WILEY | Hindawi

Research Article

Real-Time Monitoring of College Sports Dance Competition Scenes Using Deep Learning Algorithms

Fei Yang,[1,2] GeMuZi Wu,[2] and HongGang Shan [2]

[1]Graduate School, Namseoul University, Cheonan, Republic of Korea
[2]Department of Sports Art, Hebei Institute of Physical Education, Hebei, China

Correspondence should be addressed to HongGang Shan; 2008006@hepec.edu.cn

In order to improve the real-time detection effect, therefore, a research on real-time scene detection of sports dance competition based on deep learning is proposed. The collected scene image is grayed by using the weighted average method, and the best image interpolation is calculated by using the deep learning method, so as to realize the smooth processing of sawtooth and mosaic information generated by panoramic mapping. After selecting the cube model, the processed scene information is projected to the visual plane to construct the panorama of the competition scene. Finally, combined with the three-frame difference, the changes between adjacent image frames are calculated to obtain the moving target. The test results show that the motion detection accuracy of professional dancers can reach more than 75.0% and that of amateur dancer can reach more than 64.2%.

1. Introduction

With the development of modern technology, digital media technology can integrate text, images, sound, video, etc. through computers. The logical relationship between them is established through the processes of sampling quantization, editing and modification, and encoding and compression. With its incomparable advantages of convenience, accuracy, high efficiency, convenient storage, and easy modification, the computer constantly refreshes various records. In recent years, with the rapid development of multimedia technology and the continuous upgrading of software, the wide application of computer in the field of design and performance has won broad prospects. Multimedia technology has not only brought a revolution to the stage visual performance in a new form but also brought great changes to people's aesthetic ideas. With the wide application of multimedia technology, the art of dance beauty is breaking through the limitations and closure of the traditional manual era. The stage is designed as a complex of space-time art, including the temporality and hearing of literature and music, as well as the spatiality and vision of painting and architecture [1–3]. The function of dance art should not only expand the visual scope of the audience and editors but also expand people's thinking ability. The important role of computer technology in dance beauty creation and graphic design provides sports dance practitioners with new ideas. Computers can become a new "stage language." People also have enough reasons to believe that the combination of computer and sports dance will stimulate more creative passion and artistic spirit. Due to the late start of sports dance in China, there is an urgent need to improve the competitive level, teachers, teaching, and scientific research [4, 5]. Sports dance provides a positive and healthy form of sports for college students to establish friendship, cultivate personality, exchange feelings, strengthen physique, cultivate sentiment, shape body shape, and improve skills, which is deeply loved by college students. According to the results of a questionnaire in a college, a survey was conducted on the mental health of 2000 students participating in sports dance before and after training. The results showed that sports dance had a significant effect on shaping healthy personality,

