

## 基于FMCW雷达的多维参数手势识别算法

王 勇 吴金君\* 田增山 周 牧 王沙沙

(重庆邮电大学通信与信息工程学院 重庆 400065)

**摘 要:** 该文提出一种基于调频连续波(FMCW)雷达多维参数的卷积神经网络手势识别方法。通过对雷达信号进行时频分析,估计手势目标的距离、多普勒和角度参数,构建出手势动作的多维参数数据集。同时,为了进行手势特征提取和精确分类,提出多分支网络结构和高维特征融合的方案,设计出具有端到端结构的RDA-T多维参数卷积神经网络。实验结果表明,结合手势动作的距离、多普勒和角度信息进行多维参数学习,所提方法有效解决了单维参数手势识别方法中手势描述信息量低的问题,且手势识别准确率相较于单参数方法提高了5%~8%。

**关键词:** FMCW雷达; 手势识别; 深度学习; 卷积神经网络

中图分类号: TN958; TN98

文献标识码: A

文章编号: 1009-5896(2019)04-0822-08

DOI: [10.11999/JEIT180485](https://doi.org/10.11999/JEIT180485)

## Gesture Recognition with Multi-dimensional Parameter Using FMCW Radar

WANG Yong WU Jinjun TIAN Zengshan ZHOU Mu WANG Shasha

(School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

**Abstract:** A multi-parameter convolutional neural network method is proposed for gesture recognition based on Frequency Modulated Continuous Wave (FMCW) radar. A multidimensional parameter dataset is constructed for gestures by performing time-frequency analysis of the radar signal to estimate the distance, Doppler and angle parameters of the gesture target. To realize feature extraction and classification accurately, an end-to-end structured Range-Doppler-Angle of Time (RDA-T) multi-dimensional parameter convolutional neural network scheme is further proposed using multi-branch network structure and high-dimensional feature fusion. The experimental results reveal that using the combined gestures information of distance, Doppler and angle for multi-parameter learning, the proposed scheme resolves the problem of low information quantity of single-dimensional gesture recognition methods, and its accuracy outperforms the single-dimensional methods in terms of gesture recognition by 5%~8%.

**Key words:** FMCW radar; Gesture recognition; Deep learning; Convolutional Neural Network (CNN)

收稿日期: 2018-05-21; 改回日期: 2018-08-30; 网络出版: 2018-09-13

\*通信作者: 吴金君 xnwujj@foxmail.com

基金项目: 国家自然科学基金(61771083, 61704015), 长江学者和创新团队发展计划基金(IRT1299), 重庆市科委重点实验室专项经费基金, 重庆市基础与前沿研究计划基金(cstc2017jcyjAX0380, cstc2015jcyjBX0065), 重庆市高校优秀成果转化基金(KJZH17117), 重庆市教委科学技术研究项目(KJ1704083)

Foundation Items: The National Natural Science Foundation of China (61771083, 61704015), The Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), The Special Fund of Chongqing Key Laboratory (CSTC), The Fundamental and Frontier Research Project of Chongqing (cstc2017jcyjAX0380, cstc2015jcyjBX0065), The University Outstanding Achievement Transformation Project of Chongqing (KJZH17117), The Scientific and Technological Research Foundation of Chongqing Municipal Education Commission (KJ1704083)

## 1 引言

目前,常用的人机交互方式依赖于键盘、鼠标等输入设备,无法实现简单、高效的信息互动。而手势可以直观、有效地表达信息,并且其信息可以传达到视觉系统。近年来,由于电子产品的普及和发展,手势识别技术已成为了国内外的研究热点<sup>[1-4]</sup>。

传统的手势识别方法主要利用光学摄像头<sup>[1,2]</sup>和深度摄像头<sup>[3,4]</sup>进行数据采集。光学图像和深度图像可以描述手势纹理、轮廓、形状等信息,但在强光或昏暗光线下效果较差。而雷达传感器可以解决由于光线不佳而导致手势识别精度低下的问题,并可保护用户隐私。因此,调频连续波(FMCW)雷达在手势识别的应用上具有非常可观的研究前景<sup>[5,6]</sup>。文献<sup>[7]</sup>中讨论了FMCW信号的混频和调制原理。文献<sup>[8]</sup>中采用OMP(Orthogonal Matching Pursuit)算法分析FMCW雷达回波信号,得到微多普勒时频轨迹后,用K-means对数据进行聚类,再采用基于修正Hausdorff距离的邻近分类器对手势分类。Zhang等人<sup>[9]</sup>利用短时傅里叶变换(Short-Time Fourier Transform, STFT)对雷达回波信号进行时频分析,随后将多普勒频偏的正负比率和手势动作时长人工构造为特征,并利用支持向量机进行分类。Kim等人<sup>[10]</sup>利用STFT求出雷达信号的多普勒谱,并将多普勒谱图像作为卷积神经网络的输入数据进行手势识别研究。文献<sup>[5]</sup>将多普勒-时间谱的实部和虚部作为两个通道的输入数据,同样采用卷积层、池化层和全连接层叠加的网络结构对图像进行特征提取,并用softmax分类器完成手势分类。文献<sup>[11]</sup>介绍了用于车载手势识别FMCW雷达的距离-多普勒图的估计方法,以及将雷达传感器和深度传感器数据进行融合的方法;在文献<sup>[6]</sup>中提出图像、深度以及雷达多个传感器进行联合,通过不同传感器为手势识别系统补充信息。谷歌Solli项目<sup>[12]</sup>中采用带宽高达7 GHz的毫米波雷达进行手势识别研究,利用卷积神经网络将连续时刻的距离-多普勒像逐帧特征提取,再利用循环神经网络对时序特征进行提取后进行手势分类。AlexNet网络<sup>[13]</sup>在ILSVRC 2012比赛中获得了冠军,证明了卷积神经网络对图像数据具有强大的学习能力。后续提出的VGGNet<sup>[14]</sup>, ResNet<sup>[15]</sup>, InceptionNet<sup>[16-18]</sup>等对卷积神经网络的结构做出了改进和优化。文献<sup>[19]</sup>提出了3维卷积神经网络,将视频数据构造成具有长、宽、时间3个维度的矩阵后对其进行学习。文献<sup>[20]</sup>提出采用高精度和低精度两个分支网络进行手势识别,并利用分类结果进行决策融合的方案。文献<sup>[21]</sup>提出利用两个分支网络对空间信息和时序

