International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.180

Multihead Driven Self Attention Adversarial Learning based One Class Classification (MDSAL-OCC) for Medical Image Screening

Vijayalakshmi Pasupathy
Research Scholar
Saveetha Institute of Medical and Technical Sciences
Chennai, Tamil Nadu, India
pasramviji@gmail.com

Dr. Rashmita Khilar
Professor, Institute of Information Technology
Saveetha School of Engineering
Saveetha Institute of Medical and Technical Sciences
Chennai, Tamil Nadu, India
rashmitakhilar.sse@saveetha.com

Abstract — **One-class classification (OCC) is used to construct classification models even though the outlier samples are inadequate, weakened and not clearly defined samples. The OCC network has been dominantly employed in diverse application of machine learning. One-Class Classification (OCC) is an exceptional condition of multi- class classification in which the data used during the training phase is generated from a single positive class. The intent of a OCC network is to learn a representation and a classifier that deals with positively labeled queries. In the recent years adversarial learning One-class classification (ALOCC) method has outperformed the efficiency of the OCC performance. Since it has some limitations such as instability within the training phase and issues in reconstruction of data between inlier and outlier data. In this paper work, we have propounded a Multi head Driven Self Attention mechanism incorporated in the ALOCC network. The proposed MDSAL-OCC network outperforms other self attention network in OCC and effective in elevating the OCC accuracy. In this paper, we also presented a discussion about the comprehensive elaboration of the proposed framework of one-class classification Multi head Driven Self Attention Adversarial Learned One Class Classification (MDSAL-OCC) and also discussed about the role of MDSAL-OCC in train One Class Classification networks in detecting fetal congenital heart diseases (FCHD).**

Index Terms* — *One-class classification (OCC), Adversarial Learning One-class Classification (ALOCC), Multihead Driven Self Attention Adversarial Learned One Class Classification (MDSAL-OCC), Fetal Congenital Heart Diseases (FCHD), Deep Convolutional Neural Networks (DCNNs), Bi-Directional GAN(BiGAN), Self Attention Generative Adversarial Network (SAGAN)

I. INTRODUCTION

The model One-class classification (OCC) builds a predictive analysis model using poorly sampled data. Over the past decade, Deep Convolutional Neural Networks (DCNNs) has exhibited remarkable performance improvements in object detection. In this paper, the work is primarily focused on image detection in medical image analysis. Because of the availability of huge multi- class annotated datasets, deep networks can easily learn discriminative features and also the classifier can exploit to perform recognition. One-class classification (OCC) is defined as the concept of conceding test data which has been disseminated in a different way from the data presented during the training phase [2]. This method of classification has established significant research in various areas of domains such as biomedical field [3], fault diagnosis detection system in manufacturing systems [4], financial systems [5], and security and communication systems [6]. The data's of different classes needed to be labeled in all the applications. But in the real- world scenario, data's from unusual classes tend to be inadequate volume for effective modeling. In order to overcome the former situation, the OCC does the task of performing training only on the usual classes and latter identifies the unusual classes from the usual class distribution. Thus, OCC differ from the traditional classification network. OCC also solves imbalanced dataset issues in an effective way. Many methodologies are employed to explain one - class classification (OCC). These approaches are categorized in to different methods. The Density estimation methods emphasize on calculating the density of the data points, and threshold value setting. The boundary methods are based on the parameters such as distances and boundaries encompassed within a set of data points called target points and optimizes the volume of the data. The reconstruction method relies on building a model using prior knowledge [1].

In the contemporary situation, the end-to-end networks are widely used for one-class classification. Dimokranitou et al. [7] introduced an integrated framework comprising of Autoencoders and Generative Adversarial Networks. This work reached remarkable benchmark performance in detecting anomalies in video sequences. In 2018, Zenati et al. [13] propounded about the GAN based Anomaly detection system, which can produce excellent results than the earlier one. In the latter work, GAN models were implemented for anomaly detection in medical images and network intrusion. The GAN called BiGAN is used in this work and applied combined training for bidirectional map-ping between the image space and the latent space which achieved better performance accuracy. In a research work by Akcay et al. [8], using conditional generative adversarial network (CGAN) has been used to develop an anomaly detection model which utilizes learning of high-dimensional image space and the inference of latent space. This model contains an encoder- to- decoder and a decoder - to - encoder networks. In the generator part, mapping of the input image to a lower dimension is carried out. Later, the lower dimension vector generated by the Generator network is in turn used to reconstruct the output image. This model achieves better efficacy over several benchmark datasets, domains and previous state-of-the-art approaches.

