International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 14, Issue 6 (Nov. – Dec. 2025), PP. 220-227

Comparative Study on Video Anomaly Detection Based on Multi-scale Optical Flow Features and Classical Classifiers

Yun Xia, Fengchang Fei
** (Lecturer, Modern Economics & Management College, Jiangxi University of Finance and Economics, Nanchang, China*
*** (Associate Professor, Modern Economics & Management College, Jiangxi University of Finance and Economics, Nanchang, China)*

Abstract: *Current video surveillance systems mainly rely on manual identification of abnormal events inmonitored scenes. If intelligent detection and real-time alarm can be realized, the impact of emergency events on society will be significantly reduced. In this paper, Multi-scale Histogram of Optical Flow (MHOF) is adopted as the spatiotemporal feature representation of videos, and two classical classifiers, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN), are used for anomaly detection respectively. By performing multi-scale statistics on optical flow vectors, MHOF effectively fuses the temporal and spatial information of videos, reducing data dimensionality while suppressing noise interference. Experimental results on the public UMN dataset show that based on MHOF features, SVM achieves better and more stable detection performance in most scenarios, with the highest AUC value reaching 0.918, while the performance of KNN is greatly affected by scene characteristics. This study provides experimental basis for the adaptation of features and classifiers in video anomaly detection.*

Keywords: *Video Anomaly Detection; Multi-scale Histogram of Optical Flow (MHOF); Support Vector Machine (SVM); K-Nearest Neighbor (KNN); Spatiotemporal Features.*

Date of Submission: 15-12-2025 Date of acceptance: 31-12-2025

I. INTRODUCTION

With the in-depth advancement of China's "Digital China" and "Smart City" strategies, intelligent video surveillance systems have gradually become an important technical support in fields such as public security, traffic management, and emergency early warning. Traditional video surveillance mainly relies on manual inspection, which is not only inefficient but also prone to missing abnormal events due to visual fatigue. Therefore, researching intelligent algorithms that can automatically identify abnormal events in videos to achieve real-time detection and immediate alarm is of great practical significance for improving the response speed and reliability of security systems. Compared with traditional video surveillance systems, intelligent video surveillance systems are more suitable for smart cities. Anomaly detection in surveillance videos, as an important research branch, has been widely applied in detecting traffic violations, accidents, crimes, etc., and has become a research hotspot. Anomaly detection in surveillance videos can analyze monitored scenes through computers to detect whether special events occur. Once detected, it will automatically alarm, which helps curb the further expansion of emergencies and dangerous events, and realizes the timely discovery and handling of emergencies.

In recent years, to realize the intelligence of surveillance, researchers at home and abroad have been striving to explore how to detect abnormal events in videos more accurately and effectively. Video anomaly detection aims to automatically identify behaviors or events that significantly deviate from normal patterns in video streams, such as crowd gathering, running, fighting, traffic accidents, etc. Current mainstream research methods can be divided into two categories: one is based on traditional handcrafted features and machine learning models; the other is based on deep learning methods.

Methods based on traditional handcrafted features usually rely on feature extraction technologies such as optical flow, gradient, and texture, combined with classifiers such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest for anomaly discrimination. These methods have the advantages of small computational complexity, strong interpretability, and easy deployment, especially suitable for real-time or resource-constrained scenarios. For example, Wang et al. (2017) proposed an algorithm based on spatiotemporal motion information to detect global and local abnormal events in video streams. The algorithm uses optical flow fields encoded by covariance matrices and corresponding partial derivatives to represent

motion, combined with one-class SVM as a classifier to achieve anomaly detection . Wang et al. (2018) proposed an anomaly detection algorithm based on image descriptors and classification methods for encoding motion information . Derived from the Hidden Markov Model, this algorithm uses the histogram of optical flow directions of video frames as the feature representation of videos, and judges whether there are abnormalities in video frames by observing the similarity between video frames and normal frames. Geng et al. (2019) proposed a tourism video anomaly detection model based on salient spatiotemporal features and sparse combination . This model has good robustness and timeliness in complex motion scenes and can achieve real-time anomaly detection of videos. The model combines spatiotemporal gradients with foreground detection to extract the three-dimensional gradients of the foreground of video sequences as spatiotemporal features. This method using the foreground as video features can effectively eliminate background interference, and finally adopts sparse learning features to achieve the final anomaly detection. In addition, as a classical spatiotemporal feature representation method, Multi-scale Histogram of Optical Flow (MHOF) effectively retains the structural information of motion patterns while reducing the impact of noise by counting the distribution of optical flow directions and magnitudes (Cong et al., 2013).

