Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **6**, Number **2**, May **2021**

Pose Measurement of Excavator Based on Convolutional Neural Network

Yun Shi

College of Mechanical & Electrical Engineering
Nanjing University of Aeronautics and Astronautics
29 Yudao Street, Qinhuai District, Nanjing, China
School of Mechanical and Vehicle Engineering
West Anhui University
West Moon Island, Yunlu Bridge, Lu'an, China
710667612@qq.com

Yan-Yan Zhu

School of Electronics and Information Engineering
West Anhui University
West Moon Island, Yunlu Bridge, Lu'an, China
870751547@qq.com

Jie Fang

School of Electrical and Optoelectronics
West Anhui University
West Moon Island, Yunlu Bridge, Lu'an, China
63640193@qq.com

Zu-Song Li

National University, Philippines
Taft Ave, Manila, Philippines
West Anhui University
West Moon Island, Yunlu Bridge, Lu'an, China
67202167@qq.com

Corresponding Author: Jie Fang (63640193@qq.com)
Received March 2021; revised April 2021

Abstract. *To realize the intelligentization of excavators, this paper proposes a posture control method for an excavator working device based on a convolutional neural network (CNN). First, binocular stereo vision is used to capture the image of the fuzzy square-coded target in the working device. Second, the formation mechanism of the motion blur effect is analyzed, and a simulation generation model for a motion blur image is established. Subsequently, the CNN is used to train and recognize the fuzzy square-coded points. Finally, the pose angle of the corresponding working arm is obtained by determining the fixed geometric relationship of the square-coded points pasted on the working arm. The test results show that this method can provide a certain reference for the position and attitude control of excavator working devices.*
Keywords: Convolutional Neural Network (CNN), Excavator, Square-coded target (SCT), Blurred image, Pose measurement

FIGURE 1. Square-coded target

1. **Introduction.** In recent years, convolutional neural networks (CNNs) have become a research hotspot in the field of artificial intelligence. A CNN forms a dense high-level semantic abstraction by combining low-level features. In [1], distributed feature representations of data could be automatically identified, solving the problem of manually designing features in conventional machine learning. The authors of [2,3] conducted breakthrough research in recognition and other fields. However, CNNs have rarely been used in the pose control of excavator working devices [4].

As one of the most extensively used piece of construction machinery, excavators play an important role in construction, mining, and other related fields [5]. Tian et al. [6] summarized previous studies on the pose control of excavator working devices. Commonly used methods include magnetometers, inertial-based systems, position sensors, and lasers. Excavators are often used in complex and dangerous projects [7,8]. During excavation, the working device will inevitably encounter soil and/or rocks, which easily damage the sensors onboard. Accurately obtaining the pose of the working device is an important topic in the development of excavators [9]. Therefore, to realize the intelligentization and automation of excavators, pose control is a key factor determining the actual operation ability of intelligent excavators.

In this paper, a pose control method based on a CNN is proposed. Aiming at the ambiguity problem of square-coded targets (SCTs) in the actual operation of an excavator, the proposed method is applied to accurately identify fuzzy SCTs and improve the accuracy of the excavator's pose.

2. Excavator Pose Control System.

2.1. System Composition.

The position and attitude control system of an excavator comprises a measuring plane, a calibration board, a computer, and an industrial camera. The industrial camera does not contact the measurement plane. The measurement plane is the side of the excavator working device. A camera is installed on a triangular frame and remains in a fixed position during the excavation process.

2.2. Design and Layout of Square-Coded Targets.

In this study, SCTs are designed as an active visual feature. The center of the SCTs is the target point, and the target point is typically the object of target recognition. Each side in the SCT is equally divided into three segments, and each segment is equivalent to a binary bit, where white is considered "1," black is considered "0," and each SCT has a unique 12-bit binary code. Therefore, it can be used for the identity characterization of the target point. Figure 1 shows the structure diagram of a square-coded target. Taking the upper left corner as the starting point of the binary code of the SCT, in the clockwise direction, the binary code in the figure is 001010001100, and the ID number is 652.

The SCTs designed in this study are pasted on the sides of the boom, bucket rod, and bucket connecting rod of the excavating in the form of patches. Different SCTs can

be affixed to each component, and the target point connection of each group of SCTs is ensured to be parallel to the hinge point connection of the corresponding component. Therefore, for the boom and the stick, we control the inclination of the line of the SCTs to obtain the pose angle of the corresponding working arm; for the bucket, the pose angle can be obtained through a geometric relationship. Figure 2 shows the specific arrangement of the SCTs.

