J. Meas. Sci. Instrum., 2025, 16(2): 173-185. http://jmsi.publish.founderss.cn/
DOI: 10.62756/jmsi.1674-8042.2025017 jmsi@nuc.edu.cn

A tactile glove for object recognition based on palmar pressure and joint bending strain sensing

ZHANG Xuefeng*, ZHANG Shaojie, CHEN Xin, ZHANG Jinhua

School of Mechanical and Electrical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China

***Corresponding author:** ZHANG Xuefeng (xfzhang@xauat.edu.cn)

Received: November 20, 2024 **Revised:** January 10, 2025 **Accepted:** February 24, 2025

Abstract: With the rapid development of flexible electronics, the tactile systems for object recognition are becoming increasingly delicate. This paper presents the design of a tactile glove for object recognition, integrating 243 palm pressure units and 126 finger joint strain units that are implemented by piezoresistive Velostat film. The palm pressure and joint bending strain data from the glove were collected using a two-dimensional resistance array scanning circuit and further converted into tactile images with a resolution of 32×32. To verify the effect of tactile data types on recognition precision, three datasets of tactile images were respectively built by palm pressure data, joint bending strain data, and a tactile data combing of both palm pressure and joint bending strain. An improved residual convolutional neural network (CNN) model, SP-ResNet, was developed by light-weighting ResNet-18 to classify these tactile images. Experimental results show that the data collection method combining palm pressure and joint bending strain demonstrates a 4.33% improvement in recognition precision compared to the best results obtained by using only palm pressure or joint bending strain. The recognition precision of 95.50% for 16 objects can be achieved by the presented tactile glove with SP-ResNet of less computation cost. The presented tactile system can serve as a sensing platform for intelligent prosthetics and robot grippers.

Key words: tactile glove; object recognition; Velostat; joint bending strain sensors; palmar pressure sensors; convolutional neural network

0 Introduction

Tactile perception, one of the most fundamental sensory systems of human, plays a crucial role in understanding the environment and coordinating movement. Integrating human tactile perception into robotic technology enables precise manipulation of robotic hands or prosthetics, thus playing a significant role in fields such as human-machine interfaces[1,2], healthcare[3,4], and virtual reality[5,6]. With the rapid advancement of artificial intelligence technology, deep learning has been widely applied in the field of tactile perception research.

High-density sensor array have garnered widespread attention for their ability to emulate human skin. Lee et al.[7] developed a tactile glove to identify four objects of different shapes by only measuring the bending strains of five fingers with a precision of 91.88%. This confirmed the effectiveness of utilizing bending strain information of fingers for object recognition. Zhang et al.[8] integrated a 5×5 pressure sensor array on the fingertips of a humanoid robot hand, and realized the identification of five different objects in various shapes with a precision of 99.6%. Although both of them achieved high recognition precision relying solely on overall finger bending or fingertip pressure for tactile data acquisition, there was a limit to the number of objects that could be recognized. Consequently, the dataset with small scale and few features can only achieve high classification precision for specific category and has limited practical applications. High-quality, feature-rich data ensure that artificial intelligence algorithm models perform well under various conditions and scenarios[9]. Thus, the acquired tactile features and the number of objects that can be correctly recognized directly determine the model's generalization ability.

To make robot tactile sensing more applicable to real-world use, Sundaram et al.[10] designed a low-cost piezoresistive tactile glove equipped with 548 pressure sensor units across the palm and phalanges. They developed the ResNet_3×3 network model and achieved a real-time recognition of 26 common everyday objects through grasping and touching actions. However, the top-1 precision was only 74.81%, which indicated that using only the pressure information for object recognition was insufficient. Qiu et al.[11] introduced a tactile glove that

integrated both pressure and temperature sensor array on the palm surface. This glove could simultaneously detect the contact pressure and thermal conductivity between the palm and objects. They developed a custom CNN model that recognized objects of varying weights and temperatures with a precision of 94.9%. The results demonstrated that multimodal information from the tactile glove could improve the recognition precision effectively. In addition to studies on tactile perception of pressure and temperature, Lu et al.[12] integrated 412 piezoresistive sensors into a robotic palm for the simultaneous measurement of the contact pressure between the palm and objects, as well as the overall bending state of each finger. A multimodal CNN was developed to recognize 17 target objects with an average precision of 91.67%. However, the recognition precision significantly decreased to 72.33% when only the data from the entire finger was used for recognition. These results demonstrated that adopting multimodal tactile information and developing high performance models were the effective ways to improve the object recognition precision.

To address the issue of low recognition precision when using only the bending data of a single finger, we presented a low-cost tactile glove with straightforward procedure capable of simultaneously capturing information of contact pressure distribution on the palm and bending strains on every finger joint. Additionally, a CNN was developed to integrate pressure and strain data for high-precision object recognition. The main contributions are as follows.

1) We designed a low-cost, piezoresistive tactile glove that covered the entire palm, integrating an array of joint bending strain sensors to capture the state of each finger joint and a palm pressure sensor array to measure the distribution of contact pressure. The glove collected tactile data from 16 target objects, and three tactile image datasets were built by palm pressure data, joint bending strain data, and a tactile data combing of both palm pressure and finger joint bending strain.

2) A modified ResNet model, SP-ResNet, was developed and used to train and test tactile image data for the 16 objects. The recognition precision can achieve 95.50% by processing both the palm pressure and joint bending strain information with the powerful SP-ResNet model. This demonstrated the obvious advantage over using single-source data, either joint bending strain or pressure information alone.

1 Design of tactile glove

This section primarily discusses the design of the sensor array, manufacturing processes, and principles of the signal acquisition circuit of the tactile glove. The main performance parameters of the two type sensors have been obtained by calibrating the palm pressure sensor arrays and finger joint bending strain sensor arrays. We tested the stability of sensor performance during grasping to ensure steady acquisition of tactile information from the designed glove.

1.1 Layout of tactile sensor array

To obtain tactile information from palm pressure and finger bending strain, the designed tactile glove was covered with a pressure sensor array on palm and a strain sensor array on back of five fingers. The sensor was arranged in a matrix form on the glove, with each unit consisting of two conductive electrodes on the top and bottom, and a sensitive material sandwiched in the middle. The pressure sensor array consists of 243 units arranged on the palm, as shown in Fig.1(a) and 1(c). Specifically, the thumb metacarpophalangeal joint is covered with a 3×4 sensing array, while other phalange joints are covered with sensing array sized at 3×3, and a 12×9 sensing array is placed across the entire palm. Similarly, the strain sensor array consists of 126 units with a 15×9 configuration, as shown in Fig.1(b) and 1(d). All the strain sensor array, with a dimension of 3×3, are arranged on the back of each finger joint.

