Irregular Occurrence Detection of Video SurveillanceUsing DeepConvolutionalGAN

[1]Swapna.C ,[1]B. Padmaja Rani

Dept. of CSE. JNTUH, Hyderabad,

India
Email swapnac.jntuh@gmail.com

Abstract.

In modern intelligent surveillance systems, detecting video anomalies has received considerable attention. Regardless of the technical characteristics of current technology, extraordinary event detection in surveillance video structures is difficult and necessitates extensive human efforts. Because of the difficulties associated with anomaly detection, we focused on its evolution and then investigated various methodologies. We propose (DeepConvolution GAN) to detect anomalies and build DCGAN by using Deep Learningwith PyTorch : Generative Adversarial Network. The proposed DCGAN uses frames to increase the image quality produced by the GAN and identify anomaly accurtely. We compute the logists loss function, optimize the load of images and ROC Curve. The proposedwork has extensively evaluated on the CUHK Avenue, UPSC Ped1 dataset at the frame leveland compared with state art methods.

Keywords. DeepConvoltuion GAN, Real image, Fake image, Adam optimizer.

1. INTRODUCTION

Finding anomalous, unexpected, unpredictable, uncommon events or things that don't fitthemould of typical or frequently happening events is known as Anomaly Detection. Nowadayspersonal and private property are increasingly protected important. Real-time video surveillance plays a useful role. Due to these requirements, camera placement occurs at everynook and cranny; When an unexpected action is spotted, a video surveillance system may beused to capture the moment. The primary factor comprehends the action and then reports theimmediately notify the operator or users when an unforeseen incident happens. The effectiveness of video surveillance has improved, and the use of security andsafety measuresfor managing Private and public sectors. Real-time video surveillance systems must be developed due to the growing requirement to ensure public safety in congested regions. It continuously scans the crowd to quickly spot any unusual motions, helping to avert accidents, violent crime, and terrorist strikes [1][2]. A video is a collection of frames with changing image times. Both spatially and temporally are digitized for the image data. Quantization is done on the resulting pixel intensities. The great majority of modern, state- of-the-art techniques for detecting video anomalies rely on intricate neural network designs[21]. Although deep neural networks perform better on a variety of computer visionand machine learning tasks, such as object detection, picture classification, and video surveillance. There are flaws, such as handling losses and the quality of the created images [22, 23].

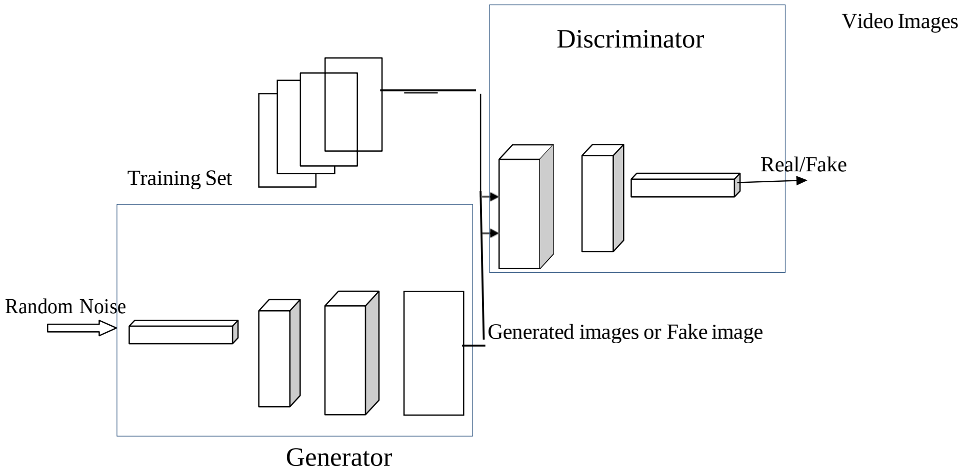
We suggest a DCGAN Methodology in response to the aforementioned domain problems and research gaps and to improve performance. The main objective is to detect anomalies effectively through well-trained Generator and Discriminators.

