Proceedings of 12th IEEE International Conference
on CYBER Technology in Automation, Control, and Intelligent Systems
July 27 - 31, 2022, Changbai Mountain, China

2022 12th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER) | 978-1-6654-7267-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/CYBER55403.2022.9907378

Multi-kernel-size Convolutional Supervised Autoencoders for Tactile Gesture Recognition

Chaoxiang Ye[1], Xiaoyu Li[1], Binhua Huang[1], Yuanzhe Su[1],
Tingting Mi[1], Zhenning Zhou[1], Zhengkun Yi[1, *], Xinyu Wu[1, 2]

Abstract— **With the development of human-computer interaction, tactile gesture recognition has been widely used in our life. To solve the problem of overfitting on the sample-limited tactile datasets, we apply the Supervised Autoencoders (SAE) to improve the generalization performance. Moreover, based on the SAE, we propose the Multi-kernel-size Convolutional Supervised Autoencoders (MCSAE) to further improve the generalization performance on the limited dataset, which provides models with more structure of receptive fields and enhances the feature extraction ability of SAE. In comparison with other state-of-the-art (SOTA) models, the SAE we apply has higher gesture recognition accuracy and MCSAE can further improve the generalization performance of SAE on the sample-limited publicly available dataset.**

Index Terms— **Gesture Recognition, Tactile Perception, Deep Learning, Autoencoders, Generalization Performance.**

I. Introduction

Gestures have been a fundamental mode of communication since the far distant past. With the development of human-computer interaction, gesture recognition has been widely used in virtual reality [1], medical care [2], industry [3], and other fields. Gestures can not only bring players a real sense of experience in video games but also improve the quality of people's lives and bring forth convenience to the deaf. Compared with visual perception, gesture recognition based on tactile perception can avoid the interference of external factors. However, the acquisition cost of tactile data is high, which requires more time and effort to experiment [4]. Therefore, it will be a challenge to improve the generalization performance of recognition models on limited tactile datasets [5].

Currently, the algorithms applied to tactile gesture recognition mainly include traditional machine learning and deep learning methods. For the traditional machine learning method, Zhihao Zhou et al. [2] achieved 98.63% accuracy using Support Vector Machines (SVM) on a dataset containing 11 gestures categories collected by yarn-based stretchable sensor arrays. Shuo Jiang et al. [6] applied Linear Discriminant Analysis (LDA) to classify American Sign Language (ASL) numbers 0-9 on a dataset collected by resistive tactile sensors and showed 94.4% accuracy. The deep learning algorithms have excellent fitting performance. Subramanian Sundaram et al. [7] used a Convolutional Neural Network (CNN) with resnet blocks to classify 8 gestures on the collected sensor array-based dataset with 89.4% accuracy. Xinan Huang et al. [8] used a three-layer neural network to classify 10 ASL gestures representing numbers 0-10 with 98.5% accuracy on 20,000 samples collected from a homemade resistive tactile sensor.

As mentioned, we have witnessed great progress in the field of tactile gesture recognition by adopting deep neural networks. However, the recent advanced models still require accessing sufficiently large datasets for training, which is often unfeasible in the tactile perception field. When trained on sample-limited datasets, the deep neural network is lack of generalization capability [9]. To solve the above problem, we apply a Supervised Autoencoders (SAE) [10] to improve the generalization performance on the tactile gesture recognition dataset, which uses Autoencoders (AE) as regularization tasks. Moreover, based on the SAE, we propose a Multi-kernel-size Convolutional Supervised Autoencoders (MCSAE) to further improve the generalization performance on the limited dataset, which provides models with more structure of receptive fields and enhances the feature extraction ability of SAE. Contributions made in this work can be summarized as follows:

1) We first apply the SAE with good general-

[1] Guangdong Provincial Key Lab of Robotics and Intelligent System, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

