An Intelligent Sewer Defect Detection Method Based on Convolutional Neural Network

Kefan Chen, Hong Hu and Chaozhan Chen
Department of Mechatronic Engineering
Harbin Institute of Technology Shenzhen Graduate School
Shenzhen, Guangdong Province, China
honghu@hit.edu.cn

Long Chen and Caiying He
Technology Center
Shenzhen Colibri Technologies Co. Ltd
Shenzhen, Guangdong Province, China
longchen@colibri.com.cn

***Abstract* - Closed circuit television (CCTV) is currently the main method of sewer pipe detection. In most practical applications, the video collected has to be checked manually since the image processing method and the machine learning method are mostly applied to experiments or to simple pipeline scenes and are rarely applied to complex and changeable pipeline detection. In this paper, a detection method based on convolutional neural network is proposed. Pipeline detection is divided into video frame abnormal detection or abnormal frame defect detection. An improved optical flow algorithm, which monitors the movement changes of the camera, is used to intelligently analyze the videos. Compared with other methods, the proposed method has a higher detection accuracy and stronger scene adaptability in real pipeline scenes.**

Keywords: Sewer pipe detect; Machine vision; CNN; Detection video

I. INTRODUCTION

With increasing urbanization, underground pipeline networks are becoming more and more dense, and some pipelines have reached the limit of their service life. If the networks are not regularly inspected and maintained, there is a great danger of damage to property and a risk to personal safety. At present, common detection methods include sonar detection, electromagnetic detection, ultrasonic guided wave detection, laser radar detection, and CCTV detection [1]. CCTV detection has become the main detection method because of its low cost, intuitive results, and high detection efficiency. An operator controls a pipeline robot equipped with a camera to collect video for detection. However, checking a large number of videos manually is inefficient, is highly subjective, and is very costly. Therefore, lots of researchers use the machine vision method instead of the human eye to view pipeline video. Pipeline defects are classified into functional defects and structural defects. Common structural defects include cracking, deformation, corrosion, and intrusion. Common functional defects include deposition, obstacles, and scum. Due to the complexity of underground pipeline systems and the diversity of defect types, existing detection methods cannot be effectively applied in practice.

Ming-Der Yang et al. [2] used wavelet transform and calculated its gray level co-occurrence matrix to describe the texture features of pipeline images, After comparing the performance of support vector machine (SVM), backward propagation neural network (BPN), and Bayesian, they found that SVM performs the best with a 60% accuracy result obtained in testing the image set. Ayan Chaki et al. [3] proposed a semi-automatic detection method that extracted the edge information of images and used K-means to remove interference edges. They then sent the segmented suspected defect image to a fuzzy multi-factor decision system for judgment. Mahmoud et al. [4] proposed a root intrusion defect detection algorithm that used image processing algorithms to extract the defect ROI and calculated the histogram of oriented gradients (HOG) features of the ROI region. They then used SVM to determine whether the ROI is defective. Jantira et al. [5] used optical flow algorithm to track the robot's motion and extracted the suspected defective video fragments according to the robot's forward speed. They then computed Haar-like features combined with multiple classifiers to judge if there were defects in the video fragments. However, this detection method assumes that the operator will slow down the robot's speed for detailed shooting when defects are found, which is untenable in most cases. Joshua et al. [6] used a multi-scale multi-directional Gabor filter to extract features from an image and used Extremely Random Tree for classification, which can obtain an accuracy of 88% in abnormal detection on a test data set.

The above-mentioned detection methods can only be applied to experimental pipelines or a specific pipeline where the scene change amplitude is relatively small. In terms of accuracy and robustness, there is still a big gap between experiments and application in real-world pipelines. In reality, the pipeline scene is complex and varied, and the camera-shooting angle is frequently changing, which requires detection methods with better robustness and stronger generalization ability. In this paper, a kind of robust defect detection method is proposed for several common types of defects with obvious features, such as obstacles, deposition, intrusion, and blur. A deep learning model is built based on a large number of detection videos provided by a pipeline robot company. Sample pipe images are shown in Fig. 1. In addition, this paper shows how the improved optical flow algorithm can be used to monitor the movement of the camera, which can effectively eliminate

the false alarm caused by a large deflection of the shooting angle. Compared with other algorithms, the algorithm presented in this paper has higher accuracy and stronger scene adaptability.