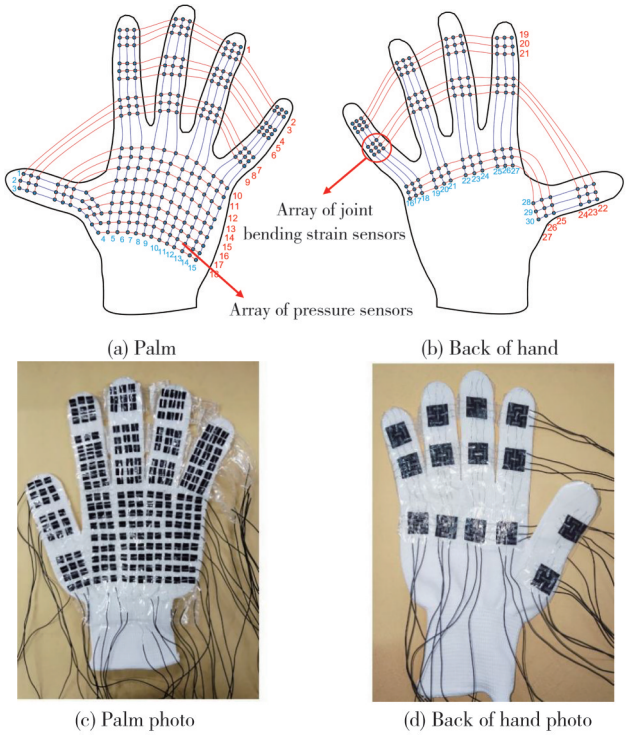


Fig. 1 Layout of tactile glove

All the sensing units in both of the pressure and strain sensor array are based on the piezoresistive effect of the

Velostat. To ensure the relative position of the sensing units in the array to the fingers, a strong acrylic adhesive was used to bond the sensor array to the glove. Each wire was arranged in a specific position to prevent short circuits between wires. When an object was grasped by this tactile glove, distinct pressure distribution and bending strain were formed according to variations in the shape, size, weight, and softness of each object. Therefore, pressure and bending strain tactile information from different sensing units of the glove can be extracted and then the accurate object recognition can be achieved by synergistic utilization of both of them.

1.2 Structural design of tactile sensor array

The performance of the tactile sensor array directly determines the precision of the extracted tactile information, thereby affects the precision of object recognition. A well-designed structure of the tactile sensor array ensures its measurement precision and working life. As shown in Fig. 2, the designed tactile sensor array consists of four parts: a protective layer, an adhesive layer, an electrode layer, and a layer of Velostat sensitive material.

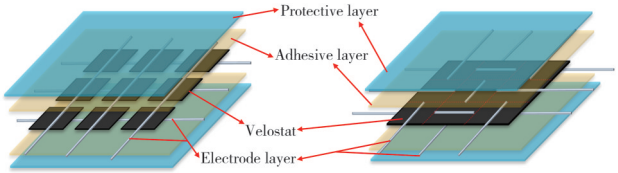


Fig. 2 Structure of sensor array

The protective layer was used to shield the sensitive layer from external damages and impacts, claiming for good mechanical properties. A piece of PE film with a thickness of 0.01 mm is adopted for the protective layer because of its high tensile strength, good flexibility, strong heat resistance, and low cost.

The electrode layer, delivering the measurement signals, was arranged in the configuration of orthogonal row and column wires. A stainless steel wire with a diameter of 0.2 mm, offering high flexibility, conductivity, and oxidation resistance for long-term use, was used for the electrodes.

The Velostat film has been widely used in the fabrication of flexible sensors due to its highly flexible and stable piezoresistive effect[13]. The imprecision in measurement results caused by the diffusion of force to nearby sensors due to the deformation of the material when pressure is applied to a sensor is termed mechanical crosstalk[14]. To minimize the influence of mechanical interference on measurement pressure precision, the Velostat was cut into squares with 5 mm × 5 mm and arranged in a 3×3 array, ensuring that there were no material connections between each units. All sensing materials were sandwiched between two orthogonally configured electrodes, thus forming a sensing array, as illustrated in Fig.2(a). In this configuration, the resistance change of each sensing unit directly reflects fluctuations in palm pressure.

The bending strain sensor array determines the bending state of each finger through changes in resistance caused by tensile and compressive deformation during bending. Essentially, the bending of the wires and the remaining layers exerts pressure on the Velostat, causing it to undergo a piezoresistive effect that results in a change in resistance. To accurately ascertain the bending state of a single joint, the sensors need to be precisely arranged at the bending points of the joints. When the finger bends, the positions of the sensing units in Fig. 2(a) shift, causing the gaps between units to align with the joint's bending point, thereby leading to measurement errors. Therefore, the bending strain sensor is configured as a 3×3 array with one piece of sensing material, effectively fixing the sensitive material to complete accurate strain measurements, as depicted in Fig.2(b).

Additionally, the structure in Fig. 2(b) features six interwoven conductive wires that apply pressure on the Velostat, enabling effective measurement of the joint's bending state. The bending state of each finger is used to enhance recognition precision through synergy with the pressure information. In the measurement of the joint's bending state with the employed 3×3 sensor array, relying on the result from any single unit alone is insufficient for comprehensive measurement of the joint's bending state. Consequently, the averaging processing is applied to ensure that each joint yields only one data point, thereby eliminating the cross-talk between units that could affect the measurement results.

1.3 Tactile data acquisition circuit

The adopted structure and arrangement of the sensor array form a matrix of variable resistors at the intersection points of the row and column electrodes. When the sensor array is interfaced with a two-dimensional array scanning method to read out the resistance of each unit, unintended crosstalk currents appear and deteriorate the precision of the measurement[14]. To read the resistance values of each sensitive unit within the resistance matrix accurately, it is necessary to suppress the crosstalk currents among the sensing array[15-19]. We implemented a scanning circuit based on the virtual grounding method to achieve the signal

acquisition circuit and suppress crosstalk, as shown in Fig.3[20]. The scanning circuit primarily consists of a 32-bit analog multiplexer (4 74HCT4051 chips), a 32-bit shift register (474HCT595 chips), and 32 operational amplifiers (LM324 chips)

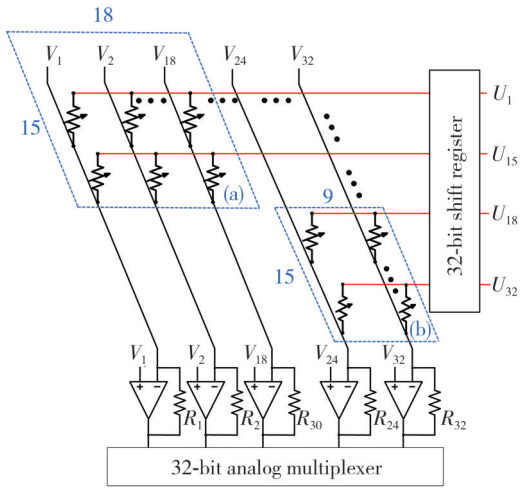


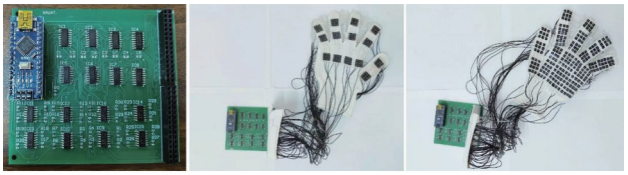
Fig. 3 Scanning circuit

We used the scanning circuit to extend the dimensions of the sampled tactile information from 30×27 to 32×32 to facilitate the convolution operations of the neural network. As shown in Fig.3, the 32-bit analog multiplexer integrates voltage nodes ($V_1=V_2=\cdots=V_{32}=5$ V) and resistors ($R_1=R_2=\cdots=R_{32}=10$ kΩ). Black dots (···) between labeled components indicate omitted intermediate elements to streamline the schematic. Similarly, the shift register truncates repetitive units ($U_1=U_2=\cdots=U_{32}=5$ V), retaining only key positions for clarity. Area (a) (rows 1—15, columns 1—18) and area (b) (rows 18—32, columns 24—32) represent the scanning distribution of the palm pressure and joint bending strain sensor array, respectively. Areas other than (a) and (b) are not accessed for scanning and are set to null.

