u-DeepHand: FMCW Radar based Unsupervised Hand Gesture Feature Learning using Deep Convolutional Auto-Encoder Network

Zhenyuan Zhang, Zengshan Tian, Ying Zhang*, *Senior Member, IEEE,* Mu Zhou, *Senior Member, IEEE,* and Bang Wang, *Member, IEEE*

***Abstract*—Recently, although radar sensors have been widely applied for Hand Gesture Recognition (HGR) tasks, conventional radar-based HGR systems still have two major challenges. Firstly, these systems rely on supervised learning approaches to learn gesture features, which normally requires a large-scale labeled dataset to address overfitting problem. However, the acquisition of such dataset is time-consuming. Secondly, the radar signature of hand movement is often influenced by micro motion caused by other body parts, which leads to distorted motion features, resulting in poor identification accuracy. To overcome these problems, we propose an unsupervised hand gesture feature learning method using deep convolutional auto-encoder network to analyze hand gesture signal collected by a Frequency Modulated Continuous Wave (FMCW) radar sensor. Firstly, via a convolutional encoder sub-network, input radar range profiles are transformed into lower dimensional representations. Then, the representations are expanded to reconstruct the corresponding input profiles by a deconvolutional decoder sub-network. In addition, to investigate the mechanisms of the proposed network and evaluate its performance, we conduct an in-depth study of the feature maps learned from various hand gesture experimental data and evaluate the corresponding classification performance. The results demonstrate that the proposed convolutional auto-encoder network is able to achieve high recognition accuracy with low training sample cost, which outperforms the state-of-the-art hand gesture recognition systems based on transfer learning VGGNet and fully connected based auto-encoder network.**

***Index Terms*—Hand gesture recognition, FMCW radar, convolutional auto-encoder network, transfer learning VGGNet, feature maps.**

I. Introduction

Over the past few years, there has been a growing interest in the research field of hand gesture recognition (HGR). This is due in part to a large number of application domains for human-computer interaction scenario, from e-Health [1], entertainment [2, 3], to driver assistance system [4-6] and mobile interactive multimedia [7]. HGR applications have made it more convenient and efficient for users to control devices remotely.

Multiple approaches have been attempted to measure hand motions, such as vision-based sensors [8-10], motion-based sensors [11-13], and radio frequency (RF) based sensors [14-24]. In the past, gesture recognition has been studied for a while within the field of computer vision. However, under poor visibility condition, vision-based sensors have limited performance and may cause privacy issue. In addition, vision-based HGRs consume significant computational resource to process image algorithms [14]. Recently, with the development of small motion sensors, the use of motion-based devices has become another popular solution to HGR. In such systems, each subject utilizes a wearable sensor, such as a magnetic sensor, a gyroscope, or an accelerometer for measuring hand motions [12, 13]. Motion-based sensors are able to realize high precision motion tracking, but require users to be equipped them all the time. Now, RF-based systems have begun to attract significant interest in Human-Computer Interaction (HCI) engineering. Authors in [15] present WiGest system, which leverages changes in Wi-Fi signal strength measured by a mobile device to recognize in-air gestures. By using a regular Wi-Fi transceiver chipset, Fan et al. in [16] propose a Wi-Fi based radar system to sense hand motion information. Since Wi-Fi signals can travel through walls, this system enables indoor gesture recognition using fewer Wi-Fi routers. However, the received Wi-Fi signals are always noisy and sensitive to layout changes in environment due to multi-path effect. Among RF-based systems, with the properties of privacy preservation, low power cost, and high-resolution ranging, radar-based systems have promising performance in short-range hand gesture sensing. A Google team proposes Soli system in [17], which is designed for fine-gained gesture sensing using a high-frequency and short-range Frequency Modulated Continuous Wave (FMCW) radar. Kim et al. utilize micro-Doppler signatures measured by a unmodulated continuous wave radar to discriminate hand gestures [18]. In this paper, we propose a FMCW radar-based system, namely u-DeepHand, to recognize hand gestures in short-range. Different from unmodulated continuous wave radar, FMCW radar provides both range and Doppler information of targets, which is capable of separating different targets with different ranges. Moreover, we discuss the impacts of two factors, including the

Z. Zhang is with the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China, and also with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30302, USA (e-mail: zhangzhenyuangm@gmail.com).

Z. Tian and M. Zhou are with the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: tianzs@cqupt.edu.cn, zhoumu@cqupt.edu.cn).

Y. Zhang is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30302 USA (e-mail: yzhang@gatech.edu).

B. Wang is with the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: wangbang@hust.edu.cn).

Manuscript received April 19, 2005; revised August 26, 2015.

testing distance between users and radar platform, and hand gesture scale, on identification accuracy.

From the point of measurement principle, the underlying phenomenology of radar data and vision data is fundamentally different. Although radar data can be effectively mapped into spectrograms by using time-frequency transformation algorithms, how well these spectrograms may be able to represent radar features is still an open question. Most of radar-based HGR systems contain two key phases: 1) feature extraction and 2) classification. In phase 1, a various of handcrafted features, such as Doppler bandwidth, envelope [19], phase response [20], and transform-based features [21] extracted by deep learning based methods [18, 22-24] are utilized for representing different gestures. In phase 2, the features extracted in phase 1 are inputted into a trained classifier to recognize hand gestures. For gesture feature representation, there is a discussion about how to define a valid criterion as the guideline for learning good intermediate features. Obviously, the good features should reflect the primary movement information contained in radar data. A supplemental criterion is that the features should be sparse, which helps to reduce the usages of computing and storage resources. In addition, a specific criterion for radar-based gesture sensing system is robust to noisy radar data, since the radar signature of hand movement is often influenced by various subtle factors such as micro motion caused by other body parts. Apart from the discussion above, a large gesture training dataset plays an important role in a HGR system. Conventional Convolutional Neural Network (CNN) based systems are typically randomly initialized [18, 23], which relies on a large training dataset to make sure gradient-based optimization algorithms converge to a global optimum [25]. Otherwise, it may converge into a local optimum instead of a global optimum. What's more, a large training dataset helps a deep learning network avoid overfitting problem. However, for HGR task, collecting large amounts of labeled data is difficult, labor-intensive, and sometimes even impossible.

To cope with these challenges, we propose u-DeepHand, an unsupervised gesture feature learning and classification system. In this system, we collect radar echo signals generated by a 24GHz FMCW radar and utilize an unsupervised deep convolutional auto-encoder network to automatically learn hand motion representation from FMCW signals. To explore the mechanisms of the proposed network, an in-depth study of learned feature maps is conducted based on various collected hand gesture data. In addition, we compare our approach against other deep learning models, including transfer learning VGGNet [26] and fully connected auto-encoder network [27]. Moreover, we investigate the impact of two main factors, including hand gesture scale and testing distance, on the accuracy of gesture classification. Although CNN-based systems and fully connected auto-encoder network based systems have been proposed in previous work, the convolutional auto-encoder network in this work is mainly motivated by unsupervised learning methods rather than just the combination of two above-mentioned networks. In this system, a deconvolutional decoder network plays a more important role in reconstructing the input range profiles using the sparse features.

The main contributions of this paper are summarized as follows:

- We present u-DeepHand, a novel FMCW radar based HGR system, to learn hand motion features and recognize hand gestures using unsupervised deep convolutional auto-encoder network, in the case of modest training dataset.
- To investigate the mechanisms of the proposed network, we make an in-depth study of learned feature maps in experimental analysis. What's more, we compare our approach against other deep learning models, including transfer learning VGGNet and fully connected auto-encoder network.
- We discuss the effects of the main factors, including hand gesture scale and testing distance, on the accuracy of gesture classification.

The rest of this paper is organized as follows. Section 2 briefly reviews the radar based HGR systems, in terms of radar systems and feature learning algorithms. Section 3 describes the proposed radar system design and its implementation, while Section 4 shows the architecture of the proposed unsupervised deep convolutional auto-encoder network. The feature learning and recognition performances are compared and contrasted with conventional methods based on experimental data in Section 5. Finally, Section 6 draws the conclusions and future work.

