arXiv:2305.02039v1 [cs.CV] 3 May 2023

采用新型无菌训练技术的深度卷积神经网络改进的静态手势分类

IEEEAccess （已接收）

乔西亚·W·史密斯
电气与计算机工程系
达拉斯德克萨斯大学
理查森，德克萨斯州75080
josiah.smith@utdallas.edu

希瓦·蒂亚加拉詹
电气与计算机工程系
达拉斯德克萨斯大学
理查森，德克萨斯州75080

Richard Willis*
电气与计算机工程系
达拉斯德克萨斯大学
理查森，德克萨斯州75080

Yiorgos Markis
电气与计算机工程系
达拉斯德克萨斯大学
理查森，德克萨斯州75080
yiorgos.makris@utdallas.edu

Murat Torlak达拉斯德克萨斯大学电气与计算机工程系，德克萨斯州75080，理查森，
torlak@utdallas.edu

摘要

本文研究了新颖的数据采集和训练技术，旨在使用卷积神经网络（CNN）和调频连续波（FMCW）毫米波（mmWave）雷达，提高非移动（静态）手势的分类准确率。近年来，非接触式手部姿态和静态手势识别在诸多应用中受到广泛关注，包括人机交互（HCI）、增强现实/虚拟现实（AR/VR），乃至医疗应用中的关节活动范围治疗。虽然目前大多数解决方案依赖光学相机或深度相机，但这些方法要求理想的光照和温度条件。毫米波雷达设备近来已成为一种颇具前景的替代方案，它采用低成本的片上系统传感器，即使在非理想成像条件下，其输出信号也包含精确的空间信息。此外，深度卷积神经网络已广泛用于图像识别，能够同时学习特征提取和分类。然而，由于难以从雷达回波信号中提取有效特征，使用毫米波雷达和 CNN 进行静态手势识别的研究仍然很少，其效果也不如动态手势分类。本文提出一种高效的数据采集方法，以及一种新颖的深度 CNN 训练技术：引入“洁净”图像，以帮助区分不同静态手势的特征，进而提高分类准确率。采用所提出的数据采集和训练方法后，静态手势的分类率在距离剖面和距离-角度剖面上分别从 85% 提高至 93%，以及从 90% 提高至 95%。

*开展这项研究时，他还是达拉斯德克萨斯大学电气与计算机工程系的本科生；目前就职于花旗环球金融有限公司，地址为纽约州纽约市格林威治街390号，邮编10013。

©2021年 IEEE。允许个人使用本材料。任何其他用途均须事先获得 IEEE 的许可，无论是在当前还是未来的任何媒介中，包括为广告或宣传目的而重印／再版本材料、创作新的合集、转售或重新分发至服务器或列表，或在其他作品中重复使用本作品的任何受版权保护部分。doi:10.1109/ACCESS.2021.3051454

关键词 卷积神经网络 · 深度学习 · 手势识别 · 毫米波雷达 ·无菌训练

1 引言

随着非接触式人机交互（HCI）传感器日益普及且备受青睐，准确分类人类手势近来受到了广泛关注。许多研究致力于利用光学相机和各种分类器，对移动（动态）手势和不移动（静态）手势进行分类 [1]。静态手势分类的应用包括增强现实/虚拟现实（AR/VR） [2], 、人机交互 [3], ，以及关节活动范围评估和治疗等医疗应用 [4]。这类光学系统可提供高分辨率二维（2-D）图像，但存在固有缺陷：需要特定的照明条件，且无法提供深度信息。一些解决方案探索了使用 RGB-D 深度相机 [5], ，但这类设备在阳光下性能不佳，因而只能在室内使用 [2]。另一方面，小型化毫米波调频连续波雷达可提供高分辨率深度信息，但不具备光学相机的横向分辨率。与光学解决方案相比，毫米波雷达更具优势，这是因为毫米波频段波长下电磁波（EM）具有半穿透性，且不受环境温度影响，因此即使在包括遮挡、雾气、室内／室外等非理想光照和温度环境中，也能进行精细测量。此外，调频连续波毫米波雷达还能同时实现手势分类与定位。人手反射的高分辨率空间信息包含在 FMCW 回波信号中。然而，由于调频连续波雷达作为飞行时间（ToF）传感器的特性以及硬件尺寸限制，商用现成雷达设备无法重建出人眼可辨识、类似人手的图像。因此，通常会采用深度卷积神经网络（CNN），在经过一定预处理后，根据雷达回波信号对动态手势进行分类 [6]。此外，从雷达回波信号中提取有效特征，是准确分类手势的关键步骤。因此，近年来利用毫米波传感器进行手势识别的研究主要集中于动态手势，关注多普勒特征和微多普勒特征 [7]–[10] ，而对静态手势的关注较少 [11]，原因是此类应用的分类率较低。此前使用 CNN 进行基于脉冲无线电超宽带（IR-UWB）[11]–[13] 和多普勒雷达的手势识别研究 [7]–[9] 已展现出良好前景。然而，除 [11], 外，基于 IR-UWB 和多普勒雷达的手势识别研究仅限于动态、移动手势。由于动作在多普勒频谱中清晰可见，脉冲无线电超宽带传感器和多普勒传感器能够轻松分类动态手势；但这两种方法在静态手势识别方面均存在不足，因为从静止的人手中提取每个手势类别的独特特征要困难得多。多普勒雷达利用多普勒频移原理提供手势的相对速度信息，却无法提供分类静止静态手势所需的空间特征。因此，对于静态手势识别问题，由于手部保持静止，多普勒雷达的表现并不理想，这使得从多普勒雷达回波信号中提取手部反射特征变得相当困难。另一方面，脉冲无线电超宽带传感器的工作方式与毫米波调频连续波雷达类似，也必须克服同样的固有难题：如后文所述，如何从低维雷达回波信号中分类高维手部姿态。本文提出了用于数据采集和 CNN 训练的新方法，旨在克服上述毫米波传感器在静态手势识别方面的局限性，为每种手势提供独特且有效的特征，从而帮助 CNN 学习。

