



Surveillance Synthesis: Elevating Crowd Monitoring Using EERGAN with LSSVM Integration

Vasudevan Iyandurai^{*ID}, Rajivkannan Athiyappan^{ID}

Computer Science and Engineering, K. S. R. College of Engineering, Tiruchengode 637215, India

Corresponding Author Email: duraiiv32@gmail.com

Copyright: ©2025The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420225>

ABSTRACT

Received: 12 December 2024

Revised: 27 March 2025

Accepted: 8 April 2025

Available online: 30 April 2025

Keywords:

object detection, feature extraction, Generative Adversarial Network (GAN), classification, prediction, support vector machine, image processing

Object detection and feature extraction in video surveillance systems is one of the most demanding tasks in computer vision for public safety and security in crowded areas. Most of the Generative Adversarial Network (GAN) models are detecting abnormal objects in public places less accurately. The Spiking Generative Adversarial Network (SPGAN) has detected objects in crowded areas, which is required for more energy efficient processing, less robustness and more latency prediction. In this research, propose novel approaches like Adversarial Variational Auto-Encoders GAN (AVAEGANs), Super Resolution Synthetic GAN (SRSGAN), and Enhanced Energy Regularized GAN (EERGAN). AVAEGANs are used to detecting objects accurately and SRSGAN model enhances quality object detection. EERGAN is conjunction with Least Square Support Vector Machine (LSSVM) and effectively detects objects with lower energy consumption (135 mJ), energy reduction (32%), and prediction latency (28ms) and enhances feature extraction accuracy through similarity metrics, utilizing benchmark data sets such as UCSD PED-1, Shanghai Tech and User collected data sets. This proposed framework EERGAN strengthens with the help of LSSVM for classification of aggressive and non-aggressive behavior detection and achieves the performance of accuracy 97.7%, precision 95.68%, recall 95.99%, and F-score 98.21%. The EERGAN produces more robust and fast object detection, ensuring public safety and security.

1. INTRODUCTION

The crowd's abnormal behavior identification is an essential technique for intelligent surveillance and although abnormal behavior in crowds and complicated backgrounds continue to be problems, the crowd counting model presented in this research is built on a GAN structure called a multiscale network. This consists of a scale module, a discriminator module, and a generation module. The generation network includes a backbone model and a multi-branch dilation convolution structure, and the discrimination network checks the intermediate results. The model also includes missed connections to maintain the structure and contextual information of input images. These GANs are also best suited for object detection and tracking. A number of other applications are being used in real-world applications for object tracking and detection, which consists of video surveillance security systems, traffic monitoring and face recognition. However, the machine learning-based object detection and tracking is problematic because all real-world photographs would vary in light, angle, changes, and occlusions. It is to be noted that convolutional neural networks (CNNs) successfully handle perspective fluctuations and illumination problems, though there is a serious variance in adversarial attack threats. The study also focuses on how to bridge the gap between the objectives of the developers and

the performance of algorithms by training machine learning models against adversarial threats [1-3].

Some scholars explained that the complexity of structural patterns across frames is the main topic of this work, which discusses the difficulties in detecting anomalous behavior in highly congested environments. To train behavioral patterns, we use the cycle GAN system, which incorporates social force and optical flow patterns. To test the system's accuracy using two models of normal and aberrant behavior. They employ geometric approaches to enhance anomalous patterns. Using the cycle GAN system, to train and evaluate the accuracy of anomalous behavior detection [4-6]. The study reveals that this technique outperforms other recent works in terms of performance and accuracy. Some research aims to improve detection performance by proposing compact models for object identification, detection, and feature extraction, incorporating a GAN-based super-resolution step [7-9]. This method solves the problem of small object pixel counts that arise from relying solely on CNN models. The model builds on the baseline architecture for ESRGAN in improving super-resolution output, incorporating a two-step spatial and channel attention mechanism [10, 11]. This attention mechanism combined with ESRGAN reduces training time and increases feature extraction efficiency to improve the performance of small object detection [12]. The previous studied explained the applications of spiked neural networks (SSN) to detecting the

moving targets in smart secure video surveillance and assigning them to appropriate categories [13, 14]. The suggested SNN classifies the moving target into different categories based on their distance from the categorization centers determined by the Hebb learning rule. It serves the visual cortex to detect motion of animals with axonal delays and utilizes spike trains to learn data, including feature parameters of targets or grey value of pixels. Simulation results demonstrate the feasibility of the proposed SNN for intelligence computing and its suitability to a variety of visual surveillance systems. Kim et al. described that the Deep Neural Networks (DNNs) are low-powered; their application in machine learning is growing. However, because of non-differential spike operation and complex neuron dynamics, they suffer a significant performance loss. In order to fix the problem that SNNs are getting worse at finding objects, this study shows two new methods: channel-wise normalization and signed neurons with uneven thresholds [15-17].

On a non-trivial dataset, the spiked-based real-time object identification approach, known as the Spiking with YOLO model, defines nearly lossless information transmission and results that are comparable to the original YOLO [18, 19]. A neural network based model for high-speed moving object filtering, detection, and tracking has been described for scientific observation and fault detection by Zhu et al., they provide a parallelized filtering and detection module that makes use of a parallel linked component labeling technique and a block-based parallel computation paradigm. Multiplexers for ADD operations and preprocessed fixed-point data for faster processing are examples of hardware optimizations. After 25 parallelization's, the accelerator achieves 19 times the acceleration and can operate more than 30,000 spike images with a dynamic power consumption of 1.618W [20]. Bhandari et al. explained that deep generative models required a lot of approximations and were centered on parametric probability distribution functions, much like Boltzmann machines. This resulted in the creation of generative stochastic networks, which replace several approximations with accurate backpropagation. Goodfellow integrated the training of two multilayer perceptron to create the Generative Adversarial Networks (GANs) model [21, 22]. However, issues with training, choosing hyper parameters, sample control, convergence, and modal collapse beset GANs. These problems were effectively tackled by GAN with boundary equilibrium (BEGAN) models [23, 24]. This research presents the development of adversarial networks and potential paths for object detection as well as abnormal behavior identification in crowded areas. This article critically examines the evolution of GANs, a potent neural network class commonly used in unsupervised learning. A generator creates fictitious data samples, and a discriminator determines which instances are real and which are fake. These two components make up GANs. Apart from convergence and stability issues, the authors discuss aspects of regularization and generalization. They go on to say how flaws can be dealt with by the model EBGAN [25-31]. In this study, Section 2 describes the literature about various deep learning models such as GAN based models, Super Resolution models, Spiking Models and Energy based models. Section3, explained about proposed methodologies such as AVAEGAN based models, SRSGAN based models, and EERGAN based models. Section-4 has illustrated that experimental evaluation and also Section-5 is demonstrated that results and discussion compare with benchmark data sets as well as real- time objects.

2. RELATED WORK

Super-resolution GAN models, energy-based, GAN-based, and spike-based models are the main models used to identify anomalous behavior. When compared to our suggested approaches, the majority of approaches fall short in terms of energy efficiency and real-time performance. This section addresses the challenges encountered in the current work on crowd surveillance, specifically in relation to systems based on feature extraction and object detection.

2.1 GAN based models

Song and Sheng [1] suggested a multi-scale GAN network that describes single-image crowd counts and also analysis anomalous behavior identification in crowded areas. The technique creates crowd density that is used to incorporating with internal GAN module such as multi-branch generator and a sectional discriminator. The addition of the multiscale GAN module improves the model's generalization capabilities. The model classifies anomalous behavior by utilizing a synthetic feature descriptor to derive the crowd movement trajectory. The algorithm slightly outperforms current algorithms in terms of accuracy and robustness when it comes to detecting anomalous behavior and crowd counting in difficult settings. For engineering applications in security surveillance, the model is appropriate. Al Jaber et al. [2] focused on classical machine learning and GANs for assessing various object tracking and identification methods in light of risks to the applications using GANs. Detection methods investigated in image classification and object localization include SVM, Adaboost, HOG model, CNN models, YOLO techniques, and GANs. The GANs are better for real-time performance with larger datasets and also increasing data generation in adversarial training. Han et al. [3] introduced an add-on encoder and a real-time GAN for identifying anomalous events in public crowds. To create better images, the network employs a discriminator and reconstructs sample images. To calculate the anomalous score using the difference in distance between two patterns. For video and image analyzed and also grouped the point wise convolution method expedites processing efficiency while guaranteeing accuracy and dependability. Alafif et al. [4] suggest a way to find strange behaviour on a large scale for security purposes using GANs and optical flow. The system extracts dynamic features based on optical flows using a transfer learning approach, and uses U-Net and Flow net to distinguish between normal and pathological behaviors. Large-scale crowd films and small-scale crowds are produced by the approach that reaches object detection accuracy. Wastupranata et al. [5] provided deep learning methods that can find abnormal human activities using a video surveillance system. Those deep learning methods can be broadly grouped into three categories, which are fully supervised, partially supervised, and unsupervised. Further, it evaluates these methods with well-known benchmark data and shows how they perform in various contexts. Besides that, it will do the further development of methods for contextual anomalous behavior detection, through the enhancement of robustness against environmental fluctuations, using varied datasets [6]. Nawaratne et al. [7] explained the unsupervised deep learning strategy for real-time video surveillance. Conclusively, it continuously learns from and discriminates between new abnormalities and normalcy through the process of fuzzy aggregation together

with active learning. They validate this further by examining its suitability for real-time video surveillance using a comparison with three benchmark datasets against contextual indicators, computing overhead, accuracy, and robustness.

