arXiv:2305.01995v1 [eess.SP] 3 May 2023

一种用于非接触式乐器接口的基于FCNN的超分辨率毫米波雷达框架

*IEEE*多媒体汇刊（已录用）

Josiah W. Smith
电气与计算机工程系
达拉斯德克萨斯大学
理查森 , TX 75080
josiah.smith@utdallas.edu

Orges Furxhi相机系统与计算成像
IMEC-USA 基西米，佛罗里达州 34744
orges.furxhi@imec-int.com

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

摘要

本文提出了一种用于非接触式人机交互（HCI）的框架，该框架采用基于深度学习超分辨率和跟踪算法的新型跟踪技术。通过利用反射雷达波形中蕴含的时空特征，我们的系统能够以前所未有的高分辨率跟踪手部位置和运动特征。与毫米波雷达深度学习领域现有研究中常见的、将样本归类到预定义手势集合中的方法不同，我们提出的成像器采用回归式全卷积神经网络（FCNN）方法，通过空间超分辨率提高定位精度。尽管所提出的技术适用于众多跟踪应用，但本文重点介绍其作为音乐交互界面的应用，以展示手势感知流程和深度学习信号处理链的稳健性。用户可以通过改变手在垂直朝向传感器上方的位置和速度来控制乐器。我们的框架采用商用多输入多输出（MIMO）雷达，而非传统光学传感器，展示了毫米波感知方式在精细运动跟踪方面的有效性，并为众多人机交互任务提供了一种优雅的解决方案。此外，我们还提供了可免费获取的软件包和用户界面，用于控制设备、实时将数据流传输至 MATLAB，并提高本文所采用的信号处理和设备接口功能的可访问性。

本文提供可下载补充材料，网址为 http://ieeexplore.ieee.org，由作者提供。该材料包括多个视频文件，用于展示所提出的框架作为乐器的应用，并将其性能与现有技术进行比较。“Music Example.mp4”文件展示了我们将所提出算法实现为音乐交互界面的演示，该界面配合伴奏音轨演奏。“Cross-Range Oscillation Tracking.mp4”和“Doppler Tracking.mp4”文件分别展示了应用我们的 FCNN-DPF 跟踪算法跟踪横向距离振荡率和多普勒速度的视频。“Simple Tracking.mp4”文件展示了现有特征提取和跟踪技术以及相应手部动作的视频。“FCNN-DPF Tracking.mp4”文件展示了我们所提出的 FCNN-DPF 特征提取和粒子滤波器跟踪算法。该材料大小为 92 MB。

©2020 IEEE。允许个人使用本材料。对于任何其他用途，包括在当前或未来的任何媒体中重新印刷或重新发布本材料以用于广告或促销目的、创建新的汇编作品、转售或再分发至服务器或列表，或在其他作品中重复使用本作品中任何受版权保护的组成部分，必须获得 IEEE 的许可。数字对象标识符:10.1109/TMM.2021.3079695

关键词 深度学习 · 人机交互（HCI） · 全卷积神经网络（FCNN） ·毫米波（mmWave） · 多输入多输出（MIMO） · 雷达感知 · 超分辨率

1 引言

基于多输入多输出（MIMO）毫米波雷达的人机交互（HCI）雷达感知，已成为解决各种感知问题的一种颇具前景的方法。毫米波的物理特性提供了一种安全的高分辨率成像方式，可应对光学传感器因光线不足、雾气或其他视线干扰而无法正常工作的情况。此外，与光学传感器相比，毫米波传感器被认为侵入性更低，也更有利于保护用户隐私。超宽带多输入多输出设备能够以小巧的外形实现厘米级空间分辨率。因此，这类成像设备可以低成本地轻松获取目标场景的精确空间信息。

毫米波传感器是一项相对现代的技术，但一些最早的电子界面是非接触式设备，用于富有表现力的身体动作音乐控制，例如无线电鼓和特雷门琴 [1]。俄罗斯物理学家列夫·特雷门于1921年展示了他的非接触式乐器；这种乐器通过差拍振荡器和电容式传感装置，根据音乐家的手与天线之间的距离进行控制 [2]。近年来，计算机视觉方法已被用于开发非接触式新型音乐接口（NMI），其中大多数依赖光学摄像解决方案。此前已有大量研究探讨了基于光学的新型音乐接口，使用微软 Kinect 和 Leap Motion 等热门传感器。

