Semi-supervised generative adversarial networks for anomaly detection

Juan Montenegro[1], and *Yeojin* Chung[1,*]

[1]Department of Data Science, Kookmin University, 77 Jungrungro, Sungbukku, Seoul, 02707 South Korea

Abstract. Advancements in security have provided ways of recording anomalies of daily life through video surveillance. For the present investigation, a semi-supervised generative adversarial network model to detect and classify different types of crimes on videos. Additionally, we intend to tackle one of the most recurring difficulties of anomaly detection: illumination. For this, we propose a light augmentation algorithm based on gamma correction to help the semi-supervised generative adversarial networks on its classification task. The proposed process performs slightly better than other proposed models.
Keywords: Generative adversarial network, Gamma correction, Computer vision, Anomaly detection

1 Introduction

The role of anomaly detection is identifying certain events, observations or data characteristics that diverge from a supposed norm [1]. The video anomalies such as anomalous activities and anomalous entities are defined as the abnormal or irregular patterns present in the video that does not follow the normal trained patterns [2]. An anomaly detection system aims to prompt a signal for a certain activity that diverges from a pattern that is considered normal. On a video, anomaly detection can subtract the anomalies from the video and categorize them according to a certain classification list [3].

Anomalous activities such as fighting, riots, traffic rule violations, and stampede as well as anomalous entities such as weapons at the sensitive place and abandoned luggage should be detected automatically in time. However, video anomalies detection is a complex task due to different reasons such as, the ambiguous nature of searched anomaly, various environmental conditions, the complex nature of human behaviours, and/or the lack of proper datasets.

For the computer vision field, one of the more recurring and greatly challenging problem to address is the one of anomaly detection. Until now, many attempts have been done to tackle the detection of violence on surveillance footage [3]. However, the present investigation focuses on the recent works to detect anomalies on videos based on the biggest crime related video anomaly detection dataset to date: the UCF crime detection dataset from the University of Central Florida of the United States.

[*] Corresponding author: ychung@kookmin.ac.kr

Numerous methods have been proposed to approach this dataset. Waqas et all.[3] proposed a deep Multiple Instance Learning (MIL) ranking loss. This method of anomaly detection treated the classification of the video as a regression problem. Adding to this, the authors also conducted experiments for the classification of the crimes committed on the videos. For this, they used two models in order to detect actions in videos: 3 Dimensional Convolutional Neural Networks (CNN) and a Tube-Convolutional Neural Network(T-CNN). Other investigation like Dubey et al. [4] proposed a deep-network with Multiple Ranking Measures (DMRMs) framework. Additionally, Doshi et all. [5] suggest the use of a framework called Multi-Objective Neural Anomaly Detector (MONAD). The proposed MONAD consists of a Generative Adversarial Network (GAN) a lightweight object detector (YOLOv3) to extract meaningful features.

The present investigation tries to tackle one of the most recurring obstacles in the UCF dataset: light issues on videos. Sensors on a camera perceive colour and luminance differently than the human eye. For example, as the number of photons captured by a camera's sensor increase, the signal will also increase, which in return makes the image brighter [6]. However, when human eyes perceive double the amount of light, it is processed as only a fraction brighter than the original amount of light met. On the "Light-based data augmentation" section, a method called "gamma correction" is introduced to counter the light issue on the videos of the UCF crime dataset.

To the investigator knowledge, there has not been any efforts to solve this problem, and so, the present study not only focuses on detecting anomalies on videos but also to regulate the low and high luminosity of the videos in question. The model is expected to perform better than previous studies and offer a state-of-the-art detection task. The contributions of the present investigation are the following:

- Use a novel framework composed of an extension of a Generative Adversarial Network for multi-label classification.
- To extensively evaluate the proposed framework on the largest publicly available video anomaly detection data set: The UCF crime detection dataset.
- Use a light-based data augmentation process to give allow the model to perform a better feature extraction for high and low light videos.

2 Related works

Over the recent years, the number of studies focused on anomalous detection algorithms has increased and recent studies have developed techniques, such as image processing [7]. From this point onwards, other studies have directed their attention to action recognition [8, 9], detection of objects [10], tracking objects [11, 12], and other type of visual focused tasks. Consequently, some recent studies [13-16] were proposed mostly using auto-encoder methods and Recurrent Neural Networks (RNN) [17, 18] for more advanced and better performance models on anomaly detection.