2.2 Super resolution models

Wang et al. [8] provided the intelligent object detection systems for generating the super-resolution synthetic image that is used to enhance the recognition of small and distant objects from low-resolution images. The edges are improved by the hierarchical self-attention module. This is applied to the images to take care of loss due to high frequency edge information and image features owing to grown resolution. The global context aware network for precise object detection in crowded places, which receives the output super-resolution images. This generates multi-scale feature maps with a cascade transformer backbone, which guarantees reliable object detection performance. The network can adapt better to complex traffic conditions recognition to an integrated cross-scale aggregation feature (CSAF) module. Multiple datasets validate the approach, revealing competitive advantages over standard optical models. Rabbi et al. [9] explained that small object detection performance in remote sensing drone images in public crowded areas is inferior to that of larger objects, especially in low-resolution and noisy images. The brand-new edge enhanced super resolution GAN (EESRGAN) is used to boost detection performance and image quality. For ESRGAN and EEN, they use relative-in-relative dense blocks, whereas the detector network employs faster region-based convolutional networks. The architecture consists of ESRGAN, EEN, and other networks that achieve better object detection and slightly improved performance. Chen et al. [10] explained the Classification orientated Super Resolution Generative Adversarial Networks (CSRGAN) to demonstrate a tiny object-detecting network. The discriminator predicts real categories, and the network reconstructs realistic super-resolved images from discriminative features. Feature-level content loss based on VGG19 enhances object classification by improving outlines. This method has shown the positive function of VGG19 and the classification-orientated enhancement of CSRGAN. Musunuri et al. [11] pointed out that video surveillance system has recently undergone rapid development with the deployment of satellites, drones, and image sensors in busy public places. Low-resolution small objects, complex scenes, and sparse data for model training always impede object detection. They indicate a new GAN-based sequential transfer learning method to improve the performance of a model when there is little data. Similar to approaches utilizing GAN model is trained the images in step-by-step manner from the most trivial to the most challenging data. This model is used for real-time remote sensing object recognition using the datasets VEDAI-VISIBLE, VEDAIIR, and DOTA and also achieved performance.

2.3 Spiking based models

Machado et al. proposed that the bioinspired hybrid spiking neural network (SNN) called the Hybrid Sensitive Motion Detector (HSMD) improves the dynamic background subtraction (DBS) technique. They compared the HSMD against other background subtraction (BS) techniques with the benchmark datasets like CDnet2012 and CDnet2014. Both datasets expose that HSMD outperforms all the tested DBS

methods and has done a little better for big crowd, night movies, object motion starting or stopping, and baseline or shadow. The HSMD is the first hybrid SNN algorithm that can perform on image and video sequences in near real time. This in turn will improve other BS algorithms in order to fine-tune the HSMD method for challenging situations. This model wants to accelerate the algorithm and reduce power consumption in real-time applications [12]. Kasabov et al. [13] proposed a new kind of Dynamic SNN to quickly learn spatiotemporal data for identifying objects in public crowded areas. They also propose here the DSNN model using the SDSP spike-time learning in free-of-control, supervised, or semi-supervised modes together with rank-order learning. Because it uses both the timing of the subsequent spikes learnt by dynamic synapses and the order of the initial input spikes, the DSNN model performs faster and more accurately than previous SNN models for detecting abnormal behavior. This is essential for the development of self-governing machine learning systems with a wide range of applications [13]. Ziegler et al. [15] have presented a new method for detection of the ball in robotic table tennis that uses spiking neural networks and the event-based camera. The method compares the accuracies and runtimes of several state-of-the-art SNN frameworks with edge devices. This model also shows that an SNN on an edge device can work in real-time in a closed-loop robotic system such as a table tennis-playing robot. Jin et al. [16] explained that SNNs are energy-efficient technology for resolving the anomaly detection crowded areas. The researchers proposed a region-based spiking neural network (R-SNN) for object detection crowded environments. The R-SNN represents positive and negative bounding box offsets using mirror output images and also achieved mean average precision of 63.1% [16, 17]. Seras et al. [18] discussed the need to strike a balance in technologies that help vehicles see and understand the world around them through performance, efficiency, and open-world learning. They show how spiking neural networks are making the technology more resilient against picture noise and gain a competitive edge in detection with as much as 85% less energy use. The presented research underlines the challenge of detecting novel items on captured photos. Lien and Chang [19] proposed a sparse spiking neural network accelerator that leverages high weight and activation map sparsity to execute models in a highly parallel fashion at a low power. The proposed method overcomes the limitation of SNN with additional time dimension information. This method processed frames per second with energy efficiency of all the frames [20].

2.4 Energy based models

Zhao et al. [21] proposed the energy-based GAN models for interpreting the data. The discriminator is now energy function for object detection that assigns lower energies in the neighborhood of the data manifold and high energies away from it. They train a generator to generate contrastive samples with low energies and the discriminator to assign high energies to those generated samples like probability-type GANs. The discriminator as an energy function can use a range of loss functions instead of relying only on the binary classifier with logistic output. The auto encoder architecture is one of the reconstruction error serves as the energy source for the entire EBGAN framework and taking the place of the discriminator. This EBGAN frame work model has been compared to ordinary GAN models for consistent of energy efficient.

Furthermore, the training of single-scale architecture can teach us how to produce high-resolution images [21]. Berthelot et al. [22] created a new technique for maintaining equilibrium. These make sure that the GAN models were always in balance during the training process of auto-encoder-based generative adversarial networks (AEGANs) for finding objects in crowded spaces. This technique delivers rapid and reliable training in crowded scenes along with excellent object quality and new approximate convergences of different metrics. The suggested method has to ensure that the visual quality and resolutions of abnormal object detection crowd areas with less energy level. The AEGANs manage the transaction between image diversity and quality. While it still needs refinement, the approach offers some partial solutions to GAN issues. This research enumerated the constraints present in different GAN models designed for object identification in crowded, public spaces. GAN are a novel method for computer vision that uses adversarial training concepts. GAN has superior feature learning and representation capabilities than conventional machine learning models [32, 33]. This article examines the analysis of recently suggested GAN based models Super Resolution GAN models, Spiking GAN models, Energy based models and their applications in smart secure video surveillance systems. These models are detecting the objects in crowded areas. But this is required more energy level [34, 35]. In this research, introduced a novel approaches like Adversarial Auto Encoder Generative Adversarial Networks (AVAEGAN), Super Resolution Synthetic Generative Adversarial Networks (SRSGAN) and Enhanced Energy regularized Generative Adversarial Networks (EERGAN). This proposed method will can detect the abnormal object in crowded areas and energy efficiently. These methods are required less energy, more robustness and more accurately level compared to exiting GAN based models and also ensure that safety and security of publics.

3. PROPOSED METHODOLOGY

This section explains the various steps that make up the suggested system working process. Furthermore, this section describes the several stages in the development of novel based advanced techniques such as AVAEGAN based models, SRSGAN based models, and EERGAN-LSSVM based models designed to detect anomalies in densely populated public areas and safeguard individuals' safety and security.

From Figure 1, the input is given as an image for preprocessing section. The trained images are stored in data bases for further processing. The proposed novel based approaches are like AVAEGAN based models, SRSGAN based models and EERGAN based models. The Adversarial Variational Auto encoders GAN (AVAEGAN) based models are mapping the training images for detecting the abnormal objects in public areas. The Super Resolution synthetic GAN (SRSGAN) based models are mapping the low resolution images into high resolution images for feature extraction in abnormal object detection accurately. Our proposed novel based approach, Enhanced Energy Regularized GAN –Least Square Support Vector Machine (EERGAN-LSSVM) based models are used to detect the objects in crowded areas for various real time processing capabilities. EERGAN-LSSVM learns about objects found at high resolution in images. This EERGAN-LSSVM model is required less energy for detecting the abnormal object detection and feature extraction. This research introduces the EERGAN-LSSVM model for detecting the abnormal objects, efficient energy processing, latency prediction, and classifying the ratio of abnormal to normal accurately. The structure of EERGAN-LSSVM propelled the development of parameter adjustment possibilities for maintaining high detection rates and also produces more robust, increasing the computational efficiency; fast object detection, and ensuring public safety and security.

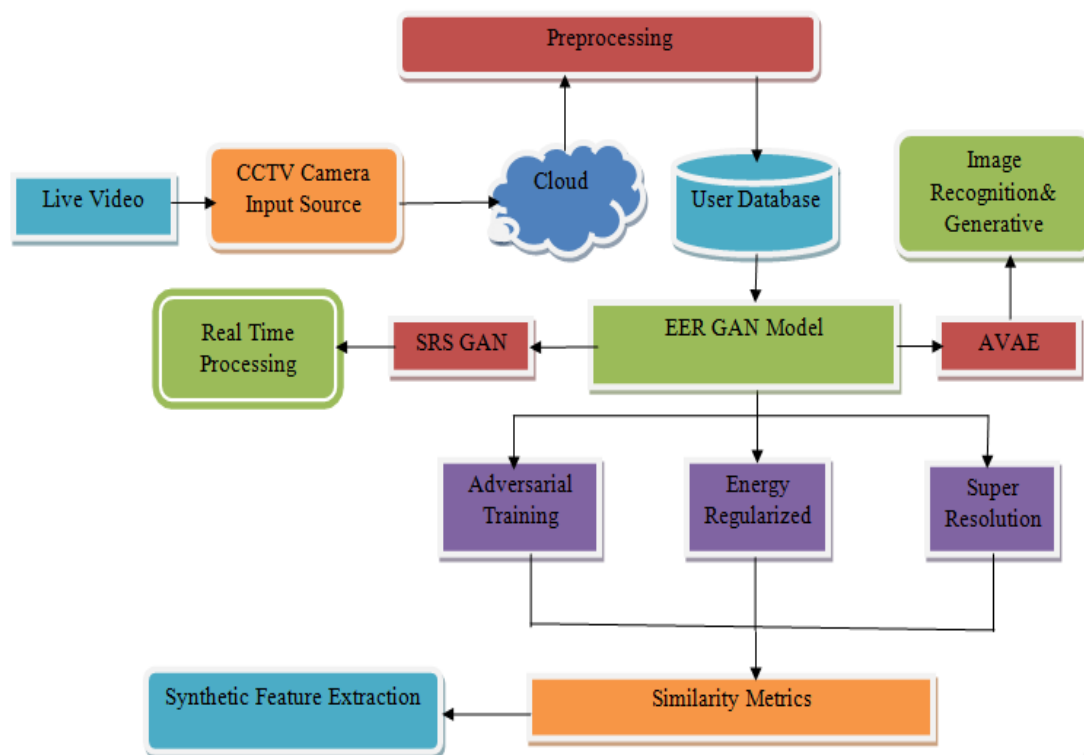


Figure 1. Proposed EERGAN model for abnormal object detection process

3.1 Adversarial variational auto encoder GAN model – AVAEGAN

In video surveillance, adversarial variational auto encoders (AVAEs) are essential for optimal reconstruction loss and close resemblance to input images. For characterizing an observation in latent space, an adversarial variational auto encoder (AVAE) offers a probabilistic method. Despite the latest improvements, the field of video surveillance still widely uses Adversarial Variational Auto encoders (AVAEs). In AVAE, the decoder model is usually referred to as the generating model and the encoder model is referred to as the recognition model for abnormal behavior detection in crowded areas [24].