strong physique, and normal interpersonal relationship. However, the current situation is that the students' increasing learning enthusiasm does not match the teaching level and venues in colleges and universities, which specifically shows that there is a general lack of teaching venues in colleges and universities all over the country, and the instruments and equipment cannot meet the needs of sports dance teaching; the quality of sports dance teachers is generally not high, and the level of teachers is uneven; the scientific research level of sports dance lags behind seriously. Especially for the construction of sports dance–related competitions, there is an obvious lag problem. For the competition environment and the participants' action, capture cannot meet the real-time requirements. At the same time, for the controversial parts, there is a lack of more powerful confirmation of video resources [6–8]. This has restrained the development of sports dance to a certain extent. On the one hand, sports dance itself has great limitations on development, which is difficult to form a good development trend through self-competition. On the other hand, due to the lack of effective professional development and management mechanism, most students have low enthusiasm to participate in the major, and the students of their major cannot find good development in the industry after graduation. In this context, it is very necessary to use modern technology to strengthen the real-time detection effect of the game scene. Reference [9] proposes a spatiotemporal sparse RPCA moving object detection algorithm. This method spatiotemporally regularizes the sparse components in the form of the graph Laplacian. Each Laplacian corresponds to a multifeature map constructed over the superpixels of the input matrix. While minimizing the RPCA objective function, the sparse component is used as the feature vector of the spatiotemporal graph Laplacian to realize the detection of scene objects. Reference [10] proposes a 3D motion detection and long-term tracking method. This method provides a new energy function optimization framework for motion pose estimation. It can independently estimate the 3D motion pose of each object. However, considering the complexity of the scene, the effect of real-time target detection is not ideal. Based on the difficulties in sports dance teaching described above, it has increasingly highlighted the necessity and urgency of digital media technology in sports dance. First of all, its advantage is that it can greatly improve the existing competition conditions [11, 12]. At present, the sports venues in most colleges and universities, especially those suitable for sports dance competitions, are very tight, with short opening hours and a large number of participants. Secondly, real-time detection of competition scenes through digital media technology is also of great value to avoid sports injuries. In some intense and confrontational movements, athletes are often vulnerable to injury, and the vast majority of injuries occur in a state beyond the control of athletes. Using digital media technology to learn movements can avoid sports injuries at this stage. If the virtual reality technology is used for practice, students can safely and boldly analyze the actions without considering these problems, which will not cause any harm to people. At the same time, they can also point out the shortcomings of students' actions, put forward suggestions, and score comprehensively, which can improve the evaluation effect of the competition. In addition, it can also break through the limitation of time and space [13–15]. Every athlete is eager to get the support of world-class coaches, and every world-class coach is also eager to promote his training concept to every corner of the world. However, due to time and space constraints, it cannot be realized. With digital media technology, all this will not be a dream. The realization of all this is inseparable from the effective detection of the game scene.

On this basis, a real-time detection method of sports dance competition scene based on deep learning is proposed. Firstly, the collected game scene image information is preprocessed. We take advantage of the information feature extraction of deep learning to improve the detection effect of game scenes. Finally, in a random test of 10 motion dancers, the recognition rate of 5 professional motion dancers was above 75%. The recognition rate of the other 5 amateur sports dancers is also above 64%. Through this research, we also hope to provide valuable help for the development of sports dance competition.

2. Image Preprocessing

When using digital media technology to collect the scene information of sports dance competition, due to the reasons of the equipment itself or the environment in the competition scene, it is very easy to cause incomplete or fuzzy information, so the effect of scene detection is not ideal. Therefore, this paper first preprocesses the collected game scene image information [16–18]. So as to ensure that the panoramic image of the game scene constructed later can meet the detection requirements. Since most real-time image acquisition devices store information in RGB color channel model in. jpg compression format, this paper first grays the image, which can effectively reduce the extraction and processing time of SIFT features, so as to meet the requirements of real-time detection of moving objects.

Image graying is the process of making the R, G, B components of image color equal. Since the value range of R, G, B is 0~255, the gray level is only 256. That is, grayscale images can only show 256 grayscales [19, 20]. The image gray processing method used in this paper is weighted average method. It has the advantage that the importance of different components can be considered. Different weights can be assigned to the three components according to their respective importance. Then, take the weighted average as the grayscale result. Make the image grayscale more in line with the needs of practical applications. According to the importance of its color or other indicators, give different weighted values to the three values of R, G, B, and make them weighted average. Since human eyes have the highest sensitivity to green and the lowest sensitivity to blue, the specific weighting formula is

$$R = G = B = 0.299R + 0.587G + 0.114B. \tag{1}$$

Because the sports dance competition scene contains moving objects, on the basis of real-time detection, it is

required to improve the speed and ensure the accuracy as much as possible. Therefore, we should try to avoid floating-point operation in practical application. Floating point arithmetic is real arithmetic. Because computers can only store integers, floating-point operations are slow and prone to errors. Combined with this demand, this paper adjusts the weighted calculation to the full integer algorithm and scales formula (1) one thousand times to realize the integer operation algorithm. At this time, the image gray processing result is

$$\text{Gray} = \frac{(299R + 587G + 114B + 500)}{1000}, \qquad (2)$$

where Gray is the game scene image after gray processing. It should be noted that the accuracy of RGB three-color channel is generally 8-bit accuracy. After scaling it a thousand times, the subsequent corresponding image operations are also 32-bit integer data operations. The division following formula (2) is integer division, and the purpose of adding 500 is to realize rounding. Because the algorithm needs 32-bit operation, the time will increase. Combined with the real-time requirements of sports dance competition scene detection, this paper further processes formula (2), and the final result is

$$\text{Gray} = \frac{(30R + 59G + 11B + 50)}{100}. \qquad (3)$$

In this way, the unified preprocessing of dance competition scene image information is realized.