In recent years, deep learning-based methods have made significant progress in anomaly detection. Especially models such as Convolutional Neural Networks (CNN), Spatiotemporal Autoencoders (ST-AE), Generative Adversarial Networks (GAN), and Vision Transformers can automatically learn high-level features from large-scale data and achieve leading performance on multiple public datasets. For example: Sun et al. (2019) proposed an end-to-end model that integrates one-class SVM into CNN. It first uses CNN to automatically extract video features, then realizes video classification through SVM . Yu et al. (2021) derived a normal event model to detect abnormal events in videos by applying an adversarial prediction method to the latent feature space jointly learned from videos and motion streams . Xia et al. (2022) proposed a multi-scale feature prediction framework for anomaly detection, which uses an autoencoder-based deep feature prediction module to capture temporal and contextual information for judging input videos . Since 2023, more studies have begun to explore multi-modal fusion, self-supervised learning, and lightweight network structures to further improve the adaptability and real-time performance of models in complex scenes (Li et al., 2023; Zhang et al., 2024).

Although deep learning methods show strong feature learning capabilities, they rely on large-scale labeled data and high-performance computing equipment, and their application in scenarios with scarce data or extremely high real-time requirements is still limited. Therefore, the combination of traditional handcrafted features and classical classifiers still has important research value and application potential in specific scenarios. At present, there is a lack of systematic comparative research on different classical classifiers under the same handcrafted feature (such as MHOF) in anomaly detection tasks, especially performance analysis combined with recent scene data and evaluation criteria.

To this end, this paper uses Multi-scale Histogram of Optical Flow as the video feature representation to systematically compare the performance differences between SVM and KNN two classical classifiers in video anomaly detection. The public UMN dataset is selected for experiments, covering various scenes such as indoor and outdoor. The detection capability and stability of classifiers are comprehensively evaluated through ROC curves and AUC values. This study aims to provide experimental basis for the selection of features and classifiers in practical systems, and provide reference for the design of lightweight anomaly detection schemes in resource-constrained scenarios.

In addition, researchers have carried out extensive research on other anomaly detections, such as crowd abnormal behavior detection , human abnormal behavior detection, traffic anomaly detection , etc. This paper mainly studies the detection of crowd group abnormal events in surveillance videos. It adopts multi-scale histogram as the video feature and uses two classical classifiers, SVM and KNN, to realize the detection of abnormal events in videos.

II. MHOF

The optical flow field reflects the motion speed of each pixel in each frame of the video. Since each frame in the video is a two-dimensional projection of a three-dimensional spatial scene, although the optical flow field is the motion speed of each pixel on the plane, it still contains the three-dimensional spatial structure information of the video scene. For each pixel in each frame of image, the corresponding optical flow vector $\left(d_{i,j}^{x}, d_{i,j}^{y}\right)$ can be obtained. However, for video processing algorithms, the data volume doubles, and noise is also likely to affect the processing results. Cong et al. (2013) proposed Multi-scale Histogram of Optical Flow (MHOF) . This model divides all optical flow vectors into 16 categories, and uses the statistical feature of optical flow vectors—histogram—as the feature of the frame. This greatly reduces the data volume of video processing and achieves the effect of suppressing noise in the optical flow field. Therefore, MHOF is a

statistical feature that can better express the scene change of the current frame and can be used to detect abnormal events in video scenes.