The AVAEGAN is a probabilistic model for latent variables to treat features of various objects and produce new samples of images. Such models generally involve an encoder and generator. The offered scenario tries to mix up AVAE losses and losses incurred by AVAEGAN so that it can formulate an overall loss function such that encoding and decoding processes are in a single unified structure. This also highlights the correlation between encoding and decoding latent variables and improves latent representations, getting data

closer and closer to being more closely resembling original data. The proposed AVAE produces good results for detecting abnormal objects in crowded environments.

(a) Dimensionality Reduction

AVAEs train to learn low-dimensional representations of high-dimensional data in order to visualize, compress, extract features, and determine the intrinsic dimensionality of data.

(b) Attribution

AVAEs focus on abnormal behavior detection, tracking, and super-resolution, helping to identify crowded objects for efficient synthetic feature extraction.

(c) Generation

In addition to producing new images and videos in busy public spaces for feature extraction, synthesis, and generative modeling, AVAEs can also learn model parameters for actual process emulating.

(d) Discriminator

AVAEGAN architecture is used to replacing the discriminator by the reconstruction error as the source of energy for the whole GAN framework and also detecting the images from abnormal crowded areas accurately and fast robustness.

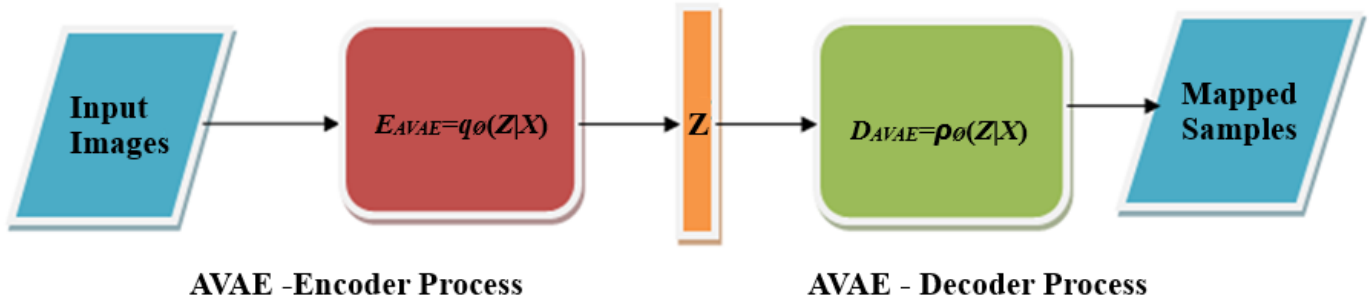


Figure 2. AVAE frameworks for encoding and decoding process

From Figure 2, the encoder, a recognition model, first maps the input images into latent semantic spaces in AVAE.

$$E_{AVAE} = q_{\phi}(Z|X) \quad (1)$$

where, ϕ is the parameter of AVAE encoder in sample objects mapping. This approximates the real posterior, $p_{\theta}(X|Z)$ which is unknown and unsolvable.

The decoder in AVAE is the second generative model, mapping the input images to the latent semantic space.

$$D_{AVAE} = p_{\theta}(X|Z) \quad (2)$$

where, θ is the parameter of AVAE decoder. The approximation for the true posterior of this AVAE decoder model, $p_{\theta}(X|Z)$ is $q_{\phi}(Z|X)$. But in some cases, it is infeasible to compute. This approximation makes the AVAE feasible in instances that may be intractable otherwise.

$$L_{AVAE} = \frac{1}{N} \sum_i (X^{(i)} - f(Z^{(i)})) \times Q \quad (3)$$

where, L_{AVAE} is the deterministic decoder loss between the sample input images and reconstructed images from N objects. Q is processing parameter of loss function; $Q = (X^{(i)} - f(Z^{(i)}))$. The reconstruction loss tells how well the AVAE

model reconstructs the input images from the latent space.

$$L_{AVAE} = \frac{1}{N} \sum_i (X^{(i)} - f(Z^{(i)})) \times Q \quad (4)$$

where, L_{AVAE} is the stochastic decoder or multivariate Gaussian form of $\log N(X^{(i)}|f(Z^{(i)}))$ different sample input and output of the images.

Consider the cross-entropy between input data and reconstructed data for calculating the reconstruction loss in the case of a discrete input space.

$$L_{AVAE} = -\frac{1}{N} \sum_i (X^{(i)} \log f(Z^{(i)}) + (1 - X^{(i)}) \log (1 - f(Z^{(i)}))) \quad (5)$$

where, $(X^{(i)} \log f(Z^{(i)}))$ is reconstruction loss that provides a natural encouragement to the generator to learn how to reconstruct the input data from the latent space. In a similar vein, the encoder gains the ability to map the input images or samples to the latent space in crowded areas, allowing the decoder to recreate the input data via back propagation.

The AdaGrad optimizer trained and optimized the proposed AVAEGAN model encoder, generator, and discriminator components separately. The learning rate for the encoder is set

to 0.0001. A batch of 16 was then used. As the network converges at 3000 epochs, this was the specified epoch size. Likewise, the other two networks were trained and optimized separately, just like the AVAEGAN model. Mostly, the AdaGrad-Adaptive Gradient optimizer is having a learning rate of 0.0001 for object detection. The batch sizes for these two networks, SRSGAN and EERGAN, were 32 and 64. SRSGAN and EERGAN converge at around 4000 and 5000 epochs, respectively; hence, the epoch size assigned for these two networks is 4000 and 5000, respectively.

3.2 SRSGAN model

SRSGAN is Super Resolution Synthetic Generative Adversarial Network that specializes in converting low-resolution images to high-resolution outputs, as well as improving image resolution in photography, playground imaging, public crowded imagery, and video up scaling. Image super-resolution (SR) approaches reconstruct a higher-resolution (HR) picture or sequence from the observed lower-resolution (LR) images.

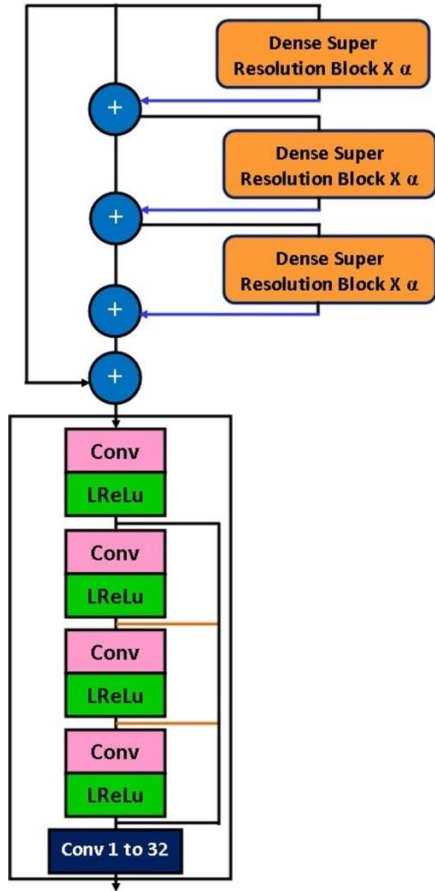


Figure 3. SRSGAN frame work for image resolution analysis

Super Resolution Synthetic Generative Adversarial Network (SRSGAN) is defined for revolutionizing the concept of combining sub-pixel efficient nets with the complete loss function of traditional GAN. SRSGAN has largely focused on improving the recognition and classification of unsafe operations using construction equipment by up scaling low-resolution surveillance video into further fine-detailed high-resolution inputs for the trained model. This directly translates into performance accuracy in object detection, synthetic feature extraction, and classifications in almost real-time

monitoring of dynamic construction sites, thus developing surveillance technology. It has the 4000 epochs and 32 batch sizes for training set.

From Figure 3, the super-resolved synthetic GAN (SRSGAN) is capable of detecting small distant objects from low-resolution (LR) images. The edge enhancer and a hierarchical self-attention module are applied to mitigate the loss of high frequency edge information and texture details in super resolved images [10, 11]. Contextual and perceptual losses are optimally combined for the best image up-scaling, while adversarial loss will train the neural network to map early on super-resolved images as natural imagery through a trained discriminator network to differentiate original photo-realistic from the super-resolved images. These enhanced improvements result in the proposed SRSGAN achieving consistent superior visual quality with better restoration of natural features.

a) Discriminator - SRSGAN

In SRSGAN, relativistic discriminator ($L_{SRSGAN(D)}^{Ra}$) calculates the likelihood of the given normal input images being more realistic than randomly selected fake image (x_f), can induce this quality. They also provide a variation in which the discriminator evaluates the chance that the given genuine object (x_r) is more efficient than fake images.

$$L_{SRSGAN(D)}^{Ra} = -ED_{xr}[\log(x_r, x_f)] - ED_{xf}[\log(1 - D_{Ra}(x_r, x_f))] \quad (6)$$

where, $-ED_{xr}[\log(x_r, x_f)]$ is the loss function of standard SRSGANs and also extract the synthetic features from abnormal objects in crowded places.

b) Generator – SRSGAN

The symmetric form of adversarial loss of generator function is represented by

$$L_{SRSGAN(G)}^{Ra} = -EG_{xr}[\log(1 - D_{Ra}(x_r, x_f))] - EG_{xf}[\log(D_{Ra}(x_r, x_f))] \quad (7)$$

where, $x_f = G(x_i)$ and x_i represents the input LR picture. It can observe here that both x_r and x_f are the participants of the adversarial loss of the generator. In this way, our generator gets gradients from the created and real object when performing adversarial training - in SRSGAN; only the generated part plays a role.

c) VGG -19 Network Model

The VGG-19 network model was noted for the deep CNNs with an identical SRSGAN architecture. This is focusing on simplicity and power and the deeper version for gaining attention. VGG-19 benefits from simplicity in the design and is also designed with small 3x3 convolution filters and better performance. The simplicity and power of VGG-19 have influenced the succeeding designs of models used in deep learning, like ResNet and Inception. Their ability to extract highly discerning synthetic features made them widely used in transfer learning and other computer vision applications. VGG-19 is a milestone model in the epochs of deep learning, balancing simplicity with depth to achieve magical accuracy using SRSGAN. This has 4000 epochs and 32 batch sizes in training set for detecting the objects and increasing the object

resolution.