DOI: rp-9788770040723.180

Later in a research work conducted by Sabokrou et al. [9], reconstruction-based OCC network which consist of encoder-decoder architecture was developed. In this architecture, two deep networks were used. In that architecture, one network acts as a novelty detector, whereas the other network enhances the inliers samples and deforms the outlier class samples using adversarial learning technique. This framework can be incorporated in different domains for the application of anomaly detection in image and video processing system. Later, the attention techniques were introduced with the aspect of diminishing the quantity of parameters, thus improving the accuracy of the classification models [14]. Self attention Generative Adversarial Network (SAGAN) [15], proposed by Zhang, elucidated about the framework consisting of the technique of self-attention incorporated in a GAN framework. It achieved best GAN efficiency specifically on class-conditional image generation in ImageNet. Zhang, Yingying, et al.[16] Proposed an improvised self-attention mechanism embedding multi-heads which outperformed the earlier methods in improving the aspect of reconstruction quality and efficiency. In unified approach comprising GAnomaly and Adversarial One Class Classification, it is very tedious to perform the reconstruction process between the inliers and outliers class labels. For better reconstruction of inliers class samples, the visual attention mechanism is incorporated in to OCC networks. Apart from that, the newly proposed self-attention technique uses features in various latent subspaces to provide observation controls to improve the inliers class labels. It can be used in broad in various anomaly detection networks such as medical image analysis, security systems, etc.

II. Related Work

A. Adversarial Learning-based Network

GANs [19][20] is a great achievement developed for the purpose of performing data augmentation used in learning models. GANs can also be used as a classification model in the case of insufficient training data. GANs are unsupervised deep networks generally used for the data generation. The supervisory information is obliquely provided by an adversarial gaming between two networks: a generator (G) and a discriminator (D). In the training phase, G generates new data and D understands the new data generated by the G and evaluates the condition that the data is a real or a fake input.

B. Attention

Attention is one of the most important emerging technique in the area of machine learning. The attention mechanism has been stimulated by the biological organization of human being aimed at differentiating various organ substructures while handling huge volumes of information [23]. Visual attention is a technique which incorporates some cognitive operations for handling issues in large volumes of data in an efficient manner by extracting valid information and excluding invalid information. Attention is a amenable technique applied on regions of interest such as specific features of an object, or the whole objects. Attention mechanism is widely used in different domains such as computer vision, natural language processing etc,. In computer vision, visual attention mechanism is incorporated in to the neural network structure for achieving better performance results. In neural networks, attention mechanism performs various tasks such as image caption generation [25, 26], image-based analysis [27,28], machine translation [29], text classification [30,31], action recognition [32], speech recognition, recommendation [34], and graph [35] in enhancing performance. Apart from this, it also serves as a tool for elucidating neural network architecture behaviour.

TABLE I. ATTENTION MECHANISM AND TYPES BASED ON VARIOUS CRITERION

Criteria	**Type**
The Softness of Attention	Soft/hard, global/local
Forms of Input Feature	Item-wise, location-wise
Input Representations	Distinctive, self, co-attention, hierarchical
Output Representations	Single-output, multi-head, multi-dimensional

C. Attention Mechanism

The different approach of attention mechanism has been classified based on the aspects such as softness related to attention, different types of input features, various input representations and output representations. The self-attention network [36] encompasses an attention mechanism involving different sequence correlated with one another for generating a prototype of a same. It finds usage in machine reading comprehension, abstractive summarization in natural language processing, or image description generation. A multihead self-attention [37] consist of a network called Transformer and employs the data from various multi-subspaces to improvise the process of feature extraction . This multihead enhanced self-attention technique [38] shows exemplary performance results thereby improving the accuracy as well the reconstruction efficiency.