II. Related work

In this section, a brief overview of previous studies related to radar-based HGR system is provided, followed by comparing and contrasting our gesture recognition algorithm with the most relevant literature.

A. Radar-Based HGR Systems

The applications of radar sensor for gesture detection and discrimination have begun to gain interests in public [7, 18-20, 23, 28-37]. In previous work, researchers mainly utilize unmodulated continuous wave radar sensors to measure the Doppler effect produced by dynamic hand gestures. For example, Kim et al. [18] explore the possibility of classifying gestures using micro-Doppler signatures and recognize seven gestures with an accuracy of 93.1%. Li, G. et al. [21] propose a sparsity-driven micro-Doppler analysis method to represent radar signals in time-frequency domain. Recently, considering the commonality of HGR platforms, several Doppler-based HGR systems have been proposed using commercial Wi-Fi signals to measure Doppler shift. Tang, M. et al. [7] extract Doppler shift from reflected Wi-Fi signals and integrate it with a 2D camera to sense 3D hand gestures. Among the Doppler-based systems mentioned above, Doppler frequency offset caused by hand movement is only a few hertz, without special requirement for high-performance Analog to Digital Converter (ADC) devices. However, an unmodulated continuous wave radar only provides the Doppler information of a single moving target. Under normal circumstances in which multiple moving targets exist, Doppler signatures of non-hand movement targets, such as other parts of body and

near pedestrians, will overlap with the signature of hand movement, thus disturbing effective hand gesture feature extraction. Therefore, some researchers have turned to use FMCW radar to compensate for the deficiencies existing in unmodulated continuous wave radar-based systems [19, 23]. Authors in [19] use a FMCW radar sensor to recognize human gestures in the case of multiple moving objects coexistence. Molchanov, P. et al. [23] present a FMCW radar-based system for sensing short-range dynamic hand gestures. In addition, to obtain a higher range-resolution, impulse radio ultra-wideband (IR-UWB) radar is used for fine-grained HGR [28].

B. Feature Learning Algorithms

Conventional radar-based HGR systems utilize predefined features of radar signals or spectrograms to analyze motion information. The predefined features mainly include Doppler bandwidth, envelope [19], phase response [20] and transform-based features [21]. However, it is time consuming and fragile to identify predefined features with relevant information in different radar-based systems. Recently, deep learning methods are becoming more popular in gesture recognition systems for learning features automatically [18, 22, 23, 38, 39]. Researchers in the Google Soli Project [38] utilize a trained combination of convolutional and recurrent neural networks to learn features of micro finger motions in fine detail. In the multi-sensor HGR system presented by NVIDIA [39], researchers employ 3D-CNN to fuse features extracted from multiple sensors to improve the system's robustness to lighting environment. However, conventional radar-based HGR systems mentioned above are based on supervised learning, which means that a good supply of sufficient training dataset is necessary for supervised training. Unfortunately, for HGR tasks, the collection of labeled training dataset is expensive and time demanding.

Recently, a few transfer learning- and unsupervised leaning-based methods have been proposed to overcome the limitation of small training dataset. In the case of insufficient high-quality training data, instead of training a deep network from scratch, transfer learning techniques try to use the knowledge transferred from some previous tasks to train target tasks [40]. For example, Alnujaim, I. et al. [26] employ transfer learning VGGNet [41] that was trained for vision images to solve the restriction of small training dataset in HGR task. The authors in [37] utilize pre-trained AlexNet [43] to learn features from radar spectrogram images for human activities recognition with a high accuracy. In addition to transfer learning based methods, unsupervised feature learning methods have attracted attention in gesture sensing community due to the fact that they make full use of arbitrary scale of unlabeled data and represent the robust and discriminative features reflecting the essential aspects of the data. In [27], to obtain the sparse representation, authors propose a stacked auto-encoder network for human gesture detection. In addition, by unsupervised feature learning techniques, Yin, J. et al. [44] try to automatically seek for discriminative, local features from noisy UWB radar signals for human identification task.

Inspired by previous studies on unsupervised deep learning models and their variants, this paper proposes a deep convolutional auto-encoder network for FMCW radar-based HGR task. The proposed framework learns effective features from a small training dataset, and then a softmax layer is applied to classify diverse gestures based on the obtained features.

III. Radar System Description

A. Radar system overview

u-DeepHand is built upon the Latern project [45], which is a FMCW radar based HGR system. Different from Latern's aim of recognizing dynamic continuous gestures, u-DeepHand focuses on learning hand motion features in the case of small training dataset.

The radar platform operates in K-band with the central frequency 24GHz, bandwidth 4GHz, and output power of transmitter 5dBm. For HGR tasks, it is important to determine radar range and velocity resolutions. According to [46], the range and velocity resolutions can be represented respectively as:

$$\Delta r = \frac{c}{2B} \tag{1}$$

$$\Delta v = \frac{\lambda}{2NT_c} \tag{2}$$

where c is the speed of light and λ is the wavelength. B and T_c are the bandwidth and sweep period of FMCW radar which are set as 4GHz and 1ms respectively. N is the number of chirps, which is set as 64. Therefore, the range and velocity resolutions are computed as 3.75cm and 9.76cm/s respectively.

In addition, for FMCW radar with high bandwidth, it is necessary to make sure that the frequency of transmitted signals varies linearly. However, in practice, the Voltage Controlled Oscillator (VCO) of radar sensors has a non-linear voltage-frequency response, which deteriorates solving range information in deramping process since the nonlinearity spreads target energy through different frequencies resulting in a more broadened response [47], as shown in Figure 1. Moreover, the spreading impact on beat signal is greater for targets with longer distance. In Latern, Residual Video Phase (RVP) correction based method [48] is utilized to compensate nonlinearity for short-range HGR application (see [45] for details).

B. Hand Gesture Data Acquisition

To demonstrate the capabilities of u-DeepHand system, we designed and evaluated 8 hand gestures frequently applied in daily life. In addition, due to the self-occlusion problem of fingers, we selected the hand gestures with main trajectories created by moving entire arm or hand. The selected gestures include (a) sliding hand from right to left, (b) sliding hand from left to right, (c) pulling, (d) pushing, (e) knocking, (f) moving hand up and down, (g) waiving hand, and (h) patting. Considering the diversity of gesture data, we invited four volunteers to make the eight gestures above. Each volunteer was trained how to make the gestures before the tests and the body of each volunteer remains still during the experiment. Each gesture lasts for about 1.5 seconds and is made for 100 repetitions for the test. In addition, a total number of 3200 samples were collected at the location with the distance

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2910810, IEEE Sensors Journal

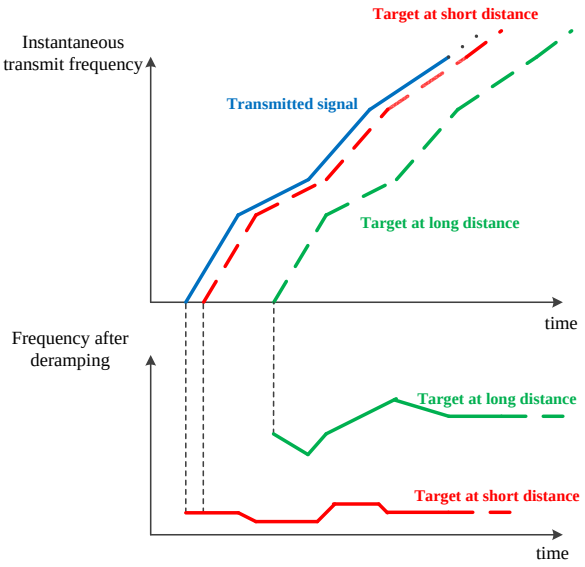

Fig. 1. Deramping of FMCW signals with frequency nonlinearity.