[2] SIAT Branch, Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen 518055

[*] Corresponding author: Zhengkun Yi, zk.yi@siat.ac.cn

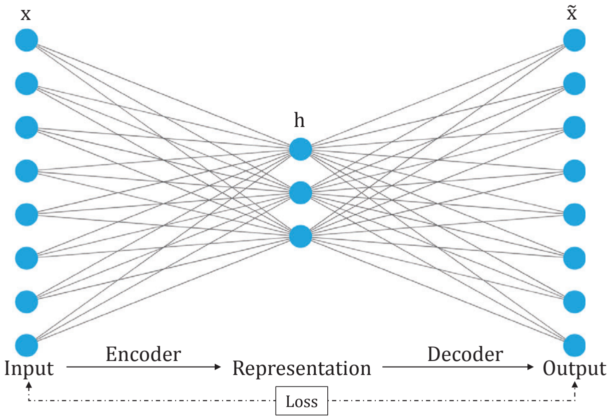

Fig. 1. The architecture of AE. AE obtains the representation of the hidden layer by encoding and decoding.

ization performance to solve the overfitting problem of the tactile gesture recognition task.

2) We propose a novel MCSAE to further improve the generalization performance by enhancing the feature extraction ability of SAE.
3) We perform extensive comparison experiments on a publicly available tactile gesture dataset to prove SAE and MCSAE have better generalization performance than other state-of-the-art (SOTA) methods.
4) We conduct ablation experiments on publicly available tactile gesture datasets to demonstrate that MCSAE is effective for improving generalization performance on the sample-limited dataset.

II. Method

All the methods use a training set $(\mathbf{x}_i, \mathbf{y}_i)$, where $\mathbf{x}_i \in \mathbb{R}^d$ is the i-th sample in the training set (N samples) and $\mathbf{y}_i \in \mathbb{R}^m$ is an m-dimensional one-hot vector, which corresponds to the class of the i-th sample. In addition, d is the dimension of each sample and m is the number of the classes.

A. Autoencoders (AE)

The AE is an unsupervised neural network widely used in data dimensionality reduction [11]. AE obtains the representation of the hidden layer by encoding and decoding, as shown in Figure 1. The encoder maps an unlabelled training sample $\mathbf{x}_i$ to a non-linear representation $\mathbf{h}_i$ of the hidden layer:

$$\mathbf{h}_i = f(\mathbf{W}_i\mathbf{x}_i + \mathbf{b}_i) \tag{1}$$

where $\mathbf{h}_i \in \mathbb{R}^h$ is representation of the hidden layer, $\mathbf{W}_i \in \mathbb{R}^{h\times d}$ is a weight matrix, and $\mathbf{b}_i \in \mathbb{R}^h$ is a bias vector for the encoding process. In addition, h is the number of hidden neurons. In the decoder, the representation $\mathbf{h}_i$ is reconstructed to the original dimention d as follows:

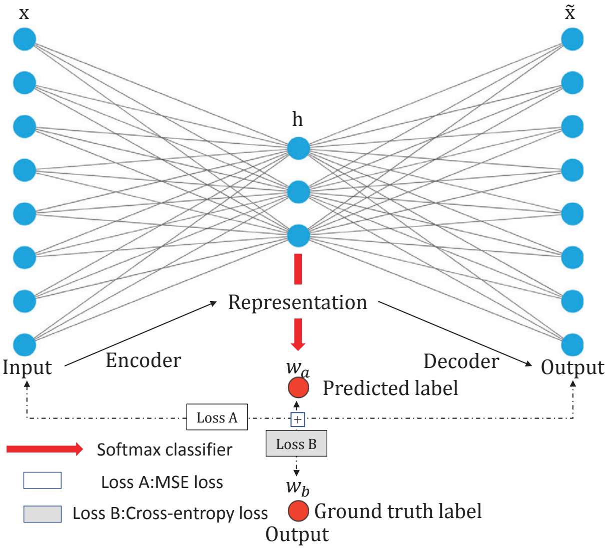

Fig. 2. The architecture of SAE. SAE is an approach to using AE as unsupervised auxiliary tasks. In addition, we train AE jointly with the classifier.

$$\tilde{\mathbf{x}}_i = f\left(\tilde{\mathbf{W}}_i\mathbf{h}_i + \tilde{\mathbf{b}}_i\right) \tag{2}$$

where $\tilde{\mathbf{x}}_i \in \mathbb{R}^d$ is the reconstructed data of the i-th sample, $\tilde{\mathbf{W}}_i \in \mathbb{R}^{d\times h}$ is a weights matrix, and $\mathbf{b}_i \in \mathbb{R}^d$ is a bias vector for the decoding process. In addition, f is an activation function of the hidden layer, e.g., sigmoid function:

$$f(x) = \frac{1}{1+\exp(-x)}. \tag{3}$$

Then, the representation of the hidden layer is obtained by minimizing the following reconstruction error loss function:

$$L = \frac{1}{N}\sum_{i=1}^{N}\|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|_2^2. \tag{4}$$