在 [3], Polfreman 使用 Kinect 跟踪站立表演者双手的三维位置，以构建多模态乐器。Trail 等人提出了一种定音打击乐超乐器，使用 Kinect 同时跟踪两根槌的尖端 [4]。由 Senturk 等人设计的 Crossole 是一种基于 Kinect 的元乐器，可将和弦进行可视化为类似填字游戏的虚拟方块 [5]。Schramm 等人使用 Kinect 分析并分类管弦乐指挥的动作 [6]。
另一种选择是广受欢迎的 Leap Motion 控制器，它能够对整只手（包括手指）进行建模，从而为音乐界面开发探索更精细的基于手部姿势的手势控制提供了可能。Han 等人使用 Leap Motion 传感器开发了两种新型音乐接口（NMI）：*Air Keys*和*Air Pad*。*Air Keys* 跟踪每根手指的动作和位置，以识别音乐家按下了哪些琴键，并演奏相应的音符。类似地，*Air Pad* 跟踪手部位置，在二维水平平面上创建一个通过按压特定区域演奏的二维虚拟鼓垫，因此需要精确的三维手部追踪 [7]。Hantrakul 和 Kaczmarek 使用 Leap Motion 控制器跟踪双手，以控制乐器数字接口（MIDI）乐器和虚拟效果 [8]。同样，Leimu 将 Leap Motion 与惯性测量单元（IMU）配对使用，展示了其相较于单独使用 Leap Motion 控制器在音乐交互界面方面的性能提升 [9]。此外，还有其他尝试，例如采用非侵入式力敏电阻，通过学习和监测音乐家的动作来增强“传统”乐器 [10]。

在光学人机交互领域， [11] 提出了一种音乐交互界面，仅使用便携式 RGB 摄像头，通过手势分类技术识别手势。Akbari 和 Cheng 开发了一套系统，利用对准钢琴琴键的光学摄像头，实时转录钢琴演奏的音乐 [12]。这些项目实现了高性能的实时音乐接口，能够持续进行高精度运动跟踪，但也受到若干关键设计约束的限制，具体而言，要求特定的光照条件和视线范围。正如本文所示，毫米波传感器通过先进的时空算法克服了这些主要障碍，同时提供了更强的隐私保护。然而，利用手部跟踪技术，通过毫米波雷达传感器实现手势音乐交互界面的研究仍然很少。尽管已有大量研究使用深度学习模型和毫米波雷达进行静态和动态手势识别 [13–16], ， Google ATAP 的 Project Soli 是唯一将毫米波雷达用作音乐交互界面的尝试，它通过手势识别和一维位置估计来控制音频合成器的参数 [17]。

本文提出的新型框架为近场毫米波手部跟踪带来了重大进展，并提供了一个易于使用的 MATLAB 软件平台，以便进一步研究实时毫米波人机交互和算法创新。通过采用一种新颖的基于深度学习的技术，突破理论限制以提高分辨率，二维定位性能较以往研究得到了显著提升 [18] 。

现有的非接触式手势控制研究（如手势雷达）通常采用机器学习和深度学习技术。Gurbuz等人 使用多频射频 (RF) 传感器，并采用支持向量机 (SVM)、随机森林、线性判别分析和 k近邻等多种机器学习技术，以高准确率识别美国手语模式 [13]。根据肌电数据，使用支持向量机分类器对足部手势进行分类 [19]。在 [14], Sang 等人 比较了多种技术的分类率。

