2024 9th International Symposium on Computer and Information Processing Technology (ISCIPT) | 979-8-3503-8840-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISCIPT61983.2024.10673176

LGMD-based Neural Network for Detecting Abnormal Velocity Targets in Moving Crowd

Zhongxiang Zhao [1,2], Bin Hu [1,2,3,*]

[1] State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang 550025, P.R. China

[2] Department of Computer Science, College of Computer Science & Technology, Guizhou University, Guiyang 550025, P.R. China

[3] Artificial Intelligence Research Institute of Guizhou University, Guizhou University, Guiyang 550025, P.R. China

* Corresponding Author: Bin Hu, Email: bhu1@gzu.edu.cn.

Abstract—**Detecting abnormal velocity targets is critical especially in public surveillance for crowd activity monitoring. Although there are some explorations on one such issue, traditional methods still perform poorly in complex scenes. In this paper, a novel visual neural network is investigated to perceive abnormal velocity targets in moving crowd, based on the latest neurophysiological achievements revealed in locusts' vision systems. The proposed neural network contains two neural counterparts, i.e., the presynaptic and the postsynaptic networks. The former one receives visual signals and processes them to capture the motion cues of different targets, and the latter one filters the salience energies to perceive abnormal velocity targets in the field of view. Numerical experiments carried out show that the proposed neural network can effectively detect abnormal velocity targets in moving crowd. This study is an important step towards dynamic visual information processing in crowd behavior analysis.**

Keywords-abnormal target detection; velocity sensitive; visual neural network; crowd behavior analysis; locust vision system; video surveillance

I. INTRODUCTION

In public areas, motion abnormal velocity targets refer to individuals moving at abnormal speeds, which may involve security threats and abnormal behaviors. With the popularization of video surveillance in public areas, computer vision and artificial intelligence technologies provide technical support for intelligent surveillance systems that perceive these targets in real time. Although traditional anomaly detection methods suffer from weak generalization ability and high resource consumption, introducing the proven reliability of biological vision systems and their neural mechanisms into computer vision[1], provides new development opportunities for crowd activity detection and analysis.

Several researchers have explained that organisms have visual neural pathways or systems that respond preferentially to some specific movement patterns, for example, Rind has demonstrated in multiple experiments the presence of Lobule Giant Movement Detector (LGMD) neurons in the locust lobular complex that respond to target incentives that are clearly approximating the movements of the compound eye. Compared to other organisms, locust LGMD neurons are able to process visual signals quickly and efficiently for decision making, leading to an increasing number of studies on locust LGMD. Utilizing this biological mechanism, the LGMD can be used in computer vision challenges such as direction detection[2] pattern recognition[3] and motion perception[4]–[8] etc., important breakthroughs have been achieved. Therefore, it is a worthwhile scientific problem to explore the bio-inspired artificial vision system for abnormal velocity target detection based on the theory of biological visual neurology and brain cognitive science. In this paper, based on the hierarchical structure of locust visual system and the visual information processing mechanism of mammalian retina, we utilize the speed-sensitive neural mechanism expressed by some neurons in the Primary Visual Cortex (V1) and Middle Temporal (MT) of primates to propose an improved LGMD-based visual system for detecting targets with abnormal velocity[9], namely Abnormal Velocity Targets Perception Neural Network (AVTPNN). The AVTPNN is divided into presynaptic and post-synaptic networks. The presynaptic network is modeled on the visual hierarchical structure of locusts, which is used to process perceived visual motion information with the help of mammalian retinal information processing; the post-synaptic network is modeled on the cellular motion energy filtering in the V1 area of the primate visual system and the difference in neuronal energy mapping in the MT area, which makes it sensitive to the targets with salient speed, and is designed for the perception and processing of abnormal velocity targets in the crowded activity scenes. A unique artificial visual neural network is designed for the perception and early warning of targets with different speeds in crowded scenes. Based on videos of moving targets with different speeds in different real-life scenarios, systematic experiments are carried out to verify the performance of the AVTPNN.

The rest of the paper is organized as follows: section II introduces the AVTPNN for moving crowd, including its internal structure and working mechanism. After that, some experiments and results are shown in Section III. Finally, the paper is concluded in Section IV.