III. Adversarial Learning One-class classification (ALOCC) network

The Adversarial Learning One-Class Classification network contains two components ‘R’ called a reconstruction network and ‘D’ called a Discriminator network. The component ‘R’ reconstruction network is an encoder-decoder part,

The encoder part GE(z) is a convolutional network used for compressing noisy input image ‘z’ to feature vector ‘s’. The noisy image z is denoted as follows:

$$z = (x \sim pt) + (\eta \sim N(0, \sigma^2 I)) \quad (1)$$

In the above representation of noisy image z, x denotes original image, pt denotes the inliers class samples distribution, The parameter ‘η’ denotes the normal distribution noise .The standard deviation present in the normal distribution noise is denoted as ‘σ’. In ALOCC network, the image noises are employed to upgrade the

International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.180

robustness of the Reconstruction network 'R'.

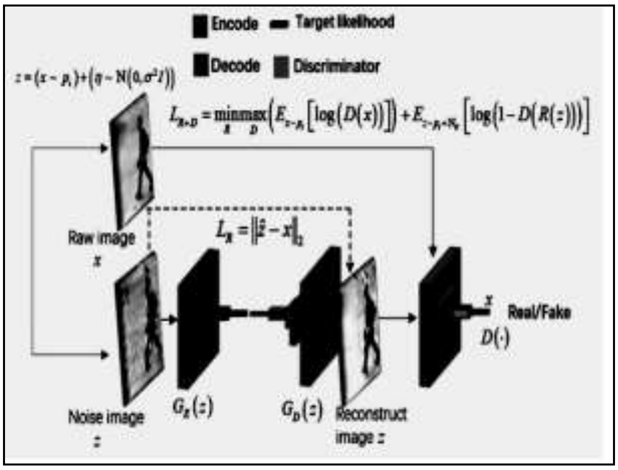

Fig. 1 Functionality of an ALOCC Network

The decoder network G_D (z) performs up scaling. In up-scaling, the vector s is converted to the reconstructed image zˆ. The discriminator D determines whether the reconstructed image zˆ is in line with the target class p_t. The component discriminator 'D' also evaluates the condition that if the input has been derived from the target class or not. The target class is denoted by p_t .

The ALOCC network, reconstruction and discriminator (R+D) are jointly adversarial learned in order to optimize a objective function represented as below,

$$L_{R+D} = min\ max\ (E_{x\sim p}t\ [\log (D(x))]) + E_{z\sim p}t + N\sigma \ [\log (1 - D(R(z)))] \quad (2)$$

The training process involves the task of integrating the loss L_{R+D} of joint network R + D and the contextual loss L_R between the original and the derived image in the ALOCC network.

The contextual loss L_R is denoted as follows,

$$L_R = ||z\hat{} - x2\ || \quad (3)$$

The overall loss function in an ALOCC network is represented as,

$$L=LR+D + \lambda\ \ LR \quad (4)$$

The trained network is represented as,

$$R(z; \theta_r) \quad (5)$$

The trained network is used to reconstruct and enhances the inliers class samples. The ALOCC uses $D(R(x))$ to represent OCC.

The complete OCC process is denoted as,

$$OCC\ (x) = \text{target } class\ if\ D(R(x)) > \tau \quad (6)$$

If D(R(x)) is higher than the τ, sample x is treated as an inlier class. If D(R(x)) is lower than the τ, sample x is treated as an outlier class.

In multi-head attention mechanism, input sequence is mapped to various subspaces based on the parameters, followed by attention mechanism employing scaled dot-product attention to its representation in each and every subspace. Then the output of each subspaces are concatenated together to generate the output. The multi-head attention mechanism imparts a unified approach to process information linearly from various representation subspaces achieved at different levels of positions [37].

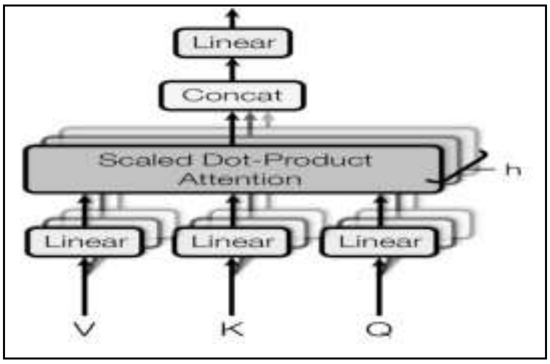

Fig. 2 Block diagram of a Multi-Head Attention mechanism.

