IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 29 January 2024, accepted 11 February 2024, date of publication 22 February 2024, date of current version 29 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3368881

RESEARCH ARTICLE

Design of Reliable Mining Algorithm for Massive Moving Image Data Trajectory

YOUMEI ZENG[1,2], BO RUN[3], AND YOUCEF DJENOURI[4,5,6], (Senior Member, IEEE)
[1]College of Humanities and Tourism, Foshan 528000, China
[2]Guangdong Dance and Drama College, Foshan 528000, China
[3]Physical Education Department, Wuhan University of Business and Technology, Wuhan 430070, China
[4]Department of Microsystems, University of South-Eastern Norway, Kongsberg, 3199 Borre, Norway
[5]Norwegian Research Center, 5838 Bergen, Norway
[6]IDEAS NCBR, 00801 Warsaw, Poland

Corresponding author: Youcef Djenouri (Youcef.djenouri@usn.no)

This work was supported in part by the 2022 Philosophy and Social Sciences Project of Hubei Provincial Department of Education under Grant 22G107, in part by Wuchang Institute of Technology University-Level Research Project 2022KY35 and Project 2023KY33, and in part by Wuchang Institute of Technology Student Innovation and Entrepreneurship Training Program under Project 2022C65.

ABSTRACT To acquire precise, dependable, and credible trajectory information from extensive motion image datasets, this study introduces a robust mining algorithm grounded in trajectory extraction for voluminous motion data. The algorithm leverages an enhanced single-stage object detection model (TFF-SSD) and employs a 3D sparse convolutional neural network for extracting point cloud features from the extensive motion data. Simultaneously, spatial semantic features are derived by integrating spatial and semantic characteristics of the vast motion data. Subsequently, a mutual attention fusion method is applied to amalgamate point cloud features and spatial semantic features into the Faster R-CNN model. This facilitates the swift identification of moving target regions, leading to the extraction of classification and coordinate information from the abundant data. Employing reference points from diverse types of extensive motion data, the algorithm selects associated feature regions for the motion data target, thereby obtaining reference point information indicative of continuous motion. Subsequent fitting of this reference point information enables the reliable mining of trajectories within the extensive motion data. Experimental outcomes demonstrate the algorithm's ability to accurately detect all crowd movements in a running race. The extracted motion data features are diverse, encompassing reference points that facilitate trajectory mining with elevated quality, reliability, and genuine data content.

INDEX TERMS Trajectory extraction, massive motion data, mutual attention fusion, target detection, faster R-CNN model, reliable mining.

I. INTRODUCTION

As the development of mobile devices and sensor technology makes it easier to collect and record human life trajectories, spatiotemporal databases have become an important storage method. Spatiotemporal databases have powerful spatiotemporal characteristics and scalability, making them ideal places for processing motion data. Motion data usually refers to the sequence of spatial positions that a person or object passes through from the starting point to the end point over a period of time [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojun Steven Li.

Compared with other data types, motion data has the characteristics of spatiotemporal characteristics, large amount of data and low quality of data. With the gradual realization of social informatization, the amount of data is also increasing exponentially. How to effectively use a large amount of useful information contained in these massive data has become an urgent demand of the whole society. In this context, data mining technology came into being, and motion data mining algorithm is the core method to extract useful information from spatio-temporal data [2], [3]. Its research involves data structure, calculation design, mathematical modeling and other fields. However, with the increasing of trajectory data, motion data mining algorithms are faced with challenges and opportunities. The main challenges include low data quality and high computational complexity.

Therefore, it is necessary to use data mining algorithms to design more efficient, accurate and practical motion data analysis models to cope with the ever-changing needs. It can find potential correlation and useful information from massive fuzzy and incomplete data sets, and can realize prediction based on the correlation between the data.

Some relevant experts and scholars have conducted research on massive data mining algorithms. For example, ColomboA et al proposed a frequent pattern data mining algorithm based on compressed bitmap [4], which realized data mining by adopting compressed bitmap structure and using less memory to store original data sets and calculate support degree. However, the algorithm adopts compressed bitmap structure, which makes the display information incomplete and can not grasp the geometric information of motion data in time, resulting in poor effect of motion data mining. ParkS et al. proposed a motion data mining algorithm based on mobile big data [5], which analyzed a large-scale motion data set and applied trajectory data mining method to understand the spatial structure of different travelers' activities. Through spatial cluster analysis and sequence pattern mining, the travel areas of travelers and the correlation between these local Spaces are revealed. Finally, the spatial model is combined with destination planning to complete data mining. However, the spatial cluster analysis and sequential pattern mining adopted by the algorithm do not consider the quality of moving data in practical application, and the detection performance of moving data is weak, resulting in unsatisfactory mining effect. In reference [6], JuL et al. proposed a data mining model based on the nearest neighbor ID3 optimization algorithm. The deficiency of traditional ID3 algorithm in sports event action data mining is analyzed. Then, aiming at the problem of missing attributes, an ID3 optimization algorithm based on the nearest neighbor is proposed to realize the improvement of the algorithm, and the missing values are filled by selecting values similar to the nearest neighbors to realize the comprehensive motion data mining. However, if the training data set of the improved ID3 algorithm is relatively small in practical application, it is easy to overfit, resulting in poor quality of data mining results. Chen Y L et al. proposed a deep learning architecture based on Graph Convolutional Network (GCN) [7], which utilizes effective information collected by inertial sensors to predict the position sequence of human lower limb joints in three-dimensional space throughout the entire motion process, thereby completing motion trajectory mining. However, in practical applications, the generalization ability of the GCN architecture is limited by the diversity of postures and motion types in the training data. When dealing with new and unseen postures and movements, the GCN architecture may not be able to effectively predict the position sequence of human lower limb joints, resulting in poor accuracy of the obtained motion trajectory mining results. Zhang J W et al. proposed a sports training trajectory data capture method based on mean shift algorithm [8]. Regard the human body model as a skeleton model with 51 degrees of freedom and 16 joints to digitize the training trajectory, and perform dimensionality reduction on the trajectory data to reduce computational complexity. To reduce the dependence of the mean shift algorithm on environmental parameters, the probability density function in the gradient iterative estimation algorithm is selected, and the color information of the target is used as a feature to capture trajectory data. However, when using this method in trajectory mining of massive motion image data, the mean shift algorithm may result in incomplete mined trajectories, affecting the final results.

Based on the shortcomings of the above algorithms, this paper combines the mutual attention fusion algorithm and the Faster R-CNN model to propose a reliable mining algorithm for massive motion data based on trajectory extraction. Point cloud and spatial data can provide different information of 3D scene, and there is a certain complementary effect between them. In the feature space, the combination of mutual attention fusion algorithm is beneficial to more accurate target detection, thus improving the authenticity of motion data mining content [9]. Faster R-CNN can quickly determine the moving target region [10], [11], which can reduce the time of target region determination and feature extraction, reduce the amount of target region data that need to be associated, greatly improve the accuracy of target region determination of complex background and texture, and make the trajectory extraction rate of massive moving data higher. Thus improve the reliability and quality of the value information mining of motion data.

II. MANUSCRIPT PREPARATION RELIABLE MINING FOR MASSIVE MOTION DATA

A. OVERVIEW

This paper proposes a reliable mining algorithm for massive motion data based on trajectory extraction, which mainly consists of two parts, namely target detection of massive motion data and mining of massive motion data. The specific process is shown in Figure 1.