- (a) Convolutional Layers - SRSGAN: 3x3 filters with a stride of 1 and padding of 1 to hold spatial resolution throughout. Block-1 contains Conv1 64x64 filters and 3x3 kernels for activating the linear function. Dense super resolution block is used to extracting the synthetic feature extraction in crowded areas with high resolution.
- (b) Activation Function-SRSGAN: ReLU-Rectified Linear Unit is used after each convolution layer and introducing non-linearity for complex objects detection crowded environments.
- (c) Max Pooling Layers-SRSGAN: Max pooling with a 2x2 filters and stride 2 reduces the spatial dimensions. But all the important synthetic features have been maintained.
- (d) Fully Connected Layers-SRSGAN: Three fully connected finalization at the end of the network for classification.
- (e) Softmax Layer-SRSGAN: The last and final layer for the class probabilities output.

3.3 Enhanced energy regularized GAN model – EERGAN

The Enhanced Energy Regularized GAN (EERGAN) is a fascinating one among our current collection of GANs. As opposed to the original GAN, which estimates reconstruction loss using a discriminator, this one employs an auto encoder. This method proves to be a highly effective method for configuring a GAN.

EERGAN Model generates Fast, stable, Energy efficient and robustness to detecting abnormal objects in public crowded areas and also helps for public safety. From Figure 4, the proposed EERGAN is used to enhance the energy level performance for object detection in public crowded areas. The noise image (z), image category (c) is added to our EERGAN Generator for updating and regularizing (z)|D(x) the energy level with detecting the objects.

The Discriminator with enhanced energy model processes the various responsibilities of object detection and also

differentiate the given image is either real or fake. This fake image has been transferred to discrepancy measures for further process and also produces robustness, energy efficient processing and accurate image an effectively and also ensure the public safety with security in crowded areas. To configure this by following the steps below: Use the source data to train an auto encoder, and then pass the generated images through it. This metric is now useful in public crowded objects due to the significant reconstruction loss caused by poorly created photos. When combined with appropriate regularization to prevent mode collapse. The EERGAN model measures the similarity between produced and real images using an energy function. Use the energy function to design a loss function that decreases during training. To introduced EERGAN to perform object detection modeling tasks over picture data from crowded settings in public spaces. This demonstrate and create well-performing and efficient models by using Enhanced Energy Regularized GAN (EERGAN), which achieves competitive detection performance compared to existing energy models at significant energy consumption savings.

The EERGAN-LSSVM model comprises a label generator and a discriminator that are optimized by means of adversarial learning. The modeling power of this model is enhanced through two modules: the EERGAN-LSSVM generator and discriminator. In addition to resolving GAN-specific problems of model collapse and convergence, EERGAN outperforms all contemporary state-of-the-art benchmarks. ERGAN learns about objects found at high resolution in images. This research introduces the EERGAN-LSSVM model for detecting the abnormal objects, efficient energy processing, latency prediction, and classifying the ratio of abnormal to normal accurately. The structure of EERGAN-LSSVM propelled the development of parameter adjustment possibilities for maintaining high detection rates and also produces more robust, increasing the computational efficiency; fast object detection, and ensuring public safety and security.

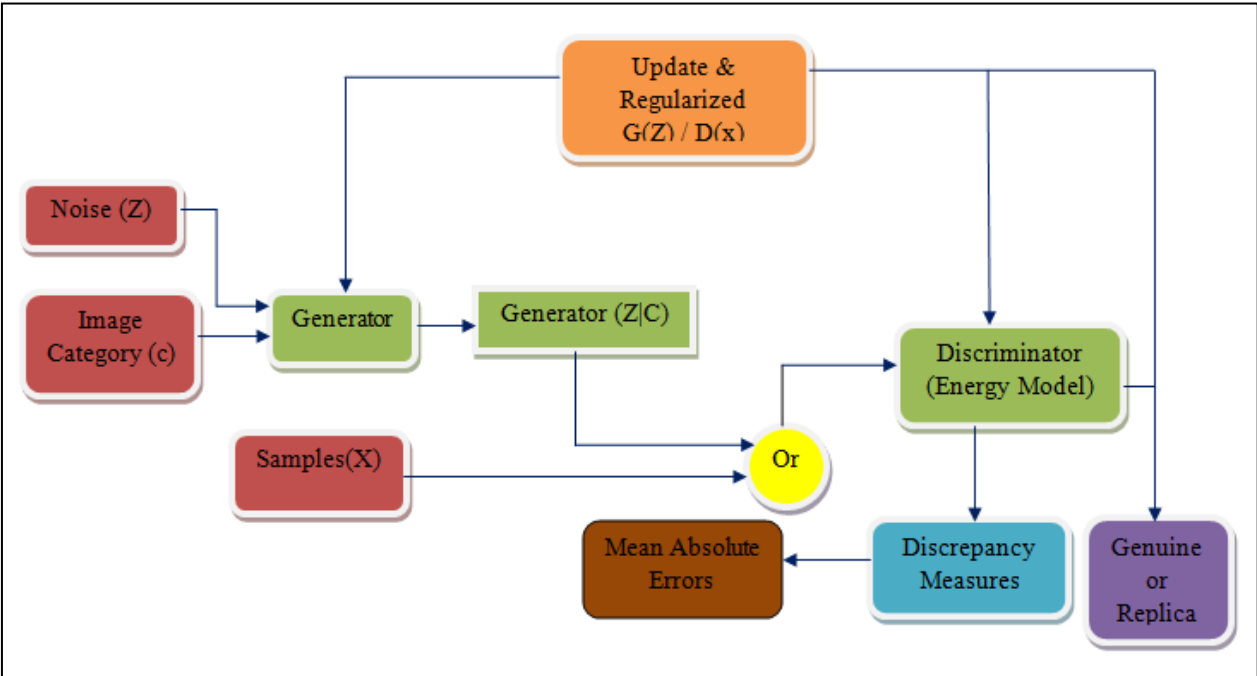


Figure 4. EERGAN frame work for object detection

Algorithm-1: EERGAN model based Adversarial Dynamic Training Process

Input: Real time object as sample input (z)

Training: D_{EERGAN} and G_{EERGAN}

Parameters: θ is parameter of both Generator and Discriminator.

Output: Detecting abnormal objects and extracting features.

```

1. for N of Adversarial Training process do
2.   for M steps do
3.     Image groups m into noise sample object
       {  $Z^{(1)}, \dots, Z^{(m)}$  } from noise  $p_g(z)$ .
4.   Collect the sample images {  $x^{(1)}, \dots, x^{(m)}$  } from
       data generating
5.   Distributing process of  $EERGAN_{data}(x)$ .
6.   Discriminator update
7.      $D(E)_{EERGAN} = \Lambda_{\theta D} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log(1 - D(G(Z^{(i)})))]$ 
8.   End for
9.   Sample batches of m noise sample images
       {  $Z^{(1)}, \dots, Z^{(m)}$  } from noise  $p_g(z)$ .
10.  Generator update
11.   $G(E)_{EERGAN} = \Lambda_{\theta G} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(Z^{(i)})))$ 
12. End for

```

(a) Preprocessing

Preprocessing is used to select the appropriate images for object detection. This gives the correct input to the proposed model for training purpose.

(b) Training

(i) Generator – EERGAN

Generator (G) is the result of minor gradient difficulties that make maximizing cross entropy useless for lowering the discriminator's likelihood of producing accurate predictions of abnormal object in crowded areas.

$$G_{EERGAN} = \Lambda_{\theta G} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(Z^{(i)}))) \quad (8)$$

where, $\log(1 - D(G(Z^{(i)})))$ is used to predict the worst on the abnormal images and probability is close to 1.

(ii) Discriminator – EERGAN

By increasing the prediction probability of distinguishing between genuine and false, discriminator (D) maximizes that log likelihood, which is equivalent to minimizing negative log likelihood.

$$D_{EERGAN} = \Lambda_{\theta D} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log(1 - D(G(Z^{(i)})))] \quad (9)$$

where, $\log D(x^{(i)})$ is used to predict well on real images in crowded areas. $\log(1 - D(G(Z^{(i)})))$ is used to predict well on the abnormal images and probability is close to 1.

(c) Equilibrium

The Equilibrium is used to demonstrate the energy balance between both Generator (G) and Discriminator (D) loss.

$$E[L(x)] = E[L(G(z))] \quad (10)$$

where, $E[L(x)]$ is the distribution loss of real image samples

(x) and $E[L(G(z))]$ is random sample of input (z). The generated samples and their predicted error in distribution should be similar if the discriminator is unable to tell them apart from genuine samples. This has to reach the great effort between both generator and discriminator [22]. Modified Eq. (10):

$$E[L(G(z))] = \gamma E[L(x)] \quad (11)$$

Adding a new hyper parameter, $\gamma \in [0,1]$ defined as will allow us to loosen the equilibrium:

$$\gamma = \frac{E[L(G(z))]}{E[L(x)]} \quad (12)$$

Because there is a natural boundary between hard and exhaustive images, the discriminator reduces image variety by balancing auto-encoding genuine images and differentiating real from created images using the γ term.

(d) Energy based boundary Equilibrium GAN

$$\begin{aligned} L_D &= L(x) - k_t \cdot L(G(z_D)) \text{ for } \theta_D \\ L_G &= L(G(z_G)) \text{ for } \theta_G \\ k_{t+1} &= k_t + \lambda_k [\gamma L(x) - L(G(z_G))] \\ &\quad \text{for Each Training Step } - t \end{aligned} \quad (13)$$

Proportional control theory is used to maintain Eq. (11) modifying k_t at each stage to maintain Eq. (10). When doing gradient descent, the equilibrium is preserved by using the proportional gain, γk . The Wasserstein distance model is affected by the addition of approximations in equation 10 and $E[L(x)]$ in Eq. (11). Unlike SRSGANs, EERGAN-LSSVM doesn't need to train D or G consistently. When training is employed by θ_G and θ_D and also updated separately depending on losses.