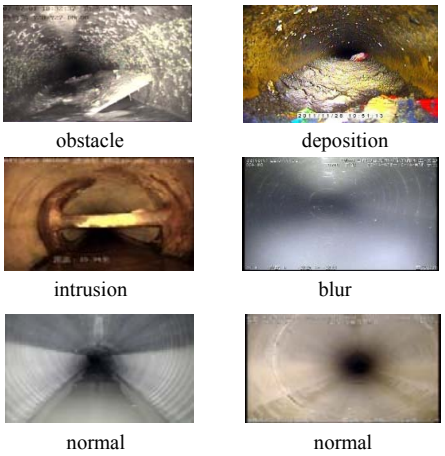

Fig. 1. CCTV images of pipe defect samples

II. Sewer Detect Framework

This paper proposes a detection method that can be directly applied to a real scene, which has high robustness and strong generalization ability. Traditional machine learning algorithms, such as SVM, NN and RF, need to select appropriate features to describe the data and performance greatly depends on the selection of features. Deep learning methods, such as convolutional neural networks (CNN), automatically extract multi-dimensional features through convolution and pool layers and display superior performance when the training data is sufficient. The proposed pipeline detection method therefore uses the CNN model as its framework, as shown in Fig. 2.

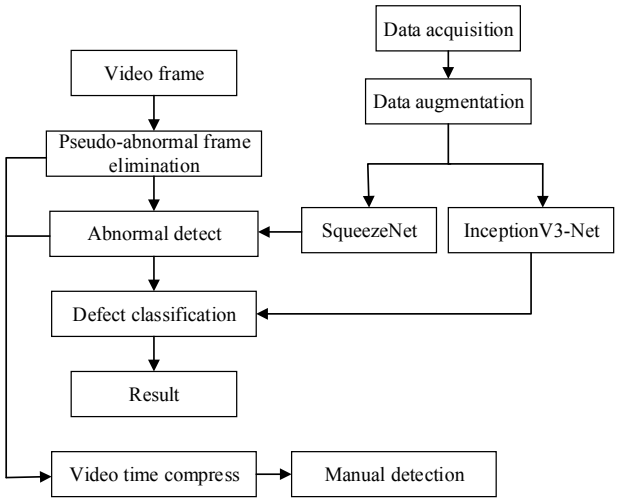

Fig. 2. Sewer detect framework

Step-1: Data acquisition. The quality of the training data largely determines the performance of the entire model. Approximately 10,000 normal images from the detection video provided by a pipeline robot company were selected, and approximately 2000 images of each defect were selected - 75% training data and 25% test data.

Step-2: Data augmentation. In order to improve the adaptability of the proposed model to scenes, scales and deflections, data augmentation must be performed on existing data. The model uses data augmentation methods such as random cropping, principal component analysis (PCA), color enhancement, and rotation transformation Random cropping enables the model to adapt to changes in scale, while PCA color enhancement enables the model to adapt to scene changes. Rotational transformation rotates the original image by a small angle, allowing the model to adapt to the small yaw of the camera angle.

Step-3: Pseudo-abnormal frame elimination. Due to changes in the topography of pipelines or artificial control of operators during the course of the robot's movement, the camera's shooting angle will change dramatically. It will generate pseudo-abnormal frames and cause false alarms. A pseudo-abnormal frame is an abnormal frame caused by a large deflection of the camera rather than the defect of the pipe itself. In the process of detecting a video frame, there is no need for alarm when there is an abnormal frame. The proposed model uses an improved bi-directional optical flow detection algorithm to monitor the movement of the camera and remove video frames whose amplitude of variation is larger than a certain threshold. This method can greatly eliminate false alarms in the detection process.

Step-4: Abnormal detection. Due to the fact that the number of normal samples is much larger than that of a single type of abnormal sample, if we take all normal images as a class and each type of defect images as a class, this will lead to unbalanced data distribution and will affect the performance of the model. Therefore, the detection is divided into two steps: abnormal detection and defect classification. During the training of the abnormal detection model, all normal images are taken as a negative sample and a total of four types of defect image are taken as a positive sample. After abnormal detection, we can eliminate video fragments that do not contain defects and compress the time of the video. Then video fragments containing defects are sent to professionals for double checking. This ensures a high defect detection rate while greatly reducing the labor intensity of the test personnel.