We could input the voltage output of each operational amplifier to the analog-to-digital conversion ports of an Arduino nano, where it was converted into digital signals. The converted digital signals are then transmitted to a computer via serial communication for further processing. Specifically, we converted the tactile information into a corresponding conductance matrix with a size of 32×32.

The circuit diagram is fabricated as a PCB for tactile information acquisition, as shown in Fig. 4(a). The hardware photos of the tactile glove prototype are shown in Fig.4(b) and (c).

The conductance adopted in this paper is a calculation result based on a voltage divider circuit and the ADC IC integrated in a Arduino PCB. When a row of the scanning circuit is grounded by a shift register, the conducting circuit for the measured resistor is shown in Fig.5.



(a) PCB (b) Back of glove (c) Palm of glove

Fig. 4 PCB design and fabrication of signal sampling circuit

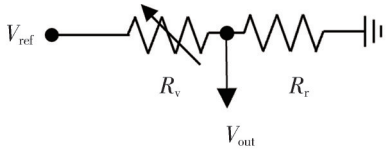


Fig. 5 Conducting circuit for measured resistor

The relationship between the measured resistance and the output voltage is

$$V_{out}=V_{ref}\frac{R_r}{R_v+R_r}, \tag{1}$$

where V_{out} is the output voltage; V_{ref} is the reference voltage (which is 5 V when there is no crosstalk); R_v is the measured resistance, and R_r is the reference resistance (set to 1 kΩ). After being sampled by the 10-bit ADC of Arduino, the V_{out} is converted into a digital value. The 10-bit sampled voltage value, V_{adc}, can be expressed as

$$V_{adc}=1\,024\times\frac{V_{out}}{V_{ref}}. \tag{2}$$

Through substitution of the equations, we obtain

$$R_v=R_r\frac{1\,024-V_{adc}}{V_{adc}}. \tag{3}$$

The conductance value $G_{velostat}$ of the sensor matrix can be calculated by

$$G_{velostat}=\frac{1}{R_v}. \tag{4}$$

The unit of the calculated parameter is 1/Ω. By the context, the conductance in this paper is not the strict physical definition, but it still has the form and meaning of the conductance as well as the unit. The same method has been employed[21].

1.4 Sensor calibration

To obtain accurate tactile signals, the pressure and bending strain sensor array should be calibrated properly before the real measurement begins. The neural networks make the best of the varieties of pressure and bending strain features from sensing units in the process of grasping an object to realize the object recognition. Therefore, the influence of differences in the initial state that are induced by the fabrication and materials can be excluded which is conductive to improving recognition precision.

1) Calibration of pressure sensor array units

As shown in Fig. 6, the pressure sensor calibration setup includes a force gauge (Zhiqu HQ-20B) and a digital multimeter (Keysight 34465A). The force gauge is used to apply specific pressure on the sensor units, and the digital multimeter records the changes in resistance with pressure. The measured resistance-pressure curve and the corresponding conductance-pressure curve are shown in Fig. 7. The sample sensing unit performs well under loads ranging from 0 N to 0.6 N.

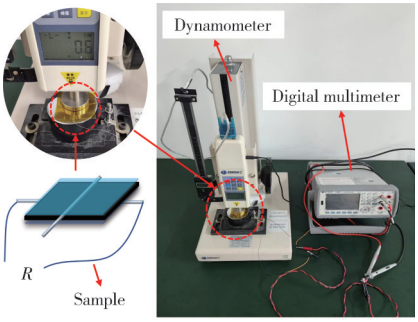
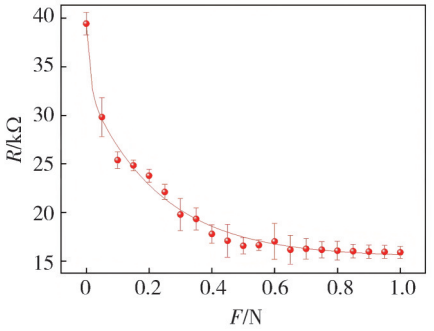
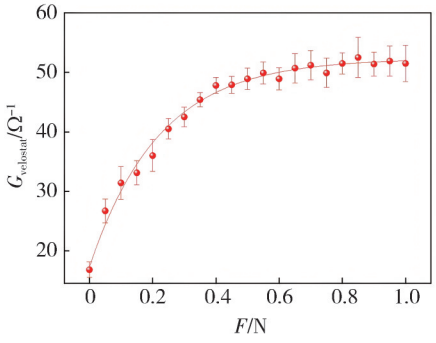


Fig. 6 Pressure sensor calibration setup



(a) Resistance-pressure



(b) Conductance-pressure

Fig. 7 Mechanical properties of sample sensing unit

For the conductance-pressure relationship of the palm pressure unit, we conduct a test over 1 000 s, during which approximately 600 cycles of 0—1 N pressure are applied, as shown in Fig. 8, indicating that the palm pressure unit exhibits good repeatability.

2) Bending performance characteristics

When a target object is grasped by the hand, every finger joint produces different bending deformations and corresponding compressions[22].

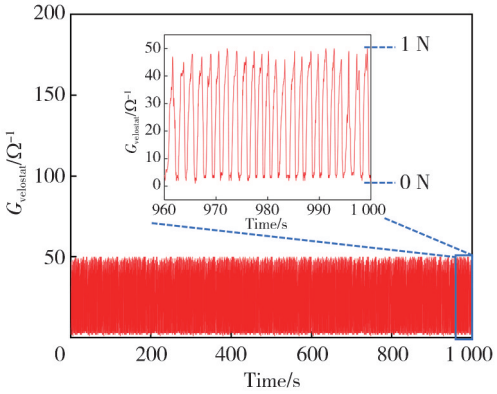
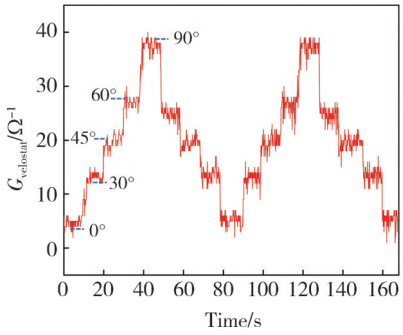