3. Build a Real-Time Panorama of Sports Dance Competition

3.1. Scene Image Interpolation Based on Deep Learning. After the above image graying, in order to have higher accuracy in the later stage of the game scene stitching, this paper carries out cylindrical mapping projection on the sequence images. However, in the actual processing process, it is found that the transformed point coordinates are often not integer coordinate values. If these coordinate values are simply calculated as integers, great errors will be caused, resulting in geometric distortion of the projected image [21, 22]. Therefore, this paper uses image interpolation technology. The algorithm implementation of image interpolation is generally divided into forward mapping method and backward interpolation method. The forward mapping algorithm is to map from the source image to the target image; the backward mapping algorithm is the reverse mapping from the target image to the source image. In the process of pixel transformation, the forward mapping algorithm may be projected to the outside of the image area, resulting in multiple calculations, and the gray value of some pixels of the target image will be repeatedly determined, which not only wastes the calculation time but also sometimes affects the real-time performance. The backward mapping algorithm generates the output image for each pixel without interval. The pixel gray value of each target image is determined by the color

values of the pixels of four source images after interpolation algorithm, and then the output image is generated [23–25]. In this paper, the nearest neighbor interpolation algorithm is used to realize this process. In the specific operation process, firstly, the gray value of the input pixel closest to the position mapped to the target image is selected as the interpolation result. After adding geometric transformation, the corresponding coordinate values of the pixels with coordinates (x, y) on the output image on the original image are (u, v). At this time, the formula of the nearest interpolation algorithm can be expressed as

$$\begin{cases} g(x, y) = f(x, y) \\ x = u + 0.5(\text{Take an integer}), \\ y = v + 0.5(\text{Take an integer}) \end{cases} \qquad (4)$$

where $g(x, y)$ represents the position information of the target image, and $f(x, y)$ represents the position information closest to the target image. This algorithm is simple and fast. However, this method will cause obvious jagged edges and mosaic in the newly generated image. However, this approach results in noticeable jagged edges in the newly generated image. Therefore, this paper uses the method of deep learning to smooth the sawtooth and mosaic information. In order to obtain the more accurate value of the output image pixel of the fuzzy part, it is not enough to only use the four nearest pixels of the input image pixel as the object of depth learning. Therefore, this paper takes the influence of the fuzzy point to the surrounding 16 nearest pixels as the learning goal. Constructing insert value learning function

$$S(x) = \sin(\pi x)/\pi x, \qquad (5)$$

where $S(*)$ represents the learning function. In this way, the values of each point between collected images can be accurately obtained. Use formula (5) to iteratively approximate the best interpolation function. The specific iterative method is

$$S(x) = \begin{cases} 1 - (\lambda + 3)|x|^2 + (\lambda + 2)|x|^3, |x| < 1 \\ -4\lambda + 8\lambda|x| - 5\lambda|x|^2 + \lambda|x|^3, 1 \le |x| \le 2, \\ 0, |x| \ge 2 \end{cases} \qquad (6)$$

where λ represents the learning coefficient. Finally, according to the actual image processing effect, the value of λ is -1. In this way, the fast interpolation processing of competition scene information is realized with a small amount of calculation and simple algorithm. At the same time, the influence of other adjacent pixels is considered to ensure that the gray level of the collected real-time scene image maintains obvious continuity, the loss of image quality is minimized, and the phenomenon of image sawtooth and mosaic is avoided.

