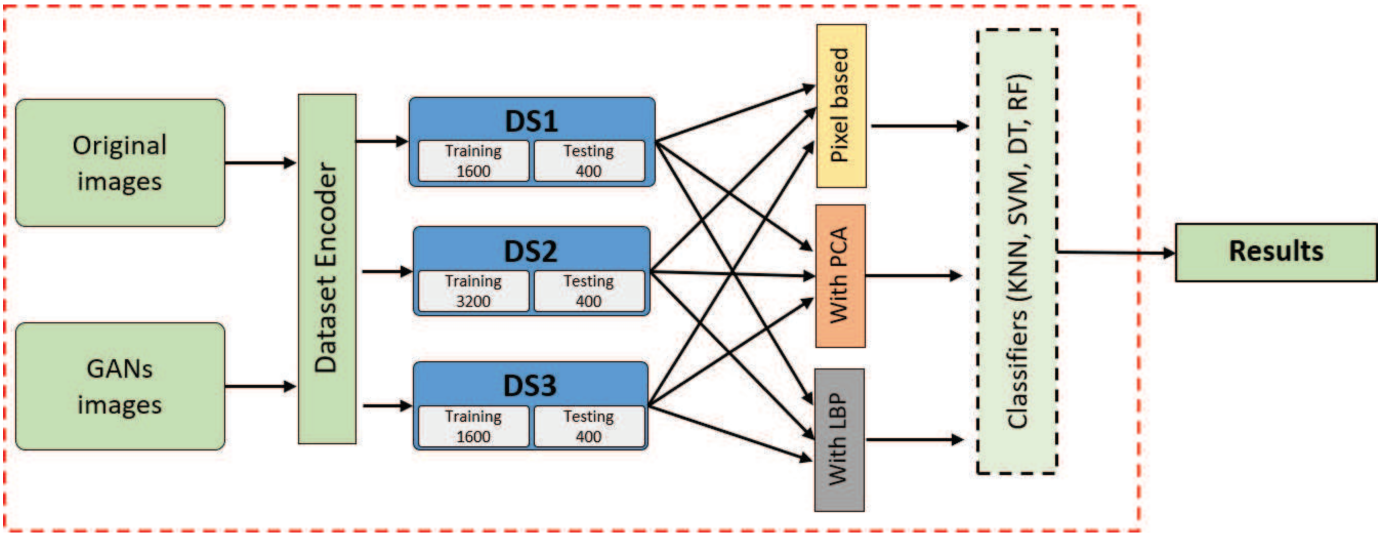


Fig. 4: The general framework of the experiments. We created three sets out of the entire dataset: DS1, DS2, and DS3. Each set is tested three times on four different ML classifiers.

extract high-level semantics using kernel principal component analysis to detect abnormality. According to their experiment, the result has proven that an improved performance is obtained over the original AED-Net [12]

Moreover, [10] asserts that anomaly detection is a challenging task because the characteristics of the anomaly are deeply connected to the scene. As a solution, [10] proposed object-oriented methodology to detect anomaly. The method simply took advantage of temporal as well as spatial data of the target object, by integrating anomaly tracking and object detection. [10] also addressed that missing semantic analysis could affect the process of anomaly detection in a negative way.

In addition to the above studies, [13] focused more on detecting mass abnormalities in digital images using Support Vector Machine (SVM) as a classifier. The study was done using 50 images that contain benign and malignant images. The features were segmented and extracted from these images by applying some techniques such as split and merge, and GLCM. The proposed technique and algorithm achieved high accuracy result, up to 94% [13]

Overall, these studies have shown good implementations of ML, DL, GANs, and CNN in solving anomaly detection problem. However, most of these studies followed a similar scheme which depends on temporal features to detect the abnormality. In contrast, our study mainly focused on static image, understanding abnormality of the same object instead of tracing abnormality across temporal features.

## III. Methodology

In this section, we explain the methodology of our experiments with more details about the datasets and GANs training process. Figure 4 summarizes the overall process of our framework.