Fig. 8 Repeatability test of palm pressure units

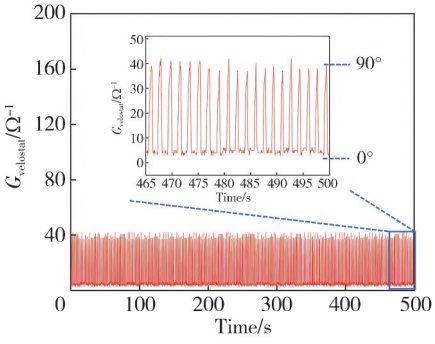
Therefore, a strain sensor array is better qualified for finger joint bending degree presentation than an individual sensing unit in the array, as described in the section of tactile sensor array structural design. To achieve a good balance between measurement precision and computational complexity, the average of conductance value of the all nine units in the array is adopted to indicate the bending degree of the finger joint. The average value can be calculated by

$$G=\frac{1}{9}\sum_{i=1}^{3}\sum_{j=1}^{3}a_{ij}, \tag{5}$$

where a_{ij} represents the conductance value of each sensing unit in a matrix of 3×3 bending strain sensor array; G represents the average of the conductance matrix of the 9 units; and i and j are the row number and the column number, respectively.



(a) Change with bending angle



(b) Repeatability of joint

Fig. 9 Joint bending strain sensor measurement

The joint angles were tested at 0°, 30°, 45°, 60°, and 90°, with a testing duration of 160 s. As shown in Fig.9(a), the bending strain sensors exhibit good repeatability. No significant noise could be observed in the collected signals, and the sensing array response to the bending angle was stable and consistent. Fig.9(b) shows approximately 300 continuous bending cycles from 0° to 90° within 500 s.

All these results demonstrated that the designed bending strain sensor array and the sampling method possessed good stability and high sensitivity.

2 Methods for tactile recognition

In order to verify the influence of different tactile information on the recognition precision of the target object, we built three distinct tactile data sets: one using palm pressure data, one using joint bending strain data, and one combining both palm pressure and finger joint bending strain data.

2.1 Data preprocessing

Different CNNs have been proven to be effective and advantageous in image recognition[23]. Therefore, the multi-frame 32×32 conductance matrices were converted into responding grayscale image, where each element of the matrix corresponds to a pixel in the grayscale image, and the depth of the pixel represents the magnitude of the element value. All the values of elements are restricted in the range of [0, 255], after the linear normalization of the conductance matrix, can be expressed as

$$X_{\text{norm}} = 255\frac{C - \min C}{\max C - \min C}, \tag{6}$$

where X_{norm} refers to the normalized matrix; C denotes the conductance matrix; and min C and max C represent the minimum and maximum values of the elements within the conductance matrix, respectively. After that these grayscale images were processed by the CNN to realize object recognitions.

Once the tactile glove with good sensing performance and an effective signal acquiring method have been implemented, we investigate the tactile sensing capability of the designed system through activities in daily living (e.g., when grasping a smartphone), as shown in Fig. 10(a). We can see that both of the magnitude and location of the bending strain and the pressure demonstrate the grasping action and shape of the object intuitively from Fig.10. The contact pressure distribution on the palm and bending strain at every finger joint are indicated with the grayscale image highlighted in blue and red, respectively.

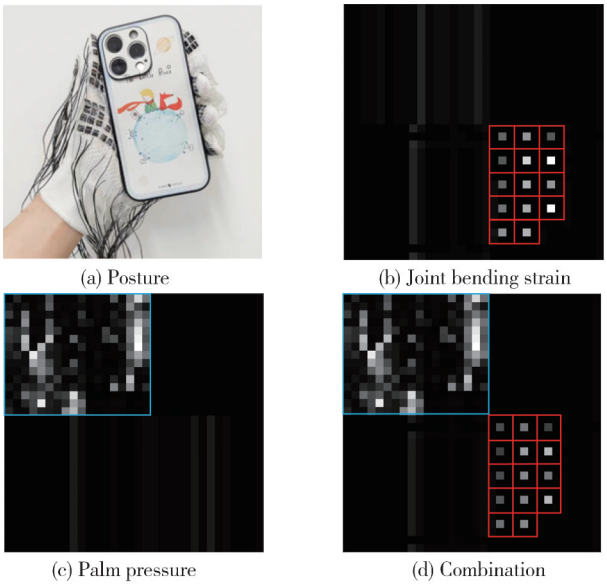


Fig. 10 Tactile images

2.2 Data augmentation

Data augmentation has been usually adopted to increase the diversity of training data and consequently improve the capabilities of feature extraction and generalization of the model[24]. In order to extract more features from the collected two-dimensional tactile grayscale images and improve recognition precision, the following data augmentation techniques were employed.

1) Each tactile image was randomly flipped vertically or horizontally with a probability of 30%.

2) Random noise with Gaussian distribution ($\mu=0$, $\sigma=0.02$) was added to each tactile image.

3) Each tactile image was randomly altered (translated, rotated, or scaled) with a probability of 30%.

2.3 SP-ResNet

We represented the acquired tactile information as single-channel grayscale images with a size of 32×32. However, the most of CNN models are commonly designed for three-channel RGB images, thus, they are not suitable for processing the acquired tactile images directly. Additionally, single-channel images have advantages in structure simplicity and low computation cost to extract features. Reasonable design of CNNs models matching the acquired tactile images will increase the recognition efficiency. Consequently, a novel CNN model, SP-ResNet, was successfully developed based on the architecture of ResNet-18[25].

As shown in Fig.11(a), the structure of the developed SP-ResNet model consists of eight layers. One of 3×3 convolution layer, max pooling layer, four layers of SeparableBlocks, and average pooling layer are arranged

in sequence and employed to extract features from a single grayscale tactile image. A fully connected layer for classification is used for the ending conclusion. To facilitate processing the collected grayscale tactile images, the adjustments of neural network configuration have been carried out. Specifically, the size of filter in the initial convolution layer and the step size are optimized to 3×3 and 1, respectively, which facilitates the convolution operations and reduces the operation calculations. The convolution layer is followed by a batch normalization layer and a nonlinear activation function, i.e., H-swish[26]. H-swish is employed to active every convolution layer to replace the activation function RELU adopted in ResNet-18. H-swish is a lightweight of the swish function that utilizes a piecewise function to approximate the swish function, reducing the number of multiplications and the number of memory accesses during function computation. As a result, the computational complexity of the model can be decreased effectively. With this configuration, a 1×32×32 grayscale tactile image is input into the SP-ResNet and transformed into a 1 000×1×1 classification vector after being sequentially processed by every layer.