信息分别进行学习的网络结构,并通过融合空间和时序特征来构建网络。

然而,现有手势识别方法存在以下问题。第一,现有的基于图像、红外、深度传感器的手势识别方法<sup>[1-4]</sup>易受光照影响,尤其是在强光侵蚀下无法正常工作,并且由于上述传感器数据可视化的特点,该方法易泄露用户隐私;第二,现有雷达手势识别研究的信号分析方法中,存在缺乏对手势运动中角度变化信息的分析和利用的问题。例如,文献<sup>[6-9]</sup>仅利用了手势动作的多普勒变化信息,而文献<sup>[10-12]</sup>仅利用了手势运动的距离和速度变化信息;第三,现有手势识别研究的数据集构建和卷积神经网络结构设计中,存在输入数据提供的信息量不足和传统卷积神经网络结构数据输入维度有限的问题。如文献<sup>[8,9]</sup>仅采用了2维卷积神经网络对多普勒-时间谱1维数据进行特征提取,文献<sup>[19]</sup>仅采用了3维卷积神经网络对距离-多普勒2维数据进行特征提取。因此,已有基于传统卷积神经网络结构的研究受限于单一的输入数据,无法拓展输入数据的维度。

针对上述问题,本文提出了基于FMCW雷达的多维参数手势识别算法。基于FMCW毫米波雷达信号分析获取手势动作数据的方法,克服了如黑暗或强光侵蚀等非正常光照的不利影响;同时,FMCW毫米波雷达利用2维快速傅里叶变换(2-Dimensional Fast Fourier Transform, 2D-FFT)估计手势目标距离、多普勒参数,并利用多重信号分类(MULTIPLE SIGNAL CLASSIFICATION, MUSIC)算法<sup>[22]</sup>估计手势目标角度参数,手势动作数据以距离、速度以及角度参数变化的形式呈现,避免了用户隐私泄露的问题;更重要的是,本文构建了多维手势参数数据集,并设计了可实现多维参数特征提取的多分支网络结构。因此,本文的创新点可归纳为:

(1)基于雷达信号处理和分析,构造出距离-时间图(Range-Time-Map, RTM)、多普勒-时间图(Doppler-Time-Map, DTM)和角度-时间图(Angle-Time-Map, ATM),将同一手势动作的3种数据同步并构建手势动作多维参数数据集,拓展了描述手势动作的信息维度,解决了输入数据单一而导致的信息量不足的问题。

(2)针对RTM, DTM和ATM 3种数据类型构建的多维参数数据集,本文设计了RDA-T(Range-Doppler-Angle of Time)多维参数卷积神经网络,克服了传统卷积神经网络结构中输入数据维度拓展困难的问题。

## 2 系统描述

本文提出的基于FMCW雷达的多维参数手势识别系统,主要包括参数估计、数据集构建、网络训练和手势分类。其中,参数估计部分研究了手势目标的距离、多普勒和角度参数;数据集构建部分讨论了多维参数网络所需的多维参数数据集;网络训练和手势分类部分分别讨论了基于多维参数网络对手势数据的训练和分类。

### 2.1 距离、多普勒和角度联合估计

本文采用FMCW毫米波雷达进行信号采集,基于对FMCW雷达中频信号的处理和分析进一步估计手势目标距离、多普勒和角度参数。本文所采用FMCW雷达信号波形为锯齿波,其发射信号和接收信号为

$$s_T(t) = A_T \cos 2\pi \left[ f_c t + \int_0^t f_T(\tau) d\tau \right] \quad (1)$$

$$s_R(t) = A_R \cos 2\pi \left[ f_c(t - \Delta t_d) + \int_0^t f_R(\tau) d\tau \right] \quad (2)$$

$$f_T(\tau) = \tau \frac{B}{T}, f_R(\tau) = (\tau - \Delta t_d) \frac{B}{T} + \Delta f_d \quad (3)$$