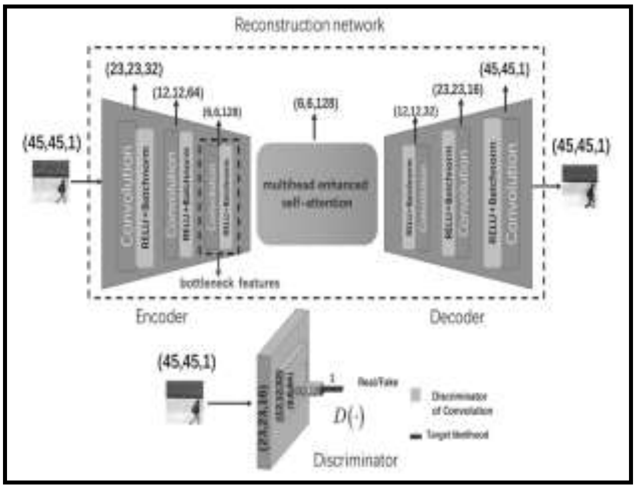

Fig. 3 Block diagram of a MDSAL-OCC network

In self-attention mechanisms, a single head performs feature learning at various subspaces, so it has to deal with issues such as handling different types of information, integrating representation in all subspaces for improving the contextual information. Multihead enhanced self-attention network can be applied to the concept of image attention [38].

IV. MULTIHEAD DRIVEN SELF ATTENTION ADVERSARIAL LEARNING BASED ONE CLASS CLASSIFICATION (MDSAL-OCC)

Multi-subspaces features captures different range of correlations in various latent spaces and merging of representation leads to obtain an improved reconstruction. But if the heads are increased, the proposed attention

DOI: rp-9788770040723.180

mechanism will minimizes the dimensionality of a single head. If the number of heads is more, it will lead to insufficient quantity of relevant feature. Hence it has been inferred that the number of heads and convolutional channels automatically affects the performance of attention mechanism. Both the factors needed to be maintained in a bias condition. A simple image seldom requires more number of heads . The reason is that the multi subspace rrepresentations cannot able to extract more information from it because of over-fitted representations. The latter in turn leads to reduce the quality of reconstruction process. The proposed MDSAL-OCC overcomes the training problems of adversarial learning such as parallelism, GAN failure, vanishing gradient problem, and unbalancing between the GAN components. The proposed attention employs more 1 × 1 convolutions and two channels of self-representation. The channels are used to determine the attention map in a multi-feature space. In this mechanism, loss function is reduced. The minimized loss function is achieved in two aspects through reconstruction loss and contextual loss. In order to improve the performance of the reconstruction network, a new adversarial-balance loss in the discriminator loss is implemented. The objective behind the adversarial-balance loss is to reduce the disimilarity between the real and reconstructed fake images.

A. MDSAL-OCC in Fetal Congenital Heart Disease Detection

Multihead driven self attention adversarial learning based one-class classification network is exceptionally satisfactory in screening fetal congenital video slices. In the traditional ALOCC network, the encoder cannot does the task of extracting features from complex fetal ultrasound heart images. In the proposed MDSAL-OCC network, feature learning is done linearly through multihead attention mechanism. Screening videos in medical image analysis to obtain FCH video slices using the proposed ALOCC network are useful for the medical practitioners to easily extract the FCH images. The proposed mechanism also provides solutions for acquiring video slices from a video dataset. The acquired video slices both normal and diseased are used for data augmentation.

The proposed MDSAL-OCC model is an integrated framework combining convolutional network incorporating multi-heads in attention network. The proposed model needs to implement a consistent mathematical relation between the convolution layers and heads. The convolutional layer plays a significant role in constructing neural networks specific to image processing domain.