II. ARTIFICIAL VISUAL NEURAL NETWORK CONSTRUCTION

Inspired by the internal characteristics of the mammalian visual system [6] and inspired by the fact that visual information processing is divided into two stages, the AVTPNN, an abnormal velocity targets detection neural network proposed in this paper, consists of five neural cell layers, namely the photoreceptor layer (P), the excitatory layer (E), the integrative layer (S), the ganglionic layer (L), and the medulla (M), and the Abnormal Velocity Targets Detection ($AVTP$) neuron. Fig. 1 shows a schematic diagram of the network with the following design details of the five neural layers and $AVTP$ neurons.

Research was supported by National Natural Science Foundation of China (Grant No.62066006), Natural Science Foundation of Guizhou Province of China (Grant No: QKHJC [2020]1Y261), and Scientific Research Project for the Introduced Talents of Guizhou University (Grant No: GDRJHZ (2019)58).

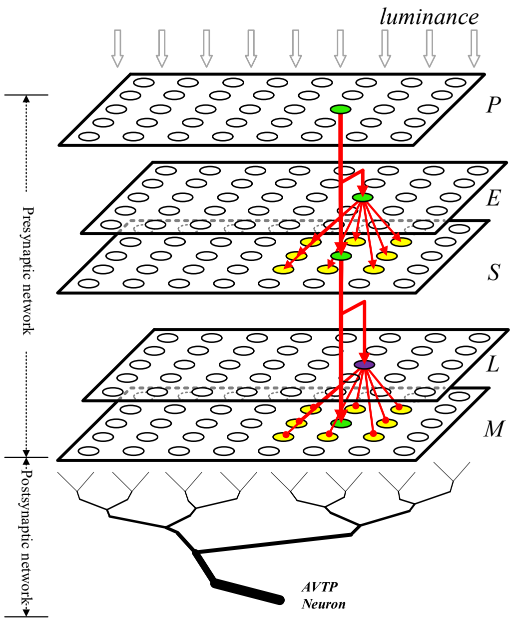

Fig.1 Structure diagram of AVTPNN.

A. P Layer

The photoreceptor layer P consists of photoreceptor cells arranged in the form of a $n_c \times n_r$ matrix, where each photoreceptor cell uniquely corresponds to a pixel point in the video image sequence. The luminance change value $P_f(x, y)$ of the pixel in the frame f can be expressed as

$$P_f(x, y) = abs(L_f(x, y) - L_{f-1}(x, y)) \tag{1}$$

where $L_f(x, y)$ and $L_{f-1}(x, y)$ denote the luminance values of the pixels in the f and $f-1$ frames, respectively; $x(x \in (0, n_c])$ and $y(y \in (0, n_r])$ denote the column and row coordinates of the cells, respectively. The collected signals are filtered by

$$\widetilde{P}_f(x, y) = \begin{cases} P_f(x, y), & if\, P_f(x, y) \geq T_p \\ 0, & otherwise \end{cases} \tag{2}$$

where T_p is the passage threshold for controlling the passage of each cell through the excitation value.

B. E Layer

The cells in the E layer are arranged in a $n_c \times n_r$ matrix, and each cell collects the delayed excitation provided by the corresponding cell in the P layer, and its output membrane potential is calculated by the following equation

$$E_f(x, y) = \widetilde{P}_{f-1}(x, y) \tag{3}$$

C. S Layer

The cells in the E layer are arranged in a $n_c \times n_r$ matrix. Modeled after bipolar cells in the mammalian retina that receive output from photoreceptor and horizontal cells in the upper layer, each cell in the S layer receives excitatory signals from corresponding cells in the P and E layers, but receives a higher pass-through weight from the P layer to the S layer to enhance the visual signal. Thus, in the S layer, the cell signal intensity $S_f(x, y)$ can be calculated from the following equation:

$$S_f(x, y) = \widetilde{P}_f(x, y)\omega_{ps} + E_f(x, y)\omega_{es} \tag{4}$$

where ω_{ps} and ω_{es} are the pass-through coefficients from the P and E layers into the S layer, respectively.