本文其余部分的安排如下。第 2 节简要概述调频连续波雷达信号模型，以及有助于建立静态手势分类问题直观认识的关键概念。第 3 节讨论测量设置，并介绍调频连续波雷达“无菌”数据的新颖性。第 4 节概述深度卷积神经网络的基础知识，并重点介绍我们在实现中采用的网络架构。第 5 节展示并讨论分类结果，最后给出结论。

2 调频连续波雷达信号模型

理解调频连续波雷达信号模型，有助于深入认识静态手势识别问题领域所面临的诸多难点。相关文献[14] 已对调频连续波信号模型进行了充分研究，本文在此作简要概述，以帮助读者了解使用毫米波调频连续波传感器进行静态手势识别时所固有的挑战。

本研究部分得到了德州仪器通过基础技术研究中心和德州模拟卓越中心提供的资助。

2.1 FMCW 拍频信号

我们将首先考虑一个单基地 FMCW 收发器，其发射器和接收器分别位于 (*x*,*y*,*z*) 空间中的点 (0,y_T,Z_0) 和 (0,y_R,Z_0)，场景中还有一个反射率为 σ、位于点 (x_0,y_0,z_0) 的静止理想点反射体。雷达收发器位于 x'-y' 平面上，与点目标相距 $z = Z_0$。

首先，FMCW 设备会生成一种称为线性调频信号的信号，该信号可建模为复正弦信号，其频率随时间线性增加，表示为

$$m(t) = e^{j2\pi(f_0 t + \frac{1}{2}Kt^2)}, \quad 0 \le t \le T, \tag{1}$$

其中，f_0 是时刻 $t = 0$的瞬时频率，K 是调频斜率，T是快时间中的调频持续时间。调频带宽可通过 $B = KT$ [15]轻松计算。

调频信号 $m(t)$由发射天线发射，经理想点反射器反射后，以发射信号经过缩放和延时的形式返回接收天线。考虑往返传播导致的振幅衰减后，接收信号可建模为

$$\hat{m}(t) = \sigma \frac{m(t-\tau)}{R_T R_R} = \frac{\sigma}{R_T R_R} e^{j2\pi(f_0(t-\tau)+\frac{1}{2}K(t-\tau)^2)}, \tag{2}$$

其中，τ是往返时延 [16]，数值 R_T和 R_R（见图1）可以是 co 计算得出的是

$$R_T = \sqrt{x_0^2 + (y_0 - y_T)^2 + (z_0 - Z_0)^2}, \tag{3}$$

$$R_R = \sqrt{x_0^2 + (y_0 - y_R)^2 + (z_0 - Z_0)^2}. \tag{4}$$

因此，往返时间延迟 τ 可计算为

$$\tau = \frac{R_T + R_R}{c}, \tag{5}$$

其中，c是光速。

现在，将接收信号 $\hat{m}(t)$ 与发射信号 $m(t)$ 解调，得到所谓的中频信号或 FMCW拍频信号，表示为

$$s(t) = \frac{\sigma}{R_T R_R} e^{j2\pi(f_0\tau + K\tau - \frac{1}{2}K\tau^2)} \tag{6}$$

式（6）的最后一个相位项称为剩余相位项，已知其可以忽略不计 [17]。最后，拍频信号可以简化为以下表达式

$$s(y_T, y_R, k) = \frac{\sigma}{R_T R_R} e^{jk(R_T + R_R)}, \tag{7}$$

其中，$k = 2\pi f/c$是与 $t \in [0, T]$对应的瞬时频率 $f = f_0 + Kt$相对应的波数。

2.2 多基地到单基地转换

式 (7) 中的结果展示了多基地天线阵列中来自单个点反射体的 FMCW拍频信号，其中发射器和接收器并未共址。为简化后续的信号处理，最好将此多基地回波信号近似为单基地版本。文献中已对这种近似进行了探讨，我们将直接采用 [17], [18] 中推导出的结果。多基地到单基地转换被称为对每个收发机对应用简单的相位调整，如下所示。

$$\hat{s}(y', k) = s(y_T, y_R, k) e^{-jk\frac{d_y^2}{4Z_0}}, \tag{8}$$

其中 d_y 是发射器和接收器之间的较小间隔，Z_0 是从雷达到目标的大致距离。

现在，单基地近似会产生一个拍频信号，其虚拟天线位置位于每个收发机对的中点。取 y' 作为这些虚拟天线位置，虚拟单基地信号可通过对公式(7)进行简化来近似为

$$\hat{s}(t) \approx \frac{\sigma}{R_0^2} e^{j2kR_0}, \tag{9}$$

其中，R_0 是每个虚拟天线单元与点反射体之间的距离，表示为

$$R_0 = \sqrt{x_0^2 + (y_0 - y')^2 + (z_0 - Z_0)^2}. \tag{10}$$

现在，目标的距离清晰地体现在拍频信号的频率中。

2.3 FMCW 距离-角度分析

考虑前文所述的理想点反射体，以及沿 y 轴排列的均匀线性单基地阵列，可以计算距离剖面和距离-角度剖面，并用它们定位点反射体。首先，正如公式 (9) 和 (11) 所示，拍频信号的频率与目标的距离直接对应。因此，可以沿拍频信号的 k 域执行快速傅里叶变换（FFT），生成距离剖面。两个目标之间可分辨的最小距离，即距离分辨率，可以很容易地计算为 $\Delta z_{min} = c/(2B)$ [19]。
同样地，如 [20], 所述，可以通过在空间 y 域上执行快速傅里叶变换来计算目标场景的角度剖面。根据奈奎斯特定理，为避免角度域中的混叠，单元之间的最大理论距离为 $\lambda/4$，其中 λ 为波长。然而，即使同时执行距离和角度快速傅里叶变换，所得的距离-角度剖面也只能描述二维强度，并且受到天线单元数量和带宽的限制。