Nonetheless, the focus of anomaly detection has been diverted to crime detection. For this area, the quantity of studies is not plenty. The reason for this is the limits of the available datasets, such as, their overwhelming difficulties (obscure, low quality, and short length videos) and small number of samples.

It was until 2018, that the UCF crimes detection dataset was published. This dataset not only presents numerous examples of crimes but also classifies them in 14 crime categories. It, in turn, presents a challenging task of detecting a crime when it occurs and classifying it correctly to the type of crime it is being committed. Due to the recentness of the UCF crime dataset, the number of studies that try to tackle this new repository is limited.

The people responsible of the publication of the UCF crime dataset also carried a study to test the capture of anomalous events. Waqas et al. [3] proposed an anomaly detection algorithm with weakly labeled videos for training. The authors introduced sparsity and smoothness limitations applied in temporal spaces when applying the ranking loss function to improve the detection of anomalies during training. The anomaly detection model was trained using the proposed deep MIL ranking loss. In the proposed approach, the method of anomaly detection was treated as a regression problem. Adding to this, the authors also conducted experiments for the classification of the crimes committed on the videos. For this they used two models: 3-dimensional CNN and T-CNN for action detection in videos.

Dubey et al. [4] proposed a DMRMs framework, which addresses context-dependency using a joint learning technique for motion and appearance features. Afterward, the authors extracted features and fused them for joint learning. However, due to the complexity of these anomalies, using only normal data was not optimal for anomaly detection. First, an extensive and annotated dataset for both normal events and abnormal events is necessary to train the proposed framework since it performs in a semi-supervised manner. A large-scale dataset requirement is an inherent problem in almost all existing visual recognition methods using deep learning. Second, the proposed framework is not robust to certain noises (occlusion, camera jitters, and illumination variations).

Doshi et al. [5] introduces a framework MONAD, which unifies deep learning-based feature extraction and analytical anomaly detection by incorporating two modules. The first module consists of a GAN-based future frame predictor and a lightweight object detector to extract meaningful features. The second module consists of a nonparametric statistical algorithm which uses the extracted features for online anomaly detection. Nevertheless, this proposed framework sacrifices the normalization (preprocessing) of new videos for in time online detection. This makes the framework more prone to mistakes. Additionally, the MONAD framework does not attempt to tackle various noises on videos (blurriness, illumination and sudden movements).

3 Proposed method

3.1 Semi-Supervised GAN

The GAN is a type of generative model that uses deep learning methods to automatically discover and learn recurrent characteristics or patterns on the input data for the GAN model to generate or output artificial examples that could plausibly been on the original dataset [19].

The SGAN model is a branch of the original GAN architecture which is characterized by involving the training of a supervised discriminator and, at the same time, also training an unsupervised discriminator and a generator model [19]. The model serves for both a supervised classification model that performs well to unobserved examples and a generator model that is in charge of outputting realistic examples of images from the main dataset. The discriminator model from the GAN is made to act as a multi-label classification model by classifying the images as real or fake and also, to identify the respective labels of certain images.

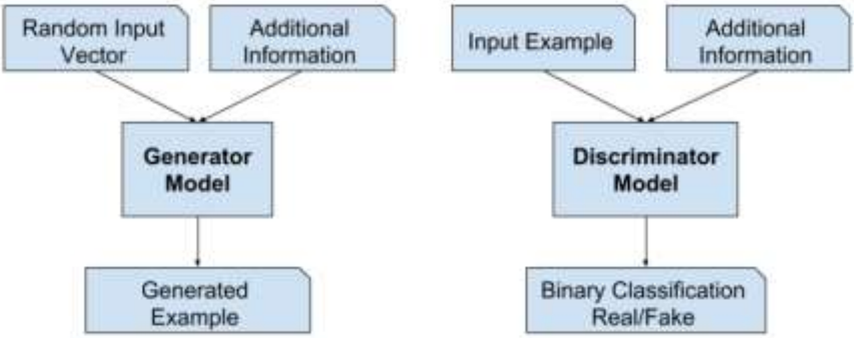

Fig. 1. Example of a Conditional Generative Adversarial Network Model Architecture [18]
Source: J. Brownlee, Generative Adversarial Networks with Python: Deep Learning Generative Models for Image Synthesis and Image Translation. Machine Learning Mastery (2021), pp. 12.