D. L Layer

The cells in the L layer are arranged in the form of a $n_c \times n_r$ matrix, and the corresponding pixel point motion velocity $L_f(x, y)$ at the moment of the f frame is calculated by the following equation

$$L_f(x, y) = \sqrt{u_f(x, y)^2 + v_f(x, y)^2} \tag{5}$$

where $u_f(x, y)$ and $v_f(x, y)$ denote the velocity in horizontal and vertical directions respectively. Here, the position change of the pixel in the image is considered to be minimal, then the gradient change of the pixel at the moment of the f th frame satisfies

$$I_{hor}u_f(x, y) + I_{ver}v_f(x, y) + I_{time} = 0 \tag{6}$$

where I_{hor} , I_{ver} and I_{time} denote the horizontal, vertical and temporal bias of the gray scale signal of the pixel point in the image, respectively. The velocity vector of the pixel point (x, y) is calculated using the temporal and spatial differentiation of the image to construct the energy function, i.e.

$$E_f(x_u, y_v) = \iint (I_{hor}u + I_{ver}v + I_{time})^2 + \alpha^2(\nabla^2 u + \nabla^2 v) \tag{7}$$

where α^2 is the scaled global smoothness term and ∇ is the Laplace operator. After using the variation algorithm, we get

$$\begin{cases} I_{hor}^2 u + I_{hor}I_{ver}v + I_{hor}I_{time} - \alpha^2 \nabla u = 0 \\ I_{ver}^2 v + I_{hor}I_{verl}u + I_{ver}I_{time} - \alpha^2 \nabla v = 0. \end{cases} \tag{8}$$

The horizontal and vertical velocities of the pixel (x, y) in the image are given by the following equation

$$\begin{cases} u^{k+1}(x, y) = \bar{u}^k - \dfrac{I_{hor}(I_{hor}\bar{u}^k + I_{ver}\bar{v}^k + I_{time})}{\alpha^2 + I_{hor}^2 + I_{ver}^2} \\ v^{k+1}(x, y) = \bar{v}^k - \dfrac{I_{ver}(I_{hor}\bar{u}^k + I_{ver}\bar{v}^k + I_{time})}{\alpha^2 + I_{hor}^2 + I_{ver}^2} \end{cases} \tag{9}$$

where the superscript k+1 indicates the next iteration of the computation and k is the result of the final computation. Here, we can only care about the magnitude of the velocity and bring the result into (5) to get the pixel point velocity magnitude.

E. M Layer

The cells in the M layer are arranged in the form of $n_c \times n_r$ matrix. With the help of primate velocity-sensitive neural properties, a set of spatiotemporal filters are constructed to separate foreground targets to obtain velocity-sensitive regions, namely

$$\widetilde{E}_f(x, y) = \chi\{E_f(x, y) : \left|\lambda - L_f(x, y)\right| \leq \mu\, and \left|\Gamma(x, y)\right| \geq \alpha\} \tag{10}$$

where χ is the schematic function to filter the noise information, λ and μ are the global velocity coefficient range threshold decision constants, and $|\Gamma(x,y)| \geq \alpha$ indicates that only exceeding the set range is retained. After the above processing, the output membrane potential signal of neural layer *M* will indicate the information of the spatiotemporal domain where the target of different speeds is located.

F. AVTP Neurons

The outputs of the excitable cells in the *G* layer all converge on the *AVTP* neurons, which summarize and normalize them, i.e., the

$$\Phi_f = 2((1+\exp(-\sum_{x=1}^{n_c}\sum_{y=1}^{n_r} E_f(x,y)/\tau)^{-1} - 1. \tag{11}$$

The excitation of *AVTP* neurons is subsequently modulated using a spike thresholding mechanism, i.e., when membrane potential excitation Φ_f continuous n_{pe} frame presence values are generated as follows

$$\hat{\Phi}_f = \begin{cases} \lceil \Phi_f \rceil, & if \sum_{i=0}^{n_{pe}} S_{f-i}^{AVTP} > n_{pe} \; and \; S_f^{AVTP} > 0 \\ 0, & otherwise. \end{cases} \tag{12}$$

Finally, the membrane potential of the *AVTP* neuron F^{AVTP} will be used as the output of the network, i.e

$$F_f^{AVTP} = \begin{cases} \hat{\Phi}_f, & if \, \hat{\Phi}_f > \gamma \; or \; \hat{\Phi}_f = 0, \\ \hat{\Phi}_f v^{(\hat{\Phi}-1)}, & otherwise. \end{cases} \tag{13}$$

Here γ is the excitation amplitude and v is the iteration coefficient. F^{AVTP} characterizes the situation of AVTPNN sensing the crowd abnormal targets, $\Phi_f^{AVTP} \neq 0$ indicates that AVTPNN senses an anomalous target in the crowd at f frame.