2. RELATED WORK

Deep learning researchers are becoming increasingly interested in generative adversarialnetworks (GANs) [3][4]. GAN's have been used GANs have been used in a variety of domains, including "computer vision" [5], "Natural language processing" [6], time series applications [7][8], and semantic-based segmentation [8]. GANs are a type of generative anddiscriminative model in machine learning. GANs are superior to other generative models like autoencoders and density functions in that they create required samples quickly, do awaywith deterministic bias, and are internally well-compatible with neural architecture [9]. GANs have shown considerable success, especially in the area of "computer vision", such as image translation [10], image creation [11] [12], high resolution of images [12][13], image reconstruction [14] and video processing.

2.1. Basic GAN Architecture

Figure 2.1 shows how to teach a DL model to use GAN to develop new data from the same dissemination as the training data. Ian Goodfellow created GAN's in2014, and he initially wrote about them in the paper "Generative Adversarial Nets". A Generator and a Discriminator, two separate models, make them up. The generator's job is to create "fake" pictures that look like the practice pictures. The Discriminator's task is to examinea frame and determine whether it is a genuine training frame or a phoney frame produced by the generator. Throughout the training, the Discriminator strives to improve while the generatoris continuously trying to outwit them by producing increasingly convincing fakes.



Figuree 1. Basic GAN Architecture

3. PROPOSED METHODOLOGY

For Generator, DCGAN was the first to use a Deepconvolutional neural network model [15].Deepconvolution is put forward as a method for observing the features of a CNN [16]. WithDCGAN, we get more stable training and high-resolution. At the beginning, DCGAN substitutes a strided convolutional layer for the discriminator and a partial strided convolutional layer for the generator. Next, to locate generated and real images centred at zero, batch normalisation is used for both Generator and Discriminator. Then, in the Discriminator for all layers, LeakyReLU activation is used to prevent the model from beingstruck. All layers in the Generator use ReLU activation except output, which uses Tanh. In this case, while the Generator gets gradients from the Discriminator, the LeakyReLU activation will stop the network from stacking in a "death state" scenario. ImageNet and Large-scale Scene Understanding (LSUN) are used to train DCGANs [17][18]. Using “stochastic gradient descent (SGD)” and a mini-batch, all models were trained. Initializationof all weights used a normal distribution with a centre at zero and a standarddeviation of 1.It uses the Adam optimizer for optimization here. A crucial turning point in the development of the GAN is DCGAN in Figure 3.1.

3.1 Proposed Framework

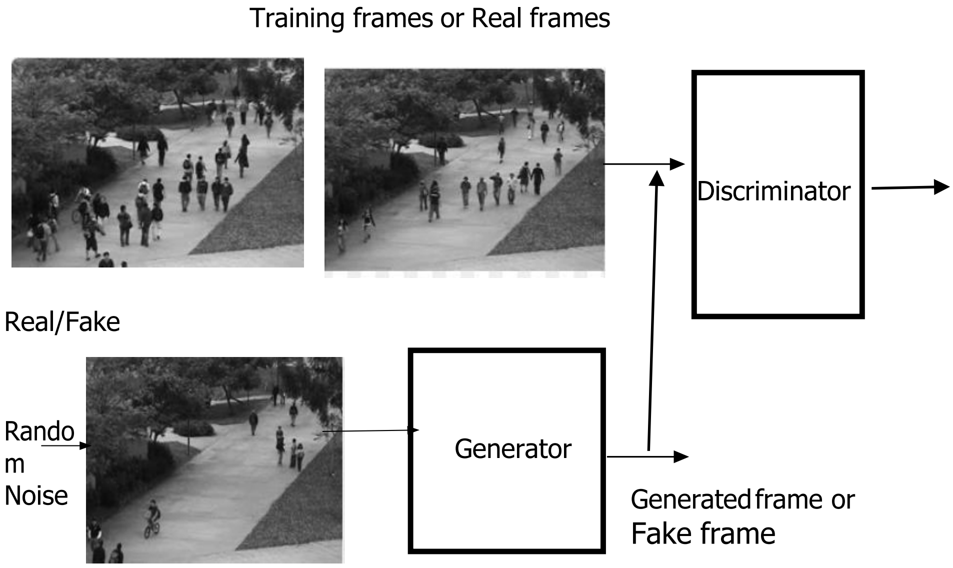


Figure 2. DCGAN Architecture

3.2 Datasets:

CHUK AVENUE DataSet:

21 test videos and 16 training videos with a frame size of 360 x 640 pixels each are includedin the CUHK Avenue dataset. The videos have a total of 30652 frames (15328 for training and 15324 for testing). There are many bizarre scenes, like people tossing objects, fleeing, and jumping on roadways. The total Dataset is divided into 240 batches. The size of a batchis 128.Sample video Frame shown in Figure 3.2.

Figure 3.2. Sample CHUK Avenue Video Frame and Sample UCSD Ped1 Test Video Frame

UCSD Ped1:

A stationary camera installed at a height and looking down on pedestrian pathways was used to collect the UCSD Anomaly Detection Dataset. The walkways had varying densities of people, from very few to many. Bikers, skateboarders, tiny carts, and pedestrians crossing awalkway or in its surrounding grass are examples of often occurring anomalies. A few incidents involving wheelchair-bound individuals were also noted. Allabnormalities are real; they weren't produced to create the dataset. They all occur spontaneously. Two separatesubsets of the data were created, one for each scene. Each scene's video recording was divided into a number of clips, each with about 200 frames.

Peds1: footage showing crowds of people moving in both directions from and toward thecamera, with some perspective distortion. Contains 34 examples of training videos and 36 examples of testing videos. Sample Video Frame Shown in Figure 3.2.

3.3 Implementation Of Discriminator Model

We use the Discriminator D classifies two classes as real and a fake. Discriminator D trainedto do classification between both real and fake classes as shown in Fig1. D computes a binarycross-entropy loss with logits loss. Here network divided into 3 blocks has layersas Conv2d,BatchNorm2d, LeakyReLU after this its flatters the results and next given to fully connectedlayer to get results and forwards from 1 block to next last send to linear layer which covers binary cross entropy with logitsloss which takes raw output.

3.4 Implementation Of Generator Model

With a random noise as an input, the Generator G learns to create an output fake frame. TheGenerator network's goal is to create realistic fake frame. We are using a random noise vector size of 64, which is used to create random noise to feed this to the Generator while training as shown in Fig1. Input is reshaped and given to the Generator as batch size, channels, height, and width. Every block contains layers as ConvTranspose2d, BatchNorm2d, ReLu given to the ConvTranspose2d, Tanh activation function to generateimages. Finally, replace random initialised weights with normal weights to make the network strong for Discriminator and Generator.

3.5 *Training Gan Model*

While training, it calculates Generator Loss and Discriminator Loss starting from 0. TheGenerator uses a real frame and noise to create a fake frame. The Discriminator is given a fake frame and a real frame to predict whether it is real or fake. It calculates fake loss by passing Discriminator prediction with a fake frame. To improve the accuracy here it calculates real loss by using Discriminator prediction with real frame. Discriminator loss byaveraging these 2 losses. The generator follows the same procedures by passing these backward and updating weights for optimization.

3.6 *Optimizing*

Real loss and fake loss are to be calculated. passing logitsloss (a linear function which takesraw input) to compute real loss while considering discriminator prediction and 1 ground truth. To compute a fake loss, consider the discriminator prediction and 0 groundtruth. TheAdam optimizer is used to optimise the discriminator and generator. Atlast computes Total DLoss and Gloss by averaging DLoss and Gloss with batch sizes. To increase the accuracy, we need to increase the epochs shown in Figure 3.6.