3.2. Panoramic Construction of Sports Dance Competition. Panorama image generally means that the angle of view of the image is greater than the normal visual angle of human eyes, that is, it is about 90 degrees in the horizontal direction
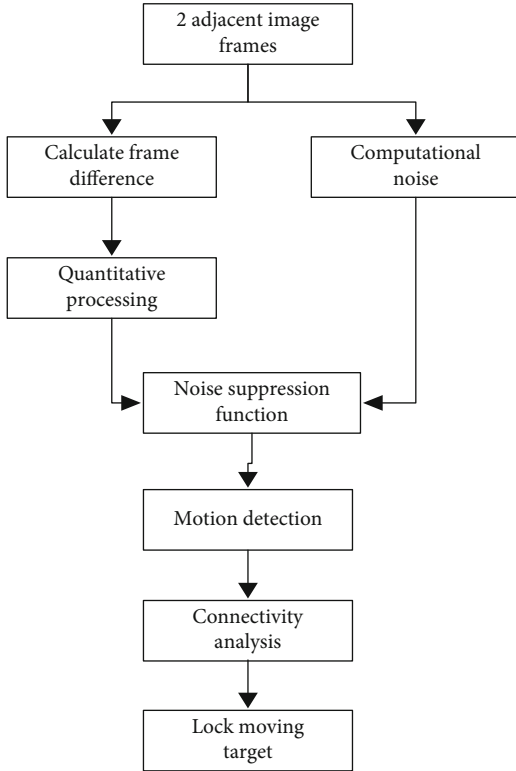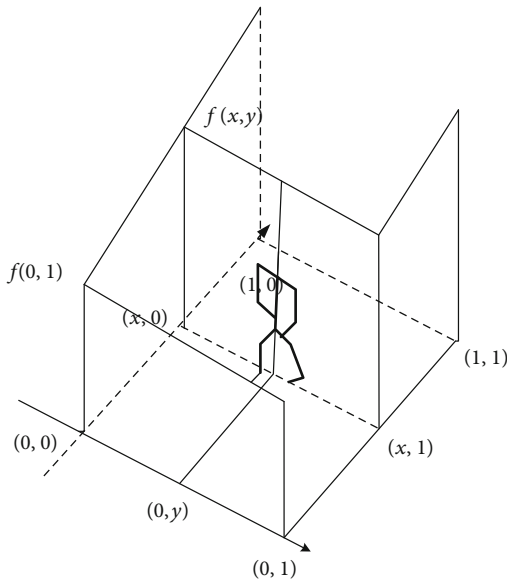
Figure 1: Scene real-time detection method.

Figure 2: Portrait fitting results.

and about 70 degrees in the vertical direction. In addition to using special equipment to obtain panoramic images, panoramic images can also be obtained through image stitching and stitching technology. Generally, first, the rotating camera is used to collect the sequence video images to ensure that the sequence images are partially overlapped. Then, the mosaic algorithm is used to splice the collected sequence video images to form a panoramic mosaic image.

On the basis of the above image processing, this paper mainly includes several steps in the process of panoramic image stitching and stitching: panoramic image generation, model selection, collecting sequence images and ensuring image overlap, image stitching. For the panorama generation model, this paper selects the cube model to project the collected scene information to the visual plane. The cube model projection allows the user to view the image 180 degrees in the vertical direction and 360 degrees in the horizontal direction. The formed cube panorama is composed of six planes, which are projected onto the six surfaces of the cube and spliced. Projection transformation formula is the key formula of panoramic mosaic image cube projection transformation and inverse transformation method. Therefore, this paper establishes the projection transformation coordinate system. The cube takes the central position point o as the origin of the coordinate system, and R is the variable length of the cube, α $(-\pi <= \alpha <= \pi)$ as azimuth and β $(-\pi/2 <= \beta <= \pi/2)$ as the pitch angle. The coordinate equation of six faces can be expressed as

Left plane

$$y = -\frac{r}{2}\left(-\frac{r}{2} \leq x, z \leq \frac{r}{2}\right). \tag{7}$$

Right plane

$$y = \frac{r}{2}\left(\frac{r}{2} \leq x, z \leq \frac{r}{2}\right). \tag{8}$$

Anterior plane

$$x = \frac{r}{2}\left(-\frac{r}{2} \leq y, z \leq \frac{r}{2}\right). \tag{9}$$

Posterior plane

$$x = -\frac{r}{2}\left(-\frac{r}{2} \leq x, z \leq \frac{r}{2}\right). \tag{10}$$

Upper plane

$$z = \frac{r}{2}\left(-\frac{r}{2} \leq x, y \leq \frac{r}{2}\right). \tag{11}$$

Lower plane

$$z = -\frac{r}{2}\left(-\frac{r}{2} \leq x, y \leq \frac{r}{2}\right). \tag{12}$$

In this way, the collected real-time scene information of the independent sports dance competition venue is transformed into a panoramic image.