The specific operation steps are as follows:

(1) Establish an improved single-stage object detection model (TFAF-SSD) to obtain point cloud features of motion data in multiple frames of images through a three-dimensional sparse convolutional neural network [12], [13]. Meanwhile, spatial features and semantic features of massive motion data were fused to establish the dependency between the two features to achieve adaptive feature fusion and obtain spatial semantic features of massive motion data. By introducing mutual attention module to adaptively align spatial semantic features and point cloud features, the enhanced point cloud features based on spatial semantic features are obtained to better capture the location, shape and other information of the target, so as to identify and classify the target more accurately, which is conducive to improving

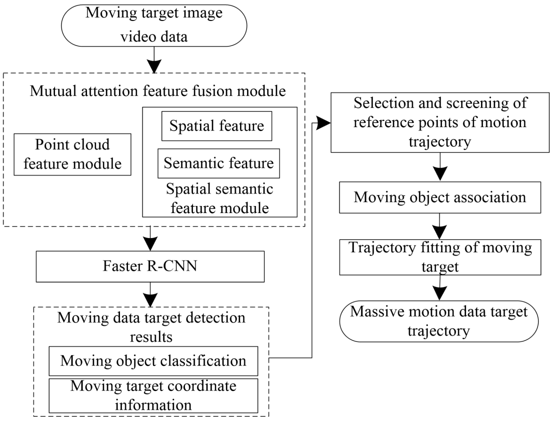


FIGURE 1. Reliable trajectory mining algorithm flow for massive motion data.

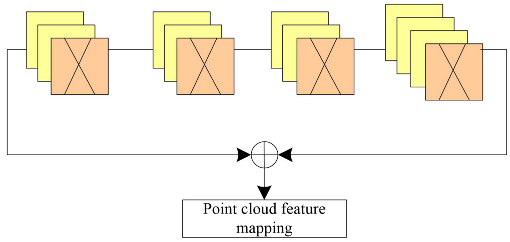


FIGURE 2. Point cloud feature extraction module.

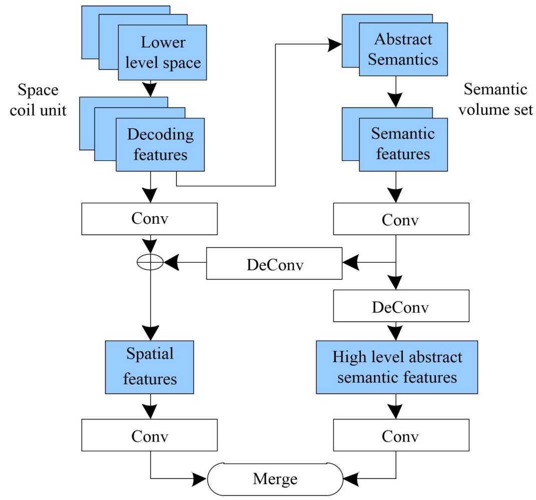


FIGURE 3. Spatial semantic feature extraction module.

the accuracy and reliability of motion data mining. Finally, on the basis of this feature, the Faster R-CNN network model is used as input to obtain the moving target classification information and coordinate information of massive data, and realize the three-dimensional target detection of massive moving data.

(2) According to the detected mass motion data feature information, the reference points of mass motion data are determined and the feature regions of mass motion data target to be associated are screened. The reference points of continuous movement of mass motion data target are obtained by associating the extracted mass data moving target regions. The extracted reference point information is fitted to realize reliable track mining of massive motion data.

B. MASSIVE TARGET DETECTION BASED ON TFF-SSD

1) POINT CLOUD DATA AND SPATIAL SEMANTIC FEATURE EXTRACTION MODULE

The feature extraction backbone network uses sparse convolutional layer to extract three-dimensional features of massive moving point cloud data. The sparse convolutional layer mainly consists of four parts, each of which includes several sub-manifold sparse convolutional layers and one sparse convolutional layer [14], [15].

Here, 3 parts are composed by 2-layer sub-manifold sparse convolutional layers and one part has 3-layer convolutional layer. At the end of each sub-manifold layer, a sparse convolutional layer is attached, which is used to perform 2x downsampling on the 3D feature graphs of massive motion data. Finally, sparse voxel features are transformed into dense feature maps, and features in z space are connected to generate bird 's-eye view feature maps as the input of the massive motion data target detection module. The point cloud feature extraction module is shown in Figure 2.

In order to detect the target in the massive moving data, it is necessary to accurately return to the target location of the moving data. In the process of determining the target location of the moving data, it is necessary to consider the low-level spatial features and high-level abstract semantic features. When extracting low-level spatial features and high-level abstract semantics from feature mapping, we can realize spatial semantic feature extraction through the convolution layer adopted by the aerial view feature extraction module [16]. The spatial semantic feature extraction module is shown in Figure 3.

2) MUTUAL ATTENTION FUSION MODULE

In order to overcome the disadvantage of not being able to extract two high-quality features at the same time, the multi-scale feature fusion module of attention mechanism is used to fuse spatial features and semantic features [17], [18], [19], and they are divided into two aerial view attention graphs. In the attention graphs, the multi-scale fusion module establishes the dependency relationship between the two features, thus achieving adaptive feature fusion. Inspired by the concept of cross-correlation in signal processing and the attention mechanism in the field of machine learning, based on the point cloud features extracted in previous section, a mutual attention module is designed for the fusion of point cloud features and spatial semantic features of massive motion data to enhance the point cloud features and lay a foundation for subsequent target detection to better capture the location, shape and other information of the target. Therefore, the target can be identified and classified more accurately, which is helpful to improve the accuracy and reliability of motion data mining.

The point cloud features and spatial semantic features in massive motion data are represented as $\mathbf{G}_{\mathrm{p}}$ and $\mathbf{G}_{\mathrm{i}}$ respectively. The steps of fusion of the two features are as follows:

(1) Calculate the correlation between$\mathbf{G}_{\mathrm{p}}$ and $\mathbf{G}_{\mathrm{i}}$

$$\mathbf{R}_{i} = \tanh(\mathbf{G}_{\mathrm{p}} \cdot (\mathbf{Q}_{\mathrm{i}}\mathbf{G}_{\mathrm{i}} + \mathbf{b}_{\mathrm{i}})) \tag{1}$$

$$\mathbf{R}_{\mathrm{p}} = \tanh(\mathbf{G}_{\mathrm{i}} \cdot (Q_{\mathrm{p}}\mathbf{G}_{\mathrm{p}} + \mathbf{b}_{\mathrm{p}})) \tag{2}$$

Among them, in order to realize the alignment of point cloud features and spatial semantic features in massive motion data at the feature space level, $\mathbf{G}_{\mathrm{p}}$ and $\mathbf{G}_{\mathrm{i}}$ are transformed by the multiplication of matrix $\mathbf{Q}_{\mathrm{p}}$ and $\mathbf{Q}_{\mathrm{i}}$ respectively, and the offset vectors bp and bi are added to realize the feature space translation. In this way, point cloud features are aligned with spatial semantic features by learning to optimize the values of $\mathbf{Q}_{\mathrm{p}}$, $\mathbf{Q}_{\mathrm{i}}$, $\mathbf{b}_{\mathrm{p}}$ and $\mathbf{b}_{\mathrm{i}}$. Then the cross-correlation function values $\mathbf{R}_{\mathrm{i}}$ and $\mathbf{R}_{\mathrm{p}}$ between $\mathbf{G}_{\mathrm{p}}$ and $\mathbf{G}_{\mathrm{i}}$ are obtained by using tanh function to exchange the two correlation values respectively.

The obtained correlation function values are used to calculate the attention scores of each component in the point cloud feature and spatial semantic feature in the massive motion data:

$$A_{\mathrm{i}} = soft\max(\mathbf{R}_{\mathrm{i}}) \tag{3}$$

$$A_{\mathrm{p}} = soft\max(\mathbf{R}_{\mathrm{p}}) \tag{4}$$

Among them, A_{i} and A_{p} are the mutual attention scores between point cloud features and spatial semantic features in massive motion data, softmax is used to convert the value of correlation function. On the one hand, the original correlation score can be normalized and converted into a probability distribution with the sum of all the weights of all elements being 1. On the other hand, softmax's internal function conversion mechanism can highlight important relevant attention scores and amplify significance feature weights, which will be more conducive to the extraction of deep semantic features from massive motion data.