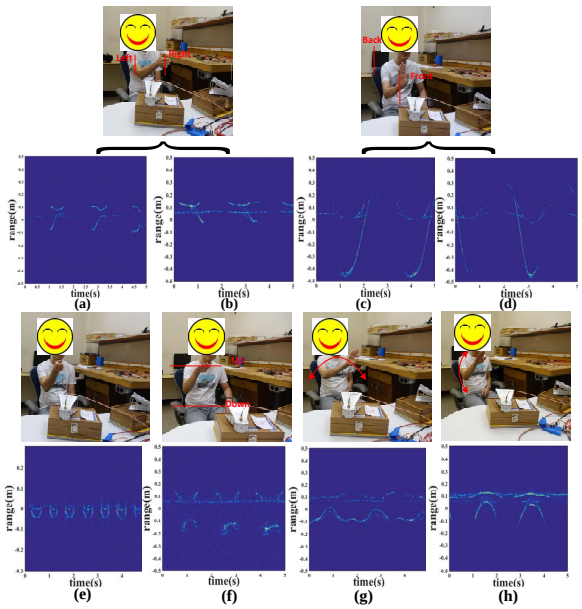

Fig. 2. Examples of the gestures and their corresponding radar range profile: (a) sliding hand from right to left, (b) sliding hand from left to right, (c) pulling, (d) pushing, (e) knocking, (f) moving hand up and down, (g) waiving hand, and (h) patting.

$d = 1.5m$, and meanwhile the hand gestures were made with the scale $r = 0.5m$. Examples of the gestures and the corresponding radar range profiles are shown in Figure 2. It is easy to be observed that different gestures can be discriminated according to their trajectory outlines.

IV. Deep Convolutional Auto-Encoder Network

A. Analysis and modeling

In some realistic scenarios, the whole upper body is in radar detection range, as shown in Figure 2. The interference caused by upper torso movement, will overlap with the signature of hand gestures, thus disturbing effective hand motion feature extraction. However, this interference can be reflected in the range profiles. Motivated by the development of unsupervised learning methods, we mainly focus on filtering out this interference caused by micro motions to improve recognition accuracy. In this paper, in the absence of interference, we denote $\mathbf{r} \in \mathbf{R}^{m \times n}$ as the original input radar range profile with the spatial size of $m \times n$ pixels. Then, the input range profile is firstly mapped to a hidden representation $\mathbf{p} \in \mathbf{R}^{i \times j}$ through the CNN function

$$\mathbf{p} = f_\theta(\mathbf{r}) \tag{3}$$

which is parameterized by $\theta = \{\mathbf{W}, \mathbf{b}\}$. $\mathbf{W}$ represents a 3×3 convolution kernel set and $\mathbf{b}$ is a bias vector set in encoder sub-network. Through deconvolutional decoder sub-network $\mathrm{g}()$, the resulting representation $\mathbf{p}$ is then transformed to a "reconstructed" output $\mathbf{z} \in \mathbf{R}^{m \times n}$

$$z = g_{\theta'}(\mathbf{p}) \tag{4}$$

which is parameterized by $\theta' = \{\mathbf{W}', \mathbf{b}'\}$. Similar to $\mathbf{W}$ and $\mathbf{b}$, $\mathbf{W}'$and $\mathbf{b}'$ represent a 3×3 deconvolution kernel set and a bias vector set in decoder sub-network respectively.Each input sample $\mathbf{r}^{(i)}$ is mapped into the corresponding representation $\mathbf{p}^{(i)}$ and then the reconstructed output $\mathbf{z}^{(i)}$. The parameters of this model are optimized to minimize the average reconstruction error,

$$\begin{aligned} \theta^*, \theta'^* &= \arg\min_{\theta,\theta'} \tfrac{1}{N} \sum_{i=1}^{N} \mathrm{L}\left(\mathbf{r}^{(i)}, \mathbf{z}^{(i)}\right) \\ &= \arg\min_{\theta,\theta'} \tfrac{1}{N} \sum_{i=1}^{N} \mathrm{L}\left(\mathbf{r}^{(i)}, g_{\theta'}\left(f_\theta\left(\mathbf{r}^{(i)}\right)\right)\right) \end{aligned} \tag{5}$$

where $\mathrm{L}()$ is the loss function representing conventional squared error $\mathrm{L}(\mathbf{r}, \mathbf{z}) = \|\mathbf{r} - \mathbf{z}\|^2$. N is the number of samples in the dataset. In this paper, considering data variation in practice, we firstly represent $\mathbf{r}^{(i)}$ and $\mathbf{z}^{(i)}$ as bit vectors ${\mathbf{r}_b}^{(i)}$ and ${\mathbf{z}_b}^{(i)}$, and then compute reconstructed cross-entropy using the loss function [49],

$$\mathrm{L_H}(\mathbf{r}, \mathbf{z}) = -\sum_{i=1}^{N} \left[{\mathbf{r}_b}^{(i)} \log {\mathbf{z}_b}^{(i)} + \left(1 - {\mathbf{r}_b}^{(i)}\right) \log\left(1 - {\mathbf{z}_b}^{(i)}\right) \right] \tag{6}$$

therefore, equation 5 with $L = L_H$ can be written as

$$\theta^*, \theta'^* = \arg\min_{\theta,\theta'} \mathrm{E}_{q^0(\mathbf{r})}\left[\mathrm{L}_H\left(\mathbf{r}, g_{\theta'}\left(f_\theta(\mathbf{r})\right)\right)\right] \tag{7}$$

where $q^0(\mathbf{r})$ denotes the empirical distribution corresponding to N training inputs.

However, in practice, the raw radar data are often influenced by many subtle factors such as micro motion caused by other body parts in the process of performing hand gestures. To improve hand recognition accuracy, the proposed network should be able to reconstruct a clean repaired data from noisy input. To get the noisy version input $\tilde{\mathbf{r}}$, we assume that the initial input $\mathbf{r}$ is corrupted by means of a stochastic mapping $\tilde{\mathbf{r}} \sim q_{\mathbf{N}}(\tilde{\mathbf{r}}|\mathbf{r})$. The noisy input $\tilde{\mathbf{r}}$ is transformed to a hidden representation $\mathbf{p} = f_\theta(\tilde{\mathbf{r}})$, and then we reconstruct $\mathbf{z} = g_{\theta'}(\mathbf{p})$, as shown in Figure 3. Different from that the parameters are trained to minimize the average reconstruction error described in equation 7, $\mathbf{z}$ is a deterministic function of

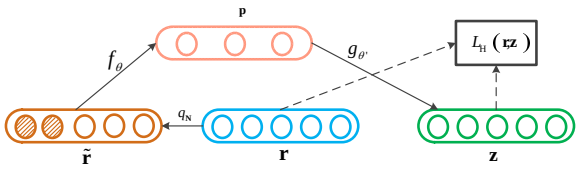

Fig. 3. A schematic representation of the process of auto-encoder network.

$\tilde{\mathbf{r}}$, which is the result of a stochastic mapping of $\mathbf{r}$. Therefore, we define a joint empirical distribution

$$q^0(\mathbf{r}, \tilde{\mathbf{r}}) = q^0(\mathbf{r})\, q_{\mathbf{N}}(\tilde{\mathbf{r}}|\mathbf{r}) \tag{8}$$

where $q^0(\mathbf{r})$ denotes the empirical distribution corresponding to N training inputs. At last, the objective function can be described as

$$\theta^*, \theta'^* = \arg\min_{\theta,\theta'} \mathrm{E}_{q^0(\mathbf{r},\tilde{\mathbf{r}})} \left[\mathrm{L}_H\left(\mathbf{r}, g_{\theta'}\left(f_\theta\left(\tilde{\mathbf{r}}\right)\right)\right)\right] \tag{9}$$

which is minimized by stochastic gradient descent algorithm [50]. According to above discussion, we can reconstruct the clean output from noisy input.