其中,  $A_T$  和  $A_R$  表示发射信号和接收信号的幅值,  $f_c$  是载波的中心频率,  $\Delta t_d$  表示从发射信号发出到接收到回波信号的飞行时延。  $f_T(\tau)$  和  $f_R(\tau)$  分别表示发射信号和接收信号频率,  $B$  是信号带宽,  $T$  是线性调频信号的扫频周期,  $\Delta f_d$  表示多普勒频移。将  $s_R(t)$  和  $s_T(t)$  混频,并通过低频滤波器得到中频信号  $s_{IF}(t)$  为

$$s_{IF}(t) = f_{LPF} \{ s_T(t) s_R(t) \} = \frac{1}{2} A_T A_R \cdot \cos 2\pi \left[ f_c \Delta t_d + \left( \frac{B}{T} \Delta t_d - \Delta f_d \right) t \right] \quad (4)$$

理论上, FMCW雷达目标距离为  $R = c \cdot \Delta t_d / 2$ , 其中  $c$  表示光速。但由于  $\Delta t_d$  难以测量, 实际应用中一般采用中频信号频点  $f_{IF}$  来代替  $\Delta t_d$  的估计。其理论依据在于, FMCW雷达目标对应的频点  $f_{IF}$  与该距离处回波信号飞行时延  $\Delta t_d$  之比恰好等于 FMCW雷达锯齿波信号的斜率  $B/T$ , 其关系为  $B/T = f_{IF} / \Delta t_d$ 。进一步地, 根据  $\Delta t_d = 2R/c$  可导出距离估计值为  $R = c T f_{IF} / 2B$ 。于是,  $s_{IF}(t)$  的相位为

$$\varphi = 2\pi [f_c \Delta t_d + (f_{IF} - \Delta f_d) t] \quad (5)$$

其中,  $\varphi$  中包含着  $f_{IF}$  和  $\Delta f_d$ , 时延  $\Delta t_d$  的影响可忽略不计。因此, 可通过频域分析对雷达手势目标的距离和多普勒进行估计。

本文每隔一帧对手势目标距离和多普勒参数估

计一次。一帧数据包含  $N = 128$  个扫频, 每个扫频时长  $T = 40$  ms。在快时间域(单个扫频时间)内, 对扫频信号FFT分析得到中频信号频谱, 通过谱峰搜索求出目标对应的频点估计距离值。在慢时间域(多个扫频)内, 多普勒的估计需要将多个快时间域的频谱估计结果进行累积, 然后在累积结果的慢时间域尺度上进行FFT分析, 从而得到距离-多普勒耦合估计结果(距离-多普勒图)。根据上文所述的距离求解结果, 可从距离-多普勒图提取出手势目标运动的多普勒谱, 通过谱峰搜索可求出目标运动所引起的多普勒频移。

假设雷达前方存在  $K$  个目标, 则雷达中频信号表示为

$$s_{IF}(t) = \sum_{k=1}^K A^{(k)} e^{j2\pi [f_c \Delta t_d^{(k)} + (f_{IF}^{(k)} - \Delta f_d^{(k)}) t]} \quad (6)$$

其中,  $k$  表示FMCW雷达量程内第  $k$  个距离单元处所对应的目标。

本文采用的雷达有  $N_T = 2$  个发射天线和  $N_R = 4$  个接收天线, 通过时分复用得到8个虚拟接收天线阵元。考虑到雷达信号受到噪声影响, 则信号模型为

$$s(m, t) = s_{IF}(m, t) + n(m, t) \quad (7)$$

其中,  $m = 1, 2, \dots, 8$  表示不同接收天线阵元,  $s_{IF}(m, t)$  和  $n(m, t)$  分别表示第  $m$  路的信号分量和噪声分量。然后, 以  $F_s$  的采样率对  $s(m, t)$  采样后得到离散信号:

$$s(m, l) = \sum_{k=1}^K A^{(k)} e^{j2\pi [f_c \Delta t_d^{(k)} - \frac{l}{F_s} \Delta f_d^{(k)} + f_{IF}^{(k)}] t} + n(l) \quad (8)$$

其中,  $l = 0, 1, 2, \dots, L - 1$ , 且  $L = T \cdot F_s$ 。于是可构造信号向量矩阵, 并使用MUSIC算法进行角度估计。

### 2.2 手势动作多维参数数据集构建

为了将距离、多普勒和角度参数进行多帧累积, 它们分别与时间耦合后得到RTM, DTM和ATM。需要从两个方面考虑和确定手势动作的观测时长: 第一, 为获取完整手势, 观测时长需要大于手势动作时长; 第二, 为了避免过多的冗余空闲时间占据观测时长, 并尽可能减少数据处理的时间开销。本文采用的雷达参数中, FMCW雷达每帧有128个扫频信号, 每帧持续时间为40 ms。实际中, 手势动作时长通常为0.5~1.0 s。本文以32帧数据(持续时长约为1.3 s)为观测时长对雷达信号进行分析。

手势目标在一帧时间内发生的距离变化可以忽略不计。在一帧数据内, 每个快时间域均可求出其对应的频谱(距离估计谱), 在帧内求平均后得到该

帧数据所对应的距离估计谱，将多帧距离估计谱进行累积后得到RTM。基于一帧数据进行2D-FFT分析得到距离-多普勒图，根据手势目标所处的距离单元提取出多普勒估计谱，将多帧多普勒估计谱进行累积后得到DTM。基于一帧数据内8路信号所构成的信号向量矩阵，利用MUSIC算法对手势目标搜索得到角度估计谱，并将多帧角度估计谱进行累积后得到ATM。