III. Numerical Experiments and Analysis

In this section, some public datasets with speed anomalous targets and self-filmed videos are selected to test and validate that the neural network can effectively perceive the targets with different speeds in crowd activities in real scenarios.

A. Test video

The most relevant and challenging datasets for this study were selected. These datasets include some common anomalous targets such as bicyclist targets, van targets, and pedestrians. These datasets include PETS 2009 [10], UCSD Pedestrian [11], ShanghaiTech [12]. These datasets record a wide range of population behaviors and abnormal transitions, and thus can be used to verify whether the AVTPNN is able to generate neural spike information when detecting abnormal velocity targets. To supplement the diversity of the test samples, some self-filmed videos are added here.

B. Experimental environment setup

The experimental procedure was realized on a Windows 10 computer with CPU/3.20GHz and RAM/16G. The experimental code was written in Visual Studio 2013 platform. The frame rate of the test video is regularized to 30f/s and the video image frames input to the neural net-work are regularized to 8-bit grayscale maps of 140×80 pixels. Based on the reported work [3], [6] and current experiments, the AVTPNN parameters are set as shown in Table I.

TABLE I. Parameter Settings of AVTPNN.

Parameter	Value	Parameter	Value
n_c	140	μ	50
n_r	80	α	120
ω_{es}	0.33	n_{pe}	1
ω_{ps}	0.67	γ	0.9
τ	11200	v	0.5
λ	0.95		

C. Experimental analysis

1) Validity testing

Common anomalous target objects such as vehicles traveling on one side of the sidewalk, bicycles, skateboarding targets, and running pedestrians [11], these types of anomalous targets are closely related to this study.

a) Target Irrelevance

Four videos, all from the UCSD dataset, were selected to depict the perception of the abnormal velocity phenomenon triggered by different kinds of motion targets, namely, a riding target, a skateboarding target, and a van target. And the latter three videos recorded the presence of multiple different-speed target motion behaviors in the scene. The four video example samples (with the motion information of the off-speed targets indicated by red arrows) are given in Fig. 2. The visualization results corresponding to the *M* layer of the neural network and the spike responses corresponding to the *AVTP* neurons are shown in Figs. 3 and Fig. 4, respectively, and the statistical results for this part of the video are given in Table II.

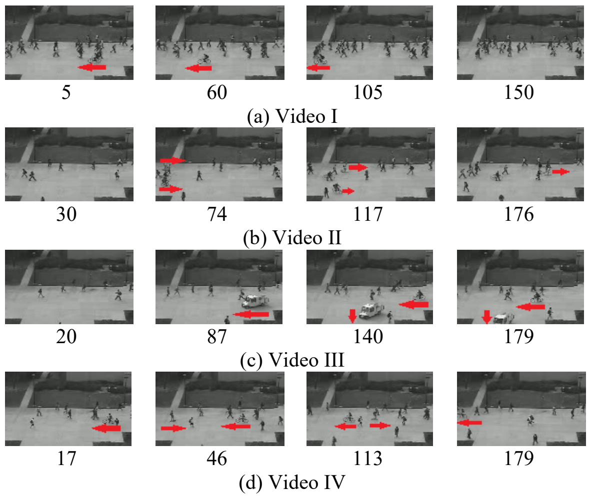

Fig. 2 Sampled frames for target-independence test.

As shown in Table II, AVTPNN can correctly detect different types of abnormal targets, and both the M layer and *AVTP* neurons respond to them. The results show that the AVTPNN can correctly perceive anomalous targets even if there are multiple anomalous targets and their regions appear in different areas of the receptive field, which is consistent with the properties of velocity-sensitive neurons in the primate visual system.

TABLE II. STATISTICAL RESULTS OF TARGET IRRELEVANCE.