B. Network architecture

Inspired by the VGGNet architecture [41], we propose a convolutional auto-encoder network, in which the desired output tries to reconstruct the corresponding input range profile. The proposed network architecture consists of two parts, a convolutional encoder sub-network and a deconvolutional decoder sub-network, as shown in Figure 4. The former transforms the input radar range profile $\mathbf{r}$ to abstract feature representation $\mathbf{p}$, whereas the latter reconstructs the initial range profile from the abstract representation. Each convolutional layer in the encoder network has a corresponding deconvolutional layer in the decoder network.

1) Convolutional Encoder Network: As shown in Figure 4, the original radar range profile is input to a stack of convolutional blocks. Each convolutional layer is described as $(h, w, in\_channels, out\_channels)$, where h and w represent the height and width of convolutional filter respectively. $in\_channels$ and $out\_channels$ are the number of input data channels and output data channels. We note that the convolutional filters with small size of 3×3 are used in this system. The reasons are listed as follows: 1) the smallest kernel to seize patterns existing in different directions are 3×3 convolutional filters; 2) the convolutional result with a stack of two 3×3 convolutional layers is equivalent to the process with a 5×5 convolutional layer (input $M \times N$ range profile and output $(M-4) \times (N-4)$ feature map), and a stack of three such layers has the same effect with a 7×7 convolutional layer (input $M \times N$ range profile and output $(M-6) \times (N-6)$ feature map). What's more, compared with using a single one activation function in convolutional layer, the utilization of two or three non-linear functions in one convolutional block makes decision more discriminative; and 3) compared with the traditional large convolution kernel in a single layer, the stacked structure decreases the number of parameters. For instance, for a three-layer 3×3 convolution

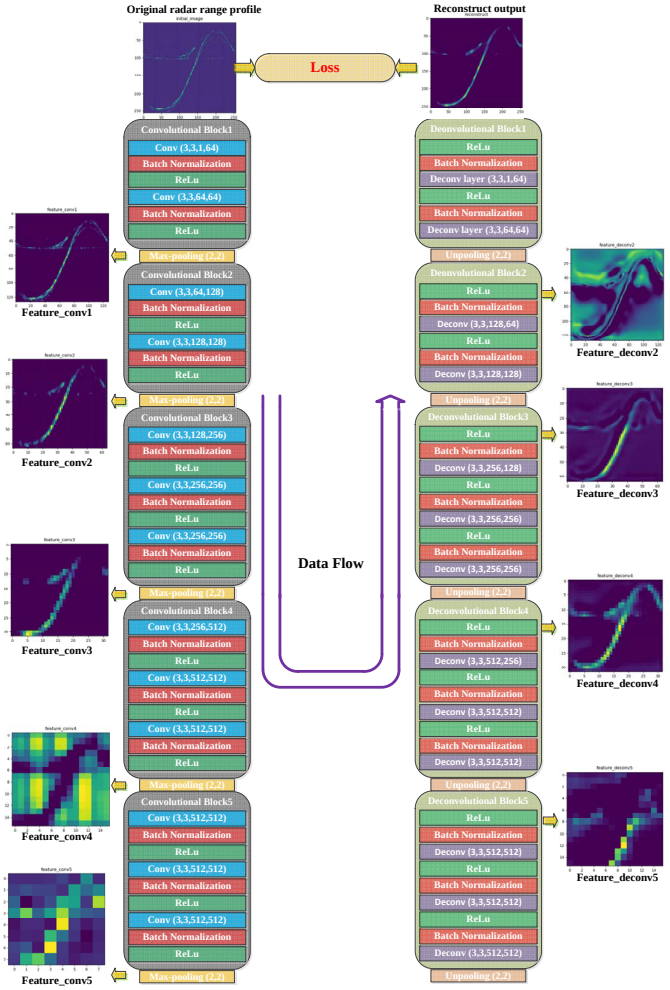

Fig. 4. Proposed network architecture.

stack, by assuming that there are C channels in the both input and output, $3 \times \left(3^2C^2\right) = 27C^2$ weights are required to parameterize the stack. However, in a single 7×7 convolutional layer, $\left(7^2C^2\right) = 49C^2$ parameters are needed. What's more, in the convolutional encoder network, the convolutional stride is fixed to 1 pixel and the padding is set to 1 for the stack of 3×3 convolutional layers. By performing several max-pooling layers with 2×2 pixel windows, the spatial pooling is achieved, which helps to reduce the dimension of learning features.

In a nutshell, the design of the convolutional encoder network follows these rules: 1) in each convolutional block, the size of feature map learned by different convolutional layers is same and 2) to preserve more discriminative information as far as possible, we increase the number of channels of feature maps in the deeper convolutional blocks. Rectified Linear Unit (ReLU) activation function is utilized in all layers in the convolutional network, which is defined as

$$f_{\mathrm{ReLU}}(x) = \max(0, x) \tag{10}$$

ReLU has advantages over traditional sigmoid and hyperbolic tangent functions, such as expediting convergence of the training course and resulting in better solutions. What's more,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2910810, IEEE Sensors Journal

batch normalization is utilized to accelerate deep network training, which allows us to use higher learning rates and being less careful about initialization [51].

2) Deconvolutional Decoder Network: By stacking convolutional layers and max-pooling layers, the convolutional encoder network is responsible for learning high-level abstract features from the input radar range profiles, spatially shrinking the feature maps layer by layer. But, in this process, the usage of pooling operation reduces the resolution of feature maps.

In this paper, deconvolutional decoder network is used to reconstruct the original input range profiles from the abstract features. The decoder network mainly contains deconvolutional layers and unpooling layers. Opposite to convolution operation, which generates relatively small feature maps from original range profiles, deconvolution focuses on restoring the original input from the small feature maps. As shown in Figure 5, it is easy to find that deconvolution can be regarded as a special convolution, which adds padding operation around the relatively small feature maps before performing convolution. The configuration of deconvolutional blocks is the same with convolutional blocks. 3×3 convolution kernel and activation function ReLU are used in the deconvolutional blocks.

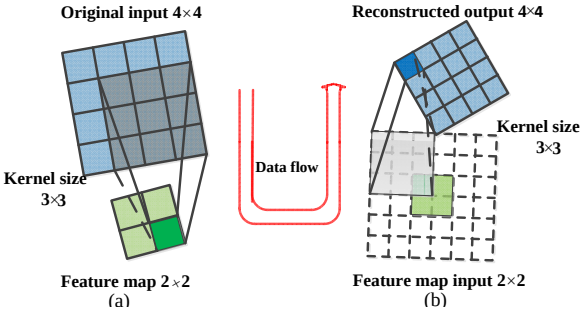

Fig. 5. Convolution VS Deconvolution. (a) Convolution: original input 4×4, kernel size 3×3, feature map size 2×2, stride 1×1 (b) Deconvolution: feature map input 2×2, kernel size 3×3, reconstructed output 4×4.

Another important step is unpooling operation. In order to map the encoded low resolution feature maps to a higher resolution feature representations, unpooling operation is used to increase the spatial span of the abstract feature maps learned from encoder network, which is in contrast to the pooling operation (shrinking the feature maps in spatial dimension). More specifically, there are two strategies used for performing unpooling operation, including max value duplication and zero padding at corresponding positions, as shown in Figure 6. Compared with the zero padding method, the max value duplication method does not require extra memory consumption to restore max-pooling indices. Thus, in our system, the max value duplication method is used to carry out the unpooling task. Corresponding to the pooling operation in the encoder network, unpooling operation is performed over 2×2 pixel windows with stride 1.