4. Scene Real-Time Detection

On the above basis, the scene detection is realized by using the inter frame image difference algorithm of the collected panoramic image. The core of the algorithm is to gray the video image of the current frame, then gray the image of the previous frame or the next frame, and then make the difference

TABLE 1: Individual test results.

Detection result	Major					Amateur				
	1	2	3	4	5	1	2	3	4	5
Correct quantity	36	38	49	43	39	38	43	58	76	59
Number of errors	12	12	5	6	12	13	25	17	17	25
Accuracy/%	75.0	76.0	90.7	87.8	76.5	74.5	64.2	77.3	81.7	70.2

between the current frame image and the previous frame or the next frame, so as to extract the moving object. Assuming that at time point t, the current frame image is I_t and the previous frame image is I_{t-1}, the moving target can realize scene change detection by comparing the image differences of three adjacent frames. This method has strong adaptability to the dynamic environment, good robustness, small amount of computation, convenient implementation, and can quickly and effectively detect the moving target from the background. The specific algorithm flow is shown in Figure 1.

According to the method shown in Figure 1, the moving object of the difference between frame images is calculated by using the three-frame difference. At the same time, the problem of image blur and edge information loss caused by mean filtering is reduced under the action of noise suppression function, so as to achieve the purpose of accurately detecting the contour information of moving objects in the scene.

5. Scene Detection and Analysis

5.1. Experimental Data. In order to test the effect of the design detection method, the experimental data are collected by Dahua DH-IPC-HF8431E camera. The experiments were carried out on the MATLAB platform. The main control chip is the gt6 stm32f407v chip made by ST company. The processor is arm series. This paper mainly analyzes the key frame data of sports dancers when walking, that is, the movement of sports dancers when stepping, so as to realize the professional evaluation of sports dancers' walking movement. This paper takes video data of 24 professional sports dancers and 13 and nonprofessional sports dancers as experimental samples and selects effective experimental data, including 2004 key frame pictures of professional sports dancers and 958 key frame pictures of nonprofessional sports dancers. 1000 key frame data of professional sports dancers and 700 data of nonprofessional sports dancers are used as the training data set, and the other samples are used as the test data set.

5.2. Data Processing. The human joint point data are fitted with different interpolation coefficients according to the method in this paper, and the results are shown in Figure 2.

The shape of the knee is relatively simple; (0, 0) represents the coordinate origin of the knee joint. (x, y) represents the coordinate interpolation coefficient of the knee joint, and the interpolation coefficient of 1.0 can better fit, while the shape of the ankle and hip is relatively complex. For these two parts, 20 groups of data are randomly selected to calculate the average curve fitting degree of polynomial fitting, and the final interpolation coefficients are 0.6223 and 0.8592, respectively.

5.3. Experimental Results. In this paper, the 10-fold cross-validation method is used to test all samples, and the average accuracy is 71.9%. In order to compare the action recognition results of different sports dancers, this paper will test each sports dancer one by one. In the test set, all key frame data of 5 professional sports dancers and 5 nonprofessional sports dancers are randomly selected for test, and the results are shown in Table 1.

It can be seen from Table 1 that among the results of 10 sports dancers tested separately, the second professional sports dancer has the lowest recognition rate, 75.0% and 76.0%, respectively. The final grade of the professional sports dancer is also the worst among the five professional sports dancers; the second amateur sports dancer has the lowest recognition rate of 64.2%. Conversely, among the five amateur sports dancers, the sports dancer has the best performance and is closest to the professional level. In fact, the final performance of the amateur sports dancer is the best among the five amateur sports dancers. The evaluation of the walking posture of sports dancers is not determined by a certain moment, but by a series of moments. Therefore, the performance of a sports dancer should be evaluated by integrating the movement and posture of all key frames of a sports dancer's walk. In the random test of 10 sports dancers, the recognition rate of 5 professional sports dancers is more than 75%, and the recognition rate of 5 amateur sports dancers is also more than 64%. It shows that the detection method designed in this paper can effectively detect the scene.