B. Supervised Autoencoders (SAE)

The traditional AE for supervised learning trains the representation learning process and the supervised learning process separately. However, during the training process, we cannot guarantee that the features learned from the representation learning method AE can be well applied to the classifier. SAE is an approach to using AE as unsupervised auxiliary tasks to improve generalization performance. Le et al.[10] theoretically and empirically showed that the addition of reconstruction error improves generalization

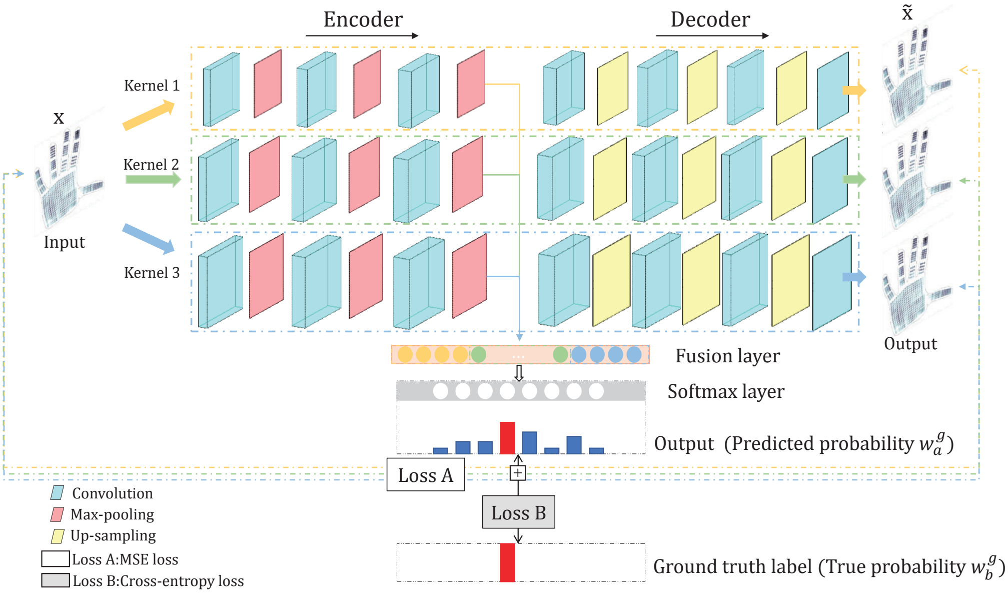

Fig. 3. The architecture of MCSAE. Three channels are used to extract useful representation by using CAE with different kernel sizes as auxiliary tasks. In adddition, we train CAE jointly with the classifier.

performance. The architecture of SAE is shown in Figure 2. For an SAE based on AE, the classifier maps the representation of the deepest hidden layer to an m-dimensional one-hot vector $\mathbf{w}_{a,i}$ as follows:

$$\mathbf{w}_{a,i} = S\left(\mathbf{W}_i^p \mathbf{h}_i + \mathbf{b}_i^p\right) \tag{5}$$

where $\mathbf{W}_i^p \in \mathbb{R}^{m \times h}$ is a weight matrix and $\mathbf{b}_i^p \in \mathbb{R}^m$ is a bias vector to predict label, and S is a softmax function to map the vector to a one-hot vector. After that, the SAE uses an auxiliary task to improve generalization performance by adding the reconstruction error loss to the classification loss. Let $\mathrm{CE}(\cdot)$ be the cross-entropy loss for classification and $\mathrm{MSE}(\cdot)$ be the mean square error loss for the reconstruction error. The total loss function of the SAE is as follows:

$$\begin{aligned} L &= \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{CE}\left(\mathbf{w}_{a,i}, \mathbf{w}_{b,i}\right) + \alpha_r \mathrm{MSE}\left(\tilde{\mathbf{x}}_\mathbf{i}, \mathbf{x}_\mathbf{i}\right)\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(-\sum_{g=1}^{m}\left(\mathbf{w}_{b,i}^g \log\left(\mathbf{w}_{a,i}^g\right)\right) + \alpha_\mathbf{r} \left\|\tilde{\mathbf{x}}_\mathbf{i} - \mathbf{x}_\mathbf{i}\right\|_2^2\right) \end{aligned} \tag{6}$$

where $\mathbf{w}_{a,i}^g$ is the predicted probability of the g-th category, $\mathbf{w}_{b,i}^g$ is the ground truth probability distribution of the g-th category, and α_r is the weight of the reconstruction loss.