### A. Dataset preparation

A custom dataset is built for this research experiment from publicly available images under (reuse) usage rights. The dataset consists of two classes “abnormal” and “normal”. An airplane was chosen to be the target object of this experiment. As mentioned earlier, the abnormal class contains 1000 images of airplanes that involve in accidents. The normal class contains the same number of images except that the airplanes are normal (i.e flying or landed). Figure 3 shows samples of each class. At this point, we have collected 2000 images which are divided equally into two classes as described above. We then generated more samples by using GANs to increase the number of images of each class to the double. As a result, a collection of 2000 original images (1000 normal + 1000 abnormal), and 2000 GANs generated images (1000 normal + 1000 abnormal) is available for this experiment.

Out of the collection of these 4000 images, we created three different sets of datasets: (**DS1**, **DS2**, and **DS3**). Each of which separated into training and testing sets to test our hypotheses. We split them based on 80% and 20% strategy. In details, **DS1**: is the original dataset, it consists of all original images for both training and testing sets. The classification results (anomaly detection) of this dataset is used as a baseline to evaluate the results obtained from DS2 and DS3. **DS2**: is a combined dataset that has both original and GANs images (1600 GANs generated + 1600 original) for the training set. The testing set contains only original images for both normal and abnormal scenes. **DS3**: is a mixed dataset which has (800 GANs generated + 800 original) images for the training. The testing set contains only original images. Figure 4 shows the overall framework of this research.

### B. Data Pre-processing

Since our dataset is built from publicly available images which come with different formats and sizes, data preprocessing is required to normalize the dataset for the later processes. We reshaped all images into 256*256, and convert them into a grayscale image to ensure the efficient training and testing process.

### C. *Testing Environment*

Training Deep Learning (DL) network such as GANs can be computationally expensive without a dedicated and stable machine for conducting the experiments. In this work, we use an Ubuntu OS with 16GB RAM and Nvidia GeForce GTX 1080 for our hardware. Python and TensorFlow are used as the main development environment tools.

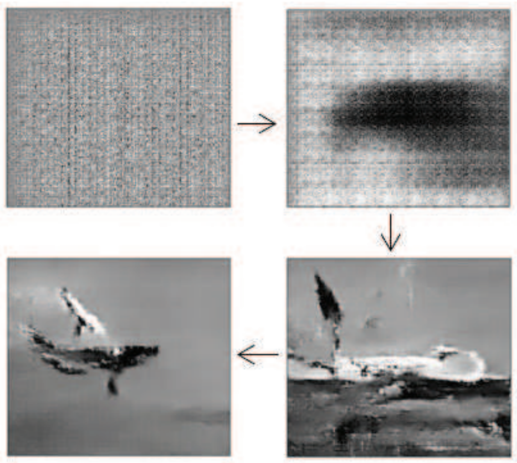

Fig. 5: GANs generator learns to generate examples of our normal class starting from noise to a low quality synthesis image.

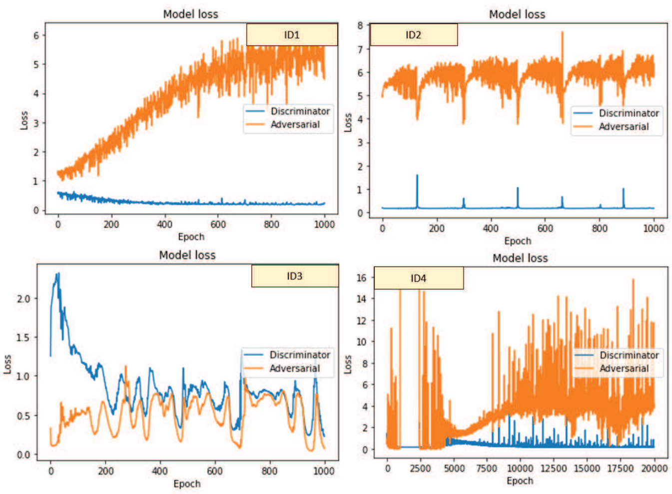


Fig. 6: GANs performance for different hyperparameters tuning based on the values described in table I.