The framework for calculating MHOF is shown in Fig. 1. First, each frame is divided into equal blocks of size M × M, and there are N blocks in one frame. At the same time, the optical flow vector of each pixel in each block is calculated to obtain the optical flow vector matrix O(o^x, o^y) of the block. Then, the class label class (i,j) of the optical flow vector $\left(o_{i,j}^x, o_{i,j}^y\right)$ of each pixel (i,j) in the block is calibrated according to equations (1) and (2), and a total of 16 categories can be divided.

$$c_{i,j} \in \begin{cases} 0 & \left\|o_{i,j}^x, o_{i,j}^y\right\| \le tr \\ 1 & \left\|o_{i,j}^x, o_{i,j}^y\right\| > tr \end{cases} \quad (1)$$

(where tr is the division threshold)

$$class_{i,j} = round\left(\theta\left(o_{i,j}^x, o_{i,j}^y\right) \Big/ (\pi/4)\right) + 8 \times c_{i,j} \quad (2)$$

where θ is the angle between $o_{i,j}^x$ and $o_{i,j}^y$.

MHOF can well reduce the number of features per frame and reduce the impact of noise points on anomaly detection.

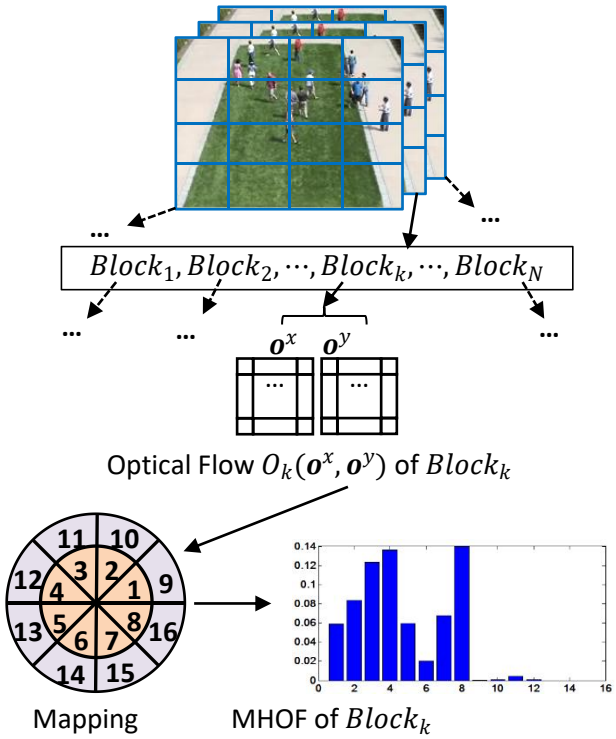


Fig. 1. Framework for Calculating MHOF

III. SVM

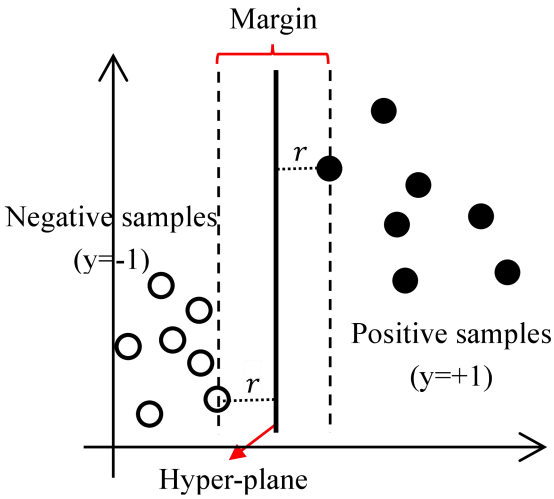
Support Vector Machines (SVM) is a supervised learning method widely used in statistical classification and regression analysis. SVM includes Support Vector Classifier and Support Vector Regressor. Due to its good accuracy and robustness, SVM is often adopted by researchers among classical machine learning algorithms. SVM was proposed by Vapnik in 1996. It is a data analysis method based on statistical learning. Since it has no high requirements on the number of samples, it not only has a good classification effect on training samples with a small number of samples but also can obtain a good classification accuracy on test samples. In addition, SVM is not very sensitive to the number of attributes of training samples, so since it was proposed, it has been frequently used in data classification and developed rapidly.