(2) Obtain the spatial semantic matrix

To multiply the attention weight and eigenvector matrix, the spatial semantic correction matrix is calculated after the attention mechanism update.

$$\mathbf{C}_{\mathrm{i}} = \mathbf{G}_{\mathrm{p}} \cdot A_{\mathrm{i}} \tag{5}$$

where "•" represents the dot product of the matrix.

The spatial semantic features modified by point cloud features are as follows.

$$\mathbf{G}_{\mathrm{i}}' = \mathbf{G}_{\mathrm{i}} \cdot \mathbf{C}_{\mathrm{i}} \tag{6}$$

(3) Calculate the point clod feature

The modified matrix of point cloud features is obtained by using the modified spatial semantic feature $\mathbf{G}'_{\mathrm{i}}$ and the attention matrix:

$$\mathbf{C}_{\mathrm{p}} = \mathbf{G}_{\mathrm{i}}' \cdot A_{\mathrm{p}} \tag{7}$$

The point cloud features modified by the spatial semantic features based on the attention mechanism are as follows:

$$\mathbf{G}_{\mathrm{p}}' = \mathbf{G}_{\mathrm{p}} + \mathbf{C}_{\mathrm{p}} \tag{8}$$

(4) Feature Fusion

Based on the above, after obtaining the corrected point cloud feature $\mathbf{G'}_{\mathrm{p}}$ and the corrected spatial semantic feature $\mathbf{G'}_{\mathrm{i}}$, the final fusion of the two features is completed using the following formula:

$$\mathbf{G'}_{\mathrm{ip}} = \alpha \times \mathbf{G'}_{\mathrm{p}} + \beta \times \mathbf{G'}_{\mathrm{i}} \tag{9}$$

In the formula, α and β represent feature weights, and $\mathbf{G'}_{\mathrm{ip}}$ represents the fusion feature of point cloud features and spatial semantic features, providing a feature basis for the subsequent three-dimensional target detection of massive motion data. The optimization of parameters $\mathbf{Q}_{\mathrm{p}}$, $\mathbf{Q}_{\mathrm{i}}$, $\mathbf{b}_{\mathrm{p}}$ and $\mathbf{b}_{\mathrm{i}}$ in the above feature fusion process can be achieved through the loss function in the subsequent detection network of massive moving data targets, so as to improve the effect of feature fusion and provide support for subsequent detection.

3) TARGET DETECTION BASED ON MASSIVE MOTION OF FASTER R-CNN

To enable reliable trajectory mining of massive motion data, subsequently, utilizing the aforementioned fused features, the Faster R-CNN model is employed for target detection, thereby determining the reference points of the massive motion data.. In general, the target detection of mass motion data can be divided into two stages, that is, the generation of the candidate region frame used to determine the target location of mass motion data, and the classification of the motion data in the region. The candidate area box refers to the given input image, the image is screened to find out all the candidate areas where there may be moving data targets, and draw the area box. The fused features are input into the trained Faster R-CNN model for recognition, and the output result is the location information of the moving data target on a single image. The output results include: moving data target frame coordinates, moving data target classification information, moving data target classification probability. The structure of the Faster R-CNN model is shown in Figure 4.

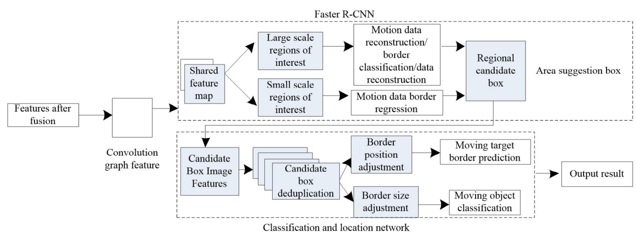


FIGURE 4. **Faster R-CNN model structure.**

The Faster R-CNN model is mainly composed of the following two parts:

(1) Regional suggestion Network (RPN)

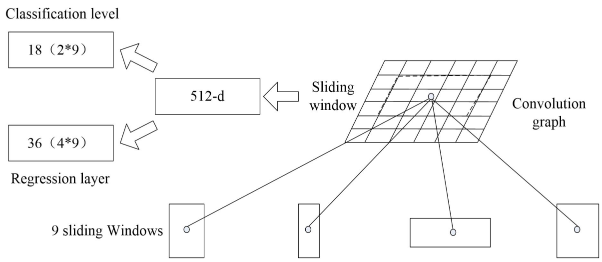


FIGURE 5. RPN model structure diagram.

Usually a full convolutional network uses anchor points to efficiently generate regions of interest with multiple scales and aspect ratios on the shared feature map, and transfers these regions to the full connection layer to achieve RPN foreground/ background classification and border regression.

RPN is a typical algorithm to generate a region candidate box, which takes an image of any size as input, and then outputs a set of suggestion boxes for all moving data targets on the image. Construct an RPN architecture for the detection of 3D objects with massive motion data [20], [21]. The RPN is built from a full convolutional network, as shown in Figure 5. The fusion features of massive motion data obtained in previous section are input into the network, and three different types of sliding Windows are used for traversal and convolution of the fused features, that is, three aspect ratios (1:1, 1:2, 2:1) sliding window to adapt to different aspect ratios of motion data targets, and each window contains 3 scales (128 × 128, 256 × 256, 512 × 512) to adapt to different sizes of motion data targets, a total of 9 different sliding Windows can be generated. The resulting Windows can correspond to moving objects with different aspect ratios and different sizes. Each area detection box is assigned a binary label, representing the two classes of the detection box, foreground and background, used to determine whether the detection box is a target or a non-target. In addition, RPN uses a four-dimensional vector (x, y, w, h) to represent the position of the border inspection detection box (center point coordinates and width and height).

In order to better generate effective motion data trajectory area detection boxes, the extracted features are input into two branches: the classification layer and the regression layer. In order to achieve effective classification, the classification layer should first consider intelligent feature selection, find the best balance between two conflicting objective functions, and minimize the number of features [22]. Therefore, the features of the classification layer are mapped to the feature vector of 18 (2 × 9) dimension, where 9 is the type of sliding window, and 2 represents the foreground or background. The features of the regression layer are mapped to 36 (4 × 9) dimensional feature vectors, corresponding to the four-dimensional vector coordinates of the detection boxes in the 9 motion data areas respectively. The purpose of the regression layer is to bring the position of the RPN through the generated detection box closer to the true value. The output of each stage of this part is upsampled to a feature map with the same size. Next, the feature maps of these three stages are stitched together into one feature map. Finally, three convolution layers of 1 × 1 convolution kernels are used to predict the class, offset and orientation of massive motion data. In the last layer of the network, a non-maximum suppression layer is added to generate 3D background classification and border regression of the trajectory target with the final detected mass motion data.

(2) Target classification and border positioning

In the classification and positioning network, the region of interest pooling technology [23], [24] is used to extract the image features of the candidate boxes output by the RPN network and deliver them to the subsequent full-connection layer to achieve the target classification of moving data and border positioning.