从隐马尔可夫模型（HMM）等传统机器学习方法，到包括卷积神经网络（CNN）、循环神经网络（RNN）和端到端网络在内的最先进深度学习模型。在毫米波雷达方面， [15] 提出了一种基于微多普勒特征的动态手势 CNN 分类器， [16]则采用一种无菌技术来提高静止、不移动手部姿态的分类率。这些先前的研究为许多分类应用提供了充分的解决方案，在这些应用中，系统会采集数据，并据此判断受试者做出的手势或手部姿态属于预定义类别中的哪一类。
然而，本文探讨的是一个截然不同的问题，并提出了一种基于深度学习的信号处理方案。我们提出的框架并非从一组类别中对样本进行分类，而是通过提升时空图像分辨率，从手部的位置和运动中提取连续的空间与时间特征。为实现这一目标，我们采用一种新型全卷积神经网络（FCNN），以保留图像的几何结构，并执行超分辨率处理，从而提升定位能力。与其他卷积神经网络技术相比，使用FCNN进行雷达信号处理更具优势，因为它能够进行数据驱动的“增强”，而不是像分类那样进行降维。因此，回归型FCNN不会造成信息损失，而是利用训练阶段学习到的额外“上下文”来增强雷达数据。增强后的数据具有多项优势，例如提高信噪比（SNR）、去除杂波、校正近场图像、抑制混叠以及获得更高分辨率的峰值。本文表明，将传统雷达信号处理算法应用于增强数据后，可以实现显著的性能提升。在 [20] 和 [21], 中，采用FCNN和U-Net，在远场平面波假设下提升雷达图像的分辨率。我们提出的新方法将基于FCNN的超分辨率与近场成像相结合；近场成像需要更复杂的球面波补偿，并且在一个小型（8通道）阵列上实现。结果表明，该方法能够显著提升手部跟踪性能。据我们所知，本文首次记录了采用FCNN方法实现近场雷达图像超分辨率以改善定位的工作。在信号处理链中引入我们的增强型全卷积神经网络后，可以实现现有技术无法达到的精细运动跟踪。此外，本文还提出了一种粒子滤波器跟踪算法，通过利用多普勒效应进一步提升跟踪的鲁棒性。与此前使用光学方案进行手势跟踪的研究相比 [3, 7, 8, 11, 22, 23],，我们的方法仅使用单个毫米波传感器即可实现精细手部跟踪，同时提供更高的深度分辨率和更出色的隐私保护。本文提出了一种新型手部跟踪方法，将时空算法、经深度学习增强的特征提取方法以及稳健的位置跟踪算法相融合，用于音乐交互界面。为促进实时毫米波手势应用的进一步开发与原型设计，完整的软件实现可向通信作者索取。据我们所知，该框架是首个公开提供的软件包，支持将实时数据流从毫米波雷达传输至MATLAB，以便开展流畅的信号处理和深度学习算法开发。

本文其余部分安排如下。第 2 节概述调频连续波（FMCW）雷达信号模型和特征提取方法。第 3 节介绍两种鲁棒跟踪算法和估计技术。第 4 节讨论系统实现，第 5 节展示结果。第 6 节讨论第 3.1 节和第 3.2 节中两种跟踪方法的性能、设计约束及其独特优势，最后给出结论。

符号说明：本文中，向量和矩阵均采用粗体表示，其中向量使用小写字母，矩阵使用大写字母。上标 ^{T}和 * 分别表示转置和共轭运算。大小为 $N \times 1$的全 1 向量和全 0 向量分别表示为 $\mathbf{1}_N$和 $\mathbf{0}_N$。变量 c 表示自由空间光速。最后，均值向量为 μ 、协方差矩阵为$\mathbf{\Sigma}$ 的多元高斯分布记作 $G(\mu,\ \mathbf{\Sigma})$。空间坐标视为连续域，以支持连续分布的目标场景；时间变量则采用离散时间建模。

2 MIMO-FMCW雷达信号的预备知识

本节概述 FMCW 雷达线性调频信号的传播模型，并考察运动目标的时空特征。如图 1 所示，成像场景由一个垂直朝向的多基地线性 MIMO 阵列构成。通过采用时分复用 MIMO（TDM-MIMO），在时间上利用正交性；其中，各发射器在不同的时刻启动。本文始终将音乐家的手建模为位于点 (y, z) 的点反射体。

2.1 MIMO-FMCW 信号模型

FMCW 线性调频信号模型已在文献中有充分记载 [24–26]，本文本节将对其进行讨论，以供参考并保持全文连贯。首先，考虑一对分别位于 y-z 平面上 $(y_T, 0)$ 和 $(y_R, 0)$ 处的发射机/接收机，以及一个反射率为 p 、位于 (y, z) 处的理想点目标。假设理想传播 on