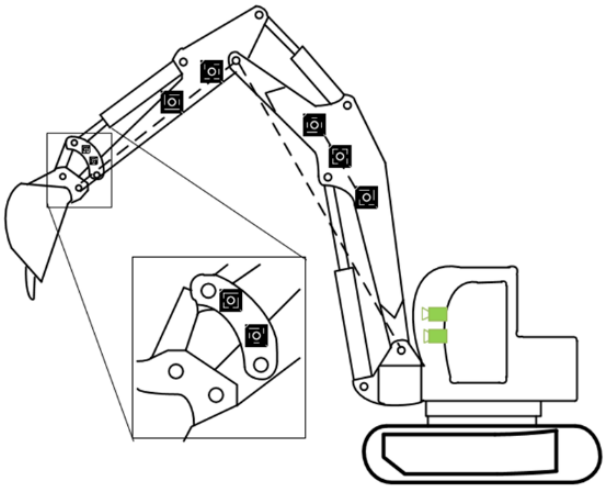

FIGURE 2. Schematic of the square-coded target layout of a pose control system

3. **Working Device Principle of Excavator.**

3.1. **Kinematics Modeling.** The linkage coordinate system of the excavator working device can be typically established using the D-H method, as shown in Figure 3. The base coordinate system is $x_0y_0z_0o_0$, and the boom, stick, bucket, and tip coordinate systems are $x_1y_1z_1o_1$, $x_2y_2z_2o_2$, $x_3y_3z_3o_3$, and $x_4y_4z_4o_4$, respectively. The z-axis in the figure is along the joint axis direction, and the coordinate axis, which is perpendicular to the paper, is not drawn. The joint space of the excavator comprises four joint angle variables: rotation θ_1, boom θ_2, stick θ_3, and bucket θ_4. The pose space is determined by the position of the tooth tip in $x_0y_0z_0o_0[x, y, z]$ and the pose angle θ_ω. Assuming that each joint angle is positive in the counterclockwise direction, the coordinate transformation of $x_iy_iz_io_i$ relative to $x_{i-1}y_{i-1}z_{i-1}o_{i-1}$ according to literature [10] can be expressed in a matrix form as follows:

$$T_i^{i-1} = \begin{vmatrix} C_i & -\cos\alpha_i S_i & \sin\alpha_i S_i & \alpha_i C_i \\ S_i & \cos\alpha_i C_i & -\sin\alpha_i S_i & \alpha_i S_i \\ 0 & S_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 0 \end{vmatrix} \tag{1}$$

$x_0y_0z_0o_0$ to $x_4y_4z_4o_4$ can be transformed into:

$$T_4^0 = T_1^0T_2^1T_3^2T_4^3 = \begin{vmatrix} C_1C_{234} & -C_1C_{234} & S_i & C_1(a_4C_{234} + a_3C_{23} + a_2C_2 + a_1) \\ S_1C_{234} & -S_1C_{234} & C_i & S_1(a_4C_{234} + a_3C_{23} + a_2C_2 + a_1) \\ S_{234} & C_{234} & 0 & a_4S_{234} + a_3S_{23} + a_2S_2 + d_1) \\ 0 & 0 & 0 & 1 \end{vmatrix} \tag{2}$$

where S_i, C_i, S_{ij}, C_{ij} are represented assin $\theta_i, \cos\theta_i, \sin(\theta_i + \theta_j)$, respectively. From Equation (2), we can obtain:

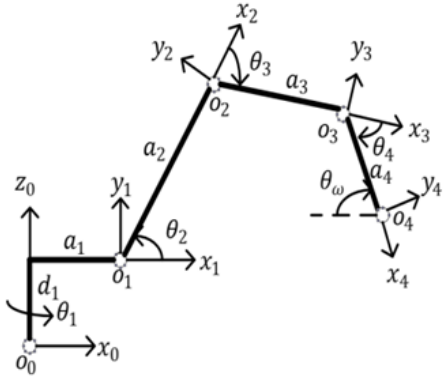

FIGURE 3. Schematic of the connecting rod of an excavator working device

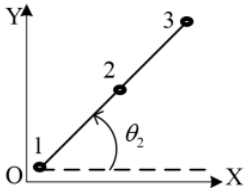

FIGURE 4. Boom inclination calculation

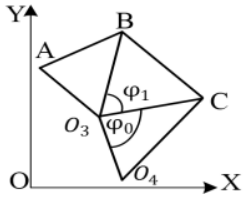

FIGURE 5. Bucket posture angle calculation

$$\begin{cases} x = C_1(a_4C_{234} + a_3C_{23} + a_2C_2 + a_1) \\ y = S_1(a_4C_{234} + a_3C_{23} + a_2C_2 + a_1) \\ z = a_4S_{234} + a_3S_{23} + a_2S_2 + d_1) \\ \theta_w = \theta_2 + \theta_3 + \theta_4 \end{cases} \tag{3}$$

3.2. **Attitude Angle Calculation.** Different SCTs are pasted on the boom, and the pose measurement system of the excavator working device extracts the plane coordinates $D_1(x_1, y_1)$, $D_2(x_2, y_2)$, and $D_3(x_3, y_3)$ of the code points, as shown in Figure 4.

The inclination angles of the three coding points on the side of the boom are:

$$\begin{cases} \theta_{12} = \arctan(y_1 - y_2)/(x_1 - x_2) \\ \theta_{13} = \arctan(y_1 - y_3)/(x_1 - x_3) \\ \theta_{23} = \arctan(y_2 - y_3)/(x_2 - x_3) \end{cases} \tag{4}$$

We take the average of the inclination angles θ_{12}, θ_{13}, and θ_{23} to reduce the error: $\theta_2 = (\theta_{12} + \theta_{13} + \theta_{23})/3$. For the boom and stick, based on the arrangement shown in Figure 2, the line of the SCTs is parallel to the line of the corresponding component; hence, the obtained inclination is the pose angle of the component. Similarly, the bucket connecting rod pose angle θ can be calculated. The bucket is often buried into the soil, making it necessary to arrange the SCTs on the bucket connecting rod and obtain the pose angle through geometric relationship conversion.

Figure 5 shows a schematic of the stick, bucket, and accessory connecting rod structure. O_3O_4C represents the bucket, the hinge point connecting the stick and the bucket is O_3, the tip of the shovel is O_4, and O_3A is on the stick. The SCTs are arranged on the connecting rod AB, and the other connecting rod is BC; thus, we have $\vec{O_3B} = \vec{AB} - \vec{AO_3}$. Assuming that the direction angle of O_3B is φ_2, we have $\varphi_2 = \arctan \frac{|\vec{AB}| \sin\theta - |\vec{AO_3}| \sin(\theta_2+\theta_3)}{|\vec{AB}| \cos\theta - |\vec{AO_3}| \cos(\theta_2+\theta_3)}$. where $\vec{AB} = (\vec{AB}\cos\theta, \vec{AB}\sin\theta)\vec{AO_3} = (|\vec{AO_3}|\cos(\theta_2+\theta_3), |\vec{AO_3}|\sin(\theta_2+\theta_3))$.

$$\theta_w = \varphi_2 - \varphi_1 - \varphi_0 \tag{5}$$

where $\varphi_1 = \arccos \frac{|O_3B|^2+|O_3C|^2+|BC|^2}{2|O_3C||BC|}$, and φ_0 is a constant value. We then solve θ_w using Equation (5), i.e., the pose angle of the bucket.

3.3. **Three-Dimensional Reconstruction.** The matching points on the left and right views have an epipolar geometric relationship [11-14], and the pixel coordinates of the matching points D_l and D_r with the same name are obtained from the left and right views using the triangulation method. We then calculate the 3D coordinates of the matching point in the camera coordinate system $D_{cl} = (x_l, y_l, f_l)$, $D_{cr} = (x_r, y_r, f_r)$. If the left camera coordinate system is selected as the world coordinate system, the rays $O_{cl}D_l$ and $O_{cr}D_r$ can be expressed as $bR_{r2l}D_{cr} + t_{r2l}$ and aD_{cl}, respectively. Theoretically, the intersection between the two rays is the 3D point to be obtained; however, because of the influence of the matching errors, the two rays generally do not intersect. In this study, the midpoint of the perpendicular common to the two rays is taken as a 3D point. Let $\vec{w}$ be a vector perpendicular to the rays $O_{cl}D_l$ and $O_{cr}D_r$, namely $\vec{w} = D_{cl} \times R_{r2l}D_{cr}$ based on the triangulation method, we have:

$$aD_{cr} - bR_{r2l}D_{cr} + c\vec{w} = t_{r2l} \tag{6}$$

From Equation (6), the parameters a, b, and c can be determined, and the coordinates of the 3D point can be expressed as:

$$D_w = aD_{cl} + c\vec{w}/2 \tag{7}$$

4. **Construction of a CNN.** In an actual operation process, the excavator will inevitably encounter rocks and soil, which can easily cause blurring of its SCTs. The method proposed in [15-17] cannot effectively identify SCTs when subjected to motion blur effects. This study uses a CNN to identify fuzzy SCTs and intends to use software to simulate actual shooting scenes to generate fuzzy SCTs and solve the bottleneck of samples in network training.

4.1. **Simulation of SCT Generation.** Using the motion blur model proposed in [18,19], this model reproduces the target motion and imaging process within the exposure time. Assuming that the code point plane is in the plane and the world coordinate system coincides with the "target point" of the SCTs, the exposure time is discretized into N times. At time "i"$(i = 1, 2, ..., N)$, the code point moves from the starting pose through the rotation transformation R^i and translation transformation t^i to the current pose. After perspective transformation imaging on the camera plane, a clear image of the SCTs at time "i" can be obtained, i.e., the motion blur image generated during the exposure time. According to the imaging principle of the camera, the SCTs at time "i" can quantitatively indicate the projection process as:

$$\lambda^{I_i}(\mu^{I_i} \quad \upsilon^{I_i} \quad 1)^T = K(R^i(X_c \quad Y_c \quad 0)^T + t^i) \tag{8}$$

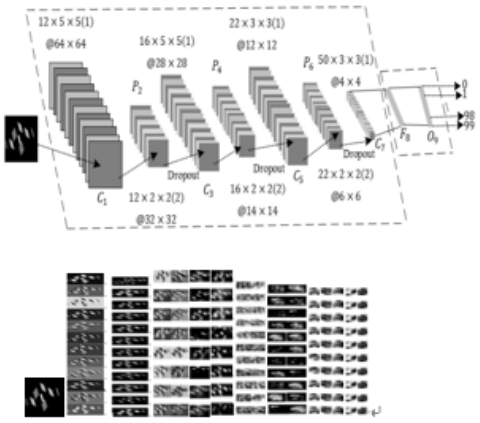

FIGURE 6. CNN structure diagram and feature diagram with the SCT

$$B = \sum_{i}^{N} \frac{I_i}{N} \tag{9}$$

where λ^{I_i} is the scale factor, and I_i is the clear image at time "i",$(\mu^{I_i}, \upsilon^{I_i})$ is the pixel coordinates of I_i, K is the projection matrix, $(X_c, Y_c, 0)$ is the coordinate of any point on the coding pattern in the world coordinate system, and B is the output blurred image. The SCTs are used as target points to be measured to simulate an actual scene, and a motion blur image simulation model is established. In an actual shooting process, noise interference from various sources is inevitable. When noise n is introduced into a generation model, the generation model that simulates the motion blur image of the SCTs can be expressed as:

$$B = f(\boldsymbol{r}, \boldsymbol{d}, \delta, n) \tag{10}$$

where δ represents the degree of blur, $\boldsymbol{r}$ represents the target pose vector of SCTs, and the unit vector $\boldsymbol{d}$ represents the direction of movement. Based on the rule selection parameters, the motion blur images of the SCTs that can simulate various spatial position orientation, target motion path, blur degree, and noise level provide sufficient sample images for training the CNN.

4.2. **Training and Identifying Code Points.** The authors of [20] conducted pioneering research on deep learning and artificial neural networks in academia, leading to many interesting studies. Krizhevsky et al. [21] achieved the best classification result in the ImageNet Large-scale Visual Recognition Challenge Competition (LSVRC), making CNNs more popular.

A CNN is typically composed of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer. In this study, the pooling layer uses the maximum pooling function, the activation function and convolutional layer of all the neurons in the fully connected layer use the ReLU function, and the output value of the fully connected layer uses Softmax logistic regression for classification.