4) LOSS FUNCTION

In order to effectively complete the task of 3D target classification and positioning of the trajectory of massive motion data, regularization technology [25] was introduced to train the loss function of Faster R-CNN to help Faster R-CNN avoid overfitting problems. The loss function is defined as the sum of classification and regression loss functions, that is, the multi-task loss function. The formula is expressed as follows:

$$L_{\mathrm{total}} = L_{\mathrm{cls}} + L_{\mathrm{reg}} \tag{10}$$

In the formula, L_{cls} and L_{reg} represent classification loss function and regression loss function respectively, and L_{cls} is represented by FocalLoss function [26], [27] shown in Equation 11.

$$L_{\mathrm{cls}} = -\lambda_2 \alpha_{\mathrm{t}} (1 - p_{\mathrm{t}})^{\gamma} \log(p_{\mathrm{t}}) \tag{11}$$

where, λ_2 represents the L2 regularization coefficient; p_{t} represents the class probability estimated by the model; α_{t} and γ indicate the parameters for FocalLoss.

For the 3D target detection of massive motion data, it is composed of 7 parameters. The real massive motion 3D data is expressed as $(x_{\mathrm{c}}^{\mathrm{g}}, y_{\mathrm{c}}^{\mathrm{g}}, z_{\mathrm{c}}^{\mathrm{g}}, l_{\mathrm{c}}^{\mathrm{g}}, w^{\mathrm{g}}, h^{\mathrm{g}}, \theta^{\mathrm{g}})$, where, $(x_{\mathrm{c}}^{\mathrm{g}}, y_{\mathrm{c}}^{\mathrm{g}}, z_{\mathrm{c}}^{\mathrm{g}})$ represents the central point of massive motion 3D data. $(l_{\mathrm{c}}^{\mathrm{g}}, w^{\mathrm{g}}, h^{\mathrm{g}})$ represents the length, width and height of massive motion 3D data, and θ^{g} represents the 3D orientation Angle of motion data. For the motion data target box, parameterize it to $(x_{\mathrm{c}}^{\mathrm{a}}, y_{\mathrm{c}}^{\mathrm{a}}, z_{\mathrm{c}}^{\mathrm{a}}, l_{\mathrm{c}}^{\mathrm{a}}, w^{\mathrm{a}}, h^{\mathrm{a}}, \theta^{\mathrm{a}})$. The residual vector $\tau^* \in (\Delta x, \Delta y, \Delta z, \Delta l, \Delta w, \Delta h, \Delta\theta)$ is defined, and the vector τ^* contains 7 object parameters of motion data to be regression, expressed as position residual $(\triangle x, \triangle y, \triangle z)$, length, width and height residual $(\Delta l, \Delta w, \Delta h)$ at three latitudes, and orientation Angle residual $\Delta\theta$, these parameters are calculated by the following Equations 11-17.

$$\Delta x = \log \frac{x_{\mathrm{c}}^{\mathrm{g}} - x_{\mathrm{c}}^{\mathrm{a}}}{d^{\mathrm{a}}} \tag{12}$$

$$\Delta y = \log \frac{y_{\mathrm{c}}^{\mathrm{g}} - y_{\mathrm{c}}^{\mathrm{a}}}{d^{\mathrm{a}}} \tag{13}$$

$$\Delta z = \log \frac{z_c^g - z_c^a}{h^a} \tag{14}$$
$$\Delta l = \log(\frac{l^g}{l^a}) \tag{15}$$
$$\Delta w = \log(\frac{w^g}{w^a}) \tag{16}$$
$$\Delta h = \log(\frac{h^g}{h^a}) \tag{17}$$
$$\Delta \theta = \theta^g - \theta^a \tag{18}$$

In which $d^a = \sqrt{(l^a)^2 + (w^a)^2}$ is the length of the diagonal line on the horizontal plane of any motion data target. In order to correctly retrieve the parameters of real 3D motion data from the motion data target to be matched, the regression loss is calculated as follows.

$$L_{reg} = \frac{1}{N_{pos}} \sum_i SmoothL1(\tau_i^*) \tag{19}$$

where τ_i^* represents the residual vector between the i th real 3D motion data and the predicted 3D motion data. The SmoothL1 function is calculated as follows.

$$SmoothL1(\tau_i^*) = \begin{cases} 0.5(\tau_i^*)^2, & if \left|\tau_i^*\right| < 1 \\ \tau_i^* - 0.5, & otherwise \end{cases} \tag{20}$$

Among them, classification loss is mainly used to solve the problem of target classification, so that the model can classify the target accurately. Regression loss is mainly used to solve the problem of target location, so that the model can accurately determine the location of the target. Therefore, the loss function of Faster R-CNN is trained based on the above content to optimize the performance of classification and positioning at the same time, so as to better complete the 3D target classification and positioning task of the trajectory of massive motion data.

C. TRAJECTORY MINING OF MASSIVE MOTION DATA

1) REFERENCE POINT SELECTION AND SCREENING

In order to realize the mining of massive moving data, the feature regions of massive moving data targets are screened according to the detection results of moving data targets obtained in section II.A (i.e. the classification results of moving data targets and border positioning results), and the target frames whose classification probability of moving data targets is less than 90% are eliminated due to insufficient features on the image.

Considering the general video surveillance image shooting Angle, as well as motion data trajectory target shape and other factors. In principle, the reference point is selected as the mapping point of the moving data target centroid on the image where the moving data target frame is located, and the relevant length-width ratio is converted to the target frame. Generally, for most camera shooting angles and positions, it is reasonable to select the center point of the target frame of the motion data trajectory. Therefore, the center position of the moving data target frame is taken as the track reference point of the moving data target [28], [29], and the coordinate information of the moving data target frame in the recognition output is divided into two parts: the reference point position information of the moving data target frame and the shape information of the moving data target frame.

2) ASSOCIATION ALGORITHM OF TARGET REGION OF MASSIVE MOTION DATA

The key to extracting trajectories of massive moving data targets is the correlation of the same moving data targets among images of different frames. Therefore, in order to realize reliable trajectory mining of massive motion data, according to the characteristics of the training results of Faster R-CNN in the improved single-stage object detection model (TFF-SSD), a method based on minimum reference point mapping distance is proposed, and the coordinate and shape information of the moving data provided by Faster R-CNN is utilized.

The selected mass data moving target area is associated to realize the association between the same moving data target in different images. In order to realize effective association of target regions with massive motion data and extract accurate trajectory of massive motion data, the threshold value $\triangle_1$ is set during association to prevent errors during association of moving data targets, but it cannot ensure the regression accuracy of target frame of moving data. For this reason, the threshold $\triangle_2$ is set to prevent the poor accuracy of the regression of the target frame of the motion data due to the overlap, occlusion and camera Angle of the motion data in the video. To sum up, the specific operation steps for the association of the target area of massive motion data are as follows:

(1) The target box of the motion data target of the trajectory to be extracted is denoted B. The target frame B contains the reference point coordinate information of the target frame and the shape information of the target frame, namely, the central position coordinate of the target frame (X, Y) and the width and height of the target frame (W, H). The target frame B is mapped to the extracted previous frame image, denoted as B', and the mapping of its reference point coordinates and shape information is denoted as X', Y', W', H' respectively. After mapping, calculate the distance between the reference point of the target frame of all motion data on the previous frame and the reference point of the mapped target frame B'. The formula is expressed as follows:

$$S_i = \sqrt{(X_i - X')^2 + (Y_i - Y')^2} \tag{21}$$

Here, S_i indicates the distance between the motion data target mapping reference point and all motion data target reference points in the frame. According to the extracted interval, adjust the setting threshold $\triangle_1$. Select the motion data target frame whose distance S_i small and less than $\triangle_1$ from n calculated distances, and record the motion data target frame M. If object frame M of motion data cannot be determined, that is, the object frame of motion data with the smallest distance to S_i is larger than $\triangle_1$, then the object frame of motion data is mapped forward. If object frame M of motion data cannot be determined for several consecutive times, extraction is considered complete or failed.