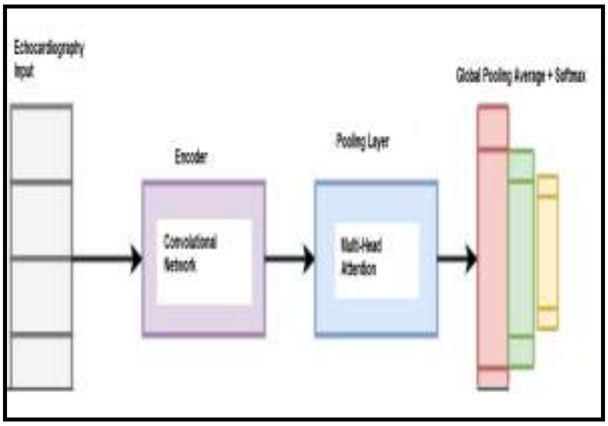

Fig. 4 Proposed MDSAL-OCC Model

In the proposed model, for processing an image of representation $W \times H \times C_{in}$

The CNN model incorporates several convolutional and sampling layers. The filters kernel sizes are represented as '$k \times k$'. Input feature is represented as C_{in} and output dimension as C_{out}. The bias vector is denoted by b. The self attention layer is implemented by the following functionalities such as D_k, denotes the key/query size, D_h, denotes the head size, N_h denotes number of heads and D denotes the output dimension .

The key is denoted as, $M_{key}^{(h)}$

The query matrix is denoted as $M_{qry}^{(h)}$,

The value matrix is denoted as, $M_{val}^{(h)}$ for each head h,

For concatenating all the heads together, the projection matrix denoted as *Mout* .

The n the layers in the models are parameterized as follows,

$$X_q = M_{qry} M_{key} M_{val}$$

Blocks	Parameters
Convolutional Blook	Input Conv. Layer 1 (Stride 1 x 1, Padding 'same') Batch Normalization ReLU 0.5 Dropout Conv. Layer 2 (Stride 1 x 1, Padding 'same') Batch Normalization ReLU Conv. Layer 3 (Stride 1 x 1, Padding 'same') Batch Normalization Softmax 0.5 Dropout Conv. Layer 4 (Stride 1 x 1, Padding 'same')
	Sequence Input
Multi-Head Block	Encoding Layers Heads (4 / 8) 0.3 Dropout
Output Block	Global Pooling Average Softmax

$$HeadsValue, H_q(h) = \sum_{[M]\times[H]} softmax(X_q^h) M_{val}^h$$

$$OutputValue, Y_q = concat(H_q^{(1)}, ..., H_q^{(N_h)}) M_{out} + b$$

Fig. 5 Summarization of the proposed model

DOI: rp-9788770040723.180

B. *Experiments*

For detecting fetal congenital heart disease, there are four views used in fetal echocardiography. The four views used in echocardiography are four chamber view, Three-vessel Trachea view, Three-vessel view and Five-chamber view. Among all the views, the four-chamber view was considered to be optimal one, since it has the capability to visualize the overall structure of the heart. The fetal echocardiography was taken during the second trimester anomaly scan. The fetal cardiac imaging is a tedious procedure to construct a consistent classification model to detect anomalies. For this model, we gathered the dataset from various diagnosis centers in and around our place of locality. The datasets were acquired from pregnant women of gestational age of 21-38 weeks. The dataset comprises of both normal and disease videos. Initial the dataset was preprocessed by converting the raw image to normalized image. Then cropping of ROI is done. Then the normalized image is converted to grayscale image. In the training phase, data synthesis using GAN is carried out. Data synthesis is done in order to adapt to GAN. The model is validated to ensure the strength of the model by increasing the training data in each class. In the training phase, epochs have been set to be 3000 and the epoch's interval has been set to be 300.