As shown in Figure 6, the CNN structure diagram and feature diagram with the SCT ID number 597 have a total of four convolutional layers C_1, C_3, C_5, C_7, three pooling layers P_2, P_4, P_6, 1 fully connected layer F_8, and 1 output layer O_9. In the figure, 12×5×5(1) @ 64×64 means that there are 12 convolution kernels, the window size of the convolution kernel is 5×5, the size of each output feature surface is 64×64, and the convolution kernel is in its input. The sliding step on the feature surface is 1. In the process of CNN training for fuzzy SCTs, it is necessary to avoid any overlap between the training and test sets

and ensure a consistent distribution. A generative model is used to simulate and generate four groups with a total of 400,000 fuzzy SCTs, of which 300,000 SCTs are used as the training set and 100,000 SCTs are used as the test set. We select 100 of the 4096 SCTs with evident characteristics and use to represent the output category.

The CNN uses the cross-entropy loss function as the optimization objective function, the optimization algorithm used in the training process is the small-batch stochastic gradient descent algorithm, and the gradient is calculated using the BP algorithm. In the training process, the batch size is set to 128, and the initial learning rate is 0.1. After each parameter is updated using the gradient descent algorithm, and the learning rate drops to an average loss of 0.05.

4.3. **Extraction of SCTs.** It is necessary to segment the local area where the motion blur SCTs are located from the actual captured image to complete the training and recognition tasks of the CNN. First, the real shot image is binarized and expanded, the contour is then extracted from the processed binary image, and finally, the target SCTs are segmented from the entire image using the smallest bounding box. The specific segmentation algorithm can be divided into four steps:

Step 1. The principle of morphology is used to binarize and then expand the input image.

Step 2. The Canny operator in [22] is used to extract the contour of the dilated binary image.

Step 3. The contour area, perimeter, smallest enclosing rectangle aspect ratio, and average gray value of the local area of the smallest bounding box are used to filter the contour, and the image edge contour is deleted.

Step 4. The minimum bounding box corresponding to the filtered contour is used to segment the local area covered by the bounding box, and the segmented image is outputted.

5. Experimental Verification.

5.1. **Experimental Environment.** The Yuchai ZH-WJ hydraulic excavator training platform is used for the test. The measurement system uses an industrial camera (model: Basler A102f) with a resolution of 1392×1040, a frame rate of 15 fps, and a USB interface. The camera calibration program is a calibration interface written in Qt by calling the OpenCV library based on C++. Under the Linux system, the code point segmentation and recognition results are performed in Python language. We employed a DELL Precision T5610 workstation with 2 GB of video memory, 24 GB of memory, and Quadro K2000 as the GPU.

5.2. **Experimental Results.** Experiments were conducted under natural lighting conditions to compare the results of the sensor control method and the control method proposed in this article. Figure 7 shows the measurement errors of the two methods. The experimental results show that the posture control method of the excavator working device based on the CNN is feasible and meets the automation and real-time requirements of the excavator.

6. Conclusion.

This study designed binary SCTs and analyzed the feature extraction, segmentation, and recognition algorithms of code points. A posture control system for an excavator working device was designed on the basis of a CNN, and the kinematics analysis and modeling of the excavator working device were carried out. The experiments conducted could help verify the feasibility of the CNN method in controlling the pose of the excavator working device.

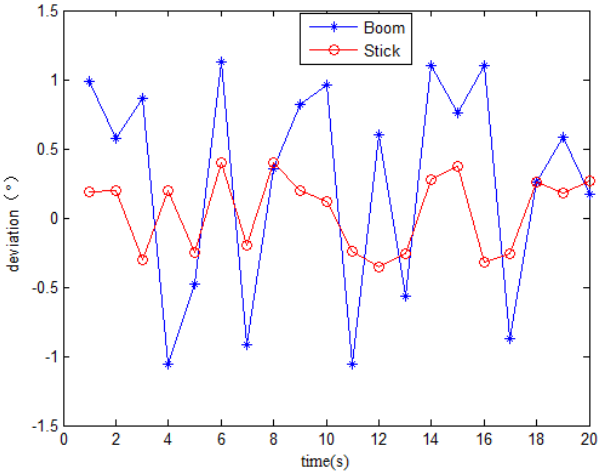

FIGURE 7. Error results

Acknowledgment. This work was supported by the School-level Research Projects of West Anhui University under grant no.WXZR201906, West Anhui University school-level quality engineering project (wxxy2020019, wxxy2020178), Anhui Provincial Quality Engineering Project (2020mooc539, 2020jyxm2150), and National Natural Science Foundation of China (project number: 61302179), a major science and technology project of Lu'an City (design and industrialization of high-speed and low-torque pulsation permanent magnet synchronous motors).