Generally speaking, samples have many attributes. If mapped to space, a sample with multiple attributes actually corresponds to a point in a multi-dimensional space. SVM performs binary classification on points in the multi-dimensional space. Since the space is multi-dimensional, SVM actually draws a multi-dimensional plane in the multi-dimensional space, which is called a hyperplane. There are many hyperplanes that can perform binary classification on points in the multi-dimensional space, and SVM finds the optimal one among them.

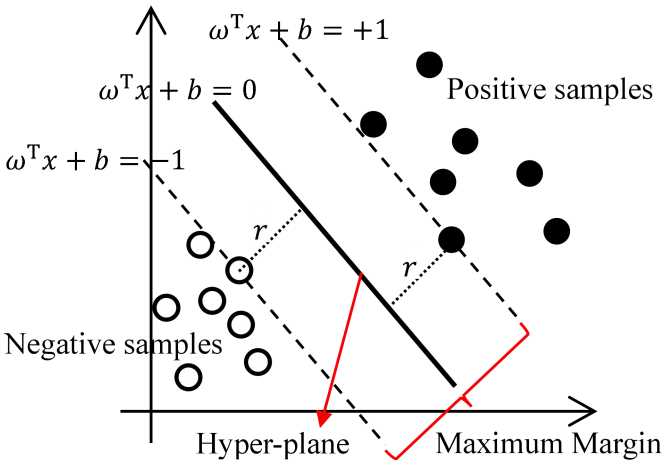
For binary classification problems, SVM maps samples to a higher-dimensional space and finds a hyperplane in this space that can perform binary classification on samples. However, there are many such planes, and how to find the optimal hyperplane is the core of the SVM algorithm. In the SVM algorithm, first, a hyperplane that can divide multi-dimensional space samples into two categories is randomly drawn, and then two sides parallel to the hyperplane are drawn on both sides of the hyperplane. Of course, these two sides are also parallel. There are many such sides, but SVM selects the side that can pass through the sample points closest to the hyperplane as the side. Therefore, in general, there are no sample points in the space enclosed by the sides on both sides of the hyperplane. In SVM, the distance from the hyperplane to one of the sides is called the margin distance. Therefore, finding the optimal hyperplane is transformed into calculating the hyperplane with the maximum margin distance, because the larger the margin distance, the smaller the total error of the classifier. This is the classification process of SVM—finding the Maximum Marginal Hyper-plane (MMH).

Since multi-dimensional space is not easy to draw, we use two-dimensional space to simply demonstrate the SVM calculation process, that is, finding the line with the maximum margin distance in two-dimensional space. As shown in Figure 2, there are many sample points in a two-dimensional coordinate system. Since it is a two-dimensional space, each sample has only two attributes, x and y. We use empty circles and solid circles to represent the two types of samples respectively. Hyperplanes (which are straight lines in two-dimensional space) are drawn in both Fig. 2(a) and (b), but we hope to find the maximum margin hyperplane, that is, the hyperplane shown in Figure 2(b). First, establish the classification function:

$$f(x)=\boldsymbol{\omega}^T\boldsymbol{x}+b \qquad (3)$$



(a)



(b)

Fig.2. Linear Separation of Samples by Hyperplaneswhere

ω is the normal vector of the hyperplane.$\{(\boldsymbol{x},\boldsymbol{y})\}$ is the sample sequence, and the hyperplane can be defined as:：

$$\omega^T x + b = 0 \quad (4)$$

The distance from a point to the plane is the length of the perpendicular line from the point to the plane, so there is a formula for calculating the length of the perpendicular line r:

$$r = \frac{\omega^T x+b}{\|\omega\|} = \frac{f(x)}{\|\omega\|} \quad (5)$$

But here r is the distance from a side to the hyperplane, and there are two sides located on both sides of the hyperplane respectively. These two sides can be expressed by the following formulas:

$$\begin{cases}\omega^T x + b = -k \\ \omega^T x + b = +k\end{cases} \quad (6)$$

Normalize k, then equation (6) is transformed into:

$$\begin{cases}\omega^T x + b = -1 \\ \omega^T x + b = +1\end{cases} \quad (7)$$

The sample point sequence $\{(\boldsymbol{x},\boldsymbol{y})\}$ should follow the following formula:

$$\begin{cases}\omega^T x_i + b \geq 1 \\ \omega^T x_i + b \leq 1\end{cases} \quad (8)$$