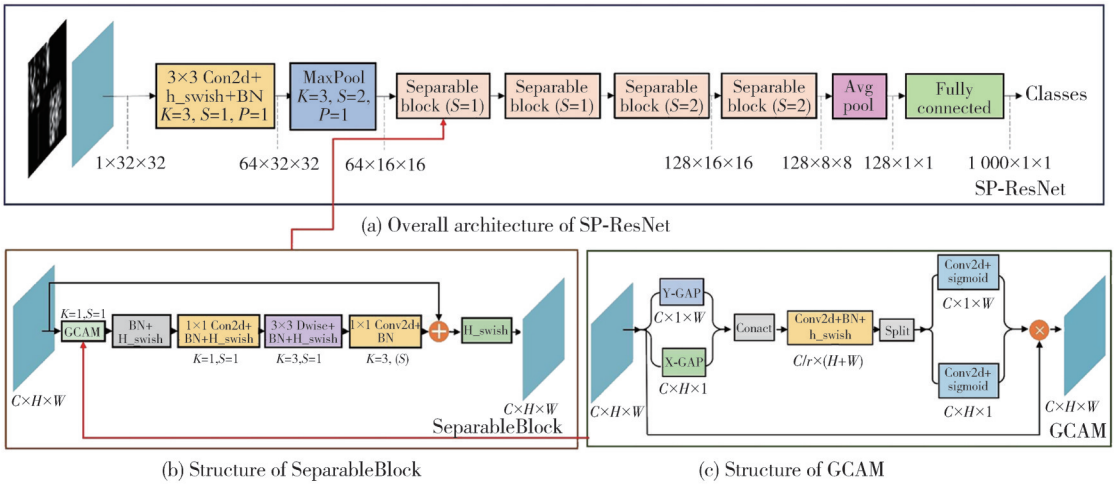


Fig. 11 Architecture of SP-ResNet developed for object recognition by tactile information

1) SeparableBlock

As shown in Fig.11 (b), each SeparableBlock contains one GCAM layer, one depth-wise separable convolution layer, and one point-wise convolution layer[27]. A point-wise convolution layer is connected to the end of GCAM layer to deepen the network structure. Point-wise convolution layer with fewer parameters demands lower computational costs, which can improve the representation capabilities with no significant increasing in the computational burden.

Standard convolutions can efficiently capture combinational features across different channels. However, they are not suitable for processing the grayscale images with only one channel. Therefore, a more lightweight depth-wise separable convolution, which decomposes the standard convolution into depth-wise and point-wise convolutions, is introduced to process the grayscale tactile images efficiently. In depth-wise convolutions, the number of input and output channels is equal and each filter operates solely on an individual input channel, avoiding the learning of cross-channel combinational features. The primary purpose of the point-wise convolution layer is to deepen the network layers while performing a dimension transformation in the second residual block. The purpose of deepening the network structure is to capture more intricate patterns and hierarchical features within the data[25]. Subsequently, through global average pooling and a fully connected layer, the feature map is transformed into a channel-dimensional classification vector.

2) GCAM

To exact precision tactile information from the single-channel grayscale images, the neural network should focus on the spatial information of the image features instead of inter channel information. The global coordinate attention mechanism (GCAM) possesses significant advantages in capturing the directional and positional feature information of images[28]. As shown in Fig.11 (c), it firstly performs global pooling along the spatial dimensions of height and width of the feature map (Y-GAP and X-GAP), and produces two 1-dimensional features. Then the features are processed through a convolutional layer that reduces the number of channels, followed by another that increases them. After being activated by a sigmoid function, two sets of attention maps that respectively correspond to spatial positional information along the height and width directions are generated. After that attention maps are element-wise

multiplied by the tensor of the original feature map. By aggregating features across different spatial directions, the long-range dependencies of each pixel in the grayscale image can be more effectively captured. Finally, the attention-weighted feature maps are merged along the channel dimension to produce the final output feature map. This mechanism is integrated into the structure of each block, allowing the model to locate features more accurately in tactile grayscale images.

The entire network model was implemented in the PyTorch framework, with an initial learning rate of 10^{-3}. The learning rate was optimized using the Adam optimizer, and the training period was set to 50 epochs. C represents the number of channels, H is the height, W is the width, S is the stride, K is the filter size, and P is the padding. In Fig. 11(b), (S) indicates the variable stride, with each stride value provided in Fig.11(a).

3 Experiments

To validate the effectiveness of the designed tactile glove and recognition method, several experiments were conducted on the object recognition. After the acquisition of tactile images of target objects, three tactile data sets were built according to different tactile information. Then the object recognition experiments and comparative experiments were performed on the data sets implemented by SP-ResNet and five other models.

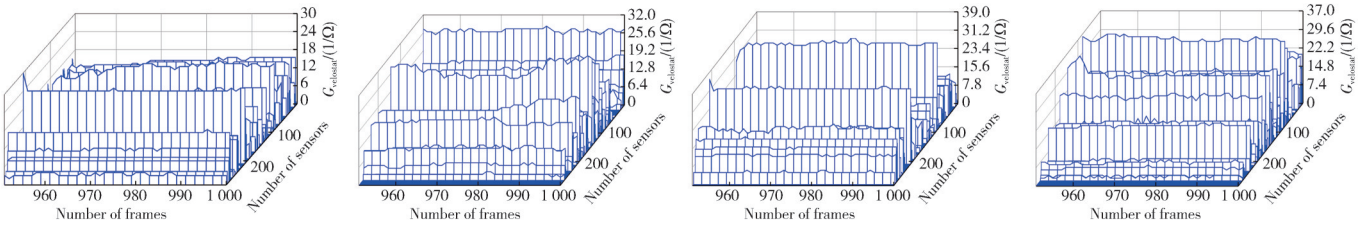
3.1 Creating tactile dataset

Before realizing accurate object recognition, the developed CNN model had to be well trained with the tactile information collected from the designed tactile glove. As shown in Fig.12, we collect tactile information from the designed tactile glove when it grasps 16 target objects. As mentioned before, palm pressure data, joint bending strain data, and a tactile data combining both palm pressure and finger joint bending strain were used to build the corresponding tactile dataset.

The objects used in the experiment had different shapes and weights. For example, an empty water bottle and a bottle full of water are similar in appearance but different in weight. While a toy and a roll of paper are the same weight but different in shape. Due to the differences in the physical properties of different objects, the tactile information collected by the tactile glove is distinctly different. Representative strategies are built for training the CNN.

Fig. 12 16 Target objects

Fig.13 shows the tactile data of the last 50 continuous time frames collected by the palm pressure and joint bending strain sensors while gripping 16 objects. This includes 270 data points from the palm pressure sensor and 14 data points from the joint bending strain sensors. After 950 frames of data collection for each object, the tactile information captured by the palm pressure and joint bending strain sensors varies, due to differences in the contours and weights of the objects. After data acquisition of 950 frames, the glove is still able to collect data normally, indicating that the tactile glove has good mechanical stability.