3) Training: The aforementioned convolutional auto-encoder network is implemented on Tensorflow designed by Google [52] and is trained on Geforce GTX 1080 Ti GPU with 11GB memory. For the proposed network architecture in Figure 4, the total number of parameters is 29188224,

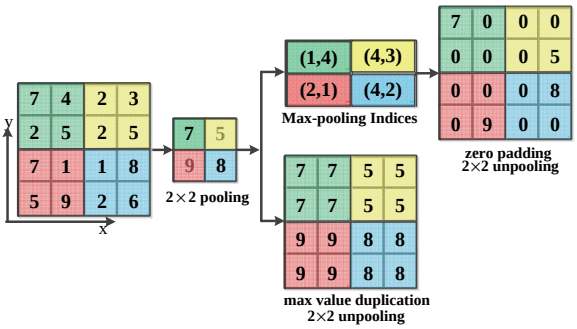

Fig. 6. Unpooling operation: zero padding and max value duplication.

wherein one half is associated with the encoder network and the other half is associated with the decoder network. As mentioned in Section III A, we use 64 chirps to get one range measurement. To get the full trajectory of hand movement, we measure the continuous range information for 5 seconds and obtain range profiles. Considering the balance between recognition accuracy, the number of training parameters, and hardware limitation, we save the profiles as JPG format and resize them to 256×256 by conducting centrally clipping on initial profiles. Then, 80% of the total 3200 samples are randomly selected for training and the rest are used for testing. In addition, 20% of the training samples are also randomly selected as a validation set. Finally, the Adam algorithm [53] is used to minimize the loss function and learning rate is set to 0.01. In addition, all weight matrices in the network are initialized in a uniform distribution in the range $[-0.1, 0.1]$. This training process is repeated 20 times and the results are averaged to obtain the loss reported in Figure 7. As shown in this figure, the both training error and validation error generally decline as the training procedure continues, and finally converge to a local minimum after 10 and 20 epochs respectively.

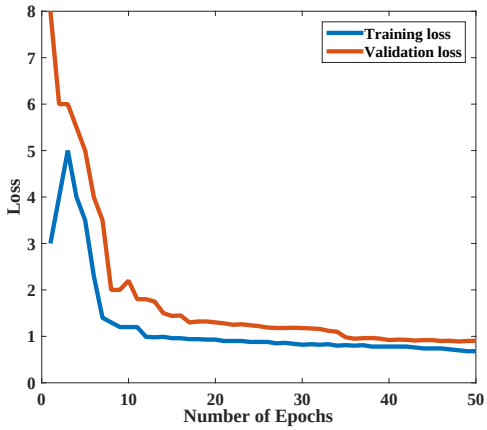

Fig. 7. Training and Validation Loss Curves.

4) Usage of learned features for classification by fine-tuning the network: Once the proposed network is trained, the convolutional network can be used as an effective feature extractor, since that the gestures' features with low spatial dimension are extracted by the internal layers in the

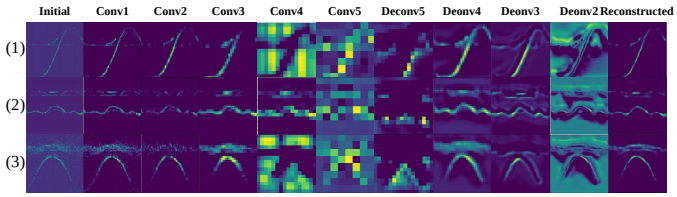

Fig. 8. Feature maps extracted from each layer in the proposed network:(1) pulling, (2) waiving hand, and (3) patting.

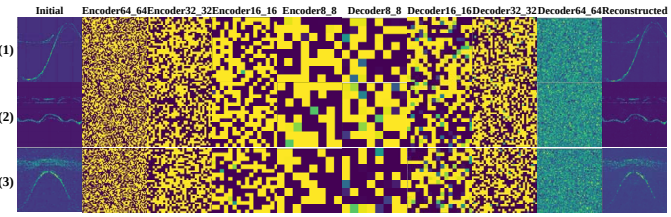

Fig. 9. Feature maps extracted from fully connected auto-encoder network(1) pulling, (2) waiving a hand, and (3) patting.

convolutional network. Different from traditional supervised learning methods that rely on a large-scale labeled training data, the proposed method does not need to use a large number of training samples to train a valid network for supervised classification. In contrast, the proposed network only needs to be fine-tuned with a limited number of training samples. For fine-tuning, the deconvolutional network is cut off, a fully connected layer with softmax function is introduced as the classifier, and a limited number of training samples is used to fine-tune this new layer.

V. Experimental Results

In this section, the performance of u-DeepHand system is verified based on the measured radar range profiles and compared with transfer learning VGGNet [26] and fully connected auto-encoder network [27].

- Transfer learning VGGNet: in [26], a pre-trained VGG16 network, which has been trained for general optical images, is fine-tuned using radar micro-Doppler spectrograms to maximize the classification accuracy of HGR problem. The network consists of 16 main layers; and thirteen of the layers are convolutional layers, followed by three fully connected layers. All of the convolution layers have the same 3×3 convolutional kernels.
- Fully connected auto-encoder network: authors in [27] propose an auto-encoder network that contains stacked fully connected hidden layers and a softmax layer for extracting sparse representations from radar range profiles. Considering GPU memory resource limitation, we design an auto-encoder network with four encoder layers and four decoder layers, according to the architecture of fully connected hidden layers described in [27], for exploring the internal mechanism in this auto-encoder network. All the radar range profiles are resized to 128×128 before processing.

A. Activation Layer Visualization

In order to investigate the 'black box' of the proposed convolutional auto-encoder network,we arbitrarily choose 3 kinds of gesture range profiles to show learning details. The learned feature maps from the encoder network and decoder network are analyzed in Figure 8. 'Initial', 'Conv1-5','Deconv5-2', and 'Reconstructed' represent the initial input radar range profiles, feature maps extracted from the first to fifth convolutional layers of encoder network, feature maps extracted from the fifth to first deconvolutional layers of decoder network, and reconstructed output respectively, as shown in Figure 4. It is easy to find that each convolutional filter focuses on the key contour features and the reconstructed outputs display more detailed features compared with the input range profiles.

To understand the features learned by different networks, we display four randomly selected feature maps of gestures extracted from transfer learning VGGNet and fully connected auto-encoder network, as shown in Figures 9 and 10 respectively. It can be observed that while transfer learning VGGNet extracts the key contour features, the contour details fade away when the layer becomes deeper, especially in the fourth layer. In addition, it is observed that different from the features extracted by convolution-based methods, the feature maps extracted by fully connected auto-encoder network do not reflect contour features. In the proposed network, the radar range profiles are simply reconstructed from the learned low-dimensional features through loop iterations. From this point of view, the convolutional auto-encoder network might be better for sparse feature representation.

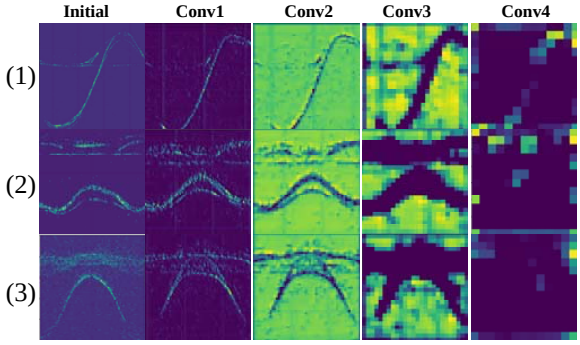

Fig. 10. Feature maps extracted from transfer learning VGGNet(1) pulling, (2) waiving hand, and (3) patting.