The resulting model is a classifier that can generalize well and offer optimal results on some classic classification tasks (such as MNIST), when trained on a small number of labeled examples, as tens, hundreds, or just one thousand images [19]. As an example, 100 images were taken as input for two models: CNN and the SGAN. The accuracy for the CNN after training was 0.895 but for the SGAN was 0.928. Additionally, the SGAN is capable of generating plausible artificial images of numbers from the MNIST dataset.

The SGAN training algorithm [20] works as follow: For each input of I (number of total iterations), the network will:

– Draw m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from the noise prior $P_g(z)$.
– Draw m examples $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$from data generating distribution $P_d(x)$.
– Perform gradient descent on the parameters of the Discriminator (D) and image classifier (C) and then, pass those values to the outputs on the combined minibatch of size $2m$.
– Draw m noise samples $\{Z^{(1)}, \dots, Z^{(m)}\}$ as for the noise prior $P_g(z)$.
– Perform gradient descent on the parameters the Discriminator (D) and image classifier (C) and then, pass those values to the outputs on the minibatch of size m**.**

3.2 Light-based data augmentation

A light-based data augmentation approach was used to help the classification process since numerous videos are presented at nighttime which makes the crime undistinctive and difficult to classify. For this, the gamma correction technique was used. The implementation process was applied as follows: The image's pixels intensity has to be converted from the range [0, 255] to [0, 1.0]. Afterwards, the gamma correction is applied on the image by following the equation:

$$O = I^{\left(\frac{1}{G}\right)}$$

Where ***I*** is the input image and *G* is the gamma value. The output, ***O,*** is scaled again to the range [0, 255]. Gamma values of less than one will shift the image towards the darker end of the spectrum while gamma values of greater than on can change the image to appear brighter. A gamma value of one will have no effect on the input image.

3.3 Anomaly detection process with Gamma correction

Fig. 2 illustrates the proposed anomaly detection process. The input images are preprocessed by the gamma correction algorithm and then fed to both the supervised and unsupervised discriminator model. The output from the supervised discriminator model results on the classification of the images into their predicted classes. The output from the unsupervised discriminator is fed to the generative model and its output is a new plausible artificial image.

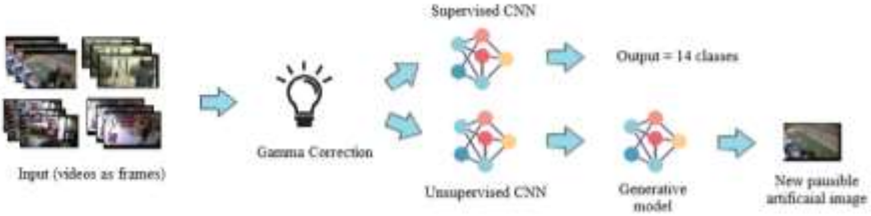

Fig. 2. Flow diagram of the proposed model
Source: Own processing.

4 Dataset

The UCF-Crime dataset was used to evaluate our method. The dataset is composed of a group of lengthy surveillance videos which can be divided in 13 real world anomalies, such as, Abuse, Arson, Road Accident, Burglary, Fighting, Robbery, Shooting, Stealing, Shoplifting, and others. It also includes footage of normal activities. The dataset is 128 hours long.

The UCF crime dataset serves two purposes: In one hand, it can be used as a general anomaly detection dataset for capturing all anomalies and differentiate them from normal events. On the other, it proposes the more challenging task of classifying any of the detected anomalies into the 13 anomalous activities categories.

5 Experiment results

In this work, the trained GANs were designed in two types of variations: The first SGAN was introduced as a normal model without any type of data augmentation process. The second SGAN was modified to deal with the light issues with the gamma correction process mentioned earlier. The output layer of both SGANs is the 14 classes from the different crimes that comprise the UCF dataset.

The videos used for training the models were trimmed in order to contain relevant footage and avoid additional visuals that could increase the noise in the model (credits, police logos, introduction sequences for news sites, etc.). A total of 3,976 extracted frames at an interval of one second were used. From this number, 284 random frames for each type of crime were used.