在本研究中，我们将采用并比较距离分析和距离-角度分析，在使用回波信号进行 CNN 训练或分类之前对其进行预处理。

2.4 建模分布式目标

对于手势识别，可以将人手在数学上建模为一个分布式目标，其反射率在空间中连续变化。了解雷达如何捕获此类目标场景，有助于直观理解手势识别问题的难度。

假设沿 y 轴排列一个简单的线性多基地阵列，如图 1 所示，经过上述转换后，分布式目标的返回信号可建模为来自目标各坐标处的回波信号与目标反射率函数 $\sigma(x,\ y,\ z)$ 相乘后叠加的结果。位置 y' 处每个虚拟单基地收发器的拍频信号可表示为

$$s(y',k) = \iiint \frac{\sigma(x,y,z)}{R^2} e^{j2kR} dxdydz. \tag{11}$$

其中，R 是位于位置 y' 处的各虚拟单基地单元到分布式目标域内各点的径向距离，表示为

$$R = \sqrt{x^2 + (y-y')^2 + (z-Z_0)^2}. \tag{12}$$

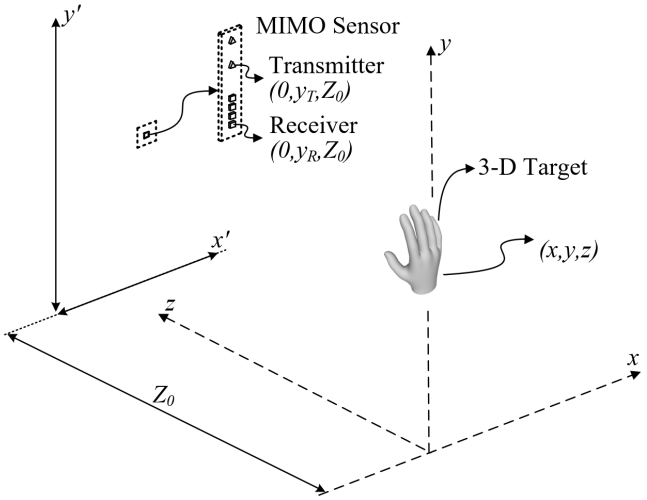


图 1：一个多输入多输出雷达传感器，其发射器和接收器天线单元分别位于 (0,y_T,Z_0) 和 (0,y_R,Z_0)，接收来自三维（3-D）目标的返回信号，该目标的反射率函数为 $\sigma(x,\ y,\ z)$。

如果在整个 x'-y' 平面上采集样本，则可以通过对（11）求逆来重建反射率函数；然而，对于手势识别这类应用，收发器单元沿 y' 轴仅覆盖很小的空间范围。该模型揭示了在调频连续波雷达上进行静态手势识别为何既可行又困难。

拍频信号中包含描述目标或所执行静态手势形状的高分辨率空间特征，这意味着不同的手部姿态或静态手势具有各自独特的回波信号。然而，无法通过解析方法将目标场景（即手部）重建为三维（3-D）图像，也无法利用传统光学图像方法轻松对手势进行分类。因此，静态手势分类需要尝试从二维雷达数据中学习三维模式（即三维空间中的手部姿态）。

此外，手势识别问题还存在另一个固有难题：人手的雷达散射截面（RCS）较小，导致信噪比（SNR）较低，后文将对此进行讨论。即使数据量很大，由于手部的雷达散射截面较小，各手势类别所特有的特征也不明显。因此，卷积神经网络难以辨别每种手势的有效特征。我们提出的方法采用“无污染”数据，旨在弥补训练数据中的这一不足，下一节将对此进行详细讨论。

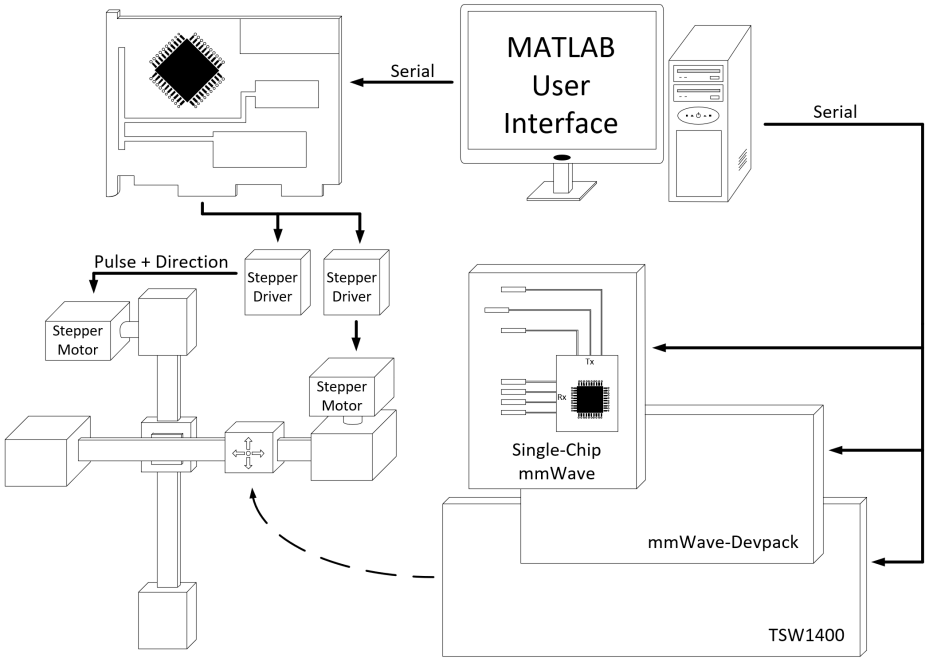


图 2：二维 x-y 矩形扫描器系统示意图。

3 测量设置

对于任何采用监督学习方法进行深度学习的问题，能否获得有效且多样化的数据，对构建准确且泛化能力良好的模型至关重要。

3.1 机械式扫描仪

为了高效地从多个角度采集数据，我们首先设计了一个二维机械扫描系统，使雷达能够在 0.5 m x 0.5 m 的方形区域内任意定位，如图 2 所示。整个系统由运行在台式电脑上的定制 MATLAB 图形用户界面（GUI）控制。AMC4030 运动控制器通过串口接收指令，并通过步进电机驱动器控制步进电机，从而将雷达精确移动到所需位置。每台步进电机均安装在线性皮带驱动导轨上，导轨的有效长度为 0.5 m。被测雷达为 IWR1443BOOST，这是德州仪器（TI）的一款商用汽车雷达。如系统示意图所示，雷达板的方向为 MIMO 阵列垂直排列。IWR1443BOOST 配有 3 个发射（Tx）天线和 4 个接收（Rx）天线，但本研究仅使用其中的 2 个 Tx 天线和 4 个 Rx 天线，它们构成一个线性 MIMO 阵列，其虚拟阵元间距为 $\lambda/4$。工作频率为 77 GHz，带宽为 4 GHz。雷达板连接至扩展板 TI mmWave-Devpack 和 TI TSW1400 数据采集卡。所有三块雷达板均通过 TI mmWave Studio 控制，后者接收来自 MATLAB GUI 的指令。因此，用户可直接通过定制 GUI 全面控制雷达的位置、设置、时序等。