Where $(\boldsymbol{x}_i, y_i) \in \{(\boldsymbol{x},\boldsymbol{y})\}$， and there is a sample point $(\boldsymbol{x}^*, y^*)$ that makes equation (8) hold with equality.$(\boldsymbol{x}^*, y^*)$ is called a support vector, and equation (5) can be transformed into:

$$r^* = \frac{\omega^T x^*+b}{\|\omega\|} = \frac{f(x^*)}{\|\omega\|} = \begin{cases}\frac{1}{\|\omega\|} & if \quad y^* = +1 \\ -\frac{1}{\|\omega\|} & if \quad y^* = -1\end{cases} \quad (9)$$

Then the distance d between the two sides is:

$$d = 2r^* = \frac{2}{\|\omega\|} \quad (10)$$

When the value of d is maximum, the maximum margin hyperplane is obtained, so we only need to solve the following equation:

$$\begin{aligned} &max \frac{2}{\|\omega\|} \\ &s.t.\, y_i(\omega^T x_i + b) \geq 1, i = 1,2\ldots,n \end{aligned} \quad (11)$$

IV.KNN

The k-Nearest Neighbor (KNN) method classifies test samples by calculating which of the two existing types of samples they are closest to.

KNN is one of the simple machine learning models and a very mature algorithm model. This model has simple operations but large computational complexity. The classification steps of the algorithm for test samples are as follows: First, calculate the distance between the test sample and all already classified samples. Then sort the distance values in ascending order and find the k samples closest to the test sample (k is given in advance by the tester). Finally, the test sample is classified into the class that most of these k samples belong to. In the KNN algorithm, although the test sample is calculated with all classified samples, the final classification of the test sample is only determined by the very small number of k samples closest to it. Precisely because of this feature of KNN, in classification, if there are many cross or overlapping phenomena of attributes of different classes in the test samples, KNN is very suitable for this situation compared with other classification algorithms.

The KNN algorithm is very simple to use, but it also has the following shortcomings:

(1) Sample imbalance. Sample imbalance means that there are too many samples of one class and too few or even no samples of another class in the sample set. Since the KNN algorithm determines the classification of the test sample by k samples, if there are too many samples of one class and too few samples of another class (less than k/2) in the classified samples, or all classified samples are of one class, most of the k samples closest to the test sample will be of the class with a large number of samples, which will lead to the test sample always being classified into the class with a large number of samples, resulting in classification errors.

(2) The KNN algorithm has a large computational complexity. This is because each test sample needs to calculate the distance with all classified samples, and as there are more classified samples, the computational complexity of subsequent test samples increases. Therefore, before using KNN, it is necessary to prune the attributes of samples, retain important attributes, and prune unimportant attributes. Generally, principal component analysis can be used to achieve this pruning work.

(3) Difficulty in selecting k value. The k in the KNN algorithm is generally selected manually, often based on experience. If the k value is too large, the classification will be inaccurate; if the k value is too small, it will fall into local optimality. This makes the size of k affect the entire classification result. The selection of k is related to the characteristics of the sample data, and the k value selected for different data types is not uniform. Precisely because of the selection of k value, KNN is more suitable for datasets with a large number of samples. For datasets with a small number of samples, the size of k value is likely to affect the classification result, and even classification errors are likely to occur for particularly small datasets.

V.EXPERIMENTS

(1) Experimental Conditions and Evaluation Criteria

In this paper, UMN is selected as our experimental video. There are 3 scenes in the video, as shown in Fig. 3. The entire video has 7739 frames, among which 11 abnormal events occur. Due to the large differences in lighting, background and other conditions of each scene, the 3 scenes are analyzed separately below.