(2) According to Faster R-CNN in the improved single-stage target detection model (TFAF-SSD), the shape information of the target frame in motion data, namely height H and width $\triangle W$, calculate the deformation information W_{M} and H_{M} of the mapping target box B' with respect to the target box M.

$$\Delta \boldsymbol{W}_{\mathrm{M}} = \boldsymbol{W}' - \boldsymbol{W} \tag{22}$$

$$\Delta \boldsymbol{H}_{\mathrm{M}} = \boldsymbol{H}' - \boldsymbol{H} \tag{23}$$

The threshold is set again. If $\triangle W < \triangle_2 \cap \triangle W < \triangle_2$ is not true, the motion data target B is directly mapped to the previous frame of this frame with its contained coordinates X, Y, W, H, and the operation is carried out. If $\triangle W < \triangle_2 \cap$ $\triangle W < \triangle_2$ is not true for several consecutive times, it will be regarded as the end of the track of the motion data target mining or failure. If $\triangle W < \triangle_2 \cap \triangle W < \triangle_2$ is true, the motion data trajectory target is regarded as the same target as the motion data target to be mined. The motion data target frame coordinates are recorded, and the motion data target frame coordinates are mapped to the previous frame image of the frame to continue the same operation as above.

In the correlation, only the quadratic operation is carried out, there is no complex operation of the higher power, and the time complexity of the algorithm is low when calculating in the computer.

The purpose of setting the threshold $\triangle_1$ is to prevent errors in the association of moving data objects, because during the identification process of moving data objects, It may occur that the target of motion data on some frames cannot be identified or the probability is too low due to insufficient classification features and is deleted. In this way, the wrong target frame of motion data is selected as the association when the target frame of motion data is associated. Setting the threshold $\triangle_1$ can effectively solve this problem. The size of the threshold is mainly related to the pixels occupied by the target frame, the video interval extracted and the speed of the moving data target. During the experiment, it is adjusted according to the image video quality, video scene and other factors.

The significance of setting the threshold $\triangle_2$ is to prevent the particularly poor accuracy of the regression of the moving data object frame due to the overlap, occlusion and camera Angle of the moving data object in the video. By setting the threshold $\triangle_2$, part of the motion data target boxes with large deviations can be deleted, that is, the reference points deviating from the actual trajectory can be deleted, which makes the trajectory extraction of massive motion data more accurate, as shown in Figure 6.

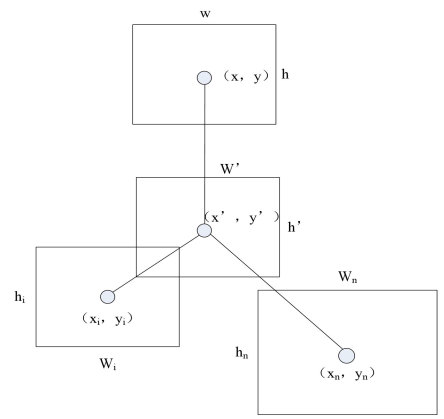


FIGURE 6. Diagram of reference points association of mass motion trajectory.

3) TRAJECTORY FITTING OF MASSIVE MOTION DATA

If the motion data target frame is close enough to the edge of the image, that is, when $X + W$ and $Y + H$ are both greater than or equal to the width and height of the image or the association of the motion data target, the end condition is reached, and the mining of massive motion data is regarded as the end. Map all previous recording points to the first frame of the computed image.

Regardless of the order in which reference points appear, the height and width of the image are regarded as a rectangular coordinate system, and the set of mapped points is fitted into a curve [30] using least squares, and the fitted curve is regarded as the motion track of massive data, as shown in Figure 7.

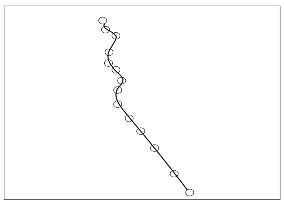


FIGURE 7. Trajectory fitting diagram of massive motion data.

III. EXPERIMENTAL ANALYSIS

A. EXPERIMENTAL DATA AND ENVIRONMENT

In order to verify the validity of the algorithm's reliable mining of massive motion data, the sports competition video of a certain road section in the city is taken as the experimental object. The video images are taken from the sports competition video taken from the height of 60 to 130 m above the ground on the main urban road of the city, and the point cloud scene is captured by a laser scanner (VelodyneHDL-64E). Based on the historical movement track data of 527 people in a road section of the city, the 35-minute images on August 12, 2015 were selected to compare the crowd movement data of the road section.

The experimental platform was a workstation with Windows10 operating system, the GPU model was NVIDIAGTXTitanX, the memory was 36 GB, and the CPU was Intel Corei7-6700-K. The VS2013+CUDA8, 0+CuDNN5, and 0+Anaconda2 environments are configured. The deep learning framework of Caffe for windows is built, and Python3.8 is used as the programming language.

B. EXPERIMENTAL INDEXES

To verify the superiority of the algorithm proposed in this article, t-tests in statistical analysis were used to complete the validation. Due to the central point of the loss function in massive motion 3D data in object detection, the length, width, and height of the massive motion 3D data, as well as the 3D orientation angle of the motion data, seven parameters play an important role in achieving reliable and accurate motion data trajectory mining. Therefore, taking these 7 parameters as an example, t-test is performed. The formula for calculating its t value is shown in equation (24):

$$t = \frac{\mu}{\sigma} \tag{24}$$

In the formula, μ represents the difference between the mean of the data processed by the proposed algorithm and the benchmark dataset, and σ represents the combination of standard deviations.

In order to verify the reliability of the algorithm in this article, based on the above experimental environment, the missed detection rate is taken as the experimental index, and TP is defined as the number of detected moving small targets, FN is the number of undetected moving small targets, and FP is the number of non-moving small targets (false targets caused by various disturbances) is detected as moving small targets. The calculation of MA is as follows.

$$\mathrm{MA} = \mathrm{FN}/(\mathrm{TP} + \mathrm{FN}) \tag{25}$$

where, MA represents the proportion of missed moving targets in all real moving targets, and the smaller the value, the better.

To verify the consistency of the trajectories mined by the algorithm proposed in this paper, smoothness is used to measure them, which generally involves calculating the angle changes between adjacent points in the trajectory. A smaller change value indicates a smoother trajectory, while the opposite indicates discontinuity in the trajectory. The calculation formula is as follows:

$$\omega_\alpha = \frac{\sum_{i=1}^{N} (|\theta_i - \theta_{i-1}|)}{N} \tag{26}$$

In the formula, N represents the number of points in the trajectory, $|\theta_i - \theta_{i-1}|$ Represents the angle value between the i-th point and the i-1st point.

C. TEST RESULT

For the t-test mentioned above, prepare two sets of data: one is the parameter data processed by the proposed algorithm, and the other is the benchmark dataset, ensuring that each set of data includes the center point coordinates, length, width, height, and orientation angle. Each group of data contains 100 samples, with a degree of freedom parameter of 99, a significance level of 0.05, and a t-critical value of 1.96. According to formula (24) above, the t-value of the parameter is calculated, and the t-test results of the proposed algorithm are shown in Table 1.

TABLE 1. T-test results.

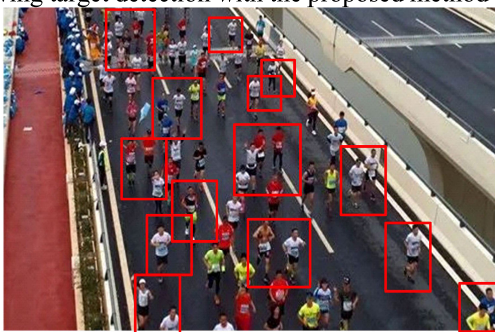
Parameter	t values
Center point coordinates	1.97
Long	1.98
Wide	1.98
High	1.98
Orientation Angle	1.97

Based on the results shown in Table 1, it can be seen that the t values of each parameter data of the algorithm in this paper are all higher than the t critical value. This indicates that the parameter data of the algorithm in this paper have significant differences from the benchmark dataset and have superiority. It can effectively complete the training of the loss function, optimize the performance of detecting target classification and localization, and better complete the 3D target classification and localization tasks of massive motion data trajectories, providing reliable support for subsequent mining of massive motion data trajectories.