这种新型机械扫描系统能够从多个位置高效采集静态手势，如图3所示。测试对象只需将手保持在正确的姿势和位置，雷达便会同时进行水平和垂直扫描，从而捕捉手势的多个不同视角。

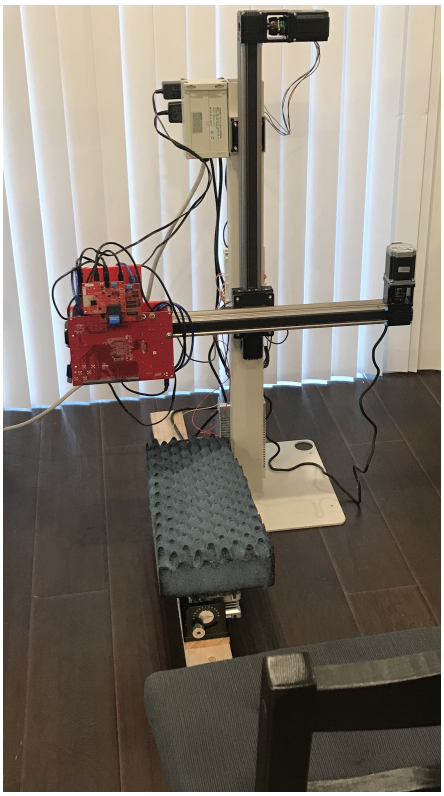


图 3：带有供测试对象就座椅子的二维 x-y 矩形扫描器系统 .

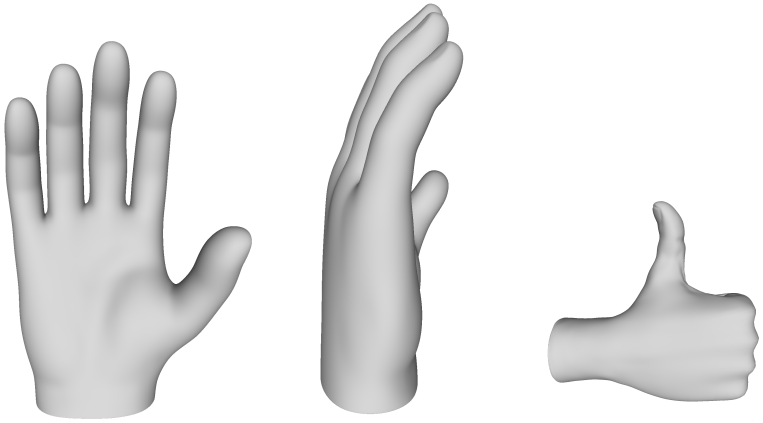
在本文的其余部分，扫描范围将限制为边长为 0.25 m 的正方形，测试对象的手将放置在距离雷达 0.25 m 至 0.55 m 的范围内。测试对象将把手伸到身体前方进行手势操作，并使手远离躯干，以避免躯干峰值的旁瓣遮挡手部。此外，用户将在距离雷达 1 m 的椅子上坐着进行手势操作。研究使用多名测试对象采集数据，这些测试对象在身高、体重、躯干尺寸、臂长、手部尺寸和性别方面各不相同，以增加数据集的多样性。由于雷达数据是在 x'-y' 平面内的各个位置采集的，因此，即使测试对象的手没有移动，数据集仍将包含许多独特的手部“视角”，仿佛手部相对于雷达位于多个不同位置。据作者所知，本文首次尝试使用二维扫描仪，通过毫米波雷达采集静态手势数据。机械扫描仪用于采集具有多种手势视角的多样化数据集；然而，人手反射率和特征显著性的问题仍然存在。接下来将进一步扩展这一新颖的数据采集技术，以尝试克服这些问题。

3.2 挑战

我们接下来使用三种不同的静态手势对所提出的系统进行测试，分别将其标记为“手掌”（palm）、“垂直”（perpendicular）和“竖起大拇指”（thumbs-up），如图 4 所示。对于“手掌”手势，用户将手放置在雷达前方，使手掌朝向雷达。“垂直”手势要求受试者的手垂直于 x'-y' 平面，且拇指背向雷达。而对于“竖起大拇指”手势，受试者需要将手背朝向雷达，除拇指向上伸展外，其余所有手指均向外张开。

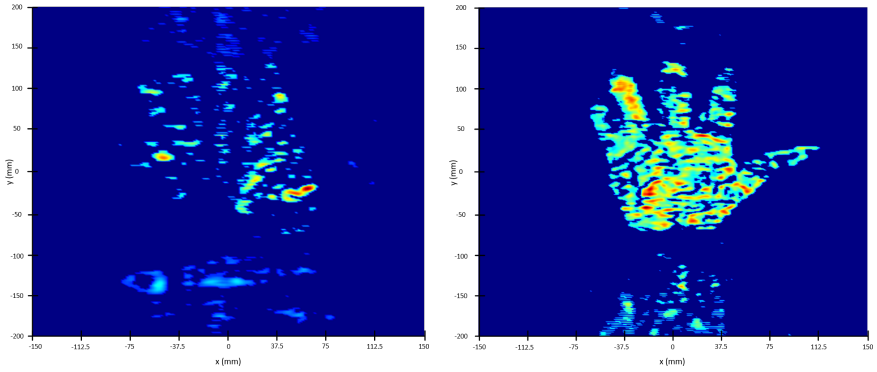
如前所述，与噪声和传播效应相比，人手的雷达散射截面过小，这给识别带来了很大困难。比较不同手势的距离剖面可以发现，它们之间的差异在人眼看来几乎无法区分。尽管在距离向快速傅里叶变换中，与人手所在距离相对应的位置存在一个峰值，但反射回雷达的手势特征并未呈现出清晰的轮廓，而且这些特征在人手上的分布位置各不相同。