(a) (b) (c)

Fig. 3 . Three Scenes in the UMN Video Data Sequence
(a) Lawn scene (b) Indoor scene (c) Square scene

For the experimental results, this paper uses FPR-TPR curves and AUC values as evaluation criteria. The FPR-TPR curve is also known as the Receiver Operating Characteristic (ROC) curve. The ROC curve is a curve plotted with FPR as the abscissa and TPR as the ordinate. The area enclosed by the ROC curve and the two lines TPR=0 and FPR=1 is called AUC (Area Under Curve). The size of AUC determines the quality of the result. FPR and TPR represent False Positive Rate and True Positive Rate respectively, and their calculation formulas are as follows:

$$FPR = \frac{False\ positive}{False\ positive + True\ negative} \quad (12)$$
$$TPR = \frac{True\ positive}{True\ positive + False\ negative}$$

False Positive (FP) refers to the number of negative samples incorrectly classified as positive samples; True Negative (TN) refers to the number of negative samples correctly classified by the classifier; True Positive (TP) refers to the number of positive samples correctly classified by the classifier; False Negative (FN) refers to the number of positive samples incorrectly classified as negative samples by the classifier.

(2) Lawn Scene

This scene has a total of 1453 frames, with 2 abnormal events occurring. Therefore, 1333 frames are normal video clips and 120 frames are abnormal event clips. The first 400 frames of this scene are selected as training samples, and the remaining 1053 frames are used as test samples. The ROC curve is shown in Fig. 4. The k value in the KNN algorithm is 5% of the number of training samples, which is 400×0.05=8.

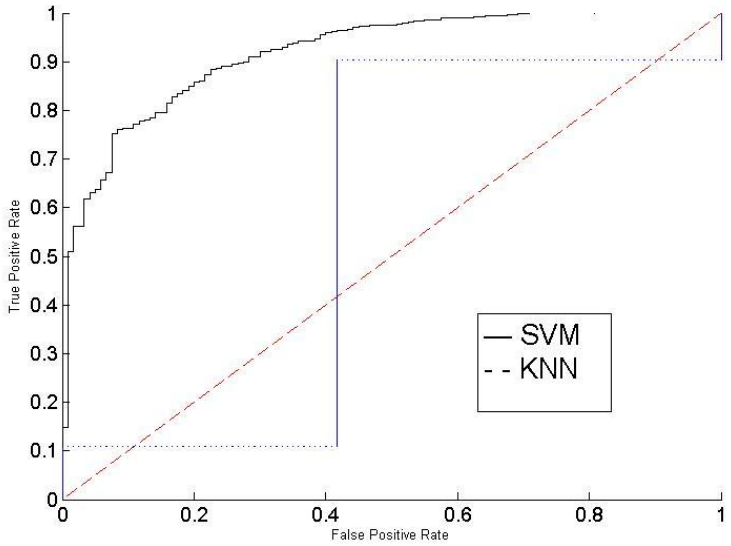


Fig. 4. ROC Curve of the Lawn Scene

As shown in Figure 4, the black solid line is the ROC curve of the SVM classifier, and the blue dashed line is the ROC curve of the KNN classifier. It can be calculated that AUC_SVM=0.917770 and AUC_KNN=0.57212. Therefore, in this scene, the anomaly detection method based on MHOF features using the SVM classifier is superior to KNN.

(3)Indoor Scene

This scene has a total of 4144 frames, with 6 abnormal events occurring. Therefore, 3715 frames are normal video clips and 429 frames are abnormal event clips. The first 400 frames of this scene are selected as training samples, and the remaining 3744 frames are used as test samples. Similarly, the k value is 8.