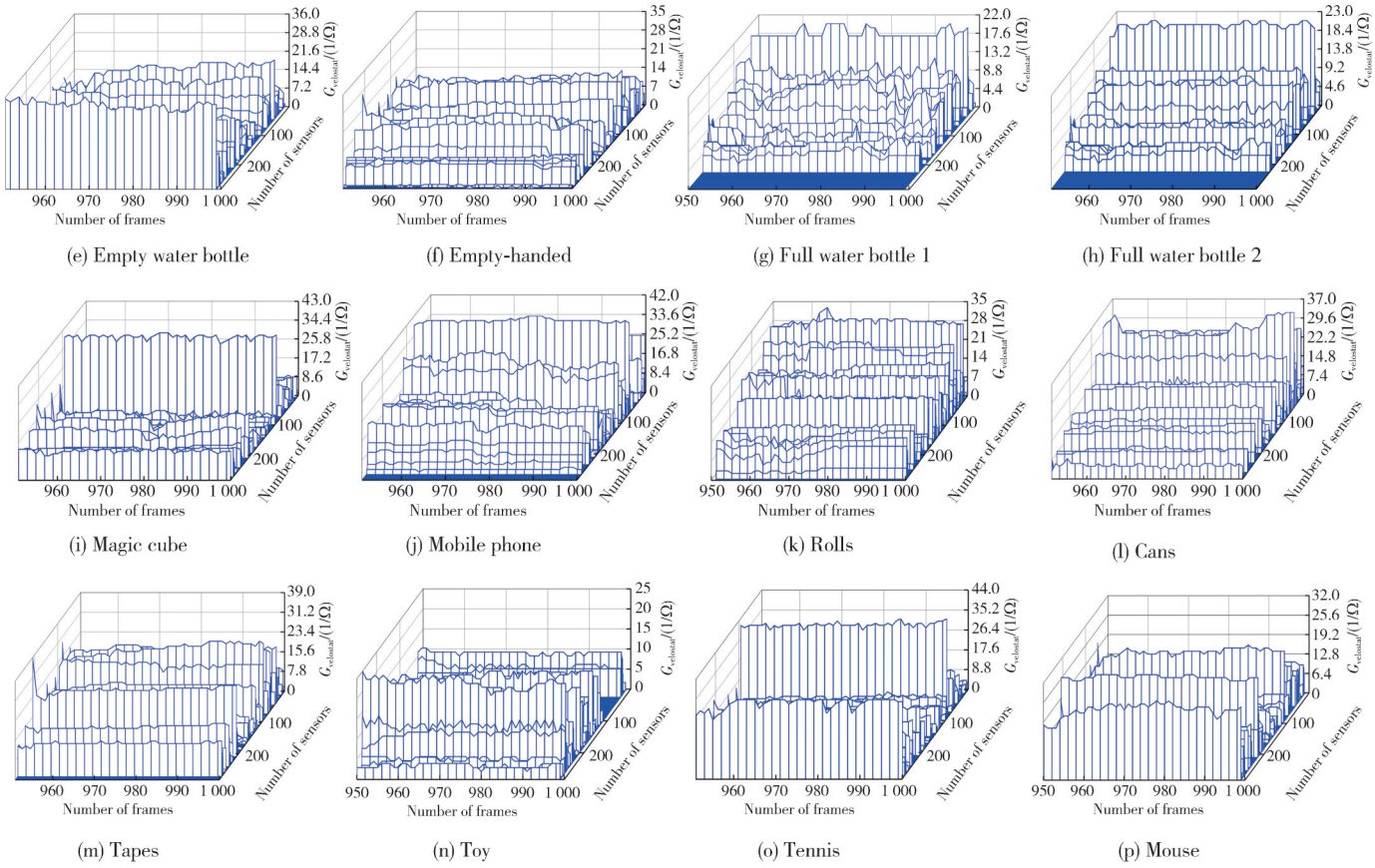


Fig. 13 Tactile sequences of 16 objects (50 consecutive frames of 284 sensor values)

3.2 Tactile object recognition

The three tactile data sets are separately adopted to train the SP-ResNet, and the corresponding results are represented by confusion matrix, as shown in Fig. 14. We can see that recognition precision of 91.17% and 88.39% is achieved by utilizing the palm pressure and bending strain tactile information, respectively. However, the precision is dramatically increased to 95.50% when tactile data combining both palm pressure and finger joint bending strain is involved together. Specifically, the palm pressure sensor array is not good at distinguishing objects with similar weight. As a result, it fails to distinguish the rolls of paper and toys. Similarly, the joint bending strain sensor array lacked the ability to distinguish objects of similar shape, for example, the two bottles. Consequently, integrating both joint bending strain and pressure tactile information effectively compensated for their individual shortcomings and enhanced classification precision.

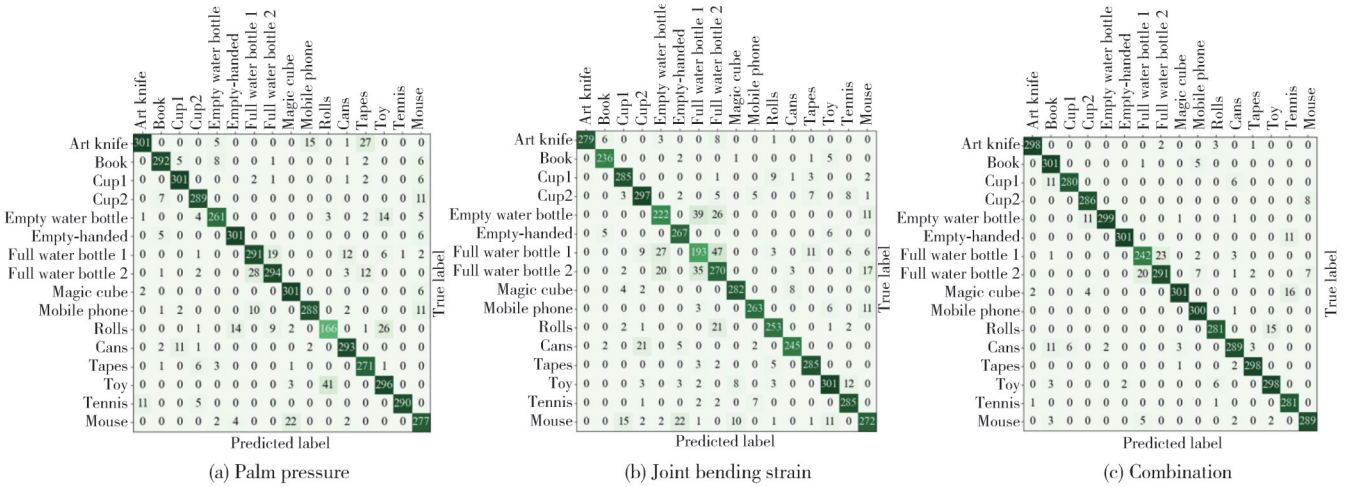


Fig. 14 Confusion matrix

3.3 Comparative experiment

To further demonstrate the advantage of the developed SP-ResNet, a publicly available tactile image data set (STAG) for 26 objects with different size, weight, and material was adopted to conduct the comparative experiment with the existing popular network models. The comparative result with ResNet_3×3 is shown in Fig. 15, which demonstrates that the recognition precision of the proposed SP-ResNet is consistently higher than that of the ResNet_3×3 across different numbers of input frames.