Step-5: Defect classification. In order to build a complete detection system, the frames that are judged as abnormal in the previous step, will be judged as a specific type of defect in this step.

III. Concrete Method

A. Camera Motion Detection Algorithm

As described earlier, in order to reduce false alarms, movement of the camera should be detected. First, we need

to find the feature points of the current video frame, and then compute the coordinate of the feature points in the next frame through a tracking algorithm. Based on the foregoing information, the movement distance and direction of the pipeline robot can be calculated.

The proposed model uses corner points as tracking feature points and the Harris feature point detection algorithm is adopted [7]. In order to define the concept of corner points in an image, the algorithm places a small window around the assumed interest point and observes the average change in the gray level in a certain direction within the window. If the displacement vector is (p, q) then the average gray level change is R:

$$R \approx \sum \left(I(x+p, y+q) - I(x, y) \right)^2 \tag{1}$$

If the values of multiple directions are high, the point is considered as a corner. The Taylor expansion of (1) is (2):

$$R = \sum \left(\left(\frac{\partial I}{\partial x} p\right)^2 + \left(\frac{\partial I}{\partial y} q\right)^2 + 2\frac{\partial I}{\partial y}\frac{\partial I}{\partial x} qp \right)^2 \tag{2}$$

Written in matrix form it is:

$$R \approx [p \quad q] \begin{bmatrix} \sum \left(\frac{\partial I}{\partial x}\right)^2 & \sum \frac{\partial I}{\partial x}\frac{\partial I}{\partial y} \\ \sum \frac{\partial I}{\partial x}\frac{\partial I}{\partial y} & \sum \left(\frac{\partial I}{\partial y}\right)^2 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} \tag{3}$$

This is a covariance matrix that represents the rate of the gray level changes in all directions. The two eigenvalues of the covariance matrix represent the maximum average gray level change and the average gray level change in the maximum vertical direction. If both are large, it is the corner point.

According to the continuity of the video frame, the coordinate of detected corner points in the next frame can be calculated by the Lucas-Kanade algorithm [8]. Assuming that the feature point gray level in each frame is constant and the displacement vector is (m, n), then (4) can be obtained by:

$$I_t(x, y) = I_{t+1}(x+m, y+n) \tag{4}$$

Where I_t and I_{t+1} are the gray level of (x, y) in current frame and that in the next instant frame. Tyler expansion can be used to get approximate equations (5):

$$I_{t+1}(x+m, y+n) \approx I_t(x, y) + \frac{\partial I}{\partial x} m + \frac{\partial I}{\partial y} n + \frac{\partial I}{\partial t} \tag{5}$$

Two items that represent intensity values are removed to get another equation:

$$\frac{\partial I}{\partial x} m + \frac{\partial I}{\partial y} n = -\frac{\partial I}{\partial t} \tag{6}$$

The L-K algorithm assumes that the displacements of all points in the neighborhood of the feature point are equal. Therefore, we can apply the optical flow constraint to all the points in the neighborhood, so we can solve the system of equations in the sense of mean square to get the displacement vector (m, n) between two adjacent frames.

However, this feature point-tracking algorithm is only suitable for scenes where the background is fixed and the foreground is variable. However, the background and foreground are changing all the time in our scene, so that it will cause false tracking. In order to solve this problem, a bi-directional optical flow method is used. That is, there is a corner point S_i in the current frame t_i, then use L-K algorithm to calculate the coordinate of S_i in the next frame t_{i+1}, call it S_{i+1}. Then take t_{i+1} as the current frame, t_i as the next frame, use L-K algorithm in a reverse way to calculate the S_i' as the corresponding corner point to S_{i+1} in t_i. If the difference between S_i and S_i' larger than a threshold δ, the corner point is thought to be a false tracking point. After the step above, most of the false tracking points will be removed, which greatly improves the tracking effect in the sewer pipe scene. The average move distance between two adjacent frames can be represented as:

$$l = \frac{\sum_{i=1}^{n} \left| S_{(t+1)i} - S_{ti} \right|}{n} \tag{7}$$

Where n is the number of the current frame. If l is very large, this frame is thought of as a pseudo-abnormal frame. And if l is very small, this frame is thought of as a static frame, which will play an important role in video time compress. The tracking effect is shown in Fig. 3.