To better investigate how different networks process noisy signals, a gesture sample containing an unexpected fluctuation caused by the vibration of other body parts, is provided to each network, as shown in Figure 11. It is observed that in the transfer learning VGGNet, the unwanted fluctuation is mistakenly modeled as an effective feature and propagated to the last layer. The feature map of the last layer in the fully connected auto-encoder network remains a fraction of the unexpected fluctuation. In contrast, this fluctuation never appears in the reconstructed output of the proposed network, which indicates that the proposed network is more robust than the other two networks.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2910810, IEEE Sensors Journal

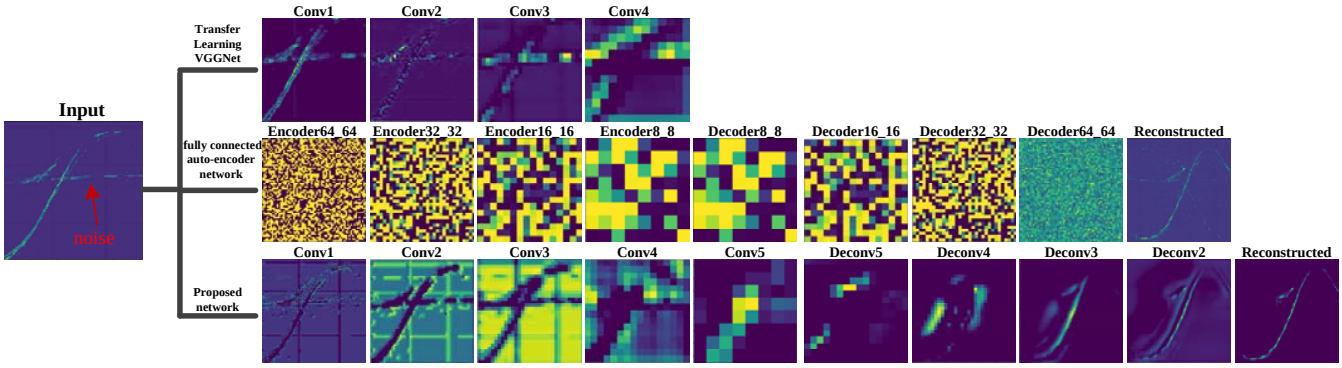

Fig. 11. Feature maps for noisy data with an unwanted fluctuation.

B. Classification Performance

We also compare the loss curves and classification performance between different networks. The overall recognition accuracy of each network is analyzed with respect to the number of epochs. In this process, for each class, the collected dataset is randomly divided into 80% for training and 20% for testing. The process is repeated 20 times and the corresponding average accuracy is obtained. From Figure 12, during the training process, it is observed that u-DeepHand and VGGNet have better convergence speeds, while the fully connected auto-encoder network has the smallest loss value. The reason is that it is easy for the fully connected layers to converge to a smaller loss value at the cost of more training parameters and training time compared with the convolutional layers. As shown in Figure 13, it can be observed that u-DeepHand has the highest recognition accuracy, yielding 4% improvement over the transfer learning VGGNet and 7% over the fully connected auto-encoder network. The results indicate that the convolutional auto-encoder is an effective method for network initialization.

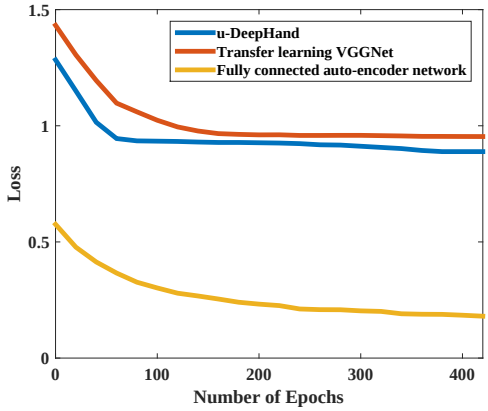

Fig. 12. Training loss curves of the different networks.

To study the impact of training sample size on recognition accuracy, we divide the whole training data into blocks with different size, varying from 100 to 2500 with 100 step size. The results in Figure 14 show that the transfer learning VGGNet yields over 10% improvement in performance when the number of training samples is less than 800, but levels off above 1000 samples. In contrast, the proposed network yields 5% performance improvement based on random initialization when the number of training samples is more than 1000. Though the transfer learning VGGNet has advantages for small training sets, the convolutional auto-encoder network yields greater accuracy with modest number of data, without the requirement of large scale of training samples for pre-training. This is due to the reason that the convolutional auto-encoder network features are directly extracted from radar range profiles, while the transfer learning VGGNet is affected by the labels of training samples. At last, for analysis of misclassification, the confusion matrix for all hand gestures is presented in Table 1.

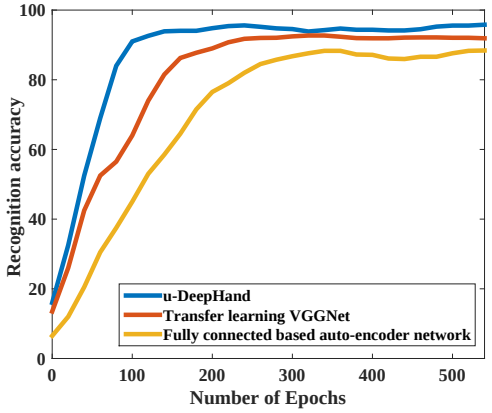

Fig. 13. Comparison of recognition accuracy of the different networks.

C. Effects of Testing Distance and Scale of Hand Gesture

Although there are many factors that may have influence on the recognition accuracy, in this work we study two main factors, including the testing distance, and the scale of hand gesture. In this experiment, we capture hand gestures with the same background in Figure 2 but different parameters.

1) Testing distance: Firstly, the samples collected with different distances are used to evaluate the impact of distance on recognition accuracy. In addition, considering that the signal amplitude depends not only on the testing distance, but also on the radar amplifier gain, we test the recognition accuracy with respect to different testing distances and am-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2910810, IEEE Sensors Journal

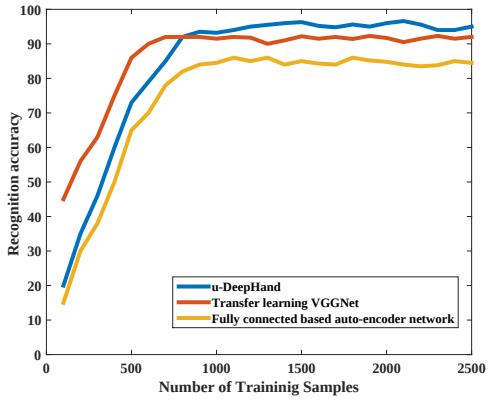

Fig. 14. Test accuracy versus training sample size for the different networks.

TABLE I
CONFUSION MATRIX FOR ALL THE HAND GESTURES. (A) SLIDING HAND FROM RIGHT TO LEFT, (B) SLIDING HAND FROM LEFT TO RIGHT, (C) PULLING, (D) PUSHING, (E) KNOCKING, (F) MOVING HAND UP AND DOWN, (G) WAIVING HAND, AND (H) PATTING.

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
(a)	94%	2%	1%	1%	0%	0%	2%	0%
(b)	4%	92%	0%	0%	1%	0%	1%	2%
(c)	0%	3%	97%	0%	0%	0%	0%	0%
(d)	0%	0%	0%	92.5%	3.5%	0%	2%	2%
(e)	1%	0%	0%	1%	96%	2%	0%	0%
(f)	3%	2%	0%	0%	1%	92%	2%	0%
(g)	2%	2%	0%	0%	0%	0%	96%	0%
(h)	0%	0%	0%	2%	2%	1%	0%	95%

plifier gains, as shown in Figure 15. It is observed that the recognition accuracy is stable when the testing distance is within 1.3m, 2.2m, and 3m for 0dB gain, 20dB gain, and 40db gain respectively. This result can be interpreted by the fact that smaller gains lead to small effective number of bits in ADC device or differential nonlinearity of converter limits the performance when signals are much lower than full-scale of an ADC device.

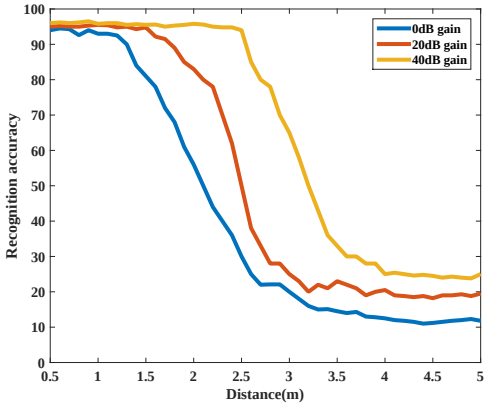

Fig. 15. Recognition accuracy with respect to measured distance under different amplifier gain.