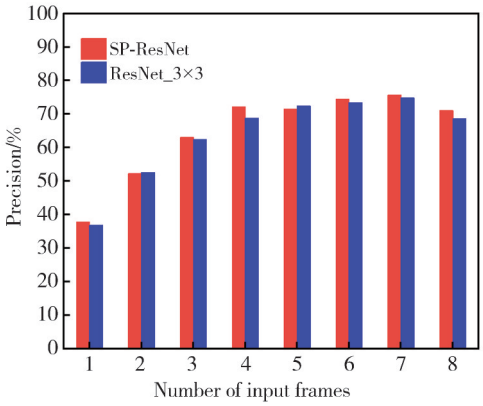


Fig. 15 Precision comparison of STAG dataset

This indicates that the proposed SP-ResNet has clear advantage for object recognition when grayscale images are employed as the tactile data. The recognition performance of the developed SP-ResNet and other popular network models are listed in Table 1.

The performances of different deep learning models have been evaluated through the following indicators: recognition precision, number of FLOPS[29], and number of parameter[30]. The designed SP-ResNet achieved the second-best recognition precision with the minimum computational cost (i. e., the number of operations and consumed memory resources). Compared to ResNet_3×3 with the same input image size of 32×32, SP-ResNet demonstrated lower FLOPs (16.332×10^6) and parameters (3.26×10^5), while maintaining a high precision of 95.50%. This good property made it more prominent in accurate object recognition. Additionally, to validate the effectiveness of incorporating GCAM, we conducted comparative tests after removing GCAM. The recognition precision of the combined tactile data decreases by 0.47%, while the model's FLOPs (12.433×10^6) and parameter (2.18×10^5) remain comparable to those before integration. These results indicate that SP-ResNet with GCAM is better suited for tactile image classification tasks.

Table 1 Comparison of performance between proposed model and five models

Model	Precision/%			Number of FLOPs/$\times10^6$	Number of parameter/$\times10^5$
	Palm pressure	Joint bending strain	Combination of pressure and strain		
Naive Bayes[31]	75.85	74.33	72.16	—	—
Decision Tree[32]	81.77	84.55	89.22	—	—
ResNet_18[25]	91.98	89.51	96.02	1 824	116.90
MobileNet[33]	91.78	89.03	95.48	588.912	42.32
ResNet_3×3[10]	91.04	87.12	94.54	72.581	6.58
SP-ResNet(deleted GCAM)	90.85	88.17	95.03	12.433	2.18
SP-ResNet(ours)	91.17	88.39	95.50	16.332	3.26

To demonstrate the effectiveness of the proposed data augmentation method in terms of recognition performance, we conducted comparative experiments. The performance of data augmentation methods A, B, and C, along with no data augmentation, is evaluated through 10 independent experimental runs for each condition, and the results are summarized in Table 2. As observed, when all data augmentation methods are applied simultaneously, the recognition precision is improved by 1.93% for palm pressure, 0.14% for joint bending strain, and 1.29% for their combination. But using only method B does not improve recognition precision. Because our grayscale images are based on data collected from sensors, which inherently contains some noise.

Table 2 precision comparison experiment of data augmentation

Method	Precision/%		
	Palm pressure	Joint bending strain	Combination of pressure and strain
No data augmentation	89.24	88.25	94.21
Only A	89.35	88.28	94.68
Only B	89.24	88.25	94.22
Only C	91.05	88.33	95.02
All	91.17	88.39	95.50

While adding noise could increase the diversity of the data, excessive noise might have led to the loss of feature information in the images, thereby preventing any improvement in precision. For methods A and C, the improvement was likely due to the nature of our data, which consisted of single-channel grayscale images. The features in the images were not symmetrically arranged, and operations such as flipping

or translation caused changes in the positions of pixels. This allowed the model to learn more complex features for different postures. Additionally, both methods increased the number of tactile images by 30%, augmenting the input data and thereby enhancing the model's generalization ability and recognition precision.

4 Discussion and future work

In practical industrial applications, multiple sensors are often used together, generating multimodal data. For such data, it is possible to convert different types of data into a unified single-channel image format to meet the input requirements of SP-ResNet, allowing the model to perform classification tasks for various applications. However, this process may have resulted in the loss of certain sensor-specific features, such as the temporal features from sound. In the future, one of our research directions will be to combine the tactile glove with vision or hearing to acquire multimodal data. For SP-ResNet, we plan to add multi-channel convolution layers and feature fusion layers to independently extract features from each type of data. This approach will enhance the model's generalization ability, ultimately expanding the industrial applications of tactile gloves.

High-resolution and multi-channel data often contained more features of objects and richer information, which allowed neural networks to perform better in recognition tasks compared to data with insufficient features. The design of SP-ResNet aimed for lightweight processing, and it was limited to handling only 32×32 grayscale images. To accommodate higher-resolution data, SP-ResNet could be improved by modifying the convolution kernel size, adding more convolution layers, or incorporating attention mechanisms to recognize finer object features. High-dimensional data could be converted into multi-channel tensors, for example, by placing the time vectors of tactile data into separate channels for feature extraction. One of the future directions of our work involve associating time-series data with the palm pressure and joint bending strain sensor array, further capturing the changes in tactile features between the glove and objects at different time points. We focus on constructing 3D tactile information with time vectors, enabling the neural network to better understand the physical properties of objects through the tactile glove.

5 Conclusions

This paper presented a low-cost object recognition system that consisted of a tactile glove, a signal acquisition circuit, and a deep learning-based CNN model. The tactile glove featured the integration of both a palm pressure sensor array and a joint bending strain sensor array to simultaneously perceive the contact pressure and bending strain of fingers during object grasping. The cooperation of the pressure and bending strain tactile information facilitated the tactile glove with enhanced recognition precision. The signal sampling circuit was designed based on the virtual grounding principle to suppress crosstalk among the two-dimensional resistive array and collect high-precision tactile images. The SP-ResNet model, optimized for single-channel grayscale tactile images, was developed for object recognition based on tactile information. Through a series of object recognition experiments, the presented object recognition system succeeded in achieving a recognition precision of 95.50% by combining both bending and pressure information. The developed tactile system with excellent sensing abilities and high signal processing capabilities demonstrated its potential applications in various fields, such as human-machine interaction and intelligent prosthetics.

Acknowledgement

This work was supported by the Key Research and Development Program of Shaanxi Province (No. 2024 GX-YBXM-178), and the Shaanxi Province Qinchuangyuan "Scientists+Engineers" Team Development (No.2022KXJ032)

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

[1] ZHU M L, HE T, LEE C K. Technologies toward next generation human machine interfaces: from machine learning enhanced tactile sensing to neuromorphic sensory systems. Applied Physics Reviews, 2020, 7(3): 031305.

[2] KOSCH T, KAROLUS J, ZAGERMANN J, et al. A survey on measuring cognitive workload in human-computer interaction. ACM Computing Surveys, 2023, 55 (13s): 1-39.

[3] XIONG Y X, SHEN Y K, TIAN L, et al. A flexible, ultra-highly sensitive and stable capacitive pressure sensor with convex microarrays for motion and health monitoring. Nano Energy, 2020, 70: 104436.