由于RTM, DTM和ATM数据集中图像样本的数值上存在较大差异性，使得卷积神经网络的训练难以收敛。因此需要对数据进行归一化处理，首先对每张图像数据进行数值缩放。

$$\bar{x}_{i,j} = \frac{x_{i,j} - \left[ \frac{\sum_{r=1}^R \sum_{l=1}^L x_{r,l}}{R \cdot L} \right]}{\max\{\mathbf{X}\} - \min\{\mathbf{X}\}} \quad (9)$$

其中，具有 $R$ 行和 $L$ 列的像素点构成了矩阵 $\mathbf{X}$ ， $x_{i,j}$ 表示原始像素值， $\bar{x}_{i,j}$ 表示更新后的像素值。数值缩放后，图像中每一个像素值均在 $[-1,1]$ 区间内。其次，对数据集中每个图像样本 $\mathbf{X}_n$ 逐样本均值消减：

$$\bar{\mathbf{X}}_n = \mathbf{X}_n - \left( \frac{\sum_{i=1}^N \mathbf{X}_i}{N} \right) \quad (10)$$

其中， $\mathbf{X}_n$ 表示原始图像样本， $\bar{\mathbf{X}}_n$ 表示均值消减后的图像样本， $N$ 表示数据集中样本个数。最后，为了使数据集中图像尺寸保持一致，需要将每张图像尺寸均缩放为 $128 \times 128$ 。

本文中手势动作包括前推、后拉、左滑、右

滑、前后推拉和左右滑动共6类，其RTM, DTM和ATM数据如图1所示。此外，为了构建多维参数数据集，需要统一类别标签和原始数据来源，保证3种数据的同步。此外，本文对雷达图像数据进行了灰度归一化，因此网络的第1层输入通道数为1。

### 2.3 基于单参数卷积神经网络的手势识别

卷积神经网络对图像数据的学习主要集中在对图像中目标轮廓、颜色、边缘以及纹理等信息的挖掘，通过对图像的多次非线性映射进行特征提取。本文将利用卷积神经网络对描述手势动作的RTM, DTM和ATM数据进行特征提取。

本文中RTM, DTM和ATM 3种图像数据之间具有较大的差异性。因此，需要针对不同参数的数据分别进行特征提取。同时，为了验证每一种参数的有效性和准确性，需要分别对RTM, DTM和ATM数据集进行训练和测试验证。本文基于对经典VGG16-Net网络改进得到的单参数卷积神经网络(简称单参数网络)。将第3~5个卷积-池化模块滤波器尺寸设置为 $3 \times 1$ ,  $1 \times 3$ 和 $1 \times 1$ ，目的在于保留多次非线性映射前提下减少网络参数量和提高计算速率。并将网络中所有卷积层和池化层输出通道数减半，以降低时间复杂度。

本文所提手势识别方法包含离线训练和在线验证两个阶段。其中，离线训练是将雷达采集处理得到的3维参数图(距离-时间图、速度-时间图和角度-时间图)输入到RDA-T网络，在服务器上通过前向传播和后向传播不断迭代的过程。而在线验证阶段无需反复迭代训练，仅需要加载RDA-T网络参数，并将新样本送入RDA-T网络进行一次前向传

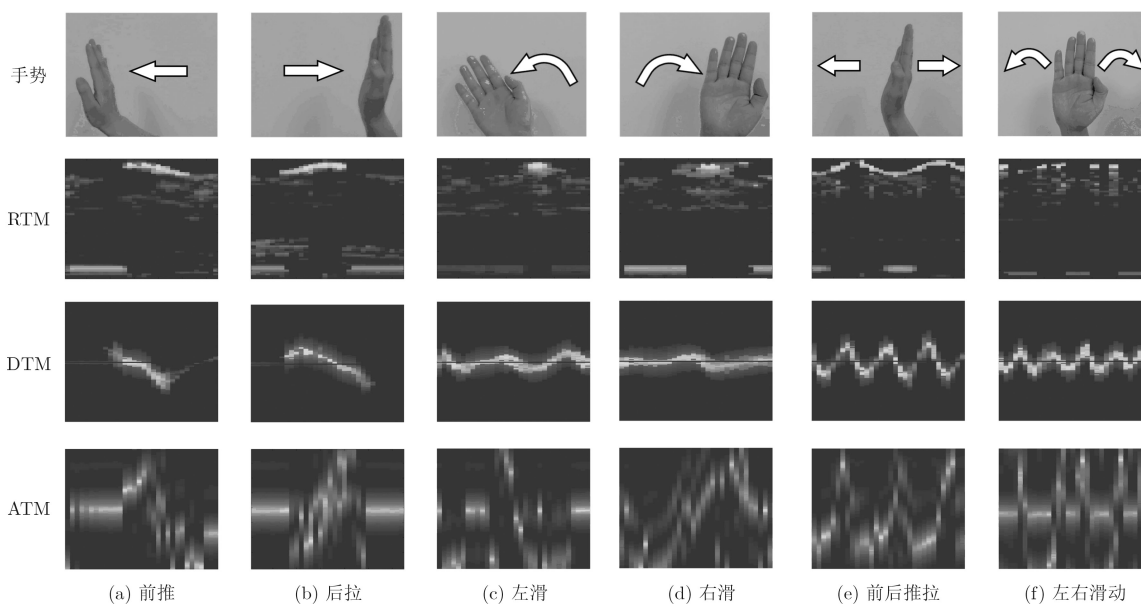


图1 数据集中6种手势对应RTM, DTM和ATM

播计算即可，而不需要进行繁重的训练任务。卷积神经网络的时间复杂度表征了新样本在该模型上进行前向传播所需的计算量，其计算公式<sup>[23]</sup>为

$$T \sim O \left( \sum_{l=1}^L M_H^{(l)} \cdot M_W^{(l)} \cdot K_H^{(l)} \cdot K_W^{(l)} \cdot C_{\text{OUT}}^{(l-1)} \cdot C_{\text{OUT}}^{(l)} \right) \quad (11)$$