2) Scale of hand gesture: We also consider how the different scales of hand gesture affects classification accuracy. In this experiment, we conduct the hand gesture experiments with three different scales $r = [0.1\text{m}, 0.3\text{m}, 0.5\text{m}]$ at the same distance $d = 1.5\text{m}$. As shown in Figure 16, the recognition accuracy is 95%, 87% and 76% for the scale of 0.5m, 0.3m, and 0.1m respectively. The reason why the recognition accuracy declines with the motion scale reducing is that reducing motion scale equates to reducing radial movement distance. The trajectory of hand gesture cannot be identified when the radial movement distance is below the range resolution. In addition, compared with hand gestures with a small scale, the larger ones contain more detailed motion information.

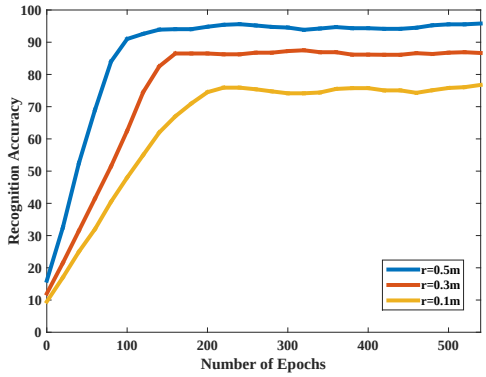

Fig. 16. Result of test recognition accuracy varying with the number of epoch with different hand gesture scales.

3) Leave-One-Out Cross Validation: To alleviate the concern of overfitting problem, we rely on Leave-One-Out (LOO) cross-validation scheme to test u-DeepHand system. In this experiment, one volunteers gesture data is taken out and the network is trained by using all the rest dataset. Firstly, to test the robustness to different volunteers, we collect the gesture data performed by the fifth volunteer under the same condition without being trained how to perform gestures. From Figure 17, the average recognition accuracy is 92% based on the previous four volunteers gesture data. However, the accuracy is reduced to 82.3% using the fifth volunteers data, which is due to feature differences existing in the process of performing gestures, such as the scale and duration time. At last, the gesture data from the fifth volunteer are included into the training set. The recognition accuracy increases to 95% after retraining the dataset, as shown in Figure 18, which indicates that the diversity of gesture data is helpful to alleviate overfitting problem.

VI. CONCLUSION

In this paper, a novel end-to-end deep convolutional auto-encoder network architecture is proposed for unsupervised feature extraction of radar range profiles. In particular, the proposed network is composed of a convolutional encoder network and a deconvolutional decoder network. They are responsible for transforming an input radar range profile to abstract feature representation and reproducing the initial input data from the encoded features respectively. In addition, to evaluate the performance of the proposed architecture, the hand gesture feature maps extracted from the proposed network as well as from the other two state-of-the-art networks, including the transfer learning VGGNet and fully connected base auto-encoder network, are analyzed in detail. It is observed that the

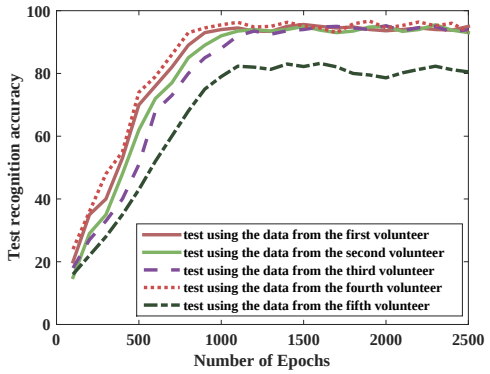

Fig. 17. Result of test recognition accuracy varying with the number of epochs in leave-one cross validation.

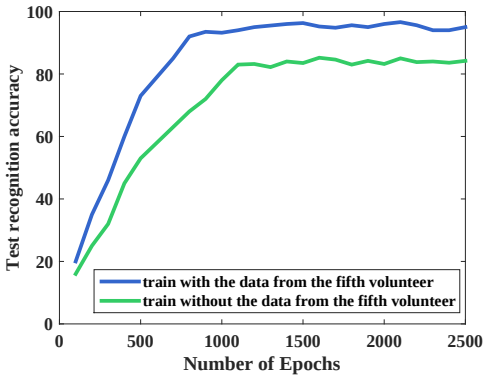

Fig. 18. Result of test recognition accuracy varying with the number of epochs with and without the data from the fifth volunteer.

convolutional auto-encoder network is better at sparse feature representation compared with the other two networks. In addition, for modest number of training datasets, unsupervised pre-training enables the convolutional auto-encoder network to learn effective representations robustly and with high accuracy. At last, the effects of testing distance and scale of hand gesture on the recognition accuracy are investigated in this paper. Experimental results demonstrate that the features learned by the proposed unsupervised network can be used for the hand gesture classification task, and a high classification accuracy is obtained. In future, considering that the continuous gestures contain not only spatial features but also temporal features, we plan to explore the feasibility of learning spatial-temporal features using unsupervised methods for continuous HGR. Then, we also plan to study multi-object tracing technology to solve multi-hand gesture recognition problem. Whats more, it is worth exploring the minimal necessary radar power for hand gesture recognition application by studying the method of measuring radar cross section for diverse hand gestures.

ACKNOWLEDGMENT

Many thanks are given to the reviewers for the careful review and valuable suggestions. This work was supported in part by the National Natural Science Foundation of China (61771083, 61704015), Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), Special Fund of Chongqing Key Laboratory (CSTC), Fundamental and Frontier Research Project of Chongqing (cstc2017jcyjAX0380, cstc2015jcyjBX0065), and University Outstanding Achievement Transformation Project of Chongqing (KJZH17117).

REFERENCES

[1] Singh G, Nelson A, et al, "Inviz: Low-power personalized gesture recognition using wearable textile capacitive sensor arrays.", *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2015, pp. 198–206.

[2] Hoshino K. "Hand Gesture Interface for Entertainment Games." *Handbook of Digital Games and Entertainment Technologies*, pp: 293–312, 2017.

[3] Lee D H, Hong K S. "Game interface using hand gesture recognition." *2010 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT)*, IEEE, 2010, pp. 1092–1097.

[4] Ohn-Bar E, Trivedi M M. "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations." *IEEE transactions on intelligent transportation systems*, vol. 15, no. 6, pp. 2368–2377, 2014.

[5] Gupta S, Molchanov P, Yang X, et al. "Towards selecting robust hand gestures for automotive interfaces." *Intelligent Vehicles Symposium (IV)*, IEEE, 2016, pp. 1350–1357.

[6] Smith K A, Csech C, Murdoch D, et al. "Gesture Recognition Using mm-Wave Sensor for Human-Car Interface." *IEEE Sensors Letters*, vol. 2, no. 2, pp. 1–4, 2018.

[7] Tang M C, Chen C L, Lin M H, et al. "A hybrid computer vision and Wi-Fi Doppler radar system for capturing the 3-D hand gesture trajectory with a smartphone." *2017 IEEE MTT-S International Microwave Symposium (IMS)*, IEEE, 2017, pp. 1251–1254.

[8] L Ge, H Liang, et al. "3d convolutional neural networks for efficient and robust hand pose estimation from single depth images." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2017, pp. 1–10.

[9] L Ge, H Liang, et al. "Robust 3d hand pose estimation in single depth images: from single-view cnn to multi-view cnns." *Proceedings of the IEEE Conference on computer vision and pattern recognition*, IEEE, 2016, pp. 3593–3601.

[10] Sridhar S, Mueller F, et al. "Real-time joint tracking of a hand manipulating an object from rgb-d input." *European Conference on Computer Vision*. Springer, 2016, pp. 294–310.