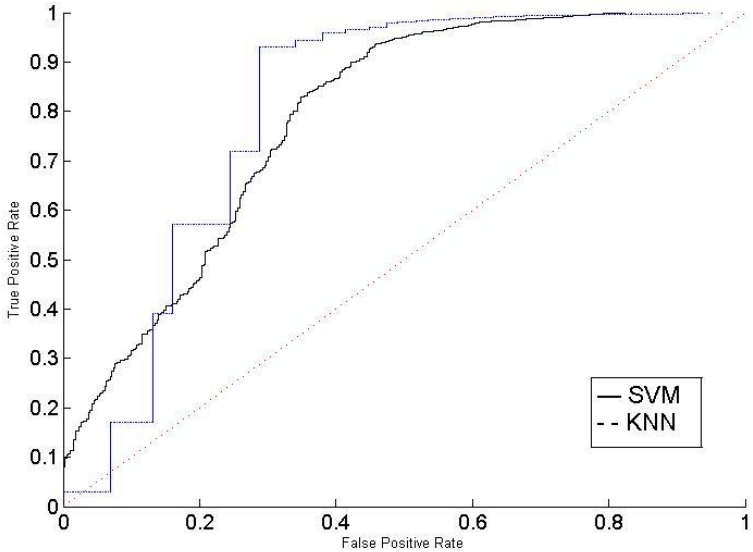


Fig. 5. ROC Curve of the Indoor Scene

As shown in Fig. 5, the black solid line is the ROC curve of the SVM classifier, and the blue dashed line is the ROC curve of the KNN classifier. It can be calculated that AUC_SVM=0.785970 and AUC_KNN=0.802416. Therefore, in this scene, the anomaly detection method based on MHOF features using the KNN classifier is superior to SVM.

(4)Square Scene

This scene has a total of 2142 frames, with 3 abnormal events occurring. Therefore, 1974 frames are normal video clips and 168 frames are abnormal event clips. The first 400 frames of this scene are selected as training samples, and the remaining 1742 frames are used as test samples. Similarly, the k value is 8.

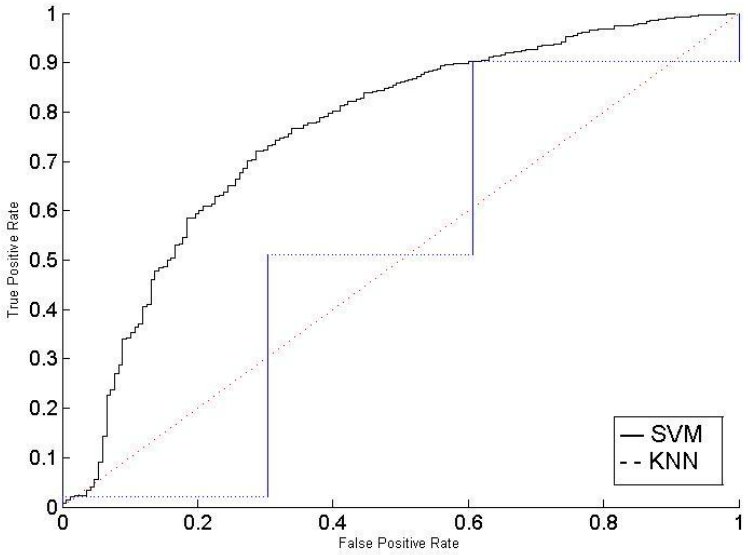


Fig. 6. ROC Curve of the Square Scene

As shown in Fig. 6, the black solid line is the ROC curve of the SVM classifier, and the blue dashed line is the ROC curve of the KNN classifier. It can be calculated that AUC_SVM=0.758062 and AUC_KNN=0.515968. Therefore, in this scene, the anomaly detection method based on MHOF features using the SVM classifier is superior to KNN.

VI.CONCLUSION

Both SVM and KNN are classifiers. It can be seen from the experimental data that the classification of SVM is relatively stable, while the classification of KNN is not very stable, which is determined by their principles. SVM has kernel functions, and kernel functions can be selected according to data characteristics. KNN only uses Euclidean distance for analysis, which is too simple and has a large computational complexity. Moreover, the selection of k is likely to affect the classification result, especially when the proportion of positive and negative samples is very different. Of course, the selection of features also has an impact on the classifier. It can be seen from the experimental results that the MHOF feature and the SVM classifier can be well combined to better realize the detection of crowd abnormal events in surveillance videos.