(e) Optimality

To use EERGANs for object detection in crowded areas and efficient energy processing, but their convergence is always challenging and results in oscillating less energy losses.

$$M_{EERGAN(global)} = L(x) + \gamma |L(x) - L(G(z_G))| \quad (14)$$

To use the equilibrium notion to produce a global measure of convergence that focuses on the closest reconstruction with the lowest instantaneous process error in detecting abnormal object and also improve the energy levels. The EERGAN-LSSVM produces more robust and fast object detection, ensuring public safety and security.

4. Experiments and Evaluation

4.1 Experimental

To demonstrate the performance of the EERGAN-LSSVM, three benchmark datasets were utilized:

a) Shanghai Tech Dataset: From Table 1, this dataset has different scenes with different densities of people, gathered from surveillance film shot in different locales. Its widespread use for abnormal event identification and crowd analysis makes it suitable for evaluating the efficacy of Spiking GANs in various contexts. The Shanghai tech dataset is a very large, huge dataset in crowded areas. It contains a total of 2000

images of crowds that are annotated. It has two classes like Class 1, and Class 2. Class 1 has 1200 images, whereas Class 2 has 800 images. 900 images are entered into training set, and the 300 images are used for testing in Class 1, while the two classes of Class 2 share a total of 555 images for training set and 245 images have used for testing. Each image in the crowd markers has a point set along the center of the head for each individual, totaling 120,275 people counted by annotation. Class 1 images had been sourced or obtained through the internet, while Class 2 was from several busy streets of Shanghai.

b) UCSD Pedestrian Dataset: The majority of the video sequences in this dataset are of pedestrian activity. It is frequently used to detect anomalies in congested environments. Because it contains annotations for a variety of activities, the dataset provides a solid foundation for developing and evaluating algorithms that recognize violent behavior. The UCSD pedestrian anomaly detection dataset was created by placing a stationary camera on pedestrian walkways, observing common anomalies like bikers, skaters, small carts, and speed-walking deficiencies. The data was partitioned into 4 Classes, with video footage broken into 4000 frames. Class

1 has 1300 images, Class 2 has 1400 images, Class 3 has 800 images and Class 4 has 500 images. 750 images have given for training, and the 550 images are used for testing in Class 1, while the four classes of Class 2 share a total of 800 images for training and 600 other images for testing. 455 images are set apart for training, and the remaining 345 are for testing in Class 3, while the four classes of Class 4 share a total of 450 images for training and 300 other images for testing. Each image in the crowd markers has a point set along the center of the head for each individual, totaling 420,485 people counted by annotation. From Class 1 to Class 4 images had been sourced or obtained through the common anomalies like bikers, skaters, small carts, and speed-walking deficiencies.

C. User Collected Data Set: Similarly, User collected Dataset has 50000 samples, 10 classes, 40000 training set and 10000 testing set for detecting normal and abnormal object event ratio range is 10:90. This study's unique dataset includes annotated video sequences of aggressive and non-aggressive behaviors in crowded contexts. This dataset was created to provide an extensive assessment setting that is specifically designed to meet the requirements of aggressive behavior detection.

Table 1. Datasets with their evaluation process

Dataset	No. of Samples	No. of Classes	Training Set	Testing Set	Aggressive to Non – Aggressive Event Ratio
Shanghai Tech	2000	2	1500	500	25:75
UCSD Pedestrian	4000	4	3000	1000	20:80
User Collected Dataset	50000	10	40000	10000	10:90

4.2 Evaluation metrics

The performance of EERGANs was assessed using several evaluation metrics:

Accuracy: This metric describes that the proportion of accurately predicted instances among the total instances of samples or images in public crowded scenes.

Precision: This metric explains that the proportion of true positive predictions among the total predicted positives images or samples in public crowded scenes.

Recall: This metric describes that the proportion of true positive predictions among the total actual positive's samples or images in public crowded scenes.

F1-score: F1-Score is providing a balance between the various data sets mean of precision and recall.

Energy Consumption: This Metric describe that the amount of energy consumed by the model during inference and detecting the object in crowded areas.

Prediction Latency: This metric explains that the time taken by the model to make a prediction of images in crowded areas.

These metrics were selected to explain a overall evaluation of both the predictive performance and the efficiency of the proposed novel based EERGAN models.

4.3 Experimental results

The experimental results are described in the following tables.

From Table 2, the results demonstrate that Enhanced Energy Regularized GAN (EERGAN) outperform AVAEGAN and SRSGAN across all performance metrics. Specifically, EERGANs achieve an accuracy of 97.7%, a precision of 95.68%, a recall of 95.68%, and an F1-score of 98.21%.

From Table 3, the efficiency metrics further highlight the advantages of Enhanced Energy Regularized GAN (EERGAN). EERGAN consume 135% less energy compared to Spiking Generative Adversarial Network (SPGAN) and 180 % less energy compared to LSHWGAN. Additionally, Enhanced Energy Regularized GAN (EERGAN) exhibits a prediction latency of 28 ms and Energy reduction of 32%, making them highly suitable for real-time aggressive behaviour detection.

From Table 4 and Table 5, the EERGAN achieved an more accuracy 97.7 %, outperforming SPGAN 96.7%, AVAEGAN 96.8 %, SRSGAN 96.9% and executing more precision (95.68), recall (95.99), F-Score (98.21), reflecting a robust capability in distinguishing aggressive from non-aggressive behaviors. EERGAN-LSSVM exhibits energy reduction of 32% and prediction latency of 28ms for detecting abnormal objects in crowded areas.

Table 2. Performance metrics comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LSGAN	91.3	92.34	92.55	93.15
LSHWGAN	94.5	94.62	93.65	96.81
Spiking GAN	96.7	95.12	94.83	97.78
AVAEGAN	96.8	95.26	94.91	97.81
SRSGAN	96.9	95.31	94.95	97.91
EERGAN	97.7	95.68	95.99	98.21

Table 3. Efficiency metrics comparison

Model	Energy Consumption (mJ)	Energy Reduction (%)	Prediction Latency (ms)
LSGAN	200	-	50
LSHWGAN	180	10%	45
Spiking GAN	140	30%	35
EERGAN	135	32%	28

Table 4. Performance metrics comparison with different GAN models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
YOLO	80.15	80.18	80.21	80.23
Faster R- CNN	80.17	80.28	80.22	80.25
DCNN	80.18	80.20	80.24	80.28
HDCNN	80.36	80.78	80.37	80.39
SGAN	83.45	81.23	80.34	80.22
DCGAN	84.32	83.04	82.23	83.2
LSGAN	91.3	92.34	92.55	93.15
LSHWGAN	94.5	94.62	93.65	96.81
Spiking GAN	96.7	95.12	94.83	97.78
AVAEGAN	96.8	95.26	94.91	97.81
SRSGAN	96.9	95.31	94.95	97.91
EERGAN	97.7	95.68	95.99	98.21

Table 5. Efficiency metrics comparison with different GAN models

Model	Energy Consumption (mJ)	Energy Reduction (%)	Prediction Latency (ms)	GPU Memory (GB)	Training Time(hrs/epochs) Approximately
YOLO	320	-	78	8	8
Faster R- CNN	310	-	76	12	12
DCNN	300	-	74	8	7
HDCNN	280	-	70	9	8
SGAN	240	-	65	12	12
DCGAN	220	-	57	10	10
LSGAN	200	-	50	8	6
LSHWGAN	180	10	45	7	5
Spiking GAN	140	30	35	6	4
EERGAN	135	32	28	5	3

Table 6. Default hardware specification

Default Energy Consumption (mJ)	Model Types	Sample Models
5-100 mJ	Edge based AI Models	EfficientDet Model
50-200 mJ	Light Weight CNN & GAN Models	DVDGAN, LSGAN, ACGAN
200-500 mJ	Mid-size CNN & GAN Models	YOLO 1 to YOLO 5, Faster R-CNN, LSTM and GAN Models,

(a) Energy consumption (mJ)

Energy consumption consists of some set of parameters for object detection, such as system architecture, hardware specification, and input resolution. This can be used to total energy for an object detection model in an image or frame.

(b) Energy reduction

Energy Reduction is used to represent the indication of energy savings for proposed model and this model can compare with various states of art models for energy consumption.

$$E_R = \frac{E(\text{ExistingModel}) - E(\text{ProposedModel})}{E(\text{ExistingModel})} \times 100 \quad (15)$$

From Eq. (15), For example:

EERGAN model is compared with LSGAN

$$E_{\text{EERGAN}} = \frac{\text{LSGAN} - \text{EERGAN}}{\text{LSGAN}} \times 100$$

$$E_{\text{EERGAN}} = \frac{200 - 135}{200} \times 100$$

$$E_{\text{EERGAN}} = \frac{65}{200} \times 100$$

$$E_{\text{EERGAN}} = 32\%$$

Spiking GAN model is compared with LSGAN

$$E_{\text{Spiking GAN}} = \frac{200 - 140}{200} \times 100$$

$$E_{\text{Spiking GAN}} = \frac{60}{200} \times 100$$

$$E_{\text{Spiking GAN}} = 30\%$$

(c) Prediction latency

Prediction latency indicates that total time taken by the proposed model for processing of a single object or image execution and defines the prediction. This is measured by milliseconds (ms). It has following hardware specification

tools such as Nvidia for GPU, CPU and TPU.

Prediction Latency = Total Time Taken per Execution

From the Table 6, **YOLO = (320mJ, 78 ms)** is slower than remaining models for detecting objects.

EERGAN = (135mJ, 28 ms) is faster than YOLO model for detecting abnormal real time objects in crowded environments.

5. ANALYSIS AND DISCUSSION

The Novel approach EERGAN with LSSVM manages to get the best out of both architectures for an aggressive behavior-detecting model, and since it is event-driven, this enhances energy-efficient processing by EERGAN, and EERGAN-LSSVM generative capability improves the learning capacity as well as data production quality.