[4] CHEN J, LI L, ZHU Z, et al. Bioinspired design of highly sensitive flexible tactile sensors for wearable healthcare

monitoring. Materials Today Chemistry, 2022, 23: 100718.
[5] SONG Z, FAN X Y, DONG J Y, et al. The third-person perspective full-body illusion induced by visual-tactile stimulation in virtual reality for stroke patients. Consciousness and Cognition, 2023, 115: 103578.
[6] LIAO X Q, SONG W T, ZHANG X Y, et al. Hetero-contact microstructure to program discerning tactile interactions for virtual reality. Nano Energy, 2019, 60: 127-136.
[7] LEE K T, CHEE P S, LIM E H, et al. Artificial intelligence (AI) -driven smart glove for object recognition application. Materials Today: Proceedings, 2022, 64: 1563-1568.
[8] ZHANG X X, LI S B, YANG J, et al. Tactile perception object recognition based on an improved support vector machine. Micromachines, 2022, 13(9): 1538.
[9] HE T X, YU S C, WANG Z Y, et al. From data quality to model quality: an exploratory study on deep learning//The 11th Asia-Pacific Symposium on Internetware, September 17-18, 2019, Fukuoka, Japan. New York: ACM, 2019: 1-6.
[10] SUNDARAM S, KELLNHOFER P, LI Y Z, et al. Learning the signatures of the human grasp using a scalable tactile glove. Nature, 2019, 569(7758): 698-702.
[11] QIU Y, WANG Z Q, ZHU P C, et al. A multisensory-feedback tactile glove with dense coverage of sensing arrays for object recognition. Chemical Engineering Journal, 2023, 455: 140890.
[12] LU X, SUN D, YIN H B, et al. 3-D tactile-based object recognition for robot hands using force-sensitive and bend sensor arrays. IEEE Transactions on Cognitive and Developmental Systems, 2023, 15(4): 1645-1655.
[13] DZEDZICKIS A, SUTINYS E, BUCINSKAS V, et al. Polyethylene-carbon composite (velostat®) based tactile sensor. Polymers, 2020, 12(12): 2905.
[14] CHEN S J, LI M, HUANG Y K, et al. Matrix-addressed flexible capacitive pressure sensor with suppressed crosstalk for artificial electronic skin. IEEE Transactions on Electron Devices, 2020, 67(7): 2940-2944.
[15] WU J F. Scanning approaches of 2-D resistive sensor arrays: a review. IEEE Sensors Journal, 2017, 17(4): 914-925.
[16] PRUTCHI D, ARCAN M. Dynamic contact stress analysis using a compliant sensor array. Measurement, 1993, 11(3): 197-210.
[17] TAKEI K, TAKAHASHI T, HO J C, et al. Nanowire active-matrix circuitry for low-voltage macroscale artificial skin. Nature Materials, 2010, 9(10): 821-826.
[18] VIDAL-VERDÚ F, OBALLE-PEINADO Ó, SÁNCHEZ-DURÁN J A, et al. Three realizations and comparison of hardware for piezoresistive tactile sensors. Sensors, 2011, 11(3): 3249-3266.
[19] DEMKOV Y N, OSTROVSKII V N. Zero-range potentials and their applications in atomic physics. Berlin: Springer Science & Business Media, 2013.
[20] SAXENA R S, BHAN R K, SAINI N K, et al. Virtual ground technique for crosstalk suppression in networked resistive sensors. IEEE Sensors Journal, 2011, 11(2): 432-433.
[21] SUPRAPTO S S, SETIAWAN A W, ZAKARIA H, et al. Low-cost pressure sensor matrix using velostat//2017 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering, November 6-7, 2017, Bandung, Indonesia. New York: IEEE, 2017: 137-140.
[22] LIN W K, WANG B, PENG G X, et al. Skin-inspired piezoelectric tactile sensor array with crosstalk-free row+column electrodes for spatiotemporally distinguishing diverse stimuli. Advanced Science, 2021, 8(3): 2002817.
[23] LI Z W, LIU F, YANG W J, et al. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(12): 6999-7019.
[24] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning. Journal of Big Data, 2019, 6(1): 60.
[25] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
[26] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3////2019 IEEE/CVF International Conference on Computer Vision, October 27-November 2, 2019, Seoul, Korea. New York: IEEE, 2019: 1314-1324.
[27] XIE C, ZHU H Y, FEI Y Q. Deep coordinate attention network for single image super-resolution. IET Image Processing, 2022, 16(1): 273-284.
[28] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 20-25, 2021, Nashville, TN, USA. New York: IEEE, 2021: 13708-13717.
[29] CHEN J R, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 17-24, 2023, Vancouver, BC, Canada. New York: IEEE, 2023: 12021-12031.
[30] DENIL M, SHAKIBI B, DINH L, et al. Predicting parameters in deep learning//27th International Conference on Neural Information Processing Systems, December 5-8, 2013, Lake Tahoe, Nevada, USA, New York: Curran Associates Inc, 2013: 2148-2156.
[31] KINGMA D P, WELLING M. Auto-encoding variational bayes. 2013. https://doi. org. /10. 48550/arXiv. 1312. 6114.
[32] CHARBUTY B, ABDULAZEEZ A. Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2021, 2(1): 20-28.
[33] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets:

efficient convolutional neural networks for mobile vision applications. 2017. https://doi.org./10.48550/arXiv.1704.04861

一种基于掌压和关节弯曲应变感知的物体识别触觉手套

张学锋*, 张少杰, 陈鑫, 张锦华

西安建筑科技大学 机电工程学院, 陕西 西安 710055

摘　要： 为了使得实现物体识别的模拟触觉系统越来越精细化和智能化，本文设计了一种用于物体识别的触觉手套。该手套集成了243个手掌压力敏感单元和126个关节应变敏感单元，传感器敏感单元利用Velostat材料的压阻效应实现手掌压力和手指关节弯曲应变的测量。为了实现物体识别，使用二维阵列扫描法采集触觉数据，并将其转换成分辨率为32×32的灰度图。通过三种数据采集的方法（仅用手掌压力、仅用关节弯曲应变以及协同使用手掌压力和关节弯曲应变），分别建立了三种触觉图像数据集。针对触觉图像分辨率的尺寸特点，基于ResNet-18（一种残差卷积神经网络）进行了轻量化改进，构建了SP-ResNet模型对触觉图像进行分类，实现了物体识别。对于所选取的16个目标物体的对比识别实验，SP-ResNet在保证高识别精度（95.50%）的同时占用的计算机资源更少。协同手掌压力和关节弯曲应变的数据采集方法相比于仅用手掌压力或关节弯曲应变方法的最佳识别效果，精度提升了4.33%。所提出的触觉手套及卷积神经网络（CNN）可以进一步提升机器人的智能化服务。

关键词： 触觉手套；物体识别；Velostat；关节弯曲应变传感器；手掌压力传感器；卷积神经网络

引用格式： ZHANG Xuefeng, ZHANG Shaojie, CHEN Xin, et al. A tactile glove for object recognition based on palmar pressure and joint bending strain sensing. Journal of Measurement Science and Instrumentation, 2025, 16(2): 173-185. DOI: 10.62756/jmsi.1674-8042.2025017