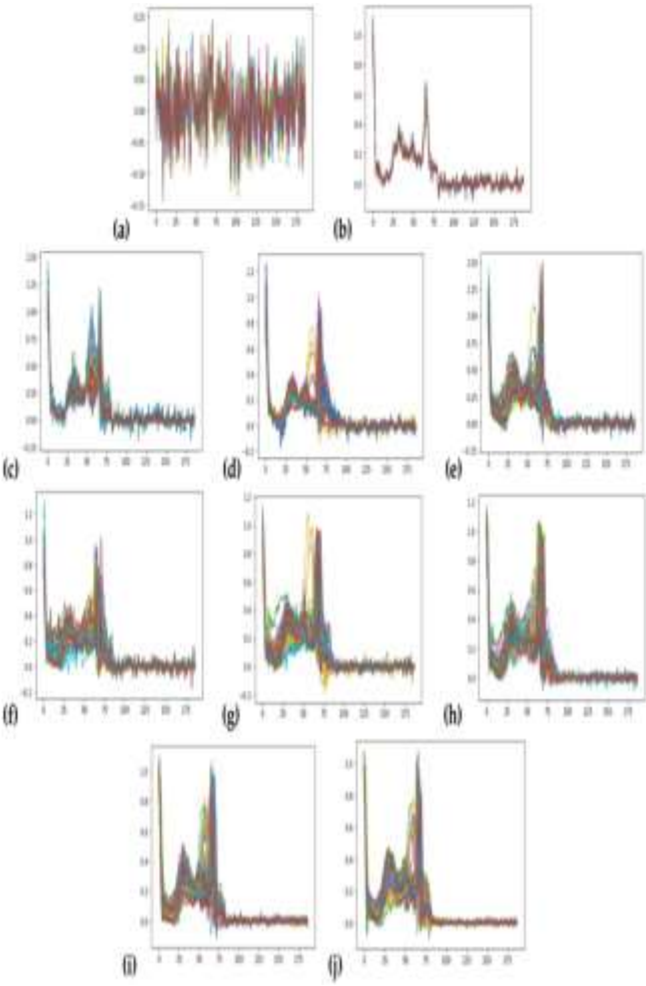

Fig. 6 Diagram depicting the GAN during Data Synthesis in 3000 Epochs.

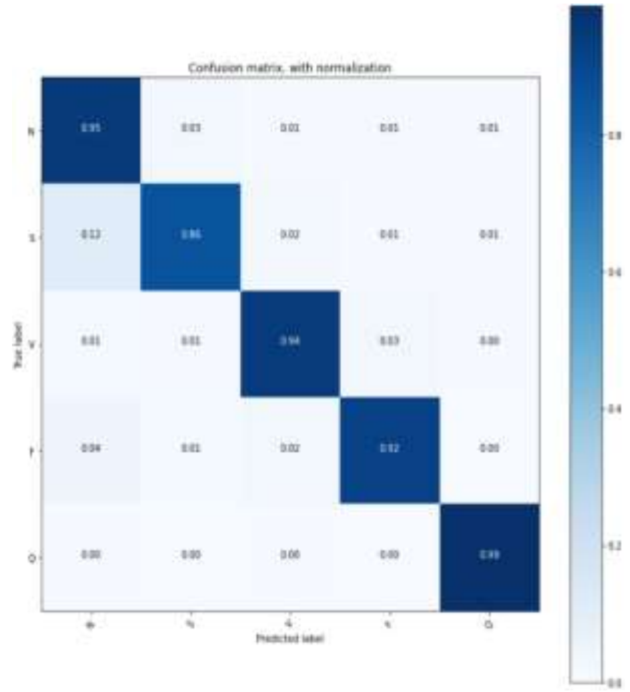

Fig. 7 Confusion matrix of our model.

The proposed model is a multi-classification model. It outperforms the general Adversarial one class classification. It achieved an accuracy performance of more than 85 % in detecting fetal congenital heart disease.

V. Conclusion

In this paper, we illustrated how the Adversarial learned One-class Classification network and described in detail about the attention and attention mechanisms. Then we presented a elaboration about the Multihead self attention network architectures used in machine vision and natural language generation and understanding. It has been presented about the various aspects of attention mechanisms provided for the generating a classification model in collaboration Adversarial Learned One-class Classification network. In conclusion, it has been explained about Multihead Driven Self Attention Adversarial Learning based One Class Classification (MDSAL-OCC) used in various fields of deep learning applications and specifically explained about the context with reference to medical image analysis domain. It has been inferred that MDSAL-OCC can attain predictive performances with equivalent model complexity (i.e., number of parameters).

References

[1] Perera, Pramuditha, Poojan Oza, and Vishal M. Patel, "One-class classification: A survey," arXiv preprint arXiv: 2101.03064, 2021.