C. Multi-kernel-size Convolutional Supervised Autoencoders (MCSAE)

Like CNN, SAE can also replace fully connected layers with convolutional layers to extract spatial information on images. Convolution shows its powerful feature extraction and integration ability in processing data with a 2D array structure. The extracted high-level features will be different by the size of the receptive field. Inspired by the inductive bias concept of SAE and the working mechanism of convolution, we propose an MCSAE to use convolutions of different kernel sizes for complementation, which provides model more structures of receptive fields and enhances the feature extraction ability of SAE. The MCSAE includes the main task CNN and the auxiliary task CAE as shown in Figure 3. The convolutional encoder maps the training data $\mathbf{x}_i$ with one channel to a non-linear hidden representation $\mathbf{h}_{j,i}^{\mathrm{k}}$ with k feature maps:

$$\mathbf{h}_{j,i}^k = \mathrm{pool}\left(f\left(\mathbf{W}_{j,i}^k \odot \mathbf{x}_i + \mathbf{b}_{j,i}^k\right)\right) \tag{7}$$

where $\mathbf{W}_{j,i}^k$ is the k-th convolutional kernel and each convolutional kernel is matched with a bias $\mathbf{b}_{j,i}^k$. In addition, $\mathrm{pool}(\cdot)$ is a max-pooling layer, $\odot$ is a convolutional operation, and $f(\cdot)$ is the sigmoid function. Then, the convolutional decoder maps the hidden representation $\mathbf{h}_{j,i}^k$ to a reconstruction data:

$$\tilde{\mathbf{x}}_{j,i} = f\left(\tilde{\mathbf{W}}_{j,i} \odot \mathrm{up}\left(\mathbf{h}_{j,i}^k\right) + \tilde{\mathbf{b}}_{j,i}\right) \tag{8}$$

where $\mathrm{up}(\cdot)$ is an up-sampling layer that uses the deconvolution operation to get a feature map with the same size as the input data and $\tilde{\mathbf{W}}_{j,i}$ is a single convolutional kernel that reconstruct the input data.

The convolution kernel sizes we choose are 1, 3, and 5. Setting the sizes of the convolution kernel to an odd number has two advantages. One is to ensure that the anchor point is just in the center, avoiding the offset of the position information. The other is to ensure that the two sides of the image are still symmetrical during padding. The convolution kernel of size 1 retains the most original data information. The convolution kernels of sizes 3 and 5 can obtain useful information about the smaller and larger receptive fields and obtain different feature maps.

The main task CNN flattens the hidden layer feature maps $\mathbf{h}_{j,i}^{k}$ obtained from different receptive fields. Then, the features are concatenated into a one-dimensional vector as a multi-kernel-size fusion feature. The kernel-fused features $\mathbf{H}_i$ is given by:

$$\mathbf{H}_i = \left[\text{ flatten } \left(\mathbf{h}_{1,i}^{k}\right);\text{ flatten } \left(\mathbf{h}_{2,i}^{k}\right);\text{ flatten } \left(\mathbf{h}_{3,i}^{k}\right)\right] \tag{9}$$

where flatten $(\cdot)$ is the flatten operation and $\mathbf{h}_{j,i}^{k}\,(j=1,2,3)$ denotes the feature maps obtained from different receptive fields. Next, the concatenated features are put into two fully connected layers and use feature weighting to reduce the number of features to eliminate redundant features. Finally, the fused features are put into the softmax classifier:

$$\mathbf{w}_{a,i} = S\left(\mathbf{W}_i^{p}\mathbf{H}_i + \mathbf{b}_i^{p}\right) \tag{10}$$