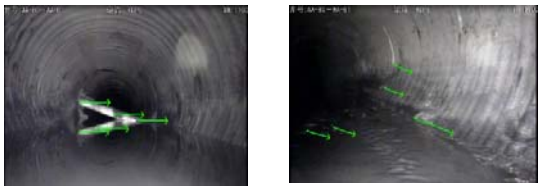

Fig. 3. Tracking effect

B. Abnormal detection

Due to the complexity of pipeline scenarios and defects, traditional machine learning models cannot achieve satisfactory results. Therefore, the proposed method uses a convolutional neural network to construct a deep learning model for abnormal detection. In order to improve the generalization ability of the model, data enhancement must be performed on the training set. PCA color enhancement can effectively improve the detection accuracy of the model by performing principal component analysis in the RGB color space of the training set to obtain three main directions of the RGB space φ_1, φ_2, φ_3 and three main eigenvalues p_1, p_2, p_3. For each pixel $I_{xy} = [I_{Rxy}, I_{Gxy}, I_{Bxy}]^T$ of each image, add the following changes:

$$I_{xynew} = I_{xyold} + [p_1, p_2, p_3][\alpha_1\varphi_1, \alpha_2\varphi_2, \alpha_3\varphi_3]^T \tag{8}$$

Where α_i is a random variable that satisfies a mean of 0 and a variance of 0.1. PCA color enhancement effect is shown in Fig. 4.

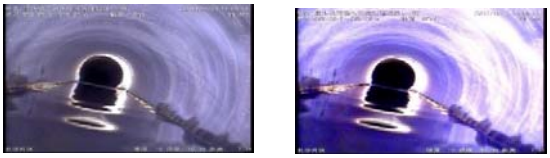

Fig. 4. PCA color enhancement effect

Considering that video detection is highly dependent on real-time performance of the algorithm, the lightweight network SqueezeNet is selected [9]. The original convolutional layer is divided into squeeze and expand layers in this model. The 1×1 convolution kernel in the squeeze layer greatly reduces the model parameters. At the same time, the output of 1×1 and 3×3 convolution layer are concatenated in the expand layer as the input of the next layer. The above structure forms the Fire Module, the core component of SqueezeNet, shown in Fig. 5.

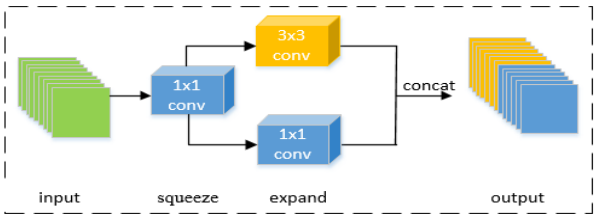

Fig. 5. Structure of Fire Module

SqueezeNet has a total of 9 fire modules, interspersed with some of the maxpool layers, and finally replaced the full connection layer with a global average layer, which can achieve almost the same accuracy as AlexNet on ILSVRC2012 data set [10]. However, the model parameters are only 1/50 of that of AlexNet. The overall structure is shown in Figure. 6.

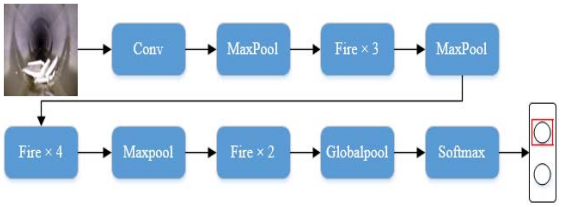

Fig. 6. Structure of SqueezeNet

The abnormal detection model can be used for time compression of the detection video. After this step, the normal frames are removed, only the abnormal video frames are left for manual secondary detection. This can greatly reduce the labor intensity while ensuring a high defect detection rate.

C. Abnormal frame defect classification

Compared to abnormal detection, the number of input images of this classifier is relatively small and the classification is more difficult. Therefore, the requirement on real-time performance is lower and that on recognition ability is higher. This paper uses GoogleNet InceptionV3 model [11]. The Inception structure uses 1×1, 3×3, 5×5 convolution kernel and sends them to the next layer after concatenation, which means that different scales of features are merged together to enhance the understanding ability of the model. At the same time, taking into account the small number of defective samples, direct training results are poor. Therefore, the model is pre-trained on the ILSVRC2012 data set, and the pipeline defect samples are used for fine-tuning to achieve the purpose of transfer learning. In addition, the optimization algorithm based on the gradient descent principle is easy to fall into the local optimal solution. Therefore, this paper uses the momentum descent method and the Adam optimization algorithm to train the model respectively [12,13], and assembles two models to achieve better detection results. The total process is shown in Fig. 7.