[2] M. Filippone, and G. Sanguinetti, "Information theoretic novelty detection", Pattern Recognit, vol. 43, no. 3, pp. 805–814, 2010, https://doi.org/10.1016/j.patcog.20 09.07.0 02.

[3] J. Mehta, and A. Majumdar, "RODEO: robust DE-aliasing autoencOder for real-time medical image reconstruction," Pattern Recognit. Vol. 63, pp. 499–510, 2017, https://doi. org/10.1016/ j.patcog.2016.09.022.

[4] D. Martinez, R.O. Fontenla, B.A. Alonso, and Principe, "Fault detection via recurrence time statistics and one-class classification",

DOI: rp-9788770040723.180

Pattern Recognit. Lett. vol. 84, pp. 8–14, 2016, https://doi.org/10.1016/j.patrec.2016.07.019 .

[5] M. Ahmed, A.N. Mahmood, and M.R. Islam, “A survey of anomaly detection techniques in financial domain,” Futur. Gener. Comput. Syst., vol. 55, pp. 278–288, 2016, https://doi.org/10.1016/j.future.2015.01.001.

[6] M. Ribeiro, A. Lazzaretti, and H. Lopes, “A study of deep convolutional auto-encoders for anomaly detection in videos,” Pattern Recognit. Lett., vol. 105, pp. 13–22, 2017, https://doi.org/10.1016/j.patrec.2017.07.016 .

[7] A. Dimokranitou, “Adversarial autoencoders for anomalous event detection in images.”,PhD thesis, Purdue University, 2017.

[8] S. Akcay, A. Atapour-Abarghouei, and T.P. Breckon, “GANomaly: semi-supervised anomaly detection via adversarial training, in: C. Jawahar,” H. Li, G. Mori, K. Schindler (Eds.), Computer Vision – ACCV 2018. Lecture Notes in Computer Science, Eds., Springer International Publishing, Cham, pp. 622–637, 2018.

[9] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli, “Adversarially learned one-class classifier for novelty detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 3379–3388, Jun. 2018.

[10] S. Akcay, A. Atapour-Abarghouei, and T.P. Breckon, “GANomaly: semi-supervised anomaly detection via adversarial training,” in: C. Jawahar, H. Li, G. Mori, K. Schindler (Eds.), Computer Vision – ACCV 2018. Lecture Notes in Computer Science, Eds., Springer International Publishing, Cham, pp. 622–637, 2018.

[11] A. Dimokranitou, “Adversarial autoencoders for anomalous event detection in images”, PhD thesis, Purdue University, 2017.

[12] V. Turchenko, E. Chalmers, and A. Luczak, “A deep convolutional auto-encoder with pooling - Unpooling layers in caffe,” 2017, arXiv:1701.04949. [Online]. Available: https://arxiv.org/abs/1701.04949.

[13] H. Zenati, C.S. Foo, B. Lecouat, G. Manek, and V.R. Chandrasekhar, Efficient GAN-Based Anomaly Detection, in: Workshop track - ICLR2018, pp. 1–7.

[14] E. Fidalgo, E. Alegre, C.V. Gonzalez, and R.L. Fernandez, “Boosting image classification through semantic attention filtering strategies,” Pattern Recognit. Lett., vol. 112, pp. 176–183, 2018, https://doi.org/10.1016/j.patrec.2018.06.033.

[15] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, “Self-attention generative adversarial networks”, arXiv:1805.08318, 2018.

[16] Zhang, Yingying, et al., “Multi-head enhanced self-attention network for novelty detection,” Pattern Recognition, vol. 107, p. 107486, 2020.

[17] S. Singh, and M. Markou, “An approach to novelty detection applied to the classification of image regions,” IEEE Trans. Knowl. Data Eng., vol. 16, pp 396–406, 2004, https://doi.org/10.1109/tkde.2004.1269665.

[18] W.W.Y. Ng, G. Zeng, J. Zhang, D.S. Yeung, and W. Pedrycz, “Dual autoencoders features for imbalance classification problem,” Pattern Recognit., vol. 60, pp. 875–889, 2016, https://doi.org/10.1016/j.patcog.2016.06.013.