为展示这一现象，本文暂时采用合成雷达孔径（SAR）方法，根据 [19], [21] 中所述的方法重建人手图像。需要注意的是，图中所示的图像



（a）“手掌”手势 （b）“垂直”手势 （c）“竖起大拇指”手势

图 4：从雷达视角观察到的三种静态手势。



(a) 距雷达 0.25 来自雷达的 m

(b) 距雷达 0.25 米处铝制剪影手部的合成孔径雷达图像

图 5：真实人手与人手铝制剪影重建的合成孔径雷达图像对比，展示了人手较低的雷达散射截面。为保持一致，真实人手和铝制剪影均放置在三脚架顶部，图中手部下方可见的反射即由此产生。

5 并非用于训练和验证卷积神经网络的数据。生成这些图像需要整个水平和垂直扫描过程中的全部数据，耗时约 5 分钟。用于训练网络的数据将在后文详细讨论。

图5a所示的人手重建图像质量较差，这是由于其雷达散射截面较低。作为对比，我们还使用手形铝制剪影重建了合成孔径雷达图像，以展示理想的手部目标，如图5b所示。在本文余下部分中，我们将使用铝制剪影采集的数据称为“纯净”数据。这项实证分析揭示了根据雷达拍频信号对手势进行分类的内在难度。即使将数千个雷达回波信号合并以构建合成孔径雷达图像，人手仍几乎不可见，手势也难以辨认。从这些图像可以推断，单个反射拍频信号中包含的人手特征并不明显，与周围环境、噪声等相比，其幅度也相当低。相反，如图5b所示，铝制剪影具有较高的信噪比，这意味着每种静态手势的特征都更加突出且一致。本文提出的新型数据采集策略是使用二维机械扫描仪，从多个视角采集“真实”人手和“纯净”铝制剪影的数据，以期提高分类准确率。

基于这些观察，我们提出了本文旨在回答的几个关键问题。与人手发出的回波信号相比，静态手势的铝制剪影所产生的雷达回波信号，是否包含更突出、更有意义且更一致、能够独特描述各个手势的特征？如果是，这些从

铝制剪影采集的“纯净”雷达数据能否用于提高卷积神经网络分类器的准确率？具体而言，同时包含人手数据和“纯净”数据的训练集，是否能为卷积神经网络提供更易于学习的特征？

4 深度卷积神经网络

在进一步探讨这些问题之前，我们首先概述所提出的分类器。在数据驱动的检测问题中，构建稳健分类器有两个基本步骤：特征提取和分类。许多方法已被用于从原始数据中提取特征，包括手工设计的特征 [10], [22], 线性预测编码 [23], 经验模态分解 [24], 主成分分析 [25], 等。同样，也采用了许多不同的分类技术，例如 k 近邻（KNN） [26], 支持向量机 [10], [11], 动态贝叶斯网络 [27], 等。然而，近年来，许多分类和回归问题的首选方法是深度卷积神经网络 [9]。深度学习的概念将特征提取和分类合并为一个步骤。现在，特征提取和分类被同时建模为一个单一的优化问题。卷积神经网络已广泛用于图像分类，也常用于基于多普勒雷达的动态手势识别 [6], [7]。与大多数传统机器学习算法不同，卷积神经网络采用多层结构，并通过相互连接的神经元模拟人脑。此外，得益于并行计算的最新进展，尤其是图形处理器（GPU）的发展，训练由多层构成的复杂架构卷积神经网络已变得越来越可行。

卷积神经网络的基本构建模块是卷积层和非线性激活函数。卷积层通过将权重数组在输入图像上进行卷积来提取特征。每次迭代中，权重都会通过反向传播算法进行更新，并结合随机梯度下降，以提高分类率并降低损失。卷积层之后会应用非线性激活函数。大多数深度卷积神经网络采用修正线性单元（ReLU），其定义为 $f(x) = \max(0, x)$，而非传统的 sigmoid 函数，以获得更好的结果 [28]。通过使用非线性激活函数，网络能够学习输入与输出之间高度非线性的复杂关系。

卷积和激活之后，通常会使用池化层，通过取局部池化区域的平均值或最大值来对数据进行下采样。卷积层、ReLU 层和池化层相互连接，构成复杂的神经元网络，最后连接全连接层，将维度缩减至已知的类别数，随后由通用感知机进行分类。

4.1 输入图像的预处理

从雷达采集输入数据，并在用于网络训练和验证之前进行预处理。首先，对式（7）中描述的复数值 MIMO 拍频信号应用式（8）中的多基地到单基地转换。随后，执行第 2.3 节所述的距离分析或距离-角度分析。下一节将比较仅采用距离分析与采用距离-角度分析所得的结果。在距离分析情况下，复数值图像的尺寸为 8xN_R；在距离和角度分析情况下，尺寸为 M_AxN_R，其中 M_A 是角度向快速傅里叶变换的频点数， N_R 是距离向快速傅里叶变换的频点数。
由于手部反射率的微小变化包含在雷达拍频信号的相位中，因此保留复数值信号的幅度和相位对于准确分类至关重要。已有研究尝试将卷积神经网络应用于雷达问题的复数值实现，例如合成孔径雷达图像分类 [29] 和增强型合成孔径雷达成像 [30], 。然而，我们考虑采用另一种方法对复数雷达回波信号进行分类。我们不只是取每个图像像素的幅值，而是将距离或距离-角度数据样本的实部和虚部分层排列，形成尺寸为 8xN_Rx2 或 M_AxN_Rx2 的图像。这样便不会丢失雷达数据幅度与相位之间固有的关系，从而提高分类率。

此外，为增强数值稳定性，大多数深度卷积神经网络的实现都会对每个通道的输入进行归一化。然而，在本例中，分别对实部层和虚部层进行归一化会严重破坏相位之间的相互依赖关系。因此，在将复值图像分离为实部层和虚部层之前，先将其归一化为零均值和单位方差。随后，不再对实值三维数组进行归一化。