Video	TP	FP	FN	FAR (%)	MAR (%)	Precision (%)
I	125	0	5	0%	3.85%	100%
II	133	0	2	0%	1.48%	100%
III	147	0	3	0%	2.00%	100%
IV	178	0	2	0%	1.11%	100%

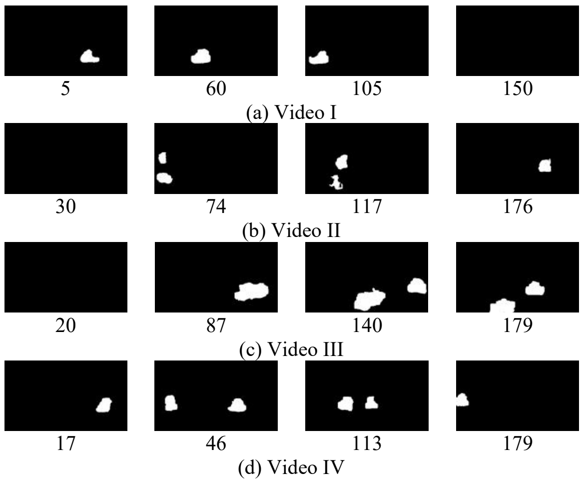

Fig. 3 Neural layer *M* visualization results.

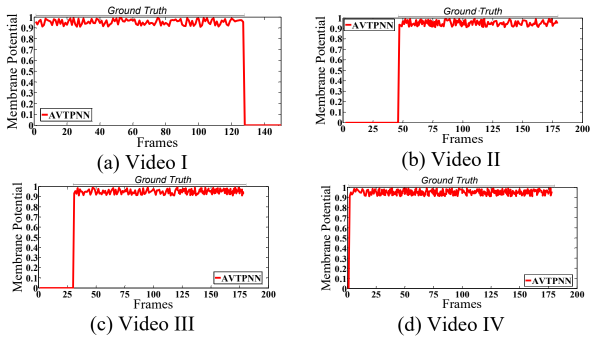

Fig. 4 Output curve of *AVTP* neuron.

b) Scene Irrelevance

The effectiveness of detecting abnormal speed targets in different scenarios is tested using selfie videos, the ShanghaiTech and the Avenue dataset, with examples of these videos shown in Fig. 5. It is significant to note that video VII is too large due to the depth distance, causing the *AVTP* neurons to be silent when the target transforms out of the abnormal velocity behavior because its range of motion is below the threshold, but responds normally when the condition is subsequently reached. The visualization results of the neural network and the peak response of *AVTP* neurons are shown in Figures 6 and 7, respectively. The statistical results of these tests are shown in Table III. The data in Table III and the output of the neural network indicate that even if there are multiple abnormal targets and their regions appear in different regions of the receptive field, AVTPNN can correctly perceive abnormal velocity targets in various complex scenes.

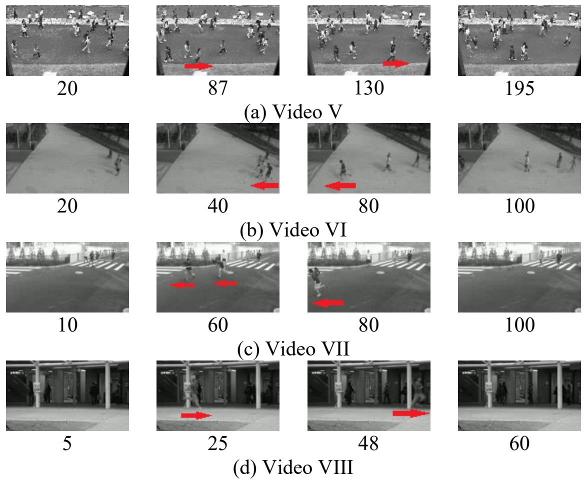

Fig. 5 Sample frames for scene irrelevance test.

TABLE III. STATISTICAL RESULTS OF SCENE IRRELEVANCE.