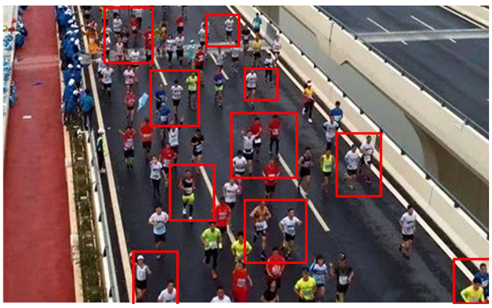
In order to demonstrate the effectiveness of track target detection, a moving video image of the city was taken as the experimental object. Considering the accuracy requirements of the method in this paper, the minimum number of rectangular frame pixels occupied by the moving data targets in this paper were all greater than 400 (20×20) pixels during track extraction, and the video resolution of the input image was no more than 680 p. With the hardware environment of section III.A, the maximum recognition speed of a single image to determine the target area is 3.5 s per image. By using the methods from Refs.4-6 to detect the movement data. The comparison results are shown in Figure 8.

According to the analysis of Figure 8, it can be clearly seen that the algorithm proposed in this paper shows high accuracy in detecting small moving targets, while compared with the other three algorithms, there are phenomena of missed recognition and recognition errors. By comparing the detection results of four algorithms, it can be concluded that the algorithm proposed in this paper has high accuracy in detecting small targets and can accurately identify moving targets. This means that it can more accurately determine the position, shape, and motion trajectory of targets, and has high efficiency. This is very valuable for many application scenarios, such as security monitoring systems, autonomous driving technology, etc., which require accurate detection and tracking of small targets. The reason why the algorithm in this article can achieve high detection accuracy is because it uses 3D sparse convolutional neural networks for point cloud feature extraction, which can more accurately describe small moving targets. It can capture the geometric information and motion details of the target, providing a more comprehensive and rich feature representation. And by fusing the spatial and

a. Moving target detection with the proposed method



b. Moving target detection results with method in Ref.4



c. Moving target detection results with method in Ref.5

d. Moving target detection results with method in Ref.6

FIGURE 8. **Comparison of the motor detection results with various methods.**

semantic features of massive motion data, the accuracy of object detection is further improved.

For small target detection results, if the moving foreground is detected at the location of the moving small target in the original image, the small moving target is considered to be detected; if multiple moving prospects are detected, the connected area of each moving foreground is calculated, and the connected area with the largest area is taken as the detection result of the small moving target, while other prospects are regarded as false targets. If the moving foreground is not detected at the location where the moving object exists in the original image, it is considered that the moving object is not detected. If the moving foreground is detected in the position where there is no moving small target in the original image, the moving foreground is considered as a false target. From this, the TP, FN and FP values of each frame are calculated, and then the TP, FN and FP values of 100 consecutive frames are counted to calculate the MA value. In addition, algorithms from Refs. 4-6 are used as the comparison algorithms of the methods in this paper, and the comparison results are shown in Table 2.

TABLE 2. **Comparison results of MA with different algorithms.**

Video	Camera1	Camera2	Camera3	Running game video
Image Size (pix)		1820×1060		320×240
Detection range (%)		0.012-0.013		0.13-0.29
Pixel number		210-2500		93-260
Algorithm in Ref.4	0.32	0.35	0.38	1
Algorithm in Ref.5	0.27	0.22	0.25	0.15
Algorithm in Ref.6	0.76	0.82	0.83	0.54
The proposed algorithm	0.13	0.11	0.12	0.06

Based on the analysis of Table 2, it can be seen that in the algorithm of Ref.4, the proportion of missing moving targets in all real moving targets in the running race video is 1, and the MA value of the algorithm of Ref.6 is also 0.54, indicating that these two algorithms cannot effectively detect moving targets under complex background. Although the MA value of the algorithm in Ref.5 is better than the above two algorithms, it is higher than that of the algorithm in this paper, and the MA value of the algorithm in this paper is the lowest in the five kinds of videos, indicating that the algorithm in this paper can effectively detect small moving targets under complex dynamic background environment and has good detection performance. This is because the algorithm in this article adopts an improved single-stage object detection model to process massive motion data, better handling complex dynamic backgrounds, and fusing the extracted features. The Faster R-CNN model is used to quickly and accurately determine the position and boundary box of the motion target, effectively avoiding missed and false detections of small moving targets in complex dynamic background environments, in order to improve the accuracy and robustness of motion target detection.

In order to further verify the target detection effect of the proposed algorithm on massive motion data, this paper applies the algorithm to KITTI test set, uses average accuracy (AP) as an evaluation index, and selects three difficulty levels, including simple, medium and difficult levels, based on different sizes, occlusion states and truncation levels of massive motion data. The frequent pattern data mining algorithm based on compressed bitmap in Ref.4, trajectory data mining algorithm based on mobile big data in Refs.5-6 are used as the comparison algorithms of the algorithms in this paper. The detection results of massive moving data of the four algorithms are compared, and the comparison results are shown in Table 3.

TABLE 3. Comparison of target detection with different algorithms.

Method	Data mode	AP/（%）		
		Easy	Intermediate	Hard
Algorithm in Ref.4	Lidar	75.24	61.27	53.62
Algorithm in Ref.5	Lidar	77.98	63.41	60.37
Algorithm in Ref.6	Lidar+Img	88.32	75.53	60.29
The proposed algorithm	Lidar+Img	88.96	76.25	69.78

From the analysis of Table 3, it can be seen that the algorithm based on literature [4] has the lowest average accuracy on the three detection difficulty levels of massive motion data, indicating that this method has the worst detection effect on moving data targets.

Although the algorithm in Ref.5 has improved the detection effect of massive moving protective objects compared with the algorithm in Ref.4, the AP value is still low. In Ref.6, when the target detection level is simple or medium, the AP value obtained by the algorithm for massive motion data detection is relatively high, but when the target detection difficulty level is difficult, the AP value is low. However, the AP values obtained by the algorithm in this paper are all the highest for the three levels of detection difficulty of simple, medium and difficult motion data, indicating that the algorithm in this paper has a stronger target detection capability of massive motion data, and can provide reliable support for massive motion data mining. This is because before conducting object detection, the algorithm in this article uses a three-dimensional sparse convolutional neural network to accurately extract the point cloud features of massive motion data, improving the expression ability of motion data, and fusing it with the spatial semantic features of massive motion data to better describe and distinguish different categories of motion targets, thereby increasing the algorithm's ability to recognize targets.

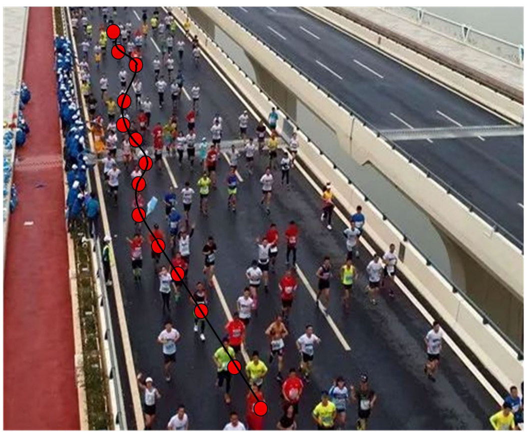


FIGURE 9. Trajectory mining of massive motion by proposed method.