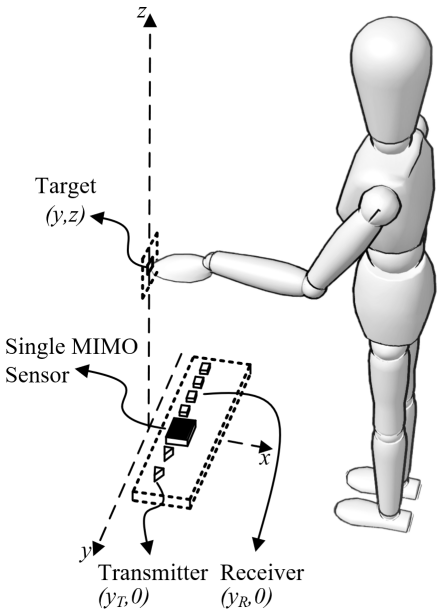

图 1：成像几何，其中线性 MIMO 阵列垂直朝向，音乐家在整个 y-z 平面内移动手部。

在无噪声信道中，连续时间信号可建模为

$$s(y_T, y_R, k) = \frac{p}{R_T R_R} e^{jk(R_T+R_R)}, \tag{1}$$

其中， R_T、 R_R 分别是发射机和接收机到点目标的距离， $k = 2\pi f/c$ 为瞬时波数。线性调频信号 $f = f_0 + Kt$ 的频率随时间线性增加，斜率为K ，起始频率为 f_0。连续时间信号（1）由雷达模数转换器（ADC）以采样频率 f_S 进行采样，可表示为离散时间形式：

$$s(y_T, y_R, n_k) = \frac{p}{R_T R_R} e^{j(k_0+\Delta n_k)(R_T+R_R)}, \tag{2}$$

其中， n_k 为波数索引， $k_0 = 2\pi f_0/c$ 为与起始频率f_0对应的起始波数， $\Delta = 2\pi K/（cf_S）$ 为波数步长。

为简化后续信号处理，期望采用 [27] 中提出的近似方法，将式（2）表示的多基地 MIMO 拍频信号近似为相应的单基地等效信号：

$$\hat{s}(y', n_k) = s(y_T, y_R, n_k) e^{-j(k_0+\Delta n_k)\frac{d_y^2}{4z_0}}, \tag{3}$$

仅当 d_y 较小时才成立，其中 z_0 是参考平面，通常取目标场景的中心。设 y' 为位于各收发对中点的虚拟阵元位置， R 为各虚拟阵元到点反射体的相应距离，则所得单基地拍频信号近似为

$$\hat{s}(y', n_k) \approx \frac{p}{R^2} e^{j2(k_0+\Delta n_k)R}. \tag{4}$$

根据式（4），目标的空间位置（y， z）以径向距离R的形式嵌入雷达拍频信号中。

2.2 多普勒雷达信号处理

利用多普勒效应，可以从式（4）所表示的拍频信号中提取目标的相对速度。如 [28], 中所述，通过以已知的脉冲重复间隔（PRI） T_{PRI} 发射一系列线性调频波形，可以将运动目标的速度识别为沿线性调频索引维度的频率分量，其表达式为

$$\hat{s}(y', n_k, n_c) = \frac{p}{R^2} e^{j(2(k_0+\Delta n_k)R + \frac{4\pi v T_{PRI}}{\lambda_0} n_c)}, \tag{5}$$

其中， R 为目标的初始距离， v 为目标的速度， λ_0 为对应于 f_0 的波长，n_c 为啁啾索引，

因此，沿时间采样的拍频信号是一个二维复正弦信号，其第一维和第二维的频率分别对应于目标的距离和速度。随后，为了提取距离和速度，传统方法会对一个矩阵执行二维快速傅里叶变换（FFT），该矩阵的行或列由连续的啁啾信号组成。