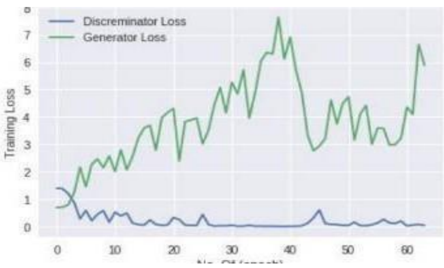


Figure 3. Generator and Discriminator Training Losses in no.of epochs

3.7 *Evaluation of DCGAN*

Frame-level evaluation and pixel-level evaluation are the two different ways that for evaluation. Without localising the anomaly, the model's accuracy is assessed at the frame- level by identifying abnormal frames. On the other hand, pixel-level analysis pinpoints anomalies and compares model output to actual image pixels. A model's detection is deemedto be accurate if it can identify at least 40% of aberrant pixels. The "True positiverate" (TPR) and "False positive rate" (FPR)-based area under the ROC curve (AUC) is acommonly used metric to assess and compare the efficacy of various models. TPR is determined as

$$\text{TPR} = \frac{TP}{P} = \frac{TP}{TP + FN}, \quad (3.1)$$

where TP represents how many frames are accurately identified as abnormal frames and FNrepresents how many frames are mistakenly classified as normal frames. Equation (3.1) canalso be used in pixel-level mode, but the number of pixels should be counted instead of frames. The FPR is calculated as follows:

$$\text{FPR} = \frac{FP}{N} = \frac{FP}{FP + TN}, \quad (3.2)$$

There is no difference between FPR and FNR at EER. FPR calculation shown in Equation(3.2). An algorithm with a lower EER is more accurate and less error-prone. The temporal complexity requirement is another crucial factor. Employing an algorithm in numerous applications is more enticing if its total execution time is swift enough.

3.8 Results using UCSD Peds1 and Chuk Avenue

Models are trained with training data, and then tested with testing data, and finally evaluatedusing ground-truth data by evaluating the system based on the criteria listed above. Figure3.8shows ROC curve.

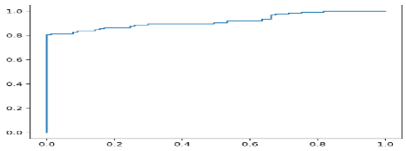


Figure 3.8. The sample ROC curve shows the "True Positive Rate" (TPR) vs the "FalsePositive Rate" (FPR).

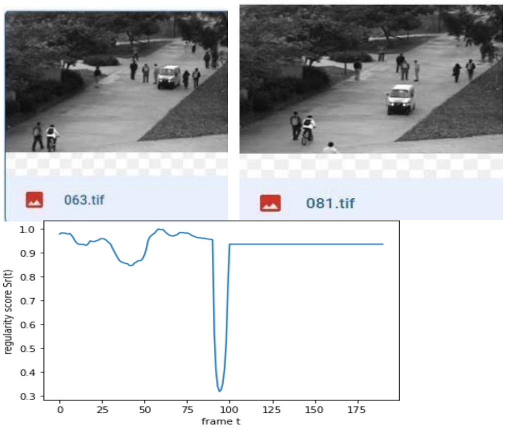


Figure 4. Anomaly frames identified by regularity scores

3.9 Results using UCSD Peds1 and Chuk Avenue

Models are trained with training data, and then tested with testing data, and finally evaluatedusing ground-truth data by evaluating the system based on the criteria listed above. Figure3.8shows ROC curve.

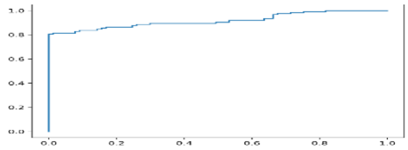


Figure 3.8. The sample ROC curve shows the "True Positive Rate" (TPR) vs the "FalsePositive Rate" (FPR).