5 分类结果

在本节中，我们利用实证结果肯定地回答第 3.2 节提出的问题，证明我们提出的“无菌”雷达数据采集技术能够有效提升使用毫米波雷达进行静态手势分类的效果。事实证明，为卷积神经网络的训练数据补充合成数据，对于许多深度学习问题都十分有效 [31]–[34]。然而，据我们所知，这项工作是

首个使用毫米波雷达捕获的合成“无菌”手部（即铝制剪影）来提高静态手势分类率的研究。

5.1 训练数据

如第 2.3 节所述，在训练之前，会对每个采集到的拍频信号应用一维距离向快速傅里叶变换或二维距离-角度快速傅里叶变换。在本文接下来的内容中，我们将分别把这些数据集称为“距离”数据集和“距离-角度”数据集。每个数据集均包含三个静态手势类别各 80000 个样本，其中 40000 个样本来自人手，另外 40000 个来自铝制剪影。通过移动安装在机械扫描仪上的雷达，在二维 x'-y' 平面内的不同位置快速采集这些样本。为了提高分类器的鲁棒性，测试对象的手部以及铝制剪影在尺寸和形状上均有所不同。研究选取了八名参与者，他们在年龄、身高、体重、手部尺寸和形状、手臂长度、躯干尺寸及性别方面各不相同。

为了制作铝制剪影，采用了一种简单的方法。从在线资源中选取不同尺寸和形状的手部图案。将这些图案从薄铝板上裁剪下来，并固定在形状相同的纸板剪影上，从而简化“无菌”数据采集的训练过程。同样，每种手势类别均制作八个铝制剪影，并用于数据采集。每个铝制剪影并不直接对应于测试对象的手部；相反，铝制剪影的尺寸和形状均独立选取。对于距离数据，沿 k-域执行快速傅里叶变换，从而得到距离剖面。选取预计放置手部的区域，其大小设定为64个距离单元。同样，对于距离-角度数据集，首先执行距离向快速傅里叶变换，随后执行大小为16的角度向快速傅里叶变换。完成数据预处理后，距离数据集图像的尺寸为64x8x2，而距离-角度图像的尺寸为64x16x2。

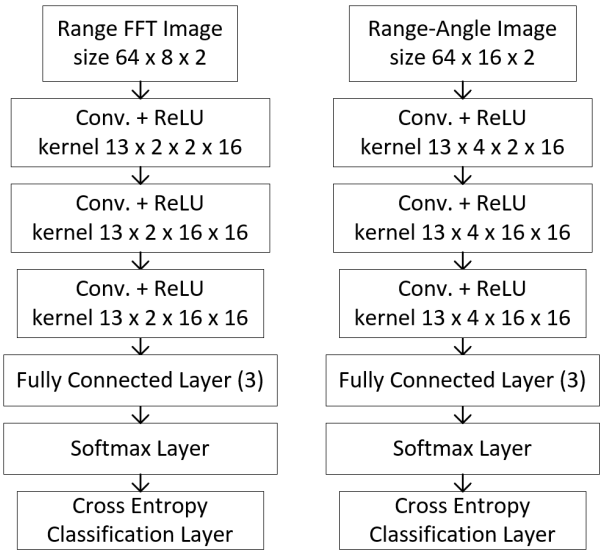


图 6：距离向快速傅里叶变换卷积神经网络和距离-角度向快速傅里叶变换卷积神经网络的网络架构。

5.2 网络架构与训练

用于对手势进行分类的网络会因数据集所采用的预处理方式而异。对于距离数据集，卷积层的卷积核尺寸均为 13x2，每层各有 16 个滤波器，且每层后接一个修正线性单元。这些层依次串联，之后连接一个具有 3 个输出神经元的全连接层、一个 softmax 层，以及一个使用交叉熵损失函数的最终分类层。距离-角度数据集采用相同架构的网络，仅将卷积层的尺寸改为 13x4，以适应更大的图像尺寸。这两个网络的关键在于其复数值分层和网络架构。将雷达距离数据的实部和虚部视为图像中不同的层，可使网络识别像素之间以及层与层之间的关系，而后者对应于每个复数值像素的相位信息。两个网络的架构均是在仔细检查观测图像域中距离和通道/角度方向的特征尺寸，并经过大量测试以优化实时实现效率和分类率后选定的。两个网络架构如图 6 所示。

训练时，每个数据集划分为 90% 的训练集和 10% 的验证集；所有“无菌”数据均包含在训练集中，而验证集则仅由随机选取的真实手势组成。在训练数据中，55.6% 为铝制剪影的采集数据，44.4% 为真实人手的采集数据。这样，“无菌”数据便可用于补充训练集，但不会纳入验证集。为确保一致性，我们留出验证集中随机选取的采集数据，以便在每次实验中重复使用，总计 8000 条。

5.3 使用“无污染”数据改进分类

为与对照组进行比较，我们首先仅使用真实人手数据训练两个网络。对于这些网络，我们使用预留的8000个采集样本作为验证数据集，使训练集与验证集的划分比例为80%比20%。在仅使用真实人手数据训练每个网络后，距离卷积神经网络和距离角度卷积神经网络的分类率分别为84.9%和90.2%。这些网络在表1中被命名为“仅人手”，因为它们仅使用来自人手的距离剖面和距离-角度剖面进行训练。
接下来，使用完整数据集训练了两个新网络，这些数据集由真实的人手数据以及铝制剪影补充的“无菌”数据组成。这些网络被称为“组合”网络，因为它们使用真实图像和“无菌”图像共同训练。需要注意的是，“组合”网络与“仅人手”网络使用相同的验证数据进行验证。每个网络的唯一差异在于所用的训练数据集。训练完成后，这些网络证实了我们的假设：使用“无菌”数据训练可提升分类率；在距离数据集和距离角度数据集上，分类率分别提高到 93.1% 和 95.4%。

	仅限人工	组合在线a
距离	84.9%	93.1%
距离-角度	90.2%	95.4%

表1：比较仅使用人手数据（HumanOnly）训练的网络与使用“无菌”数据补充真实人手数据（Combined）训练的网络之间的分类率。

这些结果对前文提出的问题给出了肯定的回答。也就是说，与仅使用非“无菌”数据训练的网络相比，同时由人手数据和“无菌”数据组成的数据集训练的网络，能够更有效地学习用于人手数据分类的有效特征。此外，我们的结果明确表明，采用本文提出的“无菌”雷达数据采集方案可以提高分类准确率。