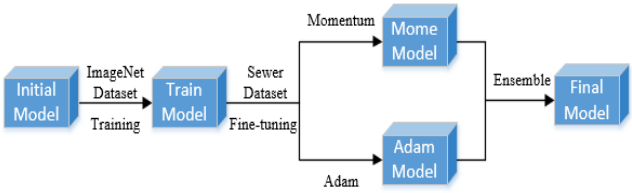

Fig. 7. Defect classification process

IV. EXPERIMENT

A. Abnormal detection

Because underground pipeline scenes do not possess public data sets, and the data set in this paper is complicated and has diverse shooting angles, it is more difficult than other data sets. Therefore, the method proposed in this paper cannot be directly compared with other methods. In order to solve this problem, we reproduce the methods with good detection effect in other papers and compare them with the data set constructed in this paper. Paper [6] used the multi-scale and multi-directional Gabor filter to extract the feature of image to compose GIST features, and used an extreme random tree as a classifier for classification. Paper [4] calculated the HOG features of the image and classifies it using SVM. After the optimization of the parameters, the final results are compared with the method proposed in this paper. The result is summarized in Table I, and the ROC curves of each model are shown in Fig. 8.

From the experimental results, it can be seen that the ROC curve of the abnormal detection method proposed in this study can completely envelope the ROC curve of other methods, which means that our method has a superior performance in all aspects [14]. And the recall rate of abnormal samples can reach 88%.

TABLE I
PERFORMANCE COMPARISON

	Recall	Precision	Accuracy
GIST+ET[6]	0.72	0.71	0.70
HOG+SVM[4]	0.75	0.76	0.75
SqueezeNet	**0.88**	**0.84**	**0.85**

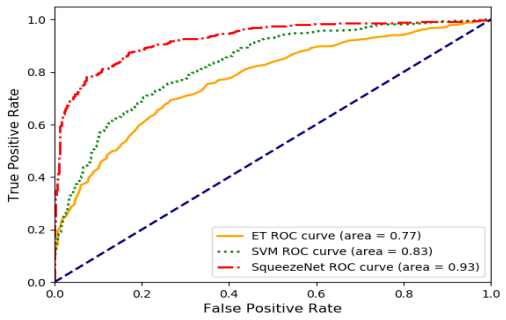

Fig. 8. ROC curve of each model

B. Video time compression

The abnormal detection is combined with the camera motion detection algorithm to compress the time of video. If the optical flow vector amplitude between adjacent frames is close to 0, then the robot is considered to be static. The video fragment with a static time of more than two seconds was removed, and only the first two seconds were retained. In this paper, many fragments selected from a large number of videos were combined to form six composite videos. Most of the components of Video 1-4 is defective fragments while most of the components of Video 5-6 is normal fragments. The time compression results are shown in Fig. 9. And statistical data is shown in Table II.

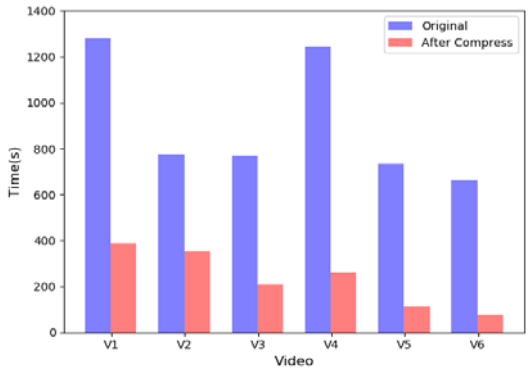

Fig. 9. Time compress results

TABLE II
STATISTICAL DATA IN TIME COMPRESS

CCTV filenames	Total time(s)	Abnormal time(s)	Defect number	Found defects	Compress rate
V1	1280	388	14	14	0.696
V2	775	353	13	13	0.545
V3	770	209	9	8	0.729
V4	1243	260	15	14	0.791
V5	734	111	3	3	0.849
V6	663	76	4	4	0.885

From the experimental results, it can be seen that using anomaly detection can greatly reduce the video time while ensuring a high defect detection rate, and can greatly improve the detection efficiency and labor intensity of the inspection personnel.