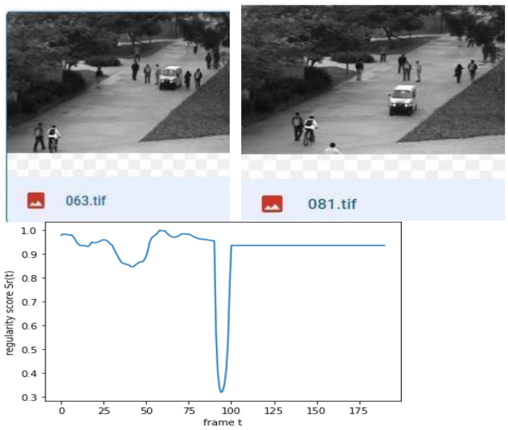


Figure 5. Anomaly frames identified by regularity scores

The model is trained with normal videos of datasets. While testing, we used both normalvideos and anomaly videos with ground truth. Regularity scores were evaluated using threshold-identified anomaly frames. The regularity score for normal frames is high; the regularity score for abnormal frames is low. .To compute ROC, EER, and AUC, find true positives,false positives, true negatives, and false negatives. A sample regularity score of the test videowith anomalous frames from the UPSC Ped1 dataset is shown in Figure 3.9.In Table 3.8, performance metrics are presented as a comparison between the proposed method and the current state of the art Methods(Unsupervised Learning Methods).

Table3.8. Compared to state-of-the-art methods

	CHUK Avenue		UCSD Ped1	
Method	EER	AUC	EER	AUC
Conv-AE[24]	25.1	70.2	27.9	81.1
MLAD[25]	38.82	52.82	23.50	82.34
Our Method	24.6	89.9	36	93

4. FUTURE DIRECTIONS

GANs were initially created to generate realistic fake frames, and they have demonstratedto function very well in the field of computer vision. Despite the fact that there are no reliableevaluation measures for the performance of GANs in these sectors, the use of GANs for time-series data and on morphing photos has not been thoroughly studied. It is recommendedthat more study be done in these areas. We required attention to creating a forgery detector to rapidly and precisely recognize images produced by AI (including usingGANs). Image classification also can be improved by using GAN's[21][22][23].

5. CONCLUSION

Due to various factors influencing outcomes, such as video noise, outliers, and resolution, finding anomalies in video surveillance is difficult. In this research, we present a methodtogenerate images by training on already existing similar images. In the dynamic systemknown as GAN, the optimization process looks for an equilibrium between two forces ratherthan a minimum. In this paper our proposed DCGAN gave more accurate generatedframesthan other methods. The input data is represented both as appearance-motion and motion- only. Consequently, it detects abnormalities in the body's shape, the skeleton, speed, and thedirection as a result. The proposed method is compared with state-of-the- art through experimental case studies. UCSD Peds1 and CHUCK Avenue datasets were used for the experiments. As compared to other methods, the proposed method has a lower EER and higher AUC. Additionally, a study of the time complexity of the proposedmethod is conducted, demonstrating that it is sufficient time efficient. An abnormality ofa frame can be detected and located within a short period of time by utilizing a personal computer. A LSTM can improve the accuracy of the system in the future.

6. REFERENCES

[1] Henry et al., "Anomaly Detection in Videos Recorded by Drones in
[2] aSurveillance Context", International Carnahan Conference on Security Technology (ICCST), Madrid, 2017.

[3] Mohammad Farhadi et al., “AAD: Adaptive Anomaly Detection throughtraffic Surveillance videos”,arXiv:1808.10044, 2018.

[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair,Courville, and Y. Bengio, “Generative adversarial nets,” in Advances in

Neural Information Processing Systems, 2014, pp. 2672– 2680.

[5] M.-Y. Liu and O. Tuzel, “Coupled generative adversarial networks,” in Advances in neural information processing systems, 2016, pp. 469–477.

[6] M. O. Turkoglu, L. Spreeuwers, W. Thong, and B. Kicanaoglu, “A layer-basedsequential framework for scene generation with GANs,” in Thirty-Third AAAI Conference on Artificial Intelligence, Honolulu, Hawaii,United States, 2019.

[7] C. Lassner, G. Pons-Moll, and P. V. Gehler, “A generative model of people in clothing,” in Proceedings of theIEEE International Conference on Computer Vision, 2017, pp. 853–862.

[8] W. Fedus, I. Goodfellow, and A. M. Dai, “MaskGAN: Better text generationvia filling in the .” arXiv preprint arXiv:1801.07736, 2018.