### D. *GANs Training*

In order to generate images of a complex scene, GANs tends to show less expected outcomes [5]. This is despite applying several improving techniques to stabilize our GANs to generate better results [10]. Figure 5 shows an example of how our GANs learns to generate samples of abnormal images from random noise during the training process.

Tuning GANs hyperparameters for our dataset yields different results every time a small change is made on one single hyperparameter value. Over the experiments, we learned that a small batch size and smaller learning rate,lr tends to give better results. Figure 6 shows examples of the loss of the generator,G and discriminator,D of different hyperparameters. Table I shows some of the corresponding parameter settings used for each graph in figure 6. Base on the empirical testing results with applying different settings, we selected lr=0.0002 for both G and D with batch size of 64 to generate our normal and abnormal scene using GANs as it yields the best result for us.

TABLE I: Examples of the different hyperparameters settings

| Test ID | D lr | GAN lr | Batch Size |
|---|---|---|---|
| ID1 | 0.0002 | 0.0002 | 64 |
| ID2 | 0.0002 | 0.0002 | 256 |
| ID3 | 0.0001 | 0.0002 | 64 |
| ID4 | 0.0003 | 0.0003 | 512 |

### E. *Interpretation of GANs generated images*

At this stage, we used the three sets of our datasets (DS1, DS2, and DS3) to interpret the outcome of the anomaly detection by applying four different classifiers: 1) K-nearest-neighbor (KNN), 2) Support Vector Machine (SVM), 3) Decision Tree (DT), and 4) Random Forest (RF). Each classifier is repeated three times with three scenarios as follow: first, we used all the pixels of the input image without applying any feature transformation techniques. Second, we performed feature reduction by applying Principle Component Analysis (PCA) on the input image to capture the most significant features from the original and GANs generated images before the classification. Third, we applied Local Binary Pattern (LBP) on the input image to extract the significant texture before the classification.

This experiment intends to prove that, although generating a complex image using GANs can be a very difficult task and yield unclear results, the inner structure of the generated images still contains useful features that are distinguishable by the classifiers, hence can be used for anomaly detection.

TABLE II: Classification Results (pixel based)

| Algorithm | DS1 | DS2 | DS3 |
|---|---|---|---|
| KNN | 0.67 | 0.71 | 0.70 |
| SVM | 0.81 | 0.94 | 0.90 |
| DT | 0.81 | 0.87 | 0.73 |
| RF | 0.81 | 0.86 | 0.85 |

## IV. Results and Discussion

After we have conducted the experiments using KNN, SVM, DT, and RF classifiers on the datasets with three different scenarios (all pixels, PCA, and LBP), the result shows that the set DS2 of our dataset always yield better results over DS1 and DS3 for all the scenarios as it can be seen in table II, III, and IV. This is because DS2 contains more training examples compare to the other sets, and also more GANs generated images.

Table II shows that SVM recorded the best result followed by RF and DT. KNN showed the lowest performance among all classifiers. Meanwhile, the dataset (DS2) shows better results compared to D1 and D3 as mentioned earlier. This could be a good indicator that our hypothesis is true because DS2 is mainly increased by the GAN generated images.

TABLE III: Classification Results after Applying PCA

| Algorithm | DS1 | DS2 | DS3 |
|---|---|---|---|
| KNN | 0.75 | 0.80 | 0.76 |
| SVM | 0.79 | 0.93 | 0.86 |
| DT | 0.85 | 0.84 | 0.69 |
| RF | 0.86 | 0.87 | 0.82 |

As for table III, it can be seen that DS2 still doing better results than others after we applied PCA. SVM and RF show better results than KNN and DT for this section. It is worth to mention that KNN becomes slightly better after using PCA.