其中， $M_H^{(l)}$ 和 $M_W^{(l)}$ 分别表示第 $l$ 层输出特征图的高和宽， $K_H^{(l)}$ 和 $K_W^{(l)}$ 分别表示第 $l$ 层卷积核(卷积滤波器)的高和宽， $C_{\text{OUT}}^{(l-1)}$ 和 $C_{\text{OUT}}^{(l)}$ 分别表示第 $l$ 层和第

$l-1$ 层的输出通道数。

根据式(11)计算后得到本文的单参数网络时间复杂度约为 $0.74 \times 10^9$  FLOPS，而VGG16-Net<sup>[14]</sup>的时间复杂度约为 $15.5 \times 10^9$  FLOPS，本文单参数网络的时间复杂度仅为VGG16-Net网络的1/20，并为本文的进一步研究提供了支撑。

## 2.4 基于多维参数RDA-T卷积神经网络的手势识别

为了将3个单参数网络融合成为一个端到端的网络，本文提出了RDA-T多维参数卷积神经网络，其结构如图2所示。

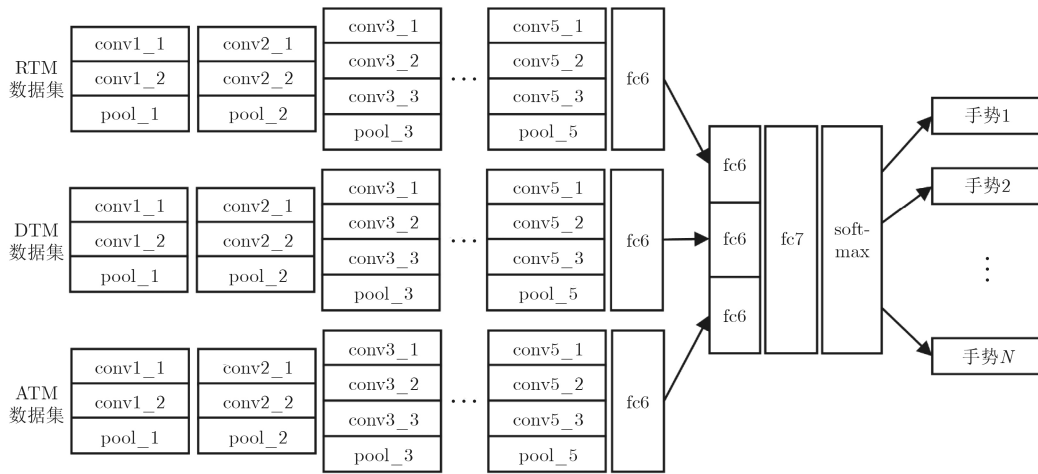


图2 RDA-T卷积神经网络结构图

RDA-T网络前半部分中，将单参数网络的前6层作为特征提取分支网络，3个分支网络具有相同的结构和平行的处理时序，每个分支网络的输出均为fc6层特征。RDA-T网络后半部分中，输入是RTM、DTM和ATM分支网络全连接层输出的fc6层特征 $\mathbf{F}_{\text{RTM}}$ 、 $\mathbf{F}_{\text{DTM}}$ 和 $\mathbf{F}_{\text{ATM}}$ ，3个分支网络融合后的特征为 $\mathbf{F}_{\text{fusion}} = [\mathbf{F}_{\text{RTM}}, \mathbf{F}_{\text{DTM}}, \mathbf{F}_{\text{ATM}}]^T$ ，第7层为全连接层。最后一层为softmax分类器，其输出为分类后的手势类别标签。

根据上述分析，RDA-T网络进行手势识别的计算复杂度也是由前向传播计算量决定。利用式(11)可知RDA-T网络的时间复杂度约为 $2.11 \times 10^9$  FLOPS，相较于文献[9]方法将时间复杂度节约了一半，相较于文献[14]将时间复杂度降低为1/21.8，大幅度降低了时间开销。因此，本文所提RDA-T手势识别算法计算复杂度低，在移动设备上进行手势识别具有广阔的应用前景。

## 3 实验与分析

### 3.1 实验环境和平台

本文实验环境为室内走廊，手势动作范围在雷

达前方0.5 m内。毫米波雷达为TI公司的AWR1642-BOOST开发板以及TSW1400高速数据采集卡，雷达频点为77 GHz，带宽为4 GHz，天线为两发四收。RDA-T网络的训练基于Tensorflow深度学习框架，采用NVIDIA公司的GTX1080显卡进行加速计算。