2.3 距离迁移算法图像重建

为实现高保真二维定位，我们采用距离迁移算法（RMA），而非传统的距离-角度快速傅里叶变换方法 [21,24],，后者的定位精度已知较低 [29]。RMA 的主要目标是重建目标场景的反射率函数， $p(y,\ z)$。对于分布式目标，拍频信号可建模为场景中每个点的后向散射信号之和，并忽略幅度项，即是

$$\hat{s}(y', n_k) = \iint p(y,z) e^{j2(k_0+\Delta n_k)R} dydz. \tag{6}$$

该目标模型假设目标在空间中分布，且其反射率仅取决于空间位置，并忽略反射率函数对频率的任何依赖。使用驻相法对 (6) 求逆，可以高效地估计反射率函数 $p(y,z)$，方法如下。

$$\hat{p}(y,z) = \text{IFT}_{\text{2D}}^{(k_y,k_z)} \left[\mathcal{S} \left[\text{IFT}_{\text{1D}}^{(y')} [\hat{s}^*(y', n_k)] \right] \right], \tag{7}$$

其中， $\mathcal{S}[\bullet]$ 表示 Stolt 插值操作 [27] ，FT[•]和 IFT[•] 分别表示正向和逆向傅里叶变换算子。为避免混叠，必须考虑图像采样准则 [26]。 y 方向和 z方向的空间分辨率受物理及设备限制，其表达式为

$$\delta_y = \frac{\lambda_c z_0}{2D_y}, \tag{8}$$

$$\delta_z = \frac{c}{2B}, \tag{9}$$

其中， λ_c 为波长　　，对应于线性调频信号扫描中心频率的波长； D_y 为孔径尺寸沿 y方向的长度，其中 z_0 为成像场景 [26]的中心。

恢复目标场景的二维反射率函数后，随后估计手部位置，具体如下：

$$\{\hat{y}, \hat{z}\} = \arg\max_{\{y,z\}} \hat{p}(y,z). \tag{10}$$

此外，还可以利用上述多普勒原理，通过对连续啁啾信号进行傅里叶分析来提取目标的速度。为了充分利用第 3.2.2 节所述的深度学习框架，并降低所需的计算复杂度，我们在执行距离迁移算法并估计手部位置后提取速度。

如式 (5) 所示，速度与波数索引相互解耦，且对应于线性调频索引维度上的缩放频率分量。因此，重建图像 $\hat{}p(y,\ z)$ 中保留了与速度对应的相位项。由此，可通过沿最近图像的线性调频索引 n_c 维度执行快速傅里叶变换来获得速度剖面。我们不对三维数组 $\hat{}p(y,\ z,\ n_c)$ 执行快速傅里叶变换，而是对与估计位置 $\hat{}y$ 对应的图像切片执行快速傅里叶变换，从而得到沿 z 方向的速度剖面，其中 n_d 为速度索引，如下所示

$$\hat{d}(z, n_d) = \text{FFT}_{\text{1D}}^{(n_c)} \left[\hat{p}(y,z,n_c) \Big|_{y=\hat{y}} \right]. \tag{11}$$

最后，可以通过视频脉冲积分，利用公式（11）估算速度。

$$\hat{v}_d = \arg\max \sqrt{\int |\tilde{d}(z, n_d)|^2 dz}. \tag{12}$$

通过该方法计算得到的速度称为多普勒速度。使用这种方法恢复的速度受到 $[-\frac{\lambda_0}{4T_{PRI}}, \frac{\lambda_0}{4T_{PRI}}]$ 之间的时间和物理约束的限制。随后，利用多普勒速度，通过多普勒验证粒子滤波器来提高跟踪性能。

3 毫米波雷达上的时空成像

本节介绍我们提出的成像器方法，该成像器能够实现高精度的 HCI 手部跟踪。本文的贡献在于，通过同时采用新型超分辨率 FCNN 和所提出的跟踪算法，提升了二维定位的算法性能。虽然我们将探讨此类算法作为新型音乐接口的应用，但我们的基于毫米波雷达的感知算法可应用于众多人机交互问题。