[9] E. Brophy, Z. Wang, and T. E. Ward, “Quick and easy time series generation with established image-based GANs,” arXiv preprint arXiv:1902.05624, 2019.

[10] N. Souly, C. Spampinato, and M. Shah, “Semi supervised semantic segmentationusing generative adversarial network,” in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 5688–5696.

[11] I. Goodfellow, “Nips 2016 tutorial: Generative adversarial networks,” arXivpreprint arXiv:1701.00160, 2016.

[12] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture forgenerative adversarial networks,” arXiv preprint arXiv:1812.04948, 2018

[13] S. Ma, J. Fu, C. Wen Chen, and T. Mei, “DA-GAN: Instance-level image translation by deep attention generative adversarial networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5657–5666.

[14] H. Dong, S. Yu, C. Wu, and Y. Guo, “Semantic image synthesis via adversarial learning,” in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 5706–5714.

[15] D. Mahapatra, B. Bozorgtabar, and R. Garnavi, “Image super-resolution using progressive generative adversarial networks for medical image analysis,”Computerized Medical Imaging and Graphics, vol. 71, pp. 30–39, 2019.

[16] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Generative image inpainting with contextual attention,” arXiv preprint arXiv:1801.07892, 2018

[17] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” arXiv preprintarXiv:1511.06434, 2015.

[18] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutionalnetworks,” in European conference on computer vision. Springer, 2014.

[19] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, “Lsun: Construction of a large -scale image dataset using deep learning with humans in the loop,” arXiv preprint arXiv:1506.03365, 2015.3911–3919.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A

large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition. IEEE, 2009, pp. 248–255
[21] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen,"Improved techniques for training GANs," in Advances in Neural Information Processing Systems, 2016, pp. 2234–2242.
[22] D. Parashar and D. K. Agrawal, "Classification of glaucoma stages using image empirical mode decomposition from fundus images," Journal of Digital Imaging,2022, doi: 10.1007/s10278-022-00648.
[23] D. Parashar and D. K. Agrawal, "Computer-aided diagnosis of glaucoma at early-stage using retinal fundus images" 2021Tenth IEEE International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), Jaipur, India, December 2021, pp. 1 -4, doi: 10.1109/IEMECON53809.2021.9689175.
[24] T.Jagadeeswari,Dr.P.Srinivas,Dr.Y.L.MalathiLatha "Deep Learning based Convolutional Neural Networks (DLCNN) on Classification Algorithm to Detectthe Brain Turnor Diseases using MRI and CT Scan Images.
[25] Yizhe Zhang, Zhe Gan, Kai Fan, Zhi Chen, Ricardo Henao, Dinghan Shen, and Lawrence Carin. Adversarial feature matching for text generation.arXiv preprintarXiv:1706.03850, 2017.
[26] MIKAEL SABUHI1 , (Student Member, IEEE), MING (CHLOE) ZHOU1 , (Student Member, IEEE), COR-PAUL BEZEMER1 , (Member, IEEE) and PETR MUSILEK1,2 (Senior Member, IEEE) Applications of Generative Adversarial Networks in Anomaly Detection: A Systematic Literature Review,IEEE Access VOLUME 4, 2016.

Biographies

Mrs. Swapna. C has completed her M.Tech in Computer Science and Engineering and presently she is on the verge of completing her Doctoral degree in the area of Machine Learning at Jawaharlal Nehru Technological University Hyderabad. She is having 10+ years of teaching experience, and her research area of interest are MachineLearning, Video Surveillance, Anomaly Detection, cand rime prediction.

Dr. Padmaja Rani received her philosophy of doctorate degree in Computer Science from Jawaharlal Nehru Technological University Hyderabad in 2009. She is currently working as Professor at the Department of Computer Science Engineering, JNTUH. HerAreas of Interest are Information Retrieval Embedded Systems, Natural Language Processing, Data Leakage Prevention, Big Data, Information Security, and Cloud Computing. She published more than 40 research papers in reputed journals