where $S(\cdot)$ is a softmax function to map the vector to predicted probability distribution $\mathbf{w}_{a,i}$. Let $\mathrm{CE}(\cdot)$ be the cross-entropy loss for the classification and $\mathrm{MSE}(\cdot)$ be the mean square error loss for the reconstruction error. The total loss function of the MCSAE is as follows:

$$\begin{aligned} L &= \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{CE}\left(\mathbf{w}_{a,i},\mathbf{w}_{b,i}\right) + \sum_{j=1}^{3}\alpha_{r,j}\,\mathrm{MSE}\left(\tilde{\mathbf{x}}_{j,i},\mathbf{x}_i,\right)\right) \\ &= \sum_{i=1}^{N}\left(-\sum_{g=1}^{m}\left(\mathbf{w}_{b,i}^{g}\log\left(\mathbf{w}_{a,i}^{g}\right)\right) + \sum_{j=1}^{3}\alpha_{r,j}\left\|\tilde{\mathbf{x}}_{j,i}-\mathbf{x}_{\mathbf{i}}\right\|_2^2\right) \end{aligned} \tag{11}$$

where $\alpha_{r,j}$ is weight of the reconstruction loss of different kernels j.

III. Experiments

A. STAG Dataset

The tactile data used in this paper was collected by Subramanian et al. [7] as shown in Figure 4. They used a scalable tactile glove (STAG) covers the whole hand with 548 sensors to obtain a 32 * 32 tactile array. They chose eight of the most representative hand positions, including seven different hand positions and one neutral hand pose as a reference. The total dataset includes 4336 frames.

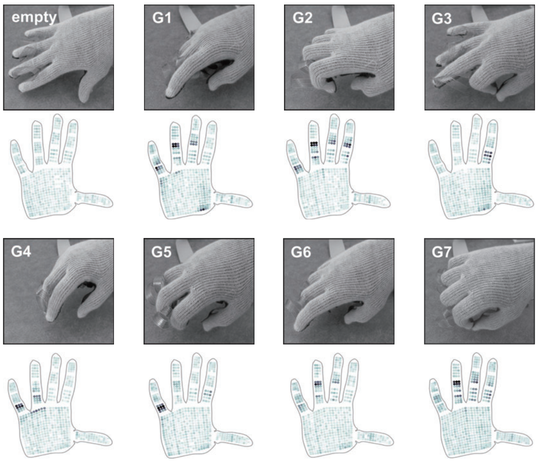

Fig. 4. STAG dataset. Eight different gestures and their corresponding sensor arrays.

B. Experimental Setups

In all experiments, we use the same dataset division method of 3080 training frames and 1256 distinct testing frames as the setting of Subramanian et al.

1) Comparison With Related Methods: This experiment aims to compare the SAE with convolution and the proposed MCSAE with the existing SOTA machine learning and deep learning methods in the field of tactile gesture recognition on the sample-limited publicly available dataset. A brief introduction of the compared methods is as follows:

SVM: SVM is a supervised learning model, which is one of the most robust prediction methods for classification tasks on the tactile dataset [12], [13]. We use the multi-class SVM method with a standard radial basis function (RBF) to classify eight different gestures on the STAG dataset, which has a better performance than SVM with other kernel functions based on our pre-experiment.

MLP: MLP is a deep learning model. We use a three-hidden-layer MLP to classify eight different gestures on the STAG dataset, which has a better performance than MLP with other hidden layer numbers based on our pre-experiment.

CNN: CNN is a deep learning model that can obtain image spatial information. We use a three-hidden-layer CNN to classify eight different gestures on the STAG dataset, which has a better performance than CNN with other hidden layer numbers based on our pre-experiment.

STAG: STAG was proposed by Subramanian et al. on the STAG dataset to classify eight different gesture. The architecture includes a convolutional layer, a batch normalization (bn) layer, a max-pooling layer, two resnet blocks, and a drop out layer.