需要注意的是，本研究并非旨在与嵌入式人机交互解决方案和现有音乐接口在计算效率方面展开竞争。相反，本文的主要贡献在于提出用于超分辨率时空手部跟踪的新算法，并提供一个可免费下载的平台，以提高可访问性并鼓励该领域的进一步研究。因此，我们将主要关注算法的开发及其定位性能。关于性能和实现问题的讨论被视为次要内容，并将在第4节和第6节中加以探讨。

3.1 经典时空特征提取技术

本节将介绍一种用于非接触式乐器接口的简单时空感知方法。虽然我们的系统通常会跟踪用户的手的二维位置和速度，但我们确定了实现对音乐交互界面精细控制所需的三个底层特征：距离、横向距离振荡和速度。根据图 1 所示的几何关系，我们将距离定义为手沿 z 轴的位置，即传感器与用户的手之间的垂直位移。同理，横向距离定义为手沿 y 轴的位置。随后，横向距离振荡指手沿横向距离方向振荡的速率。速度指手相对于距离 z轴的速度。选择这些参数，是为了使输出的音乐交互界面主要由音乐家的手的距离控制，其次由横向距离振荡和速度控制。不过，如第 4 节所述，用户也可以通过 MIDI接口根据个人偏好指定这些参数。在本文余下部分中，我们将这些参数称为从雷达拍频信号中提取的特征。

在简单手势跟踪模式下，使用公式 (10) 和 (12)，从重建图像和近期图像缓冲区中提取二维位置和速度（$\hat{}y$, $\hat{}z$, $\hat{}v_d$）。在下一节中，将从原始数据中提取的三个参数视为一个称为含噪测量向量**r**的向量。在理想场景下，带宽、天线阵列尺寸和信噪比都相当大，趋于无穷大。对于一款 8 通道车载毫米波雷达和一只人手的情况，带宽有限（4 GHz），天线阵列尺寸较小（$D_y = 2\lambda_c$），而且与噪声水平相比，人手的反射率并不高。因此，仅从重建的 RMA图像中提取最大值，会得到零散的位置和速度估计。即使在理想情况下，我们系统沿 y 和 z方向的空间分辨率也分别为 $\delta_y = 7.5$ 厘米和 $\delta_z = 3.75$ 厘米。经典直接跟踪方法还未考虑其他一些因素，包括波束方向图、残余相位误差和天线耦合。成像场景中的所有这些限制和非理想因素都会使图像质量下降，导致位置和速度估计含噪；然而，其中许多问题具有解析形式，无法直接通过经典方法解决。为了解决这些问题，我们提出了一种采用全卷积神经网络进行超分辨率和图像增强的新型数据驱动方法。

3.2 基于FCNN的超分辨率特征提取与粒子滤波器跟踪方法 ds

在本节中，我们改进了简单的跟踪技术，以克服成像场景中的噪声和基本非理想因素，从而显著提升用户体验。本节所展示的概念适用于音乐接口领域之外的许多跟踪和高分辨率成像应用。

为提高所提出音乐交互界面的跟踪鲁棒性，我们采用了广为人知的粒子滤波器 [30] 并提出了一种新颖的改进。传统方法（例如扩展卡尔曼滤波器（EKF））会使用运动模型，而我们实现的粒子滤波器则无需确定性运动模型。在本应用中选择粒子滤波器，是因为我们在实验中发现，其他传统方法的跟踪性能较差，导致定位结果时有时无，或估计过度平滑、反应迟缓。此外，粒子滤波器能够跟踪非线性动力学，而且无需预先了解运动模型或噪声参数即可实现鲁棒定位，因此具有优势。我们对粒子滤波器提出的扩展引入了一种新颖的粒子重采样和权重计算过程。这些改动并未显著改变现有的粒子滤波框架，但我们将其纳入其中，以展示其在雷达手部跟踪中的可行性，并充分记录完整的软件实现。此外，粒子滤波算法的稳定性已在其他研究中得到考察，因此本文不再讨论 [31]。