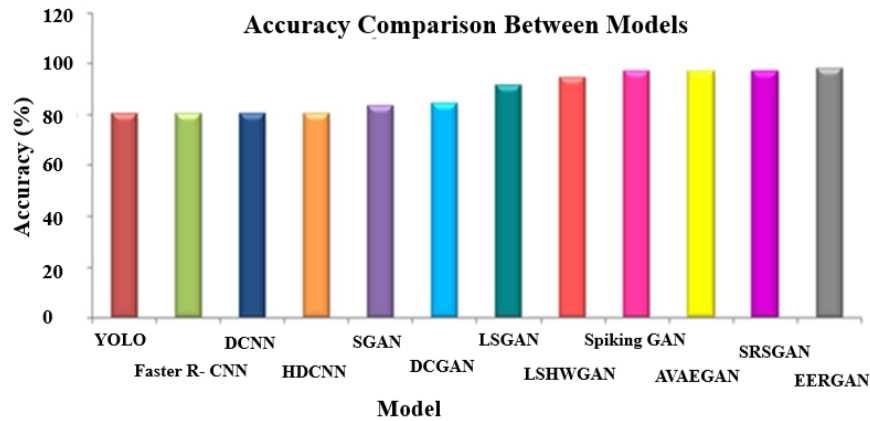


Figure 5. Accuracy comparison of EERGAN with different Models

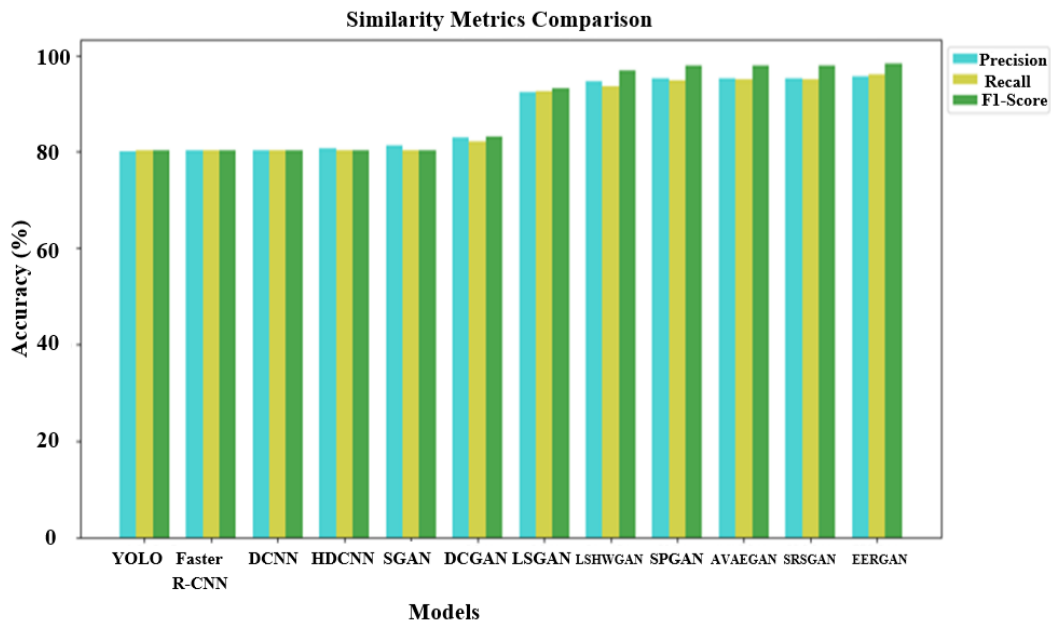


Figure 6. Comparison of the Evaluation metrics

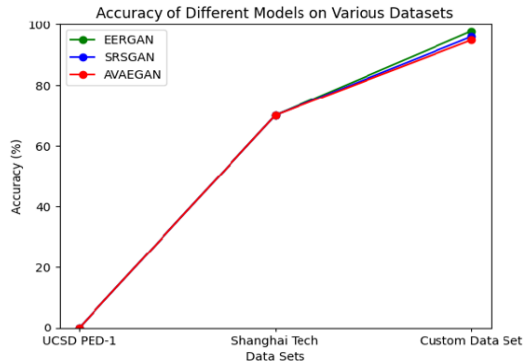


Figure 7.ROC curve with data sets Vs accuracy for different models

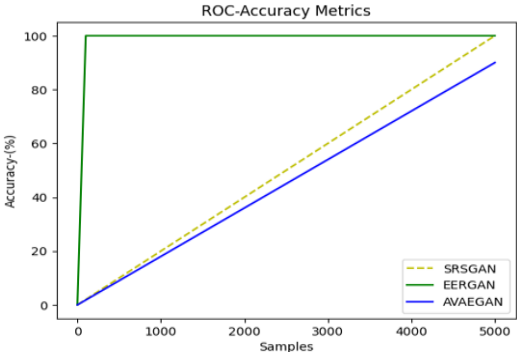


Figure 8.ROC curve accuracy of samples compared with EERGAN models

(a) Accuracy and Precision: With much greater accuracy and precision compared to the cases above, this plot shows the marked superiority of EERGAN-LSSVM in detection of aggressive behaviors with minimal false positives during surveillance applications, in which the rate should be decreased and is mainly a big issue. From Figure 5, EERGAN-LSSVM model achieved the accuracy of 97.7% for detecting abnormal object in crowded areas and SPGAN achieved 96.7%.

(b) Recall and F1-score: From Figure 6 and Figure 7 demonstrate that the high recall and F1-score show that Enhanced Energy Regularized EERGANs indeed grab all the aggressive behaviors that otherwise might fall through the cracks, thus capturing the entire surveillance of critical instances. EERGAN-LSSVM achieves an accuracy of 97.7%, precision of 95.68%, recall of 95.99%, and an F1-score of 98.21%. These metrics demonstrate a significant improvement over AVAEGAN, SRSGAN, Spiking GAN and LSHWGAN. From Figure 8, EERGAN-LSSVM model described that the false positive and negative rate. This achieved more probability than SPGAN Model.

(c) Real-time Performance: Their lower prediction latency, which is an important requirement for early action, thus substantiates the ability of the Enhanced Energy Regularized EERGANs to classify violent behavior in real-time. Enhanced Energy Regularized GAN (EERGANs) with LSSVM Models is improving sample quality, but their complexity and long run times hinder their adoption. Despite balancing these extremes, the complexity of hybrid approaches may hinder their adoption. Improvements in one area benefit other areas, such as subjective flows, diffusion models, and better variation bounds for AVAEGAN. While attention and implicit networks show promise in SRSGAN for scaling models to high-dimensional data, future generalizations will depend on unified generative models that can model continuous, irregular, and arbitrary length data.

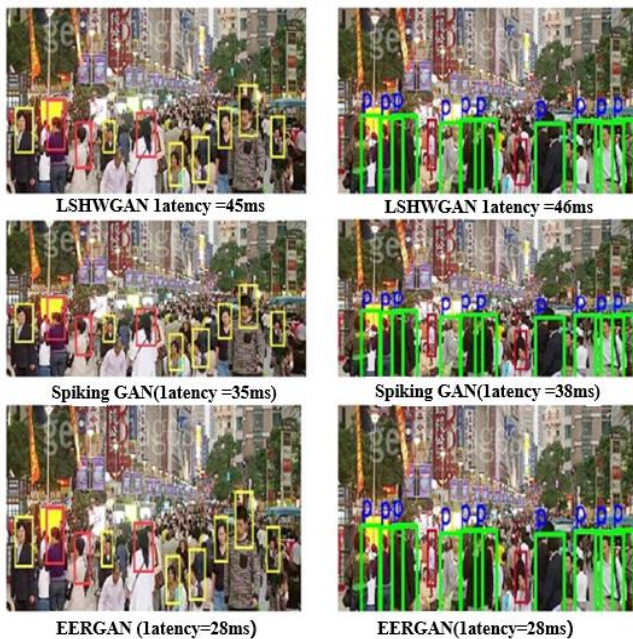


Figure 9. Latency comparisons between spiking GAN Vs EERGAN

To sum, the experiment results actually confirm the efficiency and effectiveness of EERGAN in detecting violent

behavior in crowded spaces. Bringing together significant energy-saving features along with real-time processing capabilities, besides improving predictive performance, the EERGAN model emerges as a viable contender to enhance public safety in surveillance hotspots.

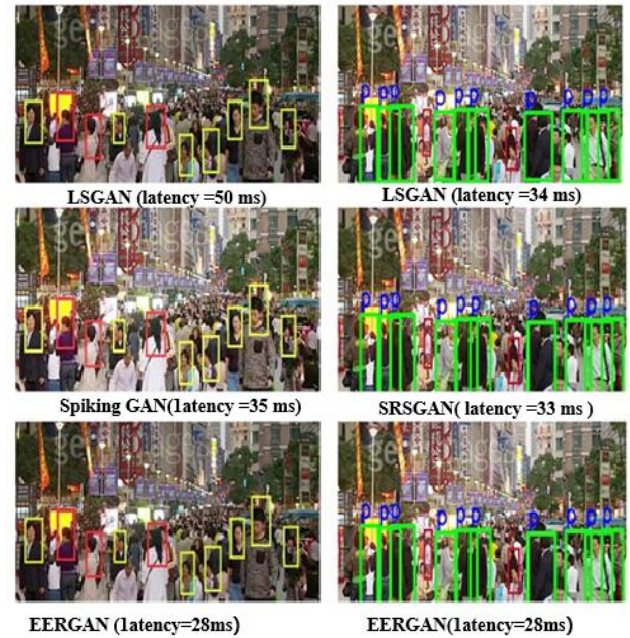


Figure 10. Latency comparisons between AVAEGAN, SRSGAN and EERGAN

From Figure 9, the LSHWGAN is used to predict the latency ranges upto 45 ms, while EERGAN prediction latency is 28 ms, similarly the LSHWGAN can also predict the latency ranges upto 46 ms, outperforming the EERGAN prediction latency is 28 ms for detecting the abnormal objects and feature extraction in real time data sets. The proposed EERGAN model is required less energy level for detecting the abnormal objects with compared to Spiking GAN models (35ms).