C. Defect classification

This step classifies the video frames identified as abnormal at the previous step and compares the results with the SVM. The result of SVM is shown in Table III and the result of Inception-V3 is shown in Table IV. At the same time, the strategy of transfer learning and multi-model ensemble was adopted to improve the detection accuracy, the comparison is shown in Table V.

TABLE III
SVM DETECTION RESULT

	Precision	Recall	F1-score
Deposition	0.57	0.49	0.52
Obstacle	0.53	0.50	0.51
Blur	0.68	0.65	0.66
Intrusion	0.75	0.76	0.75

TABLE IV
INCEPTION-V3 DETECTION RESULT

	Precision	Recall	F1-score
Deposition	0.65	0.88	0.75
Obstacle	0.86	0.60	0.71
Blur	0.85	0.98	0.91
Intrusion	0.95	0.96	0.95

TABLE V
ACCURACY COMPARISON

	Original	Transfer	Transfer+Ensemble
Accuracy	0.7531	0.7816	0.8121

V. CONCLUSION AND FUTURE WORK

This paper proposes a sewer pipe detection framework based on convolutional neural network, which is divided into camera motion detection, video frame abnormal detection, and abnormal frame defect classification. The combination of the first two can be used for video time compression, which can greatly reduce the video time while guaranteeing a high defect detection rate and reduce the labor intensity of the inspection personnel. The combination of the three can constitute a complete detection system, which has higher detection accuracy and stronger scene adaptability than other methods. However, the method proposed in this paper can only be applied to detecting defects with obvious features, and has poor detection performance for defects with non-significant features. In future research, digital image processing algorithms can be combined with the classification model based on deep learning to improve the detection effect of subtle defects.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support provided by the Shenzhen Government Fund JSGG20170412143346791 and JCY20170413105740689.

REFERENCES

[1] R. Wirahadikusumah, D.M. Abraham, T. Iseley, et al, “Assessment technologies for sewer system rehabilitation,” *Automation in Construction*, vol. 7, pp. 259-270, 1998.

[2] M.D. Yang and T.C. Su, “Automated diagnosis of sewer pipe defects based on machine learning approaches,” *Expert Systems with Application*, vol. 35, pp. 1327-1337, 2008.
[3] A. Chaki and T. Chattopadhyay, “An Intelligent Fuzzy Multifactor Based Decision Support System for Crack Detection of Underground Sewer Pipelines,” *International Conference on Intelligent Systems Design and Applications*, vol. 10, pp. 1471-1475, 2010.
[4] M.R. Halfawy and J. Hengmeechai, “Automated defect detection in sewer closed circuit television images using histograms of oriented gradients and support vector machine,” *Automation in Construction*,, vol. 38, pp. 1-13, 2014.
[5] J. Hengmeechai and M.R. Halfawy, “Integrated Vision-Based System for Automated Defect Detection in Sewer Closed Circuit Television Inspection Videos,” *American Society of Society of Civil Engineers*, vol. 29, no. 1, 2015.
[6] J. Myrans, Z. Kapelan and R. Everson, “Automated detection of faults in wastewater pipes from CCTV footage by using Random Forests,” *Procedia Engineering*, vol. 154, pp. 36-41, 2016.
[7] C. Harris and M.J. Stephens, “A combined corner and edge detector,” *Alvey Vison Conference*, vol. 5, pp. 147-152, 1988.
[8] B. Lucas and T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision,” *Int. Joint Conference in Aritificial Intelligence*, vol. 15, pp. 674-679, 1981.
[9] F.N. Iandola, S. Han, M.W. Moskewicz, et al, “Squeezenet : Alexnet – level accuracy with 50x fewer paraments and <0.5MB model size,” arXiv: 1602.07360, 2016.
[10] A. Krizhevsky, I. Sutskever and G. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1106–1114, 2012.
[11] C. Szegedy, W. Liu, Y.Q. Jia, et al, “Going deeper with convolutions,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9, 2015.
[12] N. Qian, “On the momentum term in gradient descent learning algorithms,” *Neural Networks*, vol. 12, no. 1, pp. 145-151, January 1999.
[13] N.K. Whitney, “Adaptive estimation of regression models via moment restrictions,” *Journal of Econometrics*, vol.38, no. 3, pp. 301-309, July 1988.
[14] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874.