在我们对粒子滤波器的改进中，控制输入是朝最新测量值移动的加权结果。为说明所提出的算法，考虑沿 y 和 z方向同时进行位置估计的情形。新的含噪测量向量 $\mathbf{r}$ 有两个元素，即最新的位置估计值 $\hat{}y$ 和 $\hat{}z$，它们由前一节所述的方法提取。对于二维定位，$\mathbf{X}_n$ 是一个大小为 $N \times 2$ 的矩阵，其各行是在时间索引 n处每个粒子的 (y, z) 坐标，其中 N 是粒子数量，而 $\mathbf{w}_n$ 是与各粒子对应的权重向量。二维位置的估计值（亦称估计状态）构成向量 $\mathbf{s}_n$。

执行迭代算法之前，初始粒子状态矩阵 $\mathbf{X}_0$，和初始权重向量 $\mathbf{w}_0$，分别使用感兴趣区域（ROI）内的随机位置和均匀权重进行初始化。在每次迭代中，粒子滤波器接收 $\mathbf{r} = [\hat{}y, \hat{}z]^T$（最新的位置估计值）；$\hat{}\mathbf{X}_{n-1}$（前一次粒子滤波器的位置）；$\mathbf{X}_{n-1}$（使用权重 $\mathbf{w}_{n-1}$ 采样得到）；$\mathbf{a} = [a_y, a_z]^T$（加权向量，其两个元素分别为含噪估计值 $\hat{}y$ 和 $\hat{}z$ 提供权重）；$\mathbf{s}_{n-1}$（先前估计状态的向量）；$\mathbf{\Psi} = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2]$（由分布 $G(\mathbf{0}_N, \mathbf{\Sigma}_\psi)$ 中的两个随机向量组成的矩阵）；以及 $\mathbf{\Sigma}_w$（权重分布的协方差矩阵）。

算法： 改进粒子滤波算法

输入： $\mathbf{r} = [\hat{}y, \hat{}z]^T$, $\hat{}\mathbf{X}_{n-1}$, $\mathbf{a}$, $\mathbf{s}_{n-1}$, $\mathbf{\Psi}$,
$\mathbf{\Sigma}_w$ **输出：** $\mathbf{s}_n = [\tilde{y}, \tilde{z}]^T$ $\mathbf{1}$ $\mathbf{X}_n \leftarrow \hat{}\mathbf{X}_{n-1} + \mathbf{1}_N$
$\mathbf{a}^T(\mathbf{r} - \mathbf{s}_{n-1}) + \mathbf{\Psi} \mathbf{2}$ $\mathbf{w}_n \leftarrow e^{-\frac{1}{2}(\mathbf{X}_n - \mathbf{s}_{n-1})}$
$^T\mathbf{\Sigma}_w^{-1}(\mathbf{X}_n - \bar{\mathbf{s}}_{n-1})$ $\mathbf{3}$ $\mathbf{s}^n \leftarrow \frac{1}{\mathbf{1}_N^T \mathbf{w}_n} \mathbf{X}_n^T \mathbf{w}_n$

正确处理关键步骤——（步骤1）对粒子状态进行重采样，以及（步骤2）计算新权重——对于有效实现我们提出的粒子滤波算法至关重要。
粒子重采样过程包括按照指定权重将粒子向新的测量值移动。$\mathbf{a}$、$\mathbf{r}$和$\mathbf{s}_n$ 的大小可根据粒子滤波器需要跟踪的参数数量进行调整。因此，新测量值不会主导运动跟踪，而是以加权方式影响定位过程。图2通过 $a_y = a_z = 0.5$展示了重采样过程。请注意，在计算新权重之前，需要通过添加扰动项 $\mathbf{\Psi}$来执行粒子扩散。