### 3.2 单参数网络验证

本文对单参数网络的验证中，RTM、DTM和ATM训练集和测试集的数据源相同。每种数据集中训练集和测试集分别包含900和300单位样本。每次迭代所需数据量batch\_size为32，采用随机梯度下降的优化策略，初始学习率为0.003，选择指数衰减方法为学习率衰减策略，衰减间隔为3个epoch，衰减率为0.99。

卷积神经网络训练的准确率曲线能够反映网络模型准确率随迭代步数变化情况。图3给出了单参数网络在RTM、DTM和ATM数据集上单独训练时的准确率曲线。基于同一单参数网络的训练和测试，3种数据集的效果从好到差，依次为：RTM、ATM、DTM，且测试集上准确率均能达到85%以上，达到了较好的手势识别效果。从而证明，除了描述手

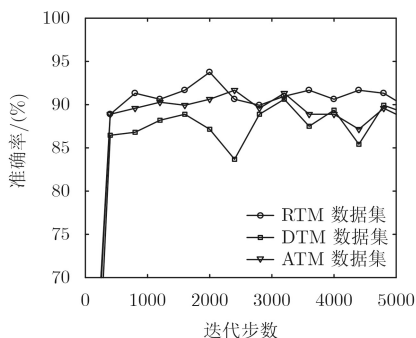


图3 RTM, DTM和ATM准确率曲线

势动作的RTM和DTM数据之外，ATM数据同样能够贡献高质量的手势动作信息。然而，单参数网络仅能利用描述手势动作的其中一种数据集对手势数据进行学习，缺乏将多维参数信息进行融合学习的能力。为此，为了达到更好的手势识别效果，本文利用多维参数数据集对RDA-T网络进行训练。

### 3.3 RDA-T网络参数寻优和性能分析

基于RDA-T网络和多维参数数据集，本节将对全连接层尺寸、初始学习率和学习衰减率进行参数讨论和分析。

全连接层不同尺寸的讨论：RDA-T网络结构的优化中，选择合适的全连接层尺寸能有效提高网络模型的性能。本文将每个分支网络fc6层尺寸分别设置为2048和4096，融合后fc7层尺寸分别设置为2048, 4096和8192，随后基于上述不同全连接层尺寸对网络进行训练和测试，测试集准确率对比如图4所示。其中，固定fc7层，增大fc6层尺寸能够提高准确率。而固定fc6层时，fc7层尺寸为4096时能够得到更好的效果。因此，RDA-T网络的训练中，适当增大全连接层尺寸能提高对手势特征的表达能力，从而降低网络模型的拟合误差，提高手势识别准确率。然而，过大的全连接层尺寸可能导致网络过拟合，降低模型泛化性能。此外，过大的全连接层尺寸会严重降低网络计算效率，因此本文将参数讨论范围限定在2048到8192以内。

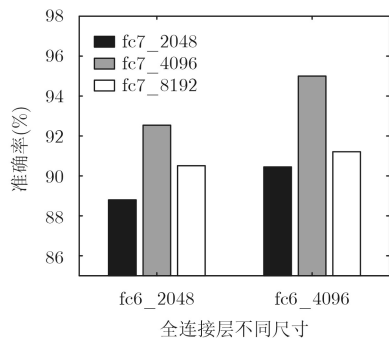


图4 fc6, fc7层不同尺寸准确率比较

不同初始学习率的讨论：为了尽可能避免RDA-T网络在训练中陷入局部极小值，本文将初始学习率分别设置为0.001, 0.003和0.009进行训练和测试，图5给出了测试集上的准确率曲线。显然，初始学习率设置为0.009时达到了较高的准确率。因此，在随后的实验和分析中本文选择0.009为初始学习率。

不同学习衰减率的讨论：RDA-T网络的训练中采用指数衰减的策略，分别将学习衰减率设置为0.99, 0.80, 0.60和0.50，图6给出了迭代步数为100, 200, 800和1200时的损失值变化情况。随着迭代步数的增加，网络的交叉熵损失值在学习衰减率为0.60时收敛情况较好。因此，本文最终采用0.60学习衰减率对RDA-T网络进行训练。

### 3.4 本文与同类方法的比较和分析

为了验证累积帧数对手势识别性能的影响，本文分别基于8帧, 16帧, 24帧和32帧的雷达数据构建了数据集，利用RDA-T网络进行训练和测试对比，准确率对比由表1给出，32帧数据集达到最高准确率。

为了比较同类算法与本文RDA-T方法的准确率性能和算法复杂度，基于RDA-T网络的多分支

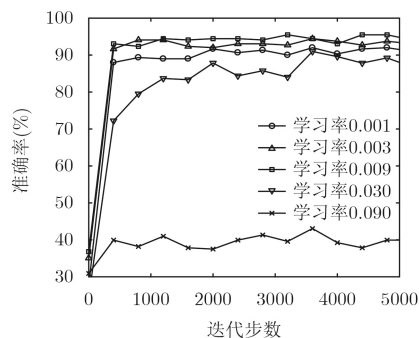


图5 RDA-T网络不同初始学习率准确率曲线

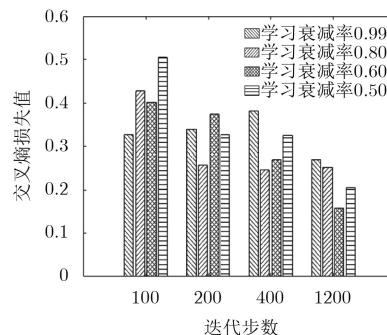


图6 不同学习衰减率交叉熵损失值比较

表1 不同帧数下的准确率比较结果(%)