From Figure 10, the Adversarial Auto Encoder Generative Adversarial Networks (AVAEGAN) model is used to predict the latency ranges upto 34ms, while Enhanced Energy Regularized Generative Adversarial Networks (EERGAN) prediction latency is 28 ms, similarly the LSGAN can also predict the latency ranges upto 50 ms, outperforming the EERGAN prediction latency is 28 ms for detecting the abnormal objects and feature extraction in real time data sets. The proposed EERGAN model is required less energy level for detecting the abnormal objects with compared to Super Resolution Synthetic Generative Adversarial Networks (SRSGAN) models (33 ms).

(d) Energy Efficiency: The significant reduction in energy consumption lists the potential of Spiking GANs for deployment in resource-constrained environments, such as battery-powered surveillance systems.

From Figure 11 and Figure 12, the experimental results indicate that Enhanced Energy Regularized Generative Adversarial Networks EERGANs achieve an accuracy of 97.7%, precision of 95.68%, recall of 95.99%, and an F1-score of 98.21%. These metrics demonstrate a significant improvement over AVAEGAN, SRSGAN, Spiking GAN and LSHWGAN. From Figures 13-15 has been demonstrated the confusion matrix for bench mark data sets and also described the normal object as a true label and abnormal as a predicted

label. From Figure 16, EERGAN-LSSVM exhibit a 32% reduction in energy consumption compared to LSGAN, SPGAN produces 30% reduction compared to LSGAN and LSHWGAN generates 10% reduction compared to LSGAN. From Figure 17, the prediction latency is 28ms, underscoring the efficiency of Enhanced Energy Regularized GANs in real-time scenarios. The integration of EERGAN with LSSVM presents a promising solution for aggressive behavior detection in crowded environments. The event-driven nature of EERGAN and the generative capabilities of spiking GANs complement each other, resulting in a framework that not only improves prediction accuracy but also enhances energy efficiency and processing speed. This novel approach addresses the limitations of traditional methods, providing a robust solution for real-time applications.

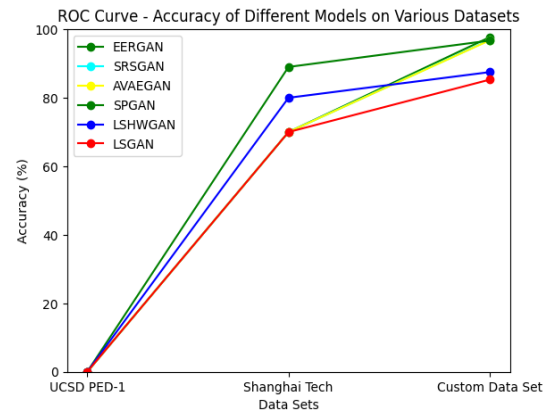


Figure 11. ROC curve accuracy of data sets compared with SOA methods Vs EERGAN model

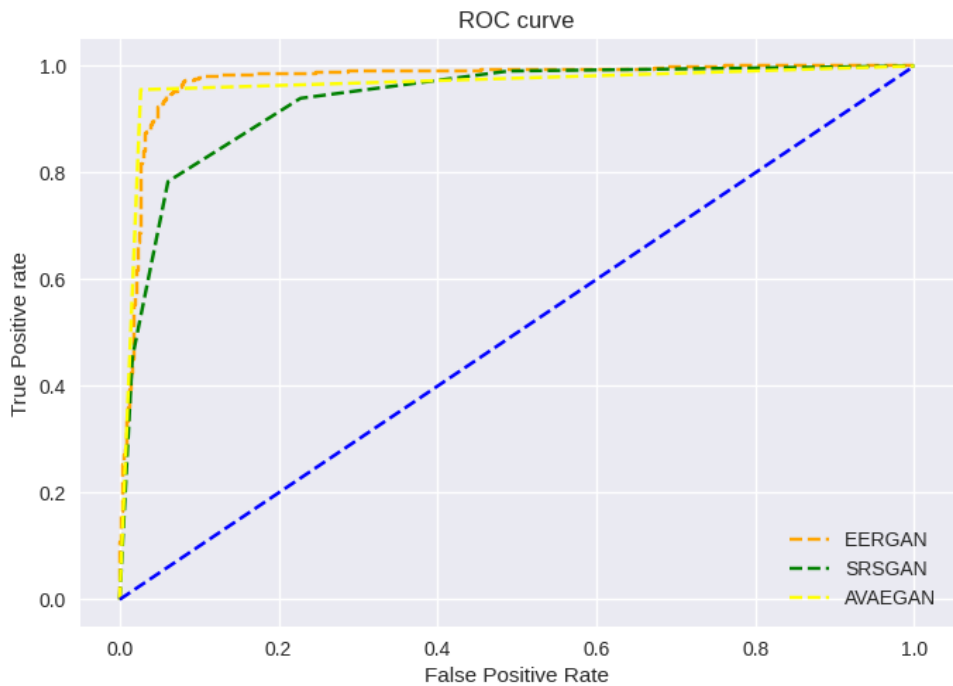


Figure 12. ROC curve for false positive and true positive rate

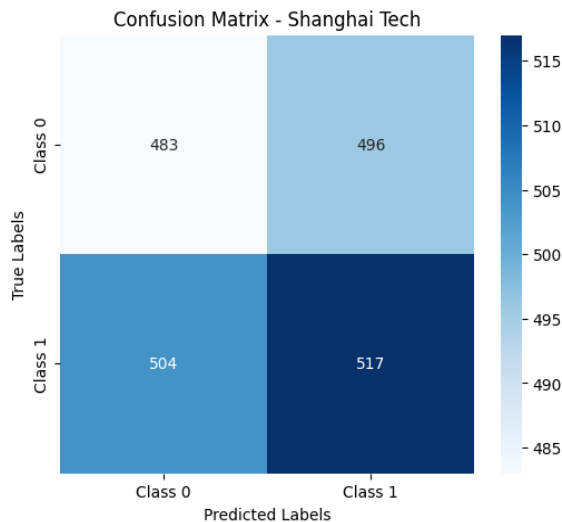


Figure 13. Confusion matrix for shanghai tech data set

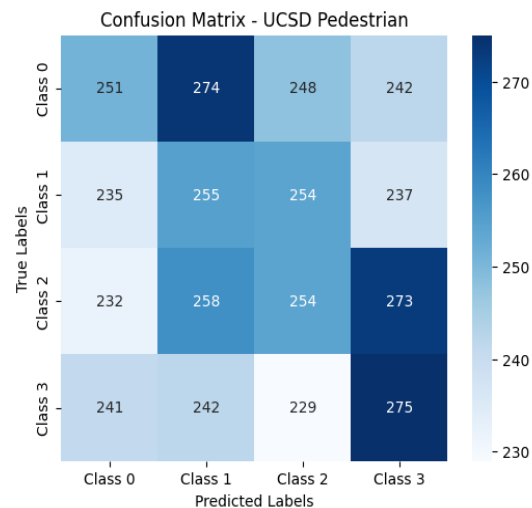


Figure 14. Confusion matrix for UCSD pedestrian data set

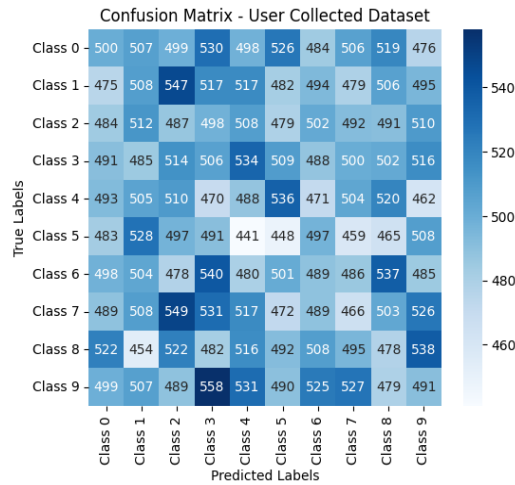


Figure 15. Confusion matrix for user collected data set

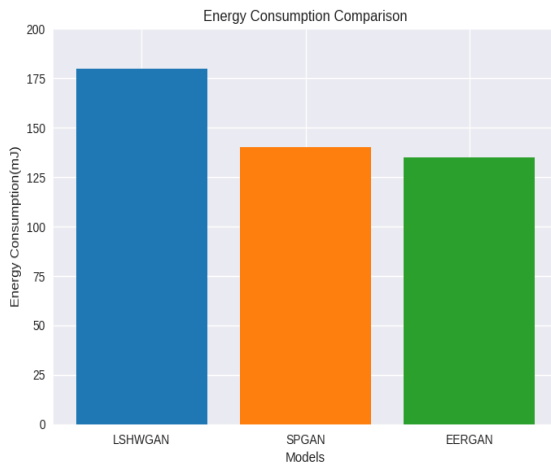


Figure 16. Comparison of the energy efficiency metrics

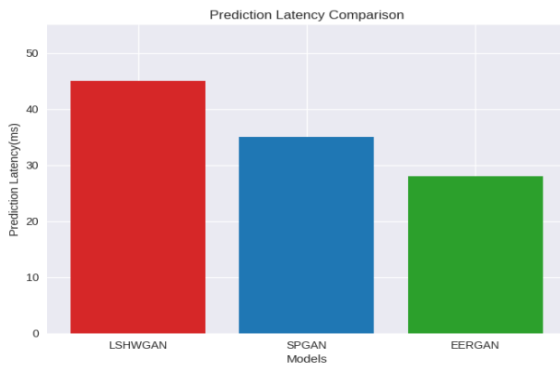


Figure 17. Comparison of the prediction latency

6. CONCLUSION

The proposed EERGAN with LSSVM validate the experimental results for detecting the objects in crowded areas energy efficiently and more accurately. The proposed Enhanced Energy Regularized Generative Adversarial Networks with LSSVM to achieve an accurately, energy efficiently and more robustness for real time processing of abnormal object detection in crowded places. The EERGAN achieved an more accuracy 97.7%, outperforming SPGAN 96.7%, AVAEGAN 96.8 %, SRSGAN 96.9% and executing

more precision (95.68), recall (95.99), F-Score (98.21), reflecting a robust capability in distinguishing aggressive from non-aggressive behaviors classification. Notably, Energy Enhanced Regularized GANs consume 25% less energy compared to SPGAN and 32 % less compared to LSHWGAN, highlighting their energy efficiency, which is crucial for resource-constrained surveillance systems. With a prediction latency of just 28 milliseconds, EERGAN-LSSVM offer superior real-time performance compared to LSGAN (50 ms) and LSHWGAN (45 ms). The confusion matrix analysis further confirms the model's high classification accuracy for both aggressive and non-aggressive instances. In future, the proposed model not only excels in predictive accuracy, robustness and energy efficient for abnormal object detection for given bench mark data set. The same model will be analyzed the advance data sets through strengthening the architecture for ensuring the public safety and security.