6. Conclusion

In this paper, the academic algorithm is applied to practice, aiming at the detection of small targets in the target detection algorithm in deep learning. High detection accuracy is achieved on the experimental data set, and the feasibility of the deep learning algorithm in the field of target detection is also verified. However, in order to be applied to the actual scene, the method still needs to be further improved, especially in the optimization of data and structure; there are some problems and areas that need to be improved:

(1) The success of the deep learning algorithm is attributed to the use of large-scale well-labeled data sets. Although the algorithm of this subject has achieved good detection on the experimental data set, the amount of data in this data set is not very large, and the generalization ability of the training model on this data set remains to be investigated. The focus of this algorithm is on panoramic image

reconstruction, which makes less contribution to the face detection algorithm. In the future, we can refer to multiscale fusion to improve the face detection network, so as to improve the effect of the face detection algorithm itself. At the same time, the image reconstruction strategy of the face detection algorithm proposed in this paper needs two feature extraction and one image reconstruction, which cannot achieve real-time effect in the detection speed. In the follow-up, we need to do research on improving the detection speed

(2) Aiming at the difficulty of small target detection in complex scenes from the perspective of media tool collection and the high missed detection rate of existing algorithms, this paper proposes a detection algorithm for scene repair based on deep learning. On the basis of the data, the annotation of various target perspectives under the panoramic perspective is added. However, due to the insufficient amount of data, the effect still needs to be improved, and the expanded data set still needs to be supplemented. At the same time, the annotation of data also needs to be further standardized

(3) Aiming at the problems existing in small target detection in practical application scenarios, this topic puts forward the construction of panoramic image, applies the theory to practice, and realizes the engineering application of the algorithm. Although it can achieve good target detection effect, the algorithm may get stuck in practical application scenarios. Therefore, the detection speed of the algorithm still needs to be improved. At the same time, in the communication of real-time image acquisition data link, the fixation of some data lines and the security of the platform need to be improved

(4) In this paper, the weighted average method is used to grayscale the collected scene images. We employ a deep learning approach to compute optimal image interpolation. In future research, advanced techniques can be introduced to detect multiple moving objects more accurately

Data Availability

The data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. Wang, "WITHDRAWN: Multimedia Technology Embedded Processor Optimizing Physical Education Teaching Innovation Under Internet Environment," *Microprocessors and Microsystems*, vol. 13, article 104086, 2021.

[2] Y. Li and S. J. Lu, "Research on physical education system model using multimedia technology," *Multimedia Tools and Applications*, vol. 79, no. 15-16, pp. 10461–10474, 2020.

[3] Z. H. U. Jingsi, "Application of the multimedia network technology in the informatization of art education," *International English education research: English version*, vol. 3, pp. 77–79, 2019.

[4] Y. Sun and J. Chen, "Human movement recognition in dance-sport video images based on chaotic system equations," *Advances in Mathematical Physics*, vol. 2021, Article ID 5636278, 12 pages, 2021.

[5] N. Keay and A. Rankin, "Infographic. Relative energy deficiency in sport: an infographic guide," *British Journal of Sports Medicine*, vol. 53, no. 20, pp. 1307–1309, 2019.

[6] A. Dc, B. Rp, and C. Rr, "A new fast and accurate heuristic for the Automatic Scene Detection Problem," *Computers & Operations Research*, vol. 136, article 105495, 2021.

[7] T. Khan, R. Sarkar, and A. F. Mollah, "Deep learning approaches to scene text detection: a comprehensive review," *Artificial Intelligence Review*, vol. 54, pp. 3239–3298, 2021.