Video	TP	FP	FN	FAR (%)	MAR (%)	Precision (%)
V	81	0	4	0%	4.71%	100%
VI	63	0	11	0%	14.86%	100%
VII	44	1	11	2.22%	20%	97.78%
VIII	43	0	1	0%	2.27%	100%

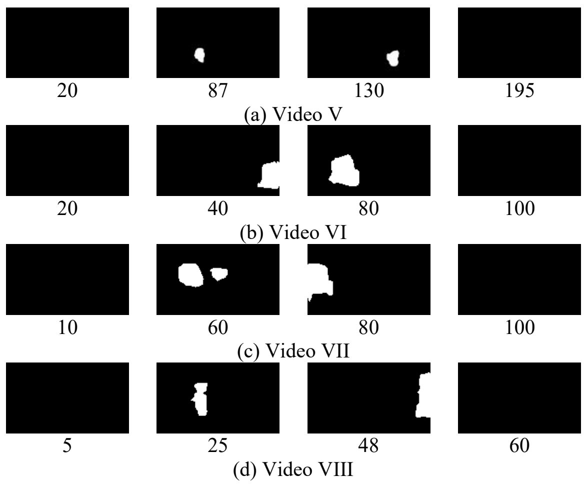

Fig. 6 Neural layer *M* visualization results.

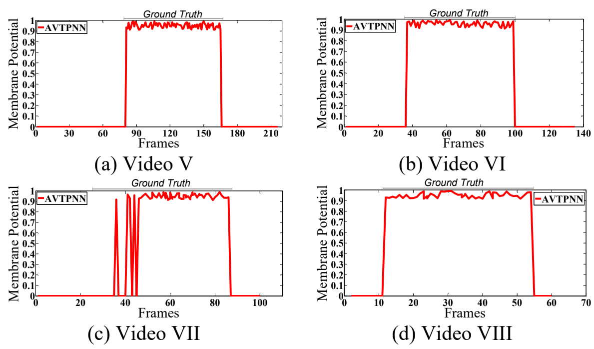

Fig. 7 Output curve of *AVTP* neuron.

D. Comparison experiments

The UCSD Pedestrian dataset ped2 sequence was selected for the [11] real-scene video test, selecting the model test that has been established in crowd anomaly detection in the past two years, and using AUC and ERR as evaluation indexes. The experimental results are shown in Table IV.

TABLE IV. STATISTIC RESULTS OF CONTRAST EXPERIMENTS.

Author	Methodologies	AUC (%) ↑	EER (%) ↓
Yang et al.[13]	MFMCPOS	91.6	17.5
Liu et al.[14]	TSSTGM	95.1	14.1
Zhang and Wang[15]	U-Net	92.4	-
Fu, Xuan and Wang[16]	MAND	95.4	12.0
Yuan and Zhang[17]	RS-BCAE	94.0	12.0
Ours	AVTPNN	97.5	4.9

As can be seen from Table IV, the AVTPNN proposed in this paper achieves the combined best in the above two evaluation indexes. The AVTPNN proposed in this article extracts clues of crowd movement changes to achieve abnormal velocity targets detection by simulating the information processing mechanism in biological vision systems. However, there are issues with poor universality and dependency in practical applications through learning and training. These models either rely on specific types of noise free video data, slow detection speed, or overly rely on the correctness of the dataset, which limits their effectiveness in complex dynamic scenes.

The above experiments show that by modeling primate velocity-sensitive neural properties, AVTPNN is able to detect sudden motion abnormal velocity targets in a crowd, i.e., whenever a localized velocity-abnormal target appears in the visual field domain, AVTPNN is able to effectively perceive the abnormal velocity region and generate responses and warnings.

IV. CONCLUSIONS

Based on the bionic visual information processing mechanism, this paper adopts a new generation of computer vision technology to study and design an Abnormal Velocity Targets Perception Neural Network for dealing with moving crowd. The proposed neural network simulates hierarchical biological visual information processing framework based on the LGMD model, which can effectively perceive and warn abnormal velocity moving targets in visual scenes. Validated with multiple public datasets and self-timed scenes, the results show that the AVTPNN can quickly and efficiently detect the presence of abnormal velocity targets in complex scenes. This artificial vision system that combines visual neurophysiology and computer vision techniques provides a powerful help for analyzing novel artificial intelligence vision systems.

ACKNOWLEDGMENT

The authors sincerely thank the anonymous reviews for their valuable comments and the editors for their support.