TABLE IV: Classification Results after Applying LBP

| Algorithm | DS1 | DS2 | DS3 |
|---|---|---|---|
| KNN | 0.78 | 0.88 | 0.87 |
| SVM | 0.82 | 0.90 | 0.90 |
| DT | 0.79 | 0.88 | 0.86 |
| RF | 0.85 | 0.91 | 0.90 |

In table IV, it is confirmed that DS2 is better than DS1 and DS3 for all the classifiers after applying LBP. With a close examination, it is clear that applying LBP results in better accuracy for DS2 and DS3 compared to PCA.

These results in table II, III, and IV confirm that the hypotheses we initially developed has good degree of positive answers. It can be seen that datasets DS2 and DS3 (which contains GAN images) have good results. A reminder that DS3 is having the same number of training data as D1. This indicates an important point of our first hypothesis, which confirms that GANs can capture important cognitive features of abnormality from a complex scene image. This is because of the higher accuracy of DS2 and DS3 as they both contain GANs generated images.

Besides, one can notice that by applying LBP in table IV, we obtained better accuracy compared to table II and table III. This may emphasize on the idea that GANs-generated images carry important texture information that helps in interpreting abnormality of a scene. Last but not least, these results also imply a positive answer for our second hypothesis that GANs can be used to augment and generate a reliable dataset for different sorts of experiments.

## V. Conclusion

The main goal of this experiment is to interpret the abnormality of a complex static scene using generative adversarial networks (GANs). We aim to understand if images generated by GANs are with important cognitive features of a complex static scene, as well as to understand if GANs can be used to generate a reliable dataset for different research needs. The outcome of this work concluded that images generated by GANs can extract important cognitive features of abnormality which is sufficient to be used in anomaly scene detection. The study also shows a positive indicator that GANs can be a good tool to build reliable datasets to generate complex scenes from a static image. Some of the future works related to this research can be a further study to investigate the GANs model in more details to understand how it actually learns and generates images, this would help in researching how machine perform intuitive learning and behave like a human.

## References


[1] A. Sharma, N. Jindal, and A. Thakur, "Comparison on generative adversarial networks–a study," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*. IEEE, 2018, pp. 391–396.

[2] D. Kastaniotis, I. Ntinou, D. Tsourounis, G. Economou, and S. Fotopoulos, "Attention-aware generative adversarial networks (ata-gans)," in *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. IEEE, 2018, pp. 1–5.

[3] V. Belenko, V. Chernenko, M. Kalinin, and V. Krundyshev, "Evaluation of gan applicability for intrusion detection in self-organizing networks of cyber physical systems," in *2018 International Russian Automation Conference (RusAutoCon)*. IEEE, 2018, pp. 1–7.

[4] S. A. Israel, J. Goldstein, J. S. Klein, J. Talamonti, F. Tanner, S. Zabel, P. A. Sallee, and L. McCoy, "Generative adversarial networks for classification," in *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE, 2017, pp. 1–4.

[5] X. Ma, R. Jin, K.-A. Sohn, J. Paik, J. Sun, and T.-S. Chung, "Improving generative adversarial networks with adaptive control learning," in *2018 IEEE Visual Communications and Image Processing (VCIP)*. IEEE, 2018, pp. 1–4.

[6] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 1975–1981.

[7] Umn crowd dataset. [Online]. Available: http://mha.cs.umn.edu/projevents.shtml#crowd

[8] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," *Behavioral and brain sciences*, vol. 40, 2017.

[9] S. Mane and S. Mangale, "Moving object detection and tracking using convolutional neural networks," in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2018, pp. 1809–1813.

[10] X. Li, W. Li, B. Liu, Q. Liu, and N. Yu, "Object-oriented anomaly detection in surveillance videos," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 1907–1911.

[11] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, "Abnormal event detection in videos using generative adversarial nets," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 1577–1581.

[12] T. Wang, Z. Miao, Y. Chen, Y. Zhou, G. Shan, and H. Snoussi, "Aed-net: An abnormal event detection network," *arXiv preprint arXiv:1903.11891*, 2019.

[13] G. Jothilakshmi and A. Raaza, "Effective detection of mass abnormalities and its classification using multi-svm classifier with digital mammogram images," in *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)*. IEEE, 2017, pp. 1–6.