	8帧数据集	16帧数据集	24帧数据集	32帧数据集
准确率	70.3	79.7	88.3	95.3

网络结构, 将其前半部分的分支网络结构分别替换CNN<sup>[9]</sup>单分支网络和VGG16-Net<sup>[14]</sup>单分支网络前半部分的网络结构, 构成CNN<sup>[9]</sup>多分支网络和VGG16-Net<sup>[14]</sup>多分支网络进行训练和测试验证, 并作为与本文RDA-T网络比较的方法。图7给出了多维参数数据集上3种方法准确率曲线, 比较可知本文方法的准确率高于CNN<sup>[9]</sup>多分支网络和VGG16-Net<sup>[14]</sup>多分支网络。

在多维参数数据集测试集上对RDA-T网络模型进行评估, 得到了表2所示的手势分类混淆矩阵。基于RTM, DTM和ATM测试数据集以及多维参数

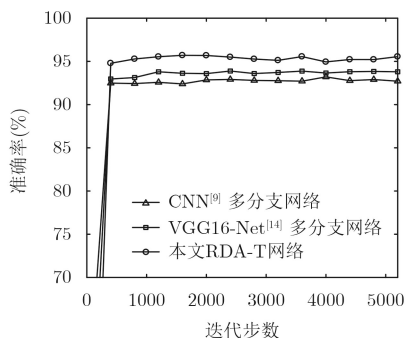


图7 多维参数数据集上不同方法准确率曲线比较

数据集, 表3给出了本文方法与其他方法准确率和算法复杂度比较。通过对单分支和多分支网络的比较可知: 一方面, 相较于单分支网络, 多分支网络均能够达到更高的准确率, 说明多维参数数据集能够提供更丰富的信息, 体现了本文构建多维参数数据集的创新性; 另一方面, 本文所提RDA-T网络的准确率达到95.3%, 相较于单参数网络结构将手势识别准确率提高了5%~8%, 体现了本文所设计RDA-T网络结构的可靠性。通过对同类方法单分支网络的比较可知, 本文基于对单参数网络结构的改进和优化, 在保证识别准确率的基础上有效降低了时间复杂度。同时, 比较同类方法的多分支网络可知, 相较于VGG16-Net<sup>[14]</sup>多分支网络, RDA-T网络将时间复杂度降低为1/20, 空间复杂度降低为1/4, 并且RDA-T网络的准确率优于同类方法, 体现了本文方法的先进性。此外, 本文方法的单个样本处理时间为ms级, 为移动设备的手势识别应用提供了理论支撑。

#### 4 结束语

本文提出了一种基于FMCW雷达多参数估计的RDA-T网络手势识别方法。首先, 本文对雷达

表2 手势分类混淆矩阵(%)

	预测类别						
	前推	后拉	左滑	右滑	前后推拉	左右滑动	
真实类别	前推	90	0	4	0	0	6
	后拉	0	100	0	0	0	0
	左滑	0	0	98	2	0	0
	右滑	0	2	0	94	0	4
	前后推拉	4	0	0	0	94	2
	左右滑动	2	0	0	0	2	96

表3 本文方法与其他方法准确率和算法复杂度对比

网络结构	数据集	平均准确率(%)	空间复杂度( $10^6$ Byte)	时间复杂度( $10^9$ FLOPS)	单个样本处理时间(ms)
CNN <sup>[9]</sup> 单分支网络	RTM	89.6			
	DTM	87.3	9.7	1.51	2.9
	ATM	84.3			
VGG16-Net <sup>[14]</sup> 单分支网络	RTM	89.3			
	DTM	86.3	136.0	15.5	3.9
	ATM	87.0			
本文单分支网络(单参数网络)	RTM	90.6			
	DTM	88.3	64.1	0.74	2.9
	ATM	89.3			
CNN <sup>[9]</sup> 多分支网络		91.6	26.6	4.50	5.6
		92.0	362.1	46.16	10.3
VGG16-Net <sup>[14]</sup> 多分支网络	多维参数数据集	92.0	362.1	46.16	10.3
本文多分支网络(RDA-T网络)		95.3	89.6	2.11	3.9

信号进行时频分析, 解算出雷达前方目标的距离、速度和角度参数; 其次, 将多个参数在时间域进行累积分别构造出RTM, DTM和ATM数据集; 然后, 针对多维参数数据集, 设计能够同时学习手势动作距离、多普勒和角度信息的RDA-T网络; 最后, 实验表明本文所提RDA-T网络相比于单参数手势识别方法的准确率提高了5%~8%, 并通过对网络结构的改进提高了网络模型的计算效率, 从而验证了提出的RDA-T网络的有效性和可靠性。