最后，与文献中的既有研究相比，我们提出的方法通过使用“无菌”数据改进了手势识别，同时为三维空间平移条件下静态手势这一困难的分类问题提供了解决方案。Kim 等。 [11] 采用了基于超宽带脉冲无线电雷达的时域手势识别方法。 [11] 分别考虑了两种场景：(1) 使用距离收发器15厘米的人手完成六种手势；(2) 使用以 10° 为增量旋转的石膏模型完成三种手势。对于场景(1)，所有采集过程中手部均保持在恒定位置。训练和测试均使用人手数据进行，并通过卷积神经网络取得了91%的分类率。在场景(2)中，通过从不同角度旋转石膏模型，采集每种手势的石膏模型数据。对于该场景，Kim 等。 记录了三种手势超过90%的分类准确率，并使用石膏模型数据进行了验证。相比之下，我们的方法在训练过程中同时纳入真实人手反射信号和“无菌”反射信号，并且仅使用人手数据进行验证，从而获得了更稳健的分类器。正如 [11], 中所述，本文提出的技术并非分别为人手数据和“无菌”数据创建两个不同的分类器，而是将人手数据与“无菌”数据结合起来构建稳健的分类器。此外，我们的方法通过从多个测试对象的多个位置采集数据，考察了更多样化的场景。

关于毫米波雷达、多普勒雷达和红外超宽带传感器在动态手势识别方面，已有大量研究成果 [6]–[10], [12], [13]；然而，这与本文所研究的问题完全不同，因为动态手势仅考虑运动。这使分类的维度降低为时间运动特征，而毫米波雷达上的静态手势识别则涉及利用低维数据对三维结构进行分类，正如第2.4节所讨论的那样。因此，我们的模型针对空间平移条件下更具挑战性的静态手势分类问题进行训练，其分类准确率优于以往的静态手势分类研究 [11]。

6 结论

本文研究了新颖的数据采集和训练技术，旨在利用毫米波调频连续波雷达和卷积神经网络提升静态手势的分类效果。针对

静态手势，我们提出了一种新颖的数据采集技术：将雷达安装在二维机械扫描仪上，从而高效采集规模庞大且多样化的雷达数据集。随后，我们探讨了静态手势识别面临的内在挑战，并发现与形状相同的理想铝制剪影相比，静态手势的雷达散射截面较低。基于这一观察，我们提出假设：如果使用真实人手和“无杂质”铝制剪影的数据共同训练卷积神经网络，那么生成的网络将优于仅使用人手数据训练的网络，因为每种静态手势特有的特征在“无杂质”数据中更加明显，因此 CNN 更容易学习这些特征。基于这一假设，我们扩展了数据采集方法，采集真实人手和铝制剪影的雷达距离剖面及距离-角度剖面，用于 CNN 训练。

本文考虑了三种静态（不移动）手势，并采用距离预处理和距离-角度预处理。仅使用人手数据和深度卷积神经网络时，经过距离预处理和距离-角度预处理的数据分类准确率分别为 85% 和 90%。然而，使用相同数据进行验证，并仅按照我们的假设更改训练数据后，分类率分别提高到 93% 和 95%。准确率的提升证明了这种新颖的“纯净”雷达数据技术带来了改进，而此前文献中从未研究过该技术。此外，本文开发的模型在更具挑战性的分类问题上，优于此前有关静态手势识别的研究 [11]。由于卷积神经网络模型依赖大量有意义的数据，因此这类“纯净”数据和合成数据的采集与生成技术可能会日益普及，因为它们已被证明适用于许多分类和回归问题。未来，我们计划将这一设想拓展至采集更多“纯净”的静态手势，乃至“纯净”的动态手势，以提升卷积神经网络的准确率和鲁棒性。

参考文献

[1] J. Baek、J. Kim 和 E. Kim，《用于手部姿态分类的不同特征分类器比较研究》，载于第*13*届国际控制、自动化与系统会议（*ICCAS*）论文集，韩国光州，2013年10月，第683–687页。[2] Y. Son 和 O. Choi，《使用 Faster R-CNN 的基于图像的手部姿态分类》，载于第*17*届国际控制、自动化与系统会议（*ICCAS*）论文集，韩国济州，2017年10月，第1569–1573页。[3] M. Matilainen、P. Sangi、J. Holappa 和 O. Silvén，《用于手部检测和姿态识别的 OUHANDS 数据库》，载于第*6*届国际图像处理理论、工具与应用会议（*IPTA*）论文集，芬兰奥卢，2016年12月，第1–5页。[4] A. Anaz、M. Skubic、J. Bridgeman 和 D. M. Brogan，《使用卷积神经网络对治疗性手势进行分类》，载于第*40*届 *IEEE* 工程医学与生物学学会国际年会（*EMBC*）论文集，美国夏威夷州檀香山，2018年7月，第3874–3877页。[5] J. Li n 和 C. Huang，《使用 RGB-D 摄像机进行深度梯度估计的三维手部姿态跟踪》，载于*IEEE ISCE* 论文集，中国台湾新竹，2013年6月，第109–110页。[6] Y. Kim 和 B. Toomajian，《利用优化的深度卷积神经网络应用多普勒雷达识别手势》，载于第*11*届欧洲天线与传播会议（*EuCAP*）论文集，法国巴黎，2017年3月，第1258–1260页。[7] B. Dekker、S. Jacobs、A. S. Kossen、M. C. Kruithof、A. G. Huizing 和 M. Geurts，《使用低功耗 FMCW 雷达和深度卷积神经网络进行手势识别》，载于第*14*届欧洲雷达会议（*EuRAD*）论文集，德国纽伦堡，2017年10月，第163–166页。[8] J. S. Suh、S. Ryu、B. Han、J. Choi、J. Kim 和 S. Hong，《用于实时手势识别的24 GHz FMCW 雷达系统及其长短期记忆网络》，载于*IEEE APMC* 论文集，日本京都，2018年11月，第860–862页。[9] Y. Kim 和 B. Toomajian，《利用卷积神经网络和微多普勒特征进行手势识别》，*IEEE Access*，第4卷，第7125–7130页，2016年10月。[10] S. Zhang、G. Li、M. Ritchie、F. Fioranelli 和 H. Griffiths，《基于雷达微多普勒特征的动态手势分类》，载于*CIE* 国际雷达会议（*RADAR*）论文集，中国广州，2016年10月，第1–4页。[11] S. Y. Kim、H. G. Han、J. W. Kim、S. Lee 和 T. W. Kim，《一种利用反射脉冲的手势识别传感器》，*IEEE Sensors J.*，第17卷，第10期，第2975–2976页，2017年3月。[12] S. K. Leem、F. Khan 和 S. H. Cho，《利用无线电传感器和卷积神经网络检测用于数字书写的空中手势》，*IEEE Trans. Instrum. Meas.*，第69卷，第4期，第1066–1081页，2020年4月。[13] J. Park 和 S. H. Cho，《使用机器学习进行人体手势识别的 IR-UWB 雷达传感器》，载于 *IEEE HPCC/SmartCity/DSS* 论文集，澳大利亚新南威尔士州悉尼，2016年12月，第1246–1249页。[14] J. W. Smith、M. E. Yanik 和 M. Torlak，《近场 MIMO-ISAR 毫米波成像》，载于*IEEE* 雷达会议（*RadarConf*）论文集，意大利佛罗伦萨，2020年9月，第1–6页。