In order to verify the effectiveness of the massive motion data mining algorithm in this paper, crowd trajectory data mining was carried out for the running race of this section of road, and the mining results are shown in Figure 9. Among them, the dark points are the associated reference points calculated according to the target box, and the black lines are the tracks of massive crowd movement data fitted by the algorithm in this paper.

As can be seen from the experimental results in Figure 9, through the mining of massive motion data by the algorithm in this paper, it can be clearly seen that the associated reference points are consistent with the fitted trajectory of massive crowd motion data, indicating that the algorithm proposed in this paper has rich feature extraction of crowd motion data, and can well determine reference points to accurately extract the crowd motion trajectory. Thus complete the motion data mining.

To further demonstrate the consistency of the massive crowd motion data trajectories fitted by the algorithm in this article, the results were quantitatively measured using smoothness. The algorithms in references [4], [5], and [6] were still compared with the algorithm in this article. Four samples were randomly selected from the KITTI test set for testing. The smoothness of the results obtained by each algorithm was calculated using the above formula (26), and the results are shown in Table 4.

TABLE 4. Smoothness results.

Sample number	Smoothness/°			
	The proposed algorithm	Algorithm in Ref.4	Algorithm in Ref.5	Algorithm in Ref.6
1	0.03	0.24	0.21	0.23
2	0.06	0.32	0.29	0.31
3	0.04	0.26	0.23	0.27
4	0.07	0.34	0.31	0.32

According to the results obtained in Table 4, for the four randomly selected samples, the smoothness of the trajectory of the massive crowd motion data fitted using the algorithm proposed in this paper is relatively low, with the highest being 0.07. However, the smoothness of the trajectory of the massive crowd motion data fitted using the algorithms proposed in reference [4], reference [5], and reference [6] have the lowest values of 0.24, 0.21, and 0.23, respectively. Comparing the results obtained from the four algorithms, it can be seen that the trajectory fitting results of the massive crowd motion data using the algorithm proposed in this paper have good continuity, better consistency, and can accurately extract crowd motion trajectories, resulting in better motion data mining effects. This is because the algorithm in this article integrates the spatial and semantic features of massive motion data. By establishing the correlation between space and semantics, it can better capture the consistency information of motion data, so that the fused features can provide a more comprehensive description and help accurately determine the position and movement trajectory of the target, effectively ensuring the consistency of the fitted trajectory results.

In order to further verify the effectiveness of massive motion data mining by the algorithm in this paper, the crowd movement trajectory of the road section is mined by the algorithm in this paper, and the crowd movement data results after mining are obtained, as shown in Table 5.

TABLE 5. Mparison results before and after data trajectory mining.

Massive data mining	Pre-excavation	Post-excavation	Valid data
Track mining time		35 min	
Number of people on the rail mining route		527	
Total data size	9.6 GB	13.6 MB	13.6 MB
Average single route data size	18.8 MB	26.3 KB	25.9 KB

It can be seen from the analysis of Table 5 that the movement tracks of 527 people in a certain section of the city were mined in the same period of time, and the total amount of crowd movement tracks data before mining was 9.6 GB, and the average amount of crowd data for a single route was 18.8 MB, indicating a large amount of data and low data quality. However, after the trajectory mining of crowd movement data by the algorithm in this paper, the gap between the massive crowd historical trajectory data obtained and the effective data is close to 0, indicating that the trajectory quality of the massive movement data mined by the algorithm in this paper is higher and the data content is more authentic. This is because the algorithm in this article fused the extracted features before object detection, and used the Faster R-CNN model to determine the moving target area and obtain the classification information of the target, filtering out effective moving targets, thereby ensuring high quality of the mined trajectory data.

IV. CONCLUSION

In this paper, a reliable mining algorithm for massive motion data based on trajectory extraction is proposed, and an improved single-stage object detection model (TFF-SSD) is established by combining mutual attention fusion method and Faster R-CNN model to detect massive motion data, and a method based on minimum reference point mapping distance is proposed. Combined with the coordinate and shape information of the motion data, the trajectory extraction of the motion data is completed, so as to realize the reliable mining of the massive motion data. The target detection and data mining of movement data of 527 people in a certain road section of the city are carried out. The results show that: The proposed algorithm can accurately detect the movement data of a certain road section in the city, and in the complex dynamic background environment, the proposed algorithm has the lowest MA value compared with other algorithms, and the average accuracy of the three detection difficulty levels is the highest, which can realize the effective detection of the moving target. The algorithm in this paper can obtain the reference point of motion data, extract the trajectory of motion data, and effectively mine the crowd motion data, and the gap between the mined data and the effective data is close to 0, with high quality. This algorithm is mainly aimed at extracting trajectories from massive motion data, but its essence is an object detection method. Therefore, this algorithm can be applied to other fields related to motion data, such as autonomous driving, robot navigation, intelligent monitoring, etc. In the future, this algorithm can be used for the detection and tracking of vehicles, pedestrians, or other moving targets in these fields, further achieving tasks such as path planning and behavior analysis. On this basis, although the three-dimensional sparse convolutional neural network used in the algorithm in this article is particularly suitable for processing large-scale point cloud data, training such a network may encounter computational resource and time challenges for large-scale data [31]. Therefore, in the future, the existing network structure will be optimized to make it more suitable for large-scale data processing, reducing the complexity and computational complexity of the model, and enhancing its application performance in other fields like inteligent transportation [32], [33].

REFERENCES

[1] X. Tang, F. Li, V. Sakthivel, and S J. J. Kirubakaran, "Deep learning approach to automated data collection and processing of video surveillance in sport activity prediction," *J. Multiple-Valued Log. Soft Comput.*, vol. 36, nos. 1–3, pp. 61–82, 2021.

[2] I. Fister and I. Fister, "Computational intelligence and data mining in sports," *Appl. Sci.*, vol. 11, no. 6, pp. 2637–2641, Mar. 2021.

[3] J. Ji and X. Chen, "Data mining method of double layer network center under uplink," *Comput. Simul.*, vol. 39, no. 6, pp. 386–390, 2022.

[4] A. J. Manacero, R. S. Lobato, M. A. Cavenaghi, A. Colombo, and R. Spolon, "Apriori-roaring: Frequent pattern mining method based on compressed bitmaps," *Int. J. Bus. Intell. Data Mining*, vol. 21, no. 1, pp. 48–65, 2022.

[5] S. Park, Y. Xu, L. Jiang, Z. Chen, and S. Huang, "Spatial structures of tourism destinations: A trajectory data mining approach leveraging mobile big data," *Ann. Tourism Res.*, vol. 84, pp. 102973.1–102973.17, Sep. 2020.

[6] L. Ju, L. Huang, and S.-B. Tsai, "Online data migration model and ID3 algorithm in sports competition action data mining application," *Wireless Commun. Mobile Comput.*, vol. 2021, no. 7, pp. 1–11, Jul. 2021.

[7] Y.-L. Chen, I.-J. Yang, L.-C. Fu, J.-S. Lai, H.-W. Liang, and L. Lu, "IMU-based estimation of lower limb motion trajectory with graph convolution network," *IEEE Sensors J.*, vol. 21, no. 21, pp. 24549–24557, Nov. 2021.

[8] J. W. Zhang and Y. Q. Liu, "A sports training trajectory data capture method based on mean shift algorithm," *Autom. Technol. Appl.*, vol. 42, no. 2, pp. 51113–51149, 2023.

[9] W. C. Zhao, W. Q. Qin, and M. Li, "Multi-attention fusion and weighted class representation for few-shot classification," *High Technol. Lett.*, vol. 28, no. 3, pp. 295–306, 2022.

[10] M. A. Jialiang, C. Bin, and S. Xiaofei, "General object detection framework based on improved Faster R-CNN," *J. Comput. Appl.*, vol. 41, no. 9, pp. 2712–2719, 2021.