### 参 考 文 献

- [1] LI Yuman, MIAO Qiguang, TIAN Kuan, *et al.* Large-scale gesture recognition with a fusion of RGB-D data based on the C3D model[C]. 2016 23rd International Conference on Pattern Recognition, Cancun, Mexico, 2016: 25–30.
  - [2] HE Yiwen, YANG Jianyu, SHAO Zhanpeng, *et al.* Salient feature point selection for real time RGB-D hand gesture recognition[C]. IEEE International Conference on Real-time Computing and Robotics, Okinawa, Japan, 2017: 103–108.
  - [3] ALMASRE M A and AL-NUAIM H. Recognizing Arabic Sign Language gestures using depth sensors and a KSVM classifier[C]. Computer Science and Electronic Engineering, Colchester, UK, 2016: 146–151.
  - [4] AUGUSTAUSKAS R and LIPNICKAS A. Robust hand detection using arm segmentation from depth data and static palm gesture recognition[C]. Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Bucharest, Romania, 2017, 2: 664–667.
  - [5] DEKKER B, JACOBS S, KOSSEN A S, *et al.* Gesture recognition with a low power FMCW radar and a deep convolutional neural network[C]. Radar Conference, Nuremberg, Germany, 2017: 163–166.
  - [6] MOLCHANOV P, GUPTA S, KIM K, *et al.* Multi-sensor system for driver's hand-gesture recognition[C]. Automatic Face and Gesture Recognition, Ljubljana, Slovenia, 2015, 1: 1–8.
  - [7] LIN J J, LI Yuanping, HSU W C, *et al.* Design of an FMCW radar baseband signal processing system for automotive application[J]. *Springerplus*, 2016, 5(1): 42–57. doi: [10.1186/s40064-015-1583-5](https://doi.org/10.1186/s40064-015-1583-5).
  - [8] LI Gang, ZHANG Rui, RITCHIE M, *et al.* Sparsity-driven micro-Doppler feature extraction for dynamic hand gesture recognition[J]. *IEEE Transactions on Aerospace & Electronic Systems*, 2018, 54(2): 655–665. doi: [10.1109/TAES.2017.2761229](https://doi.org/10.1109/TAES.2017.2761229).
  - [9] ZHANG Shimeng, LI Gang, RITCHIE M, *et al.* Dynamic hand gesture classification based on radar micro-Doppler signatures[C]. 2016 CIE International Conference on Radar (RADAR), Guangzhou, China, 2016: 1–4.
  - [10] KIM Y and TOOMAJIAN B. Hand gesture recognition using micro-Doppler signatures with convolutional neural network[J]. *IEEE Access*, 2016, 4: 7125–7130. doi: [10.1109/ACCESS.2016.2617282](https://doi.org/10.1109/ACCESS.2016.2617282).
  - [11] MOLCHANOV P, GUPTA S, KIM K, *et al.* Short-range FMCW monopulse radar for hand-gesture sensing[C]. Radar Conference, Arlington, USA, 2015: 1491–1496.
  - [12] WANG Saiwen, SONG Jie, LIEN J, *et al.* Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum[C]. Proceedings of the 29th Annual Symposium on User Interface Software and Technology, New York, USA, 2016: 851–860.
  - [13] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. Imagenet classification with deep convolutional neural networks[C]. Advances in Neural Information Processing Systems, Montreal, Canada, 2012: 1097–1105.
  - [14] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[OL]. <https://arxiv.org/abs/1409.1556>, 2014.
  - [15] HE Kaiming, ZHANG Xianyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778.
  - [16] SZEGEDY C, LIU Wei, JIA Yangqing, *et al.* Going deeper with convolutions[C]. Computer Vision and Pattern Recognition, Boston, USA, 2015: 1–9.
  - [17] IOFFE S and SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[OL]. <https://arxiv.org/abs/1502.03167>, 2015.
  - [18] SZEGEDY C, VANHOUCHE V, IOFFE S, *et al.* Rethinking the inception architecture for computer vision[C]. Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 2818–2826.
  - [19] TRAN D, BOURDEV L, FERGUS R, *et al.* Learning spatiotemporal features with 3D convolutional networks[C]. International Conference on Computer Vision, Santiago, Chile, 2015: 4489–4497.
  - [20] MOLCHANOV P, GUPTA S, KIM K, *et al.* Hand gesture recognition with 3D convolutional neural networks[C]. Computer Vision and Pattern Recognition Workshops, Boston, USA, 2015: 1–7.
  - [21] SIMONYAN K and ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]. Advances in Neural Information Processing Systems, Montreal, Canada, 2014: 568–576.
  - [22] SCHMIDT R. Multiple emitter location and signal parameter estimation[J]. *IEEE Transactions on Antennas and Propagation*, 1986, 34(3): 276–280. doi: [10.1109/TAP.1986.1143830](https://doi.org/10.1109/TAP.1986.1143830).
  - [23] HE Kaiming and SUN Jian. Convolutional neural networks at constrained time cost[C]. IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 5353–5360.
- 王 勇: 男, 1987年生, 讲师, 研究方向为无线通信、能效优化、室内定位、深度学习理论等。
- 吴金君: 男, 1994年生, 硕士生, 研究方向为手势识别和深度学习技术。
- 田增山: 男, 1968年生, 教授, 博士生导师, 研究方向为移动通信、个人通信、GPS及蜂窝网定位技术等。
- 周 牧: 男, 1984年生, 教授, 研究方向为无线定位与导航技术、信号侦察与检测技术、凸优化与深度学习理论等。
- 王沙沙: 女, 1992年生, 硕士生, 研究方向为深度学习技术和雷达信号处理。