For both the unsupervised and supervised discriminator models a 2-dimensional CNN was used. The CNN consisted of 4 convolutional layers with 16, 32, 64 and 128 filters. Each layer down sampled the images with a 2x2 stride and used a Leaky-RELU activation function with an alpha of 0.2. Batch Normalization was also used within layers. Also, according to GAN literature and best practices, the kernel initializer corresponds to a random normal distribution generator with the standard deviation of 0.02.

For the generator model, due to the output of the unsupervised discriminator model, the structure follows 4 upsampling layers[†] of Transposed 2-dimentional Convolutions. The properties of the generator model are the same as the supervised and unsupervised discriminator model. The output of the generator model is a 2-dimentional convolutional layer with Tanh activation.

The optimizer used for training both models was Adam (with a learning rate of 0.0005) and the loss functions used were the sparse categorical crossentropy for the supervised discriminator model and the mean squared error for the unsupervised one. The results of the model after 50 epochs are in Fig. 3 and it shows that the model trained is overfitted and producing fluctuating output with poor testing and training accuracy.

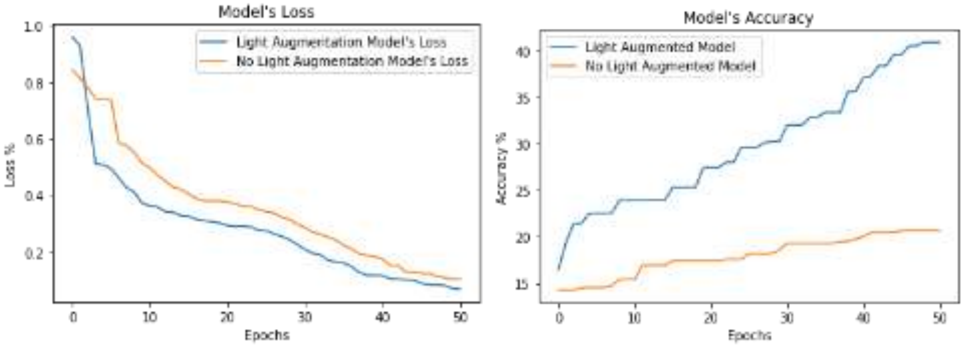

Fig. 3. Different evolutions of the accuracy score and loss of the supervised discriminator as it classifies the training video frames over iterations. The model is tested with and without the light augmentation algorithm
Source: Own processing.

According to the results, a better score on accuracy and lower loss are achieved with the proposed light augmentation algorithm. The gamma correction discussed previously. With this technique, the accuracy score is almost doubled compared to running the model without using it. The final accuracy score with light augmentation is found to be 40.9%. Meanwhile, for a non-light augmented process, the score is only 20.7%.

The light augmented model performs significantly better than the model suggested in [3] and demonstrates the effectiveness of using a gamma correction to change the luminosity inside the video frames for both the anomalous and normal videos. Light adjustment appears to be an indispensable method for a robust anomaly detection system.

Table 1. Accuracy results from [3] with the C3D and T-CNN models compared to the SGANs used

Method	C3D	t-CNN	SGAN (No Light Augmentation)	SGAN (with Light Augmentation)
Accuracy (%)	23.0	28.4	20.7	40.9

Source: Own processing.

Additionally, the use of SGANs also provides the generation of artificial images of crimes. Due to the low number of epochs the model was trained on, the images are not as convincing as they could be. Nonetheless, on the final epochs, the generated images have

[†] An upsampling layer, in this case and UpSampling2D layer, is used to double the dimensions of the input image. [19]

some small characteristics that could resemble those of real crimes images. For example, on epoch 50 there is a resemble to a frame of an arson video.

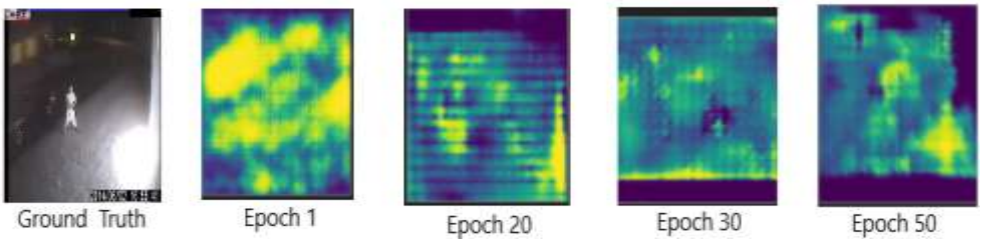

Fig. 4. Pictures generated by the generator model from the SGAN
Source: Own processing.