[11] S. Liu, S. Huang, X. Xu, J. Lloret, and K. Muhammad, "Efficient visual tracking based on fuzzy inference for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 15795–15806, 2023.

[12] S. Jung, J. Jeoung, H. Kang, and T. Hong, "3D convolutional neural network-based one-stage model for real-time action detection in video of construction equipment," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 37, no. 1, pp. 126–142, Jan. 2022.

[13] C. Wang, J. Zhang, and M. Li, "Object tracking algorithm combining convolution and transformer," *Comput. Eng.*, vol. 49, no. 4, pp. 281–288, 2023.

[14] G. Shi, Y. He, and C. Zhang, "Feature extraction and classification of cataluminescence images based on sparse coding convolutional neural networks," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 9501811.

[15] S. Liu, W. Huang, Y. Cao, D. Li, and S. Chen, "SMS-Net: Sparse multi-scale voxel feature aggregation network for LiDAR-based 3D object detection," *Neurocomputing*, vol. 501, pp. 555–565, Aug. 2022.
[16] A. Yazdinejad, H. HaddadPajouh, A. Dehghantanha, R. M. Parizi, G. Srivastava, and M.-Y. Chen, "Cryptocurrency malware hunting: A deep recurrent neural network approach," *Appl. Soft Comput.*, vol. 96, Nov. 2020, Art. no. 106630.
[17] Z. Chen, S. Tian, L. Yu, L. Zhang, and X. Zhang, "An object detection network based on YOLOv4 and improved spatial attention mechanism," *J. Intell. Fuzzy Syst.*, vol. 42, no. 3, pp. 2359–2368, Feb. 2022.
[18] L. I. Zuolong, W. Banghai, and L. U. Zeng, "Pedestrian detection method based on multi-scale feature fusion and reconstruction," *Comput. Eng. Appl.*, vol. 57, no. 4, pp. 176–182, 2021.
[19] H. Ma, D.-Y. Yin, J.-B. Liu, and R.-Z. Chen, "3D convolutional auto-encoder based multi-scale feature extraction for point cloud registration," *Opt. Laser Technol.*, vol. 149, May 2022, Art. no. 107860.
[20] Y. Wu, X. Jiang, Z. Fang, Y. Gao, and H. Fujita, "Multi-modal 3D object detection by 2D-guided precision anchor proposal and multi-layer fusion," *Appl. Soft Comput.*, vol. 108, Sep. 2021, Art. no. 107405.
[21] J. Li, R. Li, J. Li, J. Wang, Q. Wu, and X. Liu, "Dual-view 3D object recognition and detection via lidar point cloud and camera image," *Robot. Auto. Syst.*, vol. 150, Apr. 2022, Art. no. 103999.
[22] A. Yazdinejad, A. Dehghantanha, R. M. Parizi, and G. Epiphaniou, "An optimized fuzzy deep learning model for data classification based on NSGA-II," *Neurocomputing*, vol. 522, pp. 116–128, Feb. 2023, doi: 10.1016/j.neucom.2022.12.027.
[23] S. Liu, P. Chen, and Y. Zhang, "A multiscale feature pyramid SAR ship detection network with robust background interference," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 9904–9915, 2023.
[24] W. J. Liu, L. B. Dong, and H. C. Qu, "Improved R-FCN model for small-scale pedestrian detection," *Chin. J. Image Graph.*, vol. 26, no. 10, pp. 2400–2410, 2021.
[25] J. Sakhnini, H. Karimipour, A. Dehghantanha, A. Yazdinejad, T. R. Gadekallu, N. Victor, and A. Islam, "A generalizable deep neural network method for detecting attacks in industrial cyber-physical systems," *IEEE Syst. J.*, 2023, doi: 10.1109/JSYST.2023.3286375.
[26] H. Zhou, X. Xiang, X. Wang, and W. Ren, "Chained end-to-end pedestrian multi-object tracking network with multi-feature fusion," *Comput. Eng.*, vol. 48, no. 9, pp. 305–313, 2022.
[27] Z. Jing, P. Li, B. Wu, S. Yuan, and Y. Chen, "An adaptive focal loss function based on transfer learning for few-shot radar signal intra-pulse modulation classification," *Remote Sens.*, vol. 14, no. 8, pp. 45–48, Apr. 2022.
[28] S. Wang, S. Huang, S. Liu, and Y. Bi, "Not just select samples, but exploration: Genetic programming aided remote sensing target detection under deep learning," *Appl. Soft Comput.*, vol. 145, Sep. 2023, Art. no. 110570.
[29] F. You, J. Z. Liang, and S. J. Cao, "Dense pedestrian group trajectory extraction and motion semantic perception for multi-objective tracking," *Transp. Syst. Eng. Inf.*, vol. 21, no. 6, pp. 5495–5542, 2021.
[30] S. L. T. Gu, "An alpha-moving total least squares fitting method for measurement data," *Proc. Inst. Mech. Eng. B, J. Eng. Manuf.*, vol. 235, nos. 1–2, pp. 781–793, 2021.
[31] Y. Di, R. Li, H. Tian, J. Guo, B. Shi, Z. Wang, and Y. Liu, "A maneuvering target tracking based on fastIMM-extended Viterbi algorithm," *Neural Comput. Appl.*, pp. 1–10, 2023.
[32] J. Zhao, D. Song, B. Zhu, Z. Sun, J. Han, and Y. Sun, "A human-like trajectory planning method on a curve based on the driver preview mechanism," *IEEE Trans. Intell. Transp. Syst.*, 2023.
[33] H. Min, X. Lei, X. Wu, Y. Fang, S. Chen, W. Wang, and X. Zhao, "Toward interpretable anomaly detection for autonomous vehicles with denoising variational transformer," *Eng. Appl. Artif. Intell.*, 2024, Art. no. 107601.

YOUMEI ZENG was born in Hanzhong, Shaanxi, in 1982. She received the M.S. degree from Wuhan Sports Institute, in 2019, and the Ph.D. degree from the National University of Science and Technology of the Philippines, in 2023.

From 2006 to 2023, she was a Physical Education Teacher with Wuchang Institute of Technology. In October 2023, she transferred to Guangdong Dance and Drama Vocational College as a full-time Physical Education Teacher. Since 2006, she has been leading four provincial-level or above research projects, authored two works, over 20 articles, and one computer software work. Her research interests include physical education teaching and sports humanities sociology.

Dr. Zeng was awarded the title of "Excellent Teacher" by the Hubei University Sports Association, in 2020 and 2021, "Excellent Coach," in 2017 and 2019, and "Excellent Referee," in 2018 and 2022.

BO RUN was born in Xiangyang, Hubei, in 1979. He received the B.S. and M.S. degrees from Hubei University, in 2003 and 2008, respectively.

Since 2003, he has been a Physical Education Teacher with Wuhan University of Commerce. Since 2003, he has also been leading two projects at or above the school level, authored one book, and more than ten articles. His research interests include physical education teaching and training and sports psychology.

Prof. Run was awarded the title of "Excellent Coach" by the Hubei University Sports Association, in 2016 and 2018, and "Excellent Referee," in 2017 and 2019.

YOUCEF DJENOURI (Senior Member, IEEE) was born in Blida, in 1989. He received the Ph.D. degree from USTHB, in 2014.

From 2020 to 2023, he was a Research Scientist with SINTEF. Since 2023, he has been a full-time Associate Professor with the University of South-Eastern Norway. He has published more than 170 articles and have been PI/Co-PI for many Norwegian and European projects. His research interests include artificial intelligence, deep Learning, and computer vision.

Dr. Djenouri received many paper awards in prestigious conferences, including IEEE Fuzzy and PAKDD.