REFERENCES

[1]. T. Wang,M. Qiao, A. Zhu,et al. “Abnormal event detection via covariance matrix for optical flow-based feature.” Multimedia Tools and Applications,2017.
[2]. T. Wang, M. Qiao, Y. Deng, Y. Zhou, et al. “Abnormal event detection based on analysis of movement information of video sequence.” Journal for Light & Electronoptic,2018.
[3]. Y. Geng, J. Du, and M. Ling. “Abnormal event detection in tourism video based on salient spatio-temporal features and sparse combination learning.” World Wide Web 22,2019.
[4]. J. Sun, J. Shao, C. He. “Abnormal event detection for video surveillance using deep one-class learning.” Multimedia Tools and Applications 78.3,2019,3633-3647.
[5]. J. Yu, J. Kim, J. Gwak, et al. “Abnormal Event Detection using Adversarial Predictive Coding for Motion and Appearance.” 2021.
[6]. Xia, Limin, and C. Wei. “Abnormal event detection in surveillance videos based on multi-scale feature and channel-wise attention mechanism.” The Journal of Supercomputing 78.11,2022:13470-13490.
[7]. W. Ren. Abnormal crowd behavior detection using behavior entropy model, in Proc. International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), pp. 212-222, July 2012.
[8]. Cao Yihua, Yang Hua, Li Chuanzhi. Crowd anomaly detection based on spatio-temporal LBP weighted social force model. TV Technology, 36(21), 2012
[9]. Benmakrelouf, S., et al. “Abnormal behavior detection using resource level to service level metrics mapping in virtualized systems.”Future generation computer systems 102.Jan.(2020):680-700.
[10]. P. Cui. A Matrix-Based Approach to Unsupervised Human Action Categorization, IEEE Transactions on Multimedia, pp. 102-110, Vol.14, 2012.
[11]. Zhu Xudong, Liu Zhijing. Human abnormal behavior recognition based on topic Hidden Markov Model. Computer Science, 39(3), 2012.
[12]. Wang Qiao, Lei Hang, Hao Zongbo. Abnormal behavior detection based on global energy model. Application Research of Computers, 39(12), 2012.
[13]. Wu Yanping, Cui Yu, Hu Shiqiang. Abnormal behavior detection based on motion image series. Application Research of Computers, 27(7), 2010.
[14]. Guo Yingchun, Wu Peng, Yuan Haojie. Abnormal behavior detection in video frames based on self-projection and gray retrieval. Data Acquisition and Processing, 5(9), 2012.

[15]. Li Xiaodong, Ling Jie. Research on abnormal behavior detection based on video surveillance reference quantity. Computer Technology and Development, 22(9), 2012.
[16]. J. Yuan. Discriminative Video Pattern Search for Efficient Action Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33, 2011.
[17]. Y. Cong, J. S. Yuan, Y. D. Tang. Video anomaly search in crowded scenes via spatio-temporal motion context. IEEE Transactions on Information Forensics and Security, 8, 2013.
[18]. V. N. Vapnik. The Nature of Statistical Learning Theory, Springer-Verlag, New York, NY, USA, 1995.
[19]. V. N. Vapnik. Statistical Learning Theory, Wiley-Interscience Publication, USA, 1998.
[20]. Hastie, T., Tibshirani, R. Discriminant Adaptive Nearest Neighbor Classification, IEEE Transactions on Pattern Analysis and Machine Intelligence, 18, 1996.
[21]. Li, X., Wang, Y., & Zhang, Z. Lightweight spatiotemporal network for real-time video anomaly detection. IEEE Transactions on Circuits and Systems for Video Technology, 33(5),2023, 2102-2116.
[22]. Zhang, H., Chen, L., & Liu, W. Multi-modal fusion with self-supervised learning for anomaly detection in surveillance videos. Computer Vision and Image Understanding, 238,2024, 103-118.
[23]. Zhou, Y., Huang, Q., & Wang, F. Video anomaly detection via motion-appearance memory network. Proceedings of the AAAI Conference on Artificial Intelligence, 37(2), 2023,1542-1550.