REFERENCES

[1] L.W. Huang, B.T. Jiang, S.Y. Lv, Survey on Deep Learning Based Recommender Systems, *Chinese Journal of Computers*, vol. 41, no. 5, pp. 1619-1647, 2018.

[2] T.L. Liao, H.C. Chen, J.J. Yan, Design of Real-time Face Position Tracking and Gesture Recognition System based on Image Segmentation Algorithm, *Journal of Network Intelligence*, Vol. 5, No. 4, pp. 226-239, 2020.

[3] J. Tavoosi, Designing a new recurrent convolutional neural network for face detection and recognition in a color image, *Iran Journal of Computer Science*, pp. 1-10, 2021.

[4] H.L. Zou, *Excavator Manipulator's Attitude Detection System based on Vision Measurement*, Dissertation, Southwest Jiaotong University, 2015.

[5] X.J. Dai, *The Research of Excavator,S Robotics Technology*, Dissertation, Northeastern University, 2010.

[6] G.Z. Tian, Q. An, C.Y. Ji, (eds.), Real-time Motion Detection for Intelligent Agricultural Vehicle Based on Stereo Vision, *Transactions of the Chinese Society for Agricultural Machinery*, vol. 44, no. 7, pp. 210-215, 2013.

[7] Q. Zhang, *Research on Plan and Control Technology of Excavator-Robot*, Dissertation, Zhejiang University, 2002.

[8] Q. Ha, M. Santos, Q. Nguyen, (eds.), Robotic excavation in construction automation, *IEEE Robotics and Automation Magazine*, vol. 9, no. 1, pp. 20-28, 2002.

[9] Z.R. Dong, *Implementation and Joint Simulation Optimization of 3D Digital Models for Working Device of Hydraulic Excavator*, Dissertation, Jilin University, 2019.

[10] W.W. Feng, *Design and Realization of Motion Attitude Detection System*, Dissertation, Chongqing University, 2008.

[11] H. Zhang, L.Y. Zhang, H. Wei, Three-step Calibration for Binocular Stereo System, *China Mechanical Engineering*, vol. 20, no. 16, pp. 1899-1903, 2009.

[12] H. Zhang, L.Y. Zhang, J. Chen, Z.P. Zhao, Field Calibration of Binocular Stereo System Based on Planar Template and Free Snapping, *ACTA AERONAUTICAER ASTRONAUTICA SINCIA*, vol. 28, no. 3, pp. 695-701, 2007.

[13] E. Trucco and A. Verri, Introductory techniques for 3D computer vision,*Englewood Cliffs: Prentice Hall*, 1998.

[14] R. Hartley and A. Zisserman, Multiple view geometry in computer vision, *Cambridge University Press*, 2004.
[15] K. Forbes, A. Voigt, N. Bodika, An Inexpensive, Automatic and Accurate Camera Calibration Method, *Proceedings of the Thirteenth Annual South African Workshop on Pattern Recognition Prasa*, 2002. https://www.researchgate.net/publication/2550298
[16] L. Zhou, L.Y. Zhang, J.D. Zheng, (eds.), Automated reference point detection in close range photogrammetry, *Journal of Applied Sciences*, vol. 25, no. 3, pp. 288-294, 2007.
[17] L.M. Song, C.M. Chen, Z. Chen, (eds.), Detection and recognition of circular coded targets, *Optics and Precision Engineering*, vol. 21, no. 12, pp. 3239-3247, 2013.
[18] M.J. Chen, *Research on 3D Reconstruction of Motion Blurred Visual Features*, Dissertation, Nanjing University of Aeronautics and Astronautics, 2017.
[19] M.J. Chen, H.C. Zhou, L.Y. Zhang, Recognition of Motion Blurred Coded Targets Based on Convolutional Neural Network, *Journal of Computer-Aided Design & Computer Graphics*, vol. 29, no. 10, pp. 1844-1852, 2017.
[20] G. Hinton, R.R. Salakhudinov, Reducing the dimensionality of data with neural networks, *Science*, vol. 313, no. 5786, pp. 504-507, 2006.
[21] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, *International Conference on Neural Information Processing Systems. Curran Associates Inc*, pp. 1097-1105, 2012.
[22] J. Canny, A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, 1986.