2) Ablation Experiments of MCSAE: This experiment aims to analyze the effects of the multi-kernel size. Besides the proposed MCSAE, we implement and compare some other architectures with single or multiple kernel sizes:

Single size kernel: SAE with kernel size 1 (SAE-1), SAE with kernel size 3 (SAE-3), and SAE with kernel size 5 (SAE-5)

Two size kernel: MCSAE with kernel sizes 1 and 3 (MCSAE-13), MCSAE with kernel sizes 1 and 5 (MCSAE-15), and MCSAE with kernel sizes 3 and 5 (MCSAE-35)

A trial-error method is applied to set up these key parameters of all the methods. In this paper, all deep learning models are trained for 20 epochs, with a batch size of 32 samples. The Adam optimizer with a learning rate of 0.001 is applied to minimize the loss function. We use average of 10 runs as the final result. In addition, we normalize each feature between 0 and 1 through Min-Max scaling.

The experiment results are evaluated in the metric of accuracy, which is an evaluation metric of gesture recognition performance. The accuracy is defined as:

$$\text{Acc} = \frac{Z}{N} \times 100\% \tag{12}$$

where N denotes the total numbers of truth classes and Z denotes the total numbers of correctly identified classes. Moreover, we present the accuracy of the training and test set and further compare, which can evaluate the generalization performance of our proposed method.

C. Comparison With Related Methods

Table I shows the test and training accuracy of four compared methods, SAE and the proposed MCSAE on the STAG dataset. The results show that MCSAE achieves the highest test accuracy (94.43%) compared with other SOTA tactile gesture recognition methods on the sample-limited dataset. In addition, we make the following observations:

1) All methods achieve approximately 100% accuracy on the training set, which means that higher accuracy on the test set indicates better generalization performance of the model.
2) Compared with the SVM, other methods achieve higher accuracy on the test set. It indicates that the deep learning architecture has superior performance on the STAG dataset than the traditional shallow machine learning methods.
3) Compared with the MLP, CNN, and STAG, the SAE achieves higher accuracy on the test set. It appears that the SAE architecture that uses AE as a regularization task has better generalization performance on the sample-limited STAG dataset than other deep learning methods without regularization tasks.
4) Compared with the SAE, the proposed MCSAE achieves higher accuracy on the test set. Therefore, the proposed regularization strategy for input layer reconstruction based on multi-kernel-size convolution feature extraction can effectively improve the generalization performance.

TABLE I
TEST AND TRAINING ACCURACY OF FOUR COMPARED METHODS (SVM, MLP, CNN, AND STAG), SAE, AND THE PROPOSED MCSAE ON THE STAG DATASET. THE BEST PERFORMANCE OF THE TEST ACCURACY IS MARKED IN BOLD.

Method	Average Accuracy (%)	
	Test	Training
SVM	87.10	100.00
MLP	87.18	99.84
CNN	92.27	100.00
STAG	89.40	100.00
SAE	93.47	100.00
MCSAE	**94.43**	100.00

TABLE II
TEST AND TRAINING ACCURACY OF THE PROPOSED MCSAE AND RELATED ABLATION MODELS ON THE STAG DATASET. THE BEST PERFORMANCE OF THE TEST ACCURACY IS MARKED IN BOLD.

Method	Average Accuracy (%)	
	Test	Training
SAE-1	85.19	98.44
SAE-3	93.31	100.00
SAE-5	93.47	100.00
MCSAE-13	93.55	99.55
MCSAE-15	93.71	100.00
MCSAE-35	93.79	100.00
MCSAE	**94.43**	100.00

D. Ablation Experiments of MCSAE

Table II shows the test and training accuracy of MCSAE with different kernel sizes, which aims to analyze the effects of multi-kernel size. The results show that MCSAE achieves the highest test accuracy compared to other ablation models on the sample-limited dataset. In addition, we make the following observations:

1) The regularization strategy using three kernels on different channels for feature extraction has the highest performance on the test set for tactile gesture recognition. SAE with a single-kernel size has relatively poor recognition performance. Therefore, increasing certain convolution kernel numbers on different channels is beneficial to improving the generalization performance on the sample-limited dataset.
2) When the kernel size is 1, the convolutional layer will not be able to consider the spatial information of the tactile data image, and the maximum pooling downsampling operation will lose part of the information, resulting in low accuracy. However, when we use the AE with a kernel size of 1 as an auxiliary task to expand the data dimension and supplement the information required for recognition, the model has a better recognition performance than a single kernel size, which can effectively improve the generalization performance of the model.