[15] M. E. Yanik 和 M. Torlak，《利用毫米波雷达进行隐匿物品检测的近场二维合成孔径雷达成像》，载于 IEEE 无线电与无线研讨会（RWS）论文集，美国佛罗里达州奥兰多，2019年1月，第1–4页。[16] M. E. Yanik 和 M. Torlak，《基于二维合成孔径雷达数据的毫米波近场成像》，载于 SRC Techcon，美国德克萨斯州奥斯汀，2018年9月。[17] M. E. Yanik 和 M. Torlak，《利用稀疏采样孔径数据进行近场多输入多输出合成孔径雷达毫米波成像》，IEEE Access，第7卷，第31 801–31819页，2019年3月。[18] M. E. Yanik、D. Wang 和 M. Torlak，《使用多芯片级联毫米波传感器进行三维多输入多输出合成孔径雷达成像》，载于 IEEE 全球信号与信息处理会议（GlobalSIP）论文集，加拿大安大略省渥太华，2019年11月，第1–5页。[19] D. Sheen、D. McMakin 和 T. Hall，《近场三维雷达成像技术及其应用》，Appl. Opt.，第49卷，第19期，E83–E93页，2010年6月。[20] S. Rao，《毫米波传感导论：调频连续波雷达》，2017年6月26日。访问日期：2021年4月。 [在线]。网址：https://training.ti.com/node/1139153。[21] M. E. Yanik、D. Wang 和 M. Torlak，《多输入多输出合成孔径雷达毫米波成像测试平台的开发与演示》，IEEE Access，第8卷，第126 019–126038页，2020年7月。[22] Y. Kim 和 H. Ling，《基于微多普勒特征并使用支持向量机的人体活动分类》，IEEE Trans. Geosci. Remote Sens.，第47卷，第5期，第1328–1337页，2009年3月。[23] R. J. Javier 和 Y. Kim，《线性预测编码在基于微多普勒特征的人体活动分类中的应用》，IEEE Geosci. Remote Sens. Lett.，第11卷，第10期，第1831–1834页，2014年4月。[24] D. P. Fairchild 和 R. M. Narayanan，《利用人体微多普勒特征的经验模态分解对人体运动进行分类》，IET Radar, Sonar, Nav.，第8卷，第5期，第425–434页，2014年10月。[25] Y. Kim，《利用多普勒传感器和主成分分析检测眨眼》，IEEE Antennas Wireless Propag. Lett.，第14卷，第123–126页，2014年9月。[26] Q. Wan、Y. Li、C. Li 和 R. Pal，《面向智能家居应用的便携式雷达传感器手势识别》，载于第36届 IEEE 工程医学与生物学学会国际年会（EMBC）论文集，美国伊利诺伊州芝加哥，2014年8月，第6414–6417页。[27] V. Pavlovic、B. J. Frey 和 T. S. Huang，《使用混合状态动态贝叶斯网络进行时间序列分类》，载于 IEEE 计算机视觉与模式识别会议（CVPR）论文集，第2卷，美国科罗拉多州柯林斯堡，1999年6月，第609–615页。[28] X. Glorot、A. Bordes 和 Y. Bengio，《深度稀疏整流神经网络》，载于第14届国际人工智能与统计会议（ISTATS）论文集，美国佛罗里达州劳德代尔堡，2011年4月，第315–323页。[29] Z. Zhang、H. Wang、F. Xu 和 Y.-Q. Jin，《复值卷积神经网络及其在极化合成孔径雷达图像分类中的应用》，IEEE Trans. Geosci. Remote Sens.，第55卷，第12期，第7177–7188页，2017年9月。[30] J. Gao、B. Deng、Y. Qin、H. Wang 和 X. Li，《利用复值卷积神经网络增强雷达成像》，IEEE Geosci. Remote Sens. Lett.，第16卷，第1期，第35–39页，2018年9月。[31] E. Richardson、M. Sela 和 R. Kimmel，《通过从合成数据中学习进行三维人脸重建》，载于第4届国际三维视觉会议（3DV）论文集，美国加利福尼亚州斯坦福，2016年10月，第460–469页。[32] X. Wu、L. Liang、Y. Shi 和 S. Fomel，《FaultSeg3D：使用合成数据集训练端到端卷积神经网络进行三维地震断层分割》，Geophysics，第84卷，第3期，I M35–IM45页，2019年4月。[33] V. Allken、N. O. Handegard、S. Rosen、T. Schreyeck、T. Mahiout 和 K. Malde，《使用基于合成数据训练的卷积神经网络进行鱼类物种识别》，J. Mar. Sci.，第76卷，第1期，第342–349页，2019年10月。[34] T. Björklund、A. Fiandrotti、M. Annarumma、G. Francini 和 E. Magli，《使用基于合成数据训练的卷积神经网络进行自动车牌识别》，载于第19届多媒体信号处理国际研讨会（MMSP）论文集，2017年12月，第1–6页。