[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” In Advances in Neural Information Processing Systems, pp. 2672–2680, 2014.

[20] 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, pp. 2234–2242, 2016.

[21] M. Ravanbakhsh, E. Sangineto, M. Nabi and N. Sebe, “Training Adversarial Discriminators for Cross-Channel Abnormal Event Detection in Crowds,” 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1896-1904, 2019, doi: 10.1109/WACV.2019.00206.

[22] Dhanabalan, S. S., Sitharthan, R., Madurakavi, K., Thirumurugan, A., Rajesh, M., Avaninathan, S. R., & Carrasco, M. F. (2022). Flexible compact system for wearable health monitoring applications.Computers and Electrical Engineering, 102, 108130.

[23] Zhaoyang, NiuaGuoqiang, and Zhonga HuiYub, “A review on the attention mechanism of deep learning, Neurocomputing”, vol. 10 pp. 48-62, September 2021, https://doi.org/10.1016/j.neucom.2021.03.091.

[24] Itti, Laurent, Christof Koch, and Ernst Niebur, “A model of saliency-based visual attention for rapid scene analysis,”" IEEE Transactions on pattern analysis and machine intelligence 20.11, pp. 1254-1259, 1998.

[25] K. Xu, J. Ba, R. Kiros, K. Cho, A.C. Courville, R. Salakhutdinov, R.S. Zemel, and Y .Bengio, “Show, attend and tell: neural image caption generation with visual attention,” in: ICML, JMLR Workshop and Conference Proceedings, JMLR.org, vol. 37, pp. 2048–2057, 2015.

[26] J. Lu, C. Xiong, D. Parikh, and R. Socher, “Knowing when to look: Adaptive attention via a visual sentinel for image captioning”, in: CVPR, IEEE Computer Society, pp. 3242–3250, 2017.

[27] K. Song, T. Yao, Q. Ling, and T. Mei, “Boosting image sentiment analysis with visual attention,” Neurocomputing, vol. 312, pp. 218–228, 2018.

[28] X. Yan, S. Hu, Y. Mao, Y. Ye, and H. Yu, “Deep multi-view learning methods: a review”, Neurocomputing, 2021.

[29] I. Sutskever, O. Vinyals, and Q.V. Le, “Sequence to sequence learning with neural networks,” in: NIPS, pp. 3104–3112.

[30] Gomathy, V., Janarthanan, K., Al-Turjman, F., Sitharthan, R., Rajesh, M., Vengatesan, K., &Reshma, T. P. (2021). Investigating the spread of coronavirus disease via edge-AI and air pollution correlation. ACM Transactions on Internet Technology, 21(4), 1-10.

[31] Y. Li, L. Yang, B. Xu, J. Wang, and H. Lin, “Improving user attribute classification with text and social network attention”, Cogn. Comput. vol. 11, pp. 459–468, 2019.

[32] S. Song, C. Lan, J. Xing, W. Zeng, and J. Liu, “An end-to-end spatio-temporal attention model for human action recognition from skeleton data,” in: AAAI, AAAI Press, pp. 4263–4270, 2017.

[33] J. Chorowski, D. Bahdanau, K. Cho, and Y. Bengio, “End-to-end continuous speech recognition using attention-based recurrent NN: first results,” CoRR abs/ 1412.1602, 2014.

[34] S. Wang, L. Hu, L. Cao, X. Huang, D. Lian, and W. Liu, “Attention-based transactional context embedding for next-item recommendation,” in: AAAI, AAAI Press, pp. 2532–2539, 2018.

[35] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph attention networks,” in: ICLR (Poster), OpenReview.net, 2018.

[36] P. Shaw, J. Uszkoreit, and A. Vaswani, “Self-attention with relative position representations, The North American Chapter of the Association for Computational Linguistics (NAACL), 2018.

[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, in: Neural Information Processing Systems (NIPS), LongBeach, CA, pp. 5998–6008, 2017.

[38] Yingying Zhang, Yuxin Gong, Haogang Zhu XiaoBai Wenzhong Tang, Multi-head enhanced self-attention network for novelty detection, https://doi.org/10.1016/j.patcog.2020.107486, 2020.