[8] M. Mandal, V. Dhar, A. Mishra, and S. K. Vipparthi, "3DFR: a swift 3D feature reductionist framework for scene independent change detection," *IEEE Signal Processing Letters*, vol. 26, no. 12, pp. 1882–1886, 2019.

[9] S. Javed, A. Mahmood, S. Al-Maadeed, T. Bouwmans, and S. K. Jung, "Moving object detection in complex scene using spatiotemporal structured-sparse RPCA," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 1007–1022, 2019.

[10] S. Liu, Y. Wang, F. Dai, and J. Yu, "Simultaneous 3D motion detection, long-term tracking and model reconstruction for multi-objects," *International Journal of Humanoid Robotics*, vol. 16, no. 4, pp. 1950017–1952262, 2019.

[11] Z. Li, "Three-dimensional diffusion model in sports dance video human skeleton detection and extraction," *Advances in Mathematical Physics*, vol. 2021, Article ID 3772358, 11 pages, 2021.

[12] W. H. Nisbett, A. Kavuri, and M. Das, "On the correlation between second order texture features and human observer detection performance in digital images," *Scientific Reports*, vol. 10, no. 1, p. 13510, 2020.

[13] W. He, X. Y. Zhang, F. Yin, Z. Luo, J. M. Ogier, and C. L. Liu, "Realtime multi-scale scene text detection with scale-based region proposal network," *Pattern Recognition*, vol. 98, article 107026, 2020.

[14] W. Zhang, K. Wang, Y. Liu, Y. Lu, and F. Y. Wang, "A parallel vision approach to scene-specific pedestrian detection," *Neurocomputing*, vol. 394, pp. 114–126, 2020.

[15] X. Hua, X. Wang, T. Rui, H. Zhang, and D. Wang, "A fast self-attention cascaded network for object detection in large scene remote sensing images," *Applied Soft Computing*, vol. 94, article 106495, 2020.

[16] S. Pu, W. Zhao, W. Chen, S. Yang, D. Xie, and Y. Pan, "Unsupervised object detection with scene-adaptive concept learning," *Engineering*, vol. 22, no. 5, pp. 638–651, 2021.

[17] Z. L. Wang and G. H. Tian, "Integrating manifold ranking with boundary expansion and corners clustering for saliency detection of home scene," *Neurocomputing*, vol. 379, pp. 182–196, 2020.

[18] X. Wang, X. Feng, and Z. Xia, "Scene video text tracking based on hybrid deep text detection and layout constraint," *Neurocomputing*, vol. 363, pp. 223–235, 2019.

[19] Y. Jin, Y. Zhang, Y. Cen, Y. Li, V. Mladenovic, and V. Voronin, "Pedestrian detection with super-resolution reconstruction for low-quality image," *Pattern Recognition*, vol. 115, no. 3, article 107846, 2021.

[20] H. Li, Y. Dong, L. Xu, S. Zhang, and J. Wang, "Object detection method based on global feature augmentation and adaptive regression in IoT," *Neural Computing and Applications*, vol. 33, no. 9, pp. 4119–4131, 2021.

[21] S. K. Pal, D. Bhoumik, and D. B. Chakraborty, "Granulated deep learning and Z-numbers in motion detection and object recognition," *Neural Computing and Applications*, vol. 32, no. 21, pp. 16533–16548, 2020.

[22] L. Fan, T. Zhang, and W. Du, "Optical-flow-based framework to boost video object detection performance with object enhancement," *Expert Systems with Applications*, vol. 170, article 114544, 2021.

[23] H. Yang, Y. Lin, H. Zhang, Y. Zhang, and B. Xu, "Towards improving classification power for one-shot object detection," *Neurocomputing*, vol. 455, pp. 390–400, 2021.

[24] X. Zhong, C. Ding, X. Qu, and D. Tao, "Polysemy deciphering network for robust human–object interaction detection," *International Journal of Computer Vision*, vol. 129, no. 6, pp. 1910–1929, 2021.

[25] Y. Liu, Y. Gu, F. Yan, and Y. Zhuang, "Outdoor scene understanding based on multi-scale PBA image features and point cloud features," *Sensors*, vol. 19, no. 20, p. 4546, 2019.