REFERENCES

- [1] Song, B., Sheng, R. (2020). Crowd counting and abnormal behavior detection via multiscale GAN network combined with deep optical flow. *Mathematical Problems in Engineering*, 2020(1): 6692257. <https://doi.org/10.1155/2020/6692257>
- [2] Al Jaber, S.M., Patel, A., Al-Masri, A.N. (2023). Object tracking and detection techniques under GANN threats: A systemic review. *Applied Soft Computing*, 139: 110224. <https://doi.org/10.1016/j.asoc.2023.110224>
- [3] Han, Q., Wang, H., Yang, L., Wu, M., Kou, J., Du, Q., Li, N. (2020). Real-time adversarial GAN-based abnormal crowd behavior detection. *Journal of Real-Time Image Processing*, 17(6): 2153-2162. <https://doi.org/10.1007/s11554-020-01029-z>
- [4] Alafif, T., Alzahrani, B., Cao, Y., Alotaibi, R., Barnawi, A., Chen, M. (2022). Generative adversarial network based abnormal behavior detection in massive crowd videos: A Hajj case study. *Journal of Ambient Intelligence and Humanized Computing*, 13(8): 4077-4088. <https://doi.org/10.1007/s12652-021-03323-5>
- [5] Wastupranata, L.M., Kong, S.G., Wang, L. (2024). Deep learning for abnormal human behavior detection in surveillance videos - A survey. *Electronics*, 13(13): 2579. <https://doi.org/10.3390/electronics13132579>
- [6] Asl, V., Karasfi, B., Masoumi, B. (2022). Abnormal behavior detection over normal data and abnormal-augmented data in crowded scenes. *Journal of AI and Data Mining*, 10(2): 171-183. <https://doi.org/10.22044/JADM.2022.11385.2288>
- [7] Nawaratne, R., Alahakoon, D., De Silva, D., Yu, X. (2019). Spatiotemporal anomaly detection using deep learning for real-time video surveillance. *IEEE Transactions on Industrial Informatics*, 16(1): 393-402. <https://doi.org/10.1109/TII.2019.2938527>
- [8] Wang, H., Chaw, J.K., Goh, S.K., Shi, L., Tin, T.T., Huang, N., Gan, H.S. (2023). Super-resolution GAN and global aware object detection system for vehicle detection in complex traffic environments. *IEEE Access*, 12: 113442-113462. <https://doi.org/10.1109/ACCESS.2024.3442484>
- [9] Rabbi, J., Ray, N., Schubert, M., Chowdhury, S., Chao, D. (2020). Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object

- detector network. *Remote Sensing*, 12(9): 1432. <https://doi.org/10.3390/rs12091432>
- [10] Chen, Y., Li, J., Niu, Y., He, J. (2019). Small object detection networks based on classification-oriented super-resolution GAN for UAV aerial imagery. In 2019 Chinese Control and Decision Conference (CCDC), Nanchang, China, pp. 4610-4615. <https://doi.org/10.1145/3664598>
- [11] Musunuri, Y.R., Kim, C., Kwon, O.S., Kung, S.Y. (2024). Object detection using ESRGAN with a sequential transfer learning on remote sensing embedded systems. *IEEE Access*, 12: 102313-102327. <https://doi.org/10.1109/ACCESS.2024.3432532>
- [12] Machado, P., Oikonomou, A., Ferreira, J.F., McGinnity, T.M. (2021). HSMD: An object motion detection algorithm using a hybrid spiking neural network architecture. *IEEE Access*, 9: 125258-125268. <https://doi.org/10.1109/ACCESS.2021.3110933>
- [13] Kasabov, N., Dhoble, K., Nuntalid, N., Indiveri, G. (2013). Dynamic evolving spiking neural networks for on-line spatio-and spectro-temporal pattern recognition. *Neural Networks*, 41: 188-201. <https://doi.org/10.1016/j.neunet.2012.11.014>
- [14] Cai, R., Wu, Q., Wang, P., Sun, H., Wang, Z. (2012). Moving target detection and classification using spiking neural networks. In *Intelligent Science and Intelligent Data Engineering: Second Sino-foreign-interchange Workshop, IScIDE 2011, Xi'an, China*, pp. 210-217. https://doi.org/10.1007/978-3-642-31919-8_27
- [15] Ziegler, A., Vetter, K., Gossard, T., Tebbe, J., Otte, S., Zell, A. (2024). Detection of Fast-Moving Objects with Neuromorphic Hardware. *arXiv preprint arXiv:2403.10677*. <https://doi.org/10.48550/arXiv.2403.10677>
- [16] Jin, X., Zhang, M., Yan, R., Pan, G., Ma, D. (2023). R-SNN: Region-based spiking neural network for object detection. *IEEE Transactions on Cognitive and Developmental Systems*, 16(3): 810-817. <https://doi.org/10.1109/TCDS.2023.3311634>
- [17] Kim, S., Park, S., Na, B., Yoon, S. (2019). Spiking-yolo: Spiking neural network for real-time object detection. *arXiv preprint arXiv:1903.06530*
- [18] Seras, A. M., Del Ser, J., Garcia-Bringas, P. (2023). Efficient object detection in autonomous driving using spiking neural networks: Performance, energy consumption analysis, and insights into open-set object discovery. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, Bilbao, Spain, pp. 5756-5763. <https://doi.org/10.1109/ITSC57777.2023.10422244>
- [19] Lien, H.H., Chang, T.S. (2022). Sparse compressed spiking neural network accelerator for object detection. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 69(5): 2060-2069. <https://doi.org/10.1109/TCSI.2022.3153653>
- [20] Zhu, Y., Zhang, Y., Xie, X., Huang, T. (2022). An FPGA accelerator for high-speed moving objects detection and tracking with a spike camera. *Neural Computation*, 34(8): 1812-1839. https://doi.org/10.1162/neco_a_01520
- [21] Zhao, J., Mathieu, M., LeCun, Y. (2016). Energy-based generative adversarial network. *arXiv preprint arXiv:1609.03126*. <https://doi.org/10.48550/arXiv.1609.03126>
- [22] Berthelot, D., Schumm, T., Metz, L. (2017). BEGAN: Boundary equilibrium generative adversarial networks. *arXiv preprint arXiv:1703.10717*. <https://doi.org/10.48550/arXiv.1703.10717>
- [23] Bhandari, A., Tripathy, B., Adate, A., Saxena, R., Gadekallu, T.R. (2022). From beginning to BEGANing: Role of adversarial learning in reshaping generative models. *Electronics*, 12(1): 155. <https://doi.org/10.3390/electronics12010155>
- [24] Zhang, X., Shi, S., Sun, H., Chen, D., Wang, G., Wu, K. (2024). ACVAE: A novel self-adversarial variational auto-encoder combined with contrast learning for time series anomaly detection. *Neural Networks*, 171: 383-395. <https://doi.org/10.1016/j.neunet.2023.12.023>
- [25] Jolicoeur-Martineau, A. (2018). The relativistic discriminator: A key element missing from standard GAN. *arXiv preprint arXiv:1807.00734*. <https://doi.org/10.48550/arXiv.1807.00734>
- [26] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Loy, C.C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*. https://doi.org/10.1007/978-3-030-11021-5_5
- [27] Niu, L., Li, S., Li, Z. (2023). Learning kernel stein discrepancy for training energy-based models. *Applied Sciences*, 13(22): 12293. <https://doi.org/10.3390/app132212293>
- [28] Li, J., Chen, Z., Cheng, L., Liu, X. (2022). Energy data generation with wasserstein deep convolutional generative adversarial networks. *Energy*, 257: 124694. <https://doi.org/10.1016/j.energy.2022.124694>
- [29] Kheradpisheh, S.R., Ganjtabesh, M., Thorpe, S.J., Masquelier, T. (2017). STDP-based spiking deep convolutional neural networks for object recognition. *Neural Networks*, 99: 56-67. <https://doi.org/10.1016/j.neunet.2017.12.005>
- [30] Wu, Y., Deng, L., Li, G., Zhu, J., Shi, L. (2018). Spatio-temporal backpropagation for training high-performance spiking neural networks. *Frontiers in Neuroscience*, 12: 331. <https://doi.org/10.3389/fnins.2018.00331>
- [31] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Li, F.F. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp. 248-255. <https://doi.org/10.1109/CVPR.2009.5206848>
- [32] Wen, Y., Chen, J., Sheng, B., Chen, Z., Li, P., Tan, P., Lee, T.Y. (2021). Structure-aware motion deblurring using multi-adversarial optimized cyclegan. *IEEE Transactions on Image Processing*, 30: 6142-6155. <https://doi.org/10.1109/TIP.2021.3092814>
- [33] Yamazaki, K., Vo-Ho, V.K., Bulsara, D., Le, N. (2022). Spiking neural networks and their applications: A review. *Brain Sciences*, 12(7): 863. <https://doi.org/10.3390/brainsci12070863>
- [34] Mostafa, R., Baraka, H., Bayoumi, A. (2022). Lmot: Efficient light-weight detection and tracking in crowds. *IEEE Access*, 10: 83085-83095. <https://doi.org/10.1109/ACCESS.2022.3197157>
- [35] Güney, E., Bayılmış, C. (2022). An implementation of traffic signs and road objects detection using faster R-CNN. *Sakarya University Journal of Computer and Information Sciences*, 5(2): 216-224. <https://doi.org/10.35377/saucis.05.02.1073355>