[11] Kazerooni H, Fairbanks D, Chen A, and Shin G "The magic glove". *Proceedings of the IEEE Conference on Robotics and Automation*, IEEE, 2014, pp. 757–763.

[12] Chen K Y, et al. "uTrack: 3D input using two magnetic sensors." *Proceedings of the 26th annual ACM symposium on User interface software and technology*, ACM, 2013, pp. 237–244.

[13] C Xu, PH Pathak, et al. "Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch." *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, ACM, 2015, pp. 9–14.

[14] T Fan, C Ma, Z Gu, et al. "Wireless hand gesture recognition based on continuous-wave Doppler radar sensors." *IEEE Transactions on Microwave Theory and Techniques*, IEEE, vol. 64, no. 11, pp. 4012–4020, 2016.

[15] H Abdelnasser, M Youssef, KA Harras. "Wigest: A ubiquitous wifi-based gesture recognition system." *IEEE Conference on Computer Communications (INFOCOM)*, IEEE, 2015, pp. 1472–1480.

[16] T Fan, D Ye, J Hangfu, et al. "Hand gesture recognition based on Wi-Fi chipsets." *IEEE Radio and Wireless Symposium (RWS)*, IEEE, 2017, pp. 98–100.

[17] J Lien, N Gillian, et al. "Soli: Ubiquitous gesture sensing with millimeter wave radar." *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–19, 2016.

[18] Kim Y, Toomajian B. "Hand gesture recognition using micro-Doppler signatures with convolutional neural network." *IEEE Access*, vol. 4, pp. 7125-7130, 2016.

[19] ST Huang, CH Tseng, et al. "Hand-gesture sensing Doppler radar with metamaterial-based leaky-wave antennas." *2017 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*, IEEE, 2017, pp. 49–52.

[20] ST Huang, CH Tseng, et al. "Hand-gesture sensing Doppler radar with metamaterial-based leaky-wave antennas." *2017 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*, IEEE, 2017, pp. 49–52.

[21] Li, Gang, et al. "Sparsity-based dynamic hand gesture recognition using micro-Doppler signatures." *2017 IEEE Radar Conference (RadarConf)*, IEEE, 2017, pp. 0928–0931.

[22] Y Kim, B Toomajian. "Application of Doppler radar for the recognition of hand gestures using optimized deep convolutional neural networks." *2017 11th European Conference on Antennas and Propagation (EUCAP)*, IEEE, 2017, pp. 1258–1260.

[23] Molchanov P, Gupta S, Kim K, et al. "Short-range FMCW monopulse radar for hand-gesture sensing." *2015 IEEE Radar Conference (RadarCon)*, IEEE, 2015, pp. 1491–1496.

[24] KN Parashar, MC Oveneke, et al. "Micro-Doppler feature extraction using convolutional auto-encoders for low latency target classification." *2017 IEEE Radar Conference (RadarConf)*, IEEE, 2017, pp. 1739–1744.

[25] Y Bengio, Y LeCun. "Scaling learning algorithms towards AI." *Large-scale kernel machines*, vol. 34, no. 5, pp. 1–41, 2007.

[26] I Alnujaim, H Alali, F Khan, et al. "Hand Gesture Recognition Using Input Impedance Variation of Two Antennas with Transfer Learning." *IEEE Sensors Journal*, vol. 18, no. 10, pp. 4129–4135, 2018.

[27] B Jokanovic, M Amin, et al. "Radar fall motion detection using deep learning." *2016 IEEE Radar Conference (RadarConf)*, IEEE, 2016, pp. 1–6.

[28] Park J, Cho S H. "IR-UWB Radar Sensor for Human Gesture Recognition by Using Machine Learning." *IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, IEEE, 2016, pp. 1246–1249.

[29] Malysa G, Wang D, Netsch L, et al. "Hidden Markov model-based gesture recognition with FMCW radar." *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, IEEE, 2016, pp. 1017–1021.

[30] Zhang S, Li G, Ritchie M, et al. "Dynamic hand gesture classification based on radar micro-Doppler signatures." *2016 CIE International Conference on Radar (RADAR)*, IEEE, 2016, pp. 1–4.

[31] Gao X, Xu J, Rahman A, et al. "Barcode based hand gesture classification using AC coupled quadrature Doppler radar." *2016 IEEE MTT-S International Microwave Symposium (IMS)*, IEEE, 2016, pp. 1–4.

[32] Kim Y, Moon T. "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks." *IEEE geoscience and remote sensing letters*, vol. 13, no. 1, pp. 8–12, 2016.

[33] Ryu, Si-Jung, et al. "Feature-based hand gesture recognition using an FMCW radar and its temporal feature analysis." *IEEE Sensors Journal*, vol. 18, no. 18, pp. 7593–7602, 2018.

[34] Yang Le, and Gang Li. "Sparsity aware dynamic gesture recognition using radar sensors with angular diversity." *IET Radar, Sonar & Navigation*, vol. 12, no. 10, pp. 1114–1120, 2018.

[35] Li Gang, et al. "Effect of sparsity-aware timeCfrequency analysis on dynamic hand gesture classification with radar micro-Doppler signatures." *IET Radar, Sonar & Navigation*, vol. 12, no. 8, pp. 815–820, 2018.

[36] Sang, Yu, Laixi Shi, and Yimin Liu. "Micro hand gesture recognition system using ultrasonic active sensing." *IEEE Access*, vol. 6, pp. 49339–49347, 2018

[37] Wang Fu-Kang, et al. "Gesture sensing using retransmitted wireless communication signals based on Doppler radar technology." *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 12, pp. 4592–4602, 2016.

[38] Wang S, Song J, Lien J, et al. "Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum." *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, ACM, 2016, pp. 851–860.

[39] Molchanov P, Gupta S, Kim K, et al. "Multi-sensor system for driver's hand-gesture recognition." *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, IEEE, 2015, pp. 1–8.

[40] Pan S J, Yang Q. "A survey on transfer learning." *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[41] K Simonyan, A Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint*, arXiv:1409.1556, 2014.

[42] Al Hadhrami E, Al Mufti M, Taha B, et al. "Transfer learning with convolutional neural networks for moving target classification with micro-Doppler radar spectrograms." *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*, IEEE, 2018, pp. 148–154.

[43] Krizhevsky A, Sutskever I, Hinton G E. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*, pp. 1907–1105, 2012.

[44] Yin J, Tran S N, Zhang Q. "Human Identification via Unsupervised Feature Learning from UWB Radar Data." *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, 2018, pp. 322–334.

[45] Zhang Z, Tian Z, Zhou M. "Latern: Dynamic Continuous Hand Gesture Recognition Using FMCW Radar Sensor." *IEEE Sensors Journal*, vol. 18, no. 8, pp. 3278–3289, 2018.

[46] A Wojtkiewicz, J Misiurewicz, et al. "Two-dimensional signal processing in FMCW radars." *Proc. XX KKTOiUE*, pp. 475–480, 1997.

[47] GM Brooker. "Understanding millimetre wave FMCW radars." *1st international Conference on Sensing Technology*, 2005, pp. 152–157.

[48] Meta A, Hoogeboom P, et al. "Range non-linearities correction in FMCW SAR." *IEEE International Conference on Geoscience and Remote Sensing Symposium*, IEEE, 2006, pp. 403–406.

[49] Nielsen, Michael A. "Neural networks and deep learning." *USA: Determination press*, Vol. 25, pp. 59–70, 2015.

[50] Bottou L. "Large-scale machine learning with stochastic gradient descent." Proceedings of COMPSTAT'2010. Physica-Verlag HD, 2010. 177-186.

[51] S Ioffe, C Szegedy, et al. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *arXiv preprint*, arXiv:1502.03167, 2015.

[52] Abadi, Martın, et al. "Tensorflow: a system for large-scale machine learning.", OSDI. Vol. 16, pp. 265–283, 2016.

[53] Kingma D P, Ba J. "Adam: A method for stochastic optimization." *arXiv preprint*, arXiv:1412.6980, 2014.