6 Discussions

A SGAN with a light augmentation method based on gamma correction is proposed to detect and classify real world anomalies in surveillance videos according to the type of crime that is occurring on the video. The results of the light augmented SGAN on the UCF crime detection dataset show that the anomaly detection process performs significantly better than other proposed methods. Moreover, the model demonstrates the usefulness of correcting the light on the videos.

Due to computational limitations, the model was trained only for 50 epochs. Due to this fact and the complexity of the dataset, the accuracy of both models is bound to be low. Therefore, and taking into consideration past investigations, is advised that for future implementations and/or experiments, is recommended that the model should be run for more than the number of epochs used in this investigation.

References

1. I. Cohen, Anodot. [online], Available at: https://www.anodot.com/blog/what-is-anomaly-detection/(2020) (2020)
2. R. Nayak, S. Das, U. A. Pati, Comprehensive review on deep learning-based methods for video anomaly detection. *Image Vision Computer* (2021)
3. S. Waqas, C. Chen, S. Mubarak, Real-world Anomaly Detection in Surveillance Videos. *Computer Vision Foundation* (2019)
4. S. Dubey, A. Boragule, J. Gwak, M. Jeon, Anomalous Event Recognition in Videos Based on Joint Learning of Motion and Appearance with Multiple Ranking Measures. *Appl. Sci*. (2021)
5. K. Doshi, Y. Yilmaz, Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate. *Pattern Recognition* (2021)
6. A. Rosebrock, Pyimagesearch, [online], Available at: https://www.pyimagesearch.com/2015/10/05/opencv-gamma-correction/ (2015)
7. K. Hara, H. Kataoka, Y. Satoh, Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet? *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Salt Lake City, UT, USA, 18–23, pp. 6546–6555 (2018)

8. L. Wang, Y. Xu, J. Cheng, H. Xia, J. Yin, J. Wu, Human Action Recognition by Learning Spatio-Temporal Features With Deep Neural Networks. *IEEE Access, **6***, pp. 17913-17922 (2018)
9. J. Yu, D. Y. Kim, Y. Yoon, M. Jeon, Action Matching Network: Open-set Action Recognition using Spatio-Temporal Representation Matching. *Visual Computer, 36,* pp. 1457-1471 (2020)
10. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, **39***(6), pp. 1137-1149 (2017)
11. S. Sun, N. Akhtar, H. Song, A. Mian, M. Shah, Deep Affinity Network for Multiple Object Tracking. *Journal of Latex Class Files*, ***13***(9) (2018)
12. G. Ciaparrone, F. L. Sa ´nchez, S. Tabik, L. Troiano, R. Tagliaferri, F. Herrera, Deep learning in video multi-object tracking: A survey. *Neurocomputing*, 381, pp. 61-88 (2020)
13. J. Yu, K. Yow, M. Jeon, Joint Representation Learning of Appearance and Motion for Abnormal Event Detection. *Computer Science, Machine Vision and Applications* (2018)
14. D. Abati, A. Porrello, S. Calderara, R. Cucchiara, Latent Space Autoregression for Novelty Detection. *IEEE Conference on Computer Vision and Pattern Recognition* (2019)
15. L. Wang, F. Zhou, Z. Li, W. Zuo, H. Tan, Abnormal Event Detection in Videos Using Hybrid Spatio-Temporal Autoencoder. *IEEE International Conference on Image Processing* (2018)
16. D. Xu, Y. Yan, E. Ricci, N. Sebe, Detecting anomalous events in videos by learning deep representations of appearance and motion. *Computer Vision Image Understanding, 156,* 117-127 (2017)
17. Y. S. Chong, Y. H. Tay, Abnormal Event Detection in Videos Using Spatiotemporal Autoencoder. *Computer Vision and Pattern Recognition* (2017)
18. J. R. Medel, A. Savakis, Anomaly Detection in Video Using Predictive Convolutional Long Short-Term Memory Networks. *Computer Vision and Pattern Recognition* (2016)
19. J. Brownlee, Generative Adversarial Networks with Python: Deep Learning Generative Models for Image Synthesis and Image Translation. J. Brownlee. *Machine Learning Mastery* (2021)
20. A. Odena, Semi-Supervised Learning with Generative Adversarial Networks. *Machine Learning* (2016)