IV. Conclusion

In this paper, we first apply the SAE to improve generalization performance for tactile gesture recognition tasks. Moreover, we propose a novel MCSAE to further improve the generalization performance by enhancing the feature extraction ability of SAE. After that, we compare the currently SOTA models with SAE and MCSAE on the sample-limited publicly available dataset. The experimental results demonstrate that the SAE we apply has better generalization performance than compared methods, and MCSAE achieves higher accuracy on test set than SAE, which indicates the excellent generalization performance of the MCSAE we propose. In addition, we also conduct ablation tasks to show that MCSAE is effective for improving generalization performance, which has higher accuracy on the test set than other ablation models. Moving forward, we will further explore the proposed methods on other sample-limited datasets.

Acknowledgment

The work described in this paper is partially supported by the National Natural Science Foundation of China (Grant No. 62003329), and partially supported by the National Natural Science Foundation of Guangdong Province of China (Grant No. 2021A1515011316).

References

[1] M. Zhu, Z. Sun, Z. Zhang, Q. Shi, T. He, H. Liu, T. Chen, and C. Lee, "Haptic-feedback smart glove as a creative human-machine interface (hmi) for virtual/augmented reality applications," *Science Advances*, vol. 6, no. 19, p. eaaz8693, 2020.

[2] Z. Zhou, K. Chen, X. Li, S. Zhang, Y. Wu, Y. Zhou, K. Meng, C. Sun, Q. He, W. Fan *et al.*, "Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays," *Nature Electronics*, vol. 3, no. 9, pp. 571–578, 2020.

[3] J. de Gea Fernández, D. Mronga, M. Günther, T. Knobloch, M. Wirkus, M. Schröer, M. Trampler, S. Stiene, E. Kirchner, V. Bargsten *et al.*, "Multimodal sensor-based whole-body control for human–robot collaboration in industrial settings," *Robotics and Autonomous Systems*, vol. 94, pp. 102–119, 2017.

[4] Q. Li, O. Kroemer, Z. Su, F. F. Veiga, M. Kaboli, and H. J. Ritter, "A review of tactile information: Perception and action through touch," *IEEE Transactions on Robotics*, vol. 36, no. 6, pp. 1619–1634, 2020.

[5] Z. Yi, T. Xu, W. Shang, W. Li, and X. Wu, "Genetic algorithm-based ensemble hybrid sparse elm for grasp stability recognition with multimodal tactile signals," *IEEE Transactions on Industrial Electronics*, 2022.

[6] S. Jiang, L. Li, H. Xu, J. Xu, G. Gu, and P. B. Shull, "Stretchable e-skin patch for gesture recognition on the back of the hand," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 1, pp. 647–657, 2019.

[7] S. Sundaram, P. Kellnhofer, Y. Li, J.-Y. Zhu, A. Torralba, and W. Matusik, "Learning the signatures of the human grasp using a scalable tactile glove," *Nature*, vol. 569, no. 7758, pp. 698–702, 2019.

[8] X. Huang, Q. Wang, S. Zang, J. Wan, G. Yang, Y. Huang, and X. Ren, "Tracing the motion of finger joints for gesture recognition via sewing rgo-coated fibers onto a textile glove," *IEEE Sensors Journal*, vol. 19, no. 20, pp. 9504–9511, 2019.

[9] H. Li, Y. Wang, R. Wan, S. Wang, T.-Q. Li, and A. Kot, "Domain generalization for medical imaging classification with linear-dependency regularization," *Advances in Neural Information Processing Systems*, vol. 33, pp. 3118–3129, 2020.

[10] L. Le, A. Patterson, and M. White, "Supervised autoencoders: Improving generalization performance with unsupervised regularizers," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[11] D. E. Rumelhart, G. E. Hinton, and R. Williams, "Learning representations by back propagating errors. cogn," *Nature*, vol. 5, 1986.

[12] Z. Yi, T. Xu, S. Guo, W. Shang, and X. Wu, "Tactile surface roughness categorization with multineuron spike train distance," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 4, pp. 1835–1845, 2020.

[13] Z. Yi, T. Xu, W. Shang, and X. Wu, "Touch modality identification with tensorial tactile signals: A kernel-based approach," *IEEE Transactions on Automation Science and Engineering*, 2021.