REFERENCES

[1] V. J. Kok, M. K. Lim, and C. S. Chan, “Crowd behavior analysis: A review where physics meets biology,” Neurocomputing, vol. 177, pp. 342–362, Feb. 2016, doi: 10.1016/j.neucom.2015.11.021.

[2] S. Yue and F. C. Rind, “Postsynaptic organisations of directional selective visual neural networks for collision detection,” Neurocomputing, vol. 103, pp. 50–62, Mar. 2013, doi: 10.1016/j.neucom.2012.08.027.

[3] B. Hu, Z. Zhang, and L. Li, “LGMD-based visual neural network for detection crowd escape behavior,” in 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), Nov. 2018, pp. 772–778. doi: 10.1109/CCIS.2018.8691354.

[4] B. Hu, S. Yue, and Z. Zhang, “A Rotational Motion Perception Neural Network Based on Asymmetric Spatiotemporal Visual Information Processing,” IEEE Trans. Neural Networks Learn. Syst., vol. 28, no. 11, pp. 2803–2821, Nov. 2017, doi: 10.1109/TNNLS.2016.2592969.

[5] B. Hu and Z. Zhang, “Bio-plausible visual neural network for spatio-temporally spiral motion perception,” Neurocomputing, vol. 310, pp. 96–114, Oct. 2018, doi: 10.1016/j.neucom.2018.05.015.

[6] B. Hu and Z. Zhang, “Bio-inspired visual neural network on spatio-temporal depth rotation perception,” Neural Comput. Appl., vol. 33, no. 16, pp. 10351–10370, Aug. 2021, doi: 10.1007/s00521-021-05796-z.

[7] C. Liu and B. Hu, "Biologically Inspired Neural Network for Sudden Local Aggregation Perception in Crowds," Computer Engineering and Applications, vol. 58, no. 16, p. 11, 2022.

[8] B. Zhang and B. Hu, "Neural Network for Small Moving Object Pedestrian Detection Based on Situational Memory," Computer Engineering and Applications, vol. 58, no. 15, p. 15, 2022.

[9] D. C. Bradley and M. S. Goyal, “Velocity computation in the primate visual system,” Nature Reviews Neuroscience, vol. 9, no. 9. Nature Publishing Group, pp. 686–695, Sep. 13, 2008. doi: 10.1038/nrn2472.

[10] J. Ferryman and A. Shahrokni, “PETS2009: Dataset and challenge,” in 2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Dec. 2009, pp. 1–6. doi: 10.1109/PETS-WINTER.2009.5399556.

[11] A. B. Chan and N. Vasconcelos, “Modeling, Clustering, and Segmenting Video with Mixtures of Dynamic Textures,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 5, pp. 909–926, May 2008, doi: 10.1109/TPAMI.2007.70738.

[12] W. Luo, W. Liu, and S. Gao, “A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework,” in 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017, vol. 2017-Octob, pp. 341–349. doi: 10.1109/ICCV.2017.45.

[13] M. Yang, Y. Feng, A. S. Rao, S. Rajasegarar, S. Tian, and Z. Zhou, “Evolving graph-based video crowd anomaly detection,” Vis. Comput., vol. 40, no. 1, pp. 303–318, Jan. 2024, doi: 10.1007/s00371-023-02783-4.

[14] W. Liu, J. Cao, Y. Zhu, B. Liu, and X. Zhu, “Real-time anomaly detection on surveillance video with two-stream spatio-temporal generative model,” Multimed. Syst., vol. 29, no. 1, pp. 59–71, Feb. 2023, doi: 10.1007/s00530-022-00979-7.

[15] Y. Zhang and Y. Wang, "Future Frame Prediction for Abnormal Behavior Detection Algorithm Based on Attention and Feature Fusion," Journal of Detection and Control, 2023, vol. 45, no. 4, pp. 150-156.

[16] M. Fu, S. Xuan, T. Wang, et al. "Video Anomaly Detection Based on Temporal and Spatial Attention Mechanisms," Journal of Computer Technology and Development, 2023, vol. 33, no. 8, pp. 51-58.

[17] H. Yuan, W. Zhang. "Anomaly Detection Method Integrating SimAM Attention Mechanism and Bidirectional ConvLSTM," Journal of Small and Micro Computer Systems, 2023, vol. 44, no. 8, pp. 1777-1784.