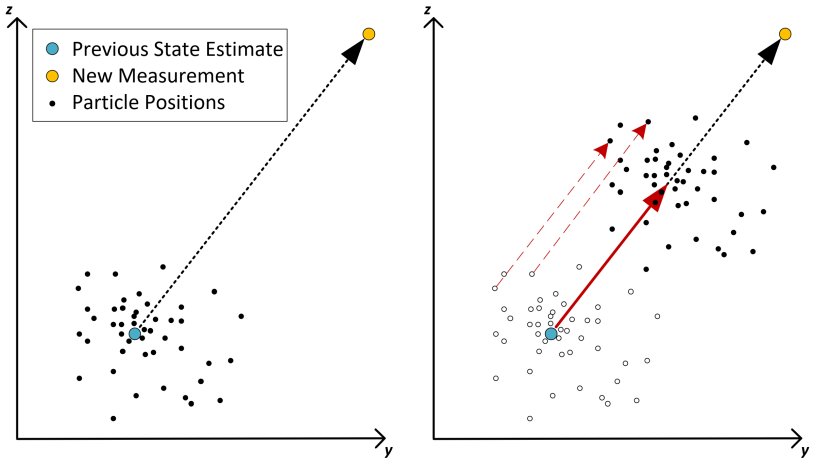

图2：改进粒子滤波算法重采样过程的可视化示例。根据权重向量$\mathbf{a}$，通过向新的测量值进行平移变换，对粒子位置进行重采样，其中$a_y = a_z = 0.5$。

新权重根据多元高斯分布计算，其中，先前估计的状态$\mathbf{s}_{n-1}$作为均值向量，预定义的协方差矩阵为 $\mathbf{\Sigma}_w$。因此，与先前估计状态距离较近的粒子将被赋予比距离较远的粒子更高的权重。这使状态估计趋于小幅变化，同时监测当前位置的移动情况。据我们所知，本节提出的重采样和权重计算步骤为粒子滤波器提供了一种全新的实现方法。对于许多

对于需要通过毫米波雷达实现精确、一致的定位和运动跟踪的应用，我们改进后的粒子滤波算法非常适用，因为它倾向于对状态进行稳态估计，同时仍会持续监测噪声传感器输入。

3.2.1 多普勒验证的实时加权

在本节中，我们提出了一种动态加权技术，通过利用位置与速度之间的依赖关系，实时更新 $\mathbf{a}$ 。我们的方法将多普勒速度估计与根据距离采样估计出的速度之间的一致性，作为衡量新测量值可靠性的指标。因此，即使存在含噪的位置估计，多普勒速度的可靠性也能提高沿距离（z）维度的目标位置跟踪性能。在通过（12）计算出多普勒速度后，利用最近的距离估计值，通过最小二乘估计器计算采样速度（$\hat{}v_s$），即

$$\hat{v}_s = \frac{N_z T_{PRI} \sum_m (\mathbf{z}^{(m)} m) - T_{PRI} \sum_m \mathbf{z}^{(m)} \sum_m m}{N_z \sum_m (\mathbf{z}^{(m)})^2 - \left(\sum_m \mathbf{z}^{(m)}\right)^2}, \tag{13}$$

其中$\mathbf{z}^{(\ \ m)}$ 是最近的 $\hat{}z$ 估计值向量 $\mathbf{z}$ 的第 m^{th} 个元素，其中 $\mathbf{z}^{(N_z-1)}$ 为最新值 .

多普勒估计速度与样本估计速度之间的差值计算为 $\Delta_v = |\hat{}v_d - \hat{}v_s|$，并用于奖励函数 (14)，以实时更新赋予新噪声测量值的权重。

$$a_z(\Delta_v) = \begin{cases} a_{z,0} \cos\left(\frac{2\pi T_{PRI} \Delta_v}{\lambda_0}\right) & \text{if } \Delta_v \le \frac{\lambda_0}{4T_{PRI}} \\ 0 & \text{if } \Delta_v > \frac{\lambda_0}{4T_{PRI}} \end{cases} \tag{14}$$

当采样速度接近多普勒速度，即 Δ_v 较小时，奖励函数接近 $a_{z,\ 0}$。因此，新的测量值得到可靠的多普勒速度的佐证，并据此加权。在粒子重采样过程中，与多普勒速度相矛盾的离群值和错误测量值所占权重较低。为实现多普勒佐证粒子滤波器，在我们提出的粒子滤波算法的每次迭代中，$\mathbf{a} = [a_y, a_z(\Delta_v)]^T$ 均依据式（14）动态更新。

3.2.2 通过改进全卷积神经网络实现改进的二维位置估计

改进粒子滤波算法提高了跟踪的一致性和平滑性；然而，仪器延迟、环境噪声/设备噪声、多基地效应以及非球形波束模式等若干问题仍未得到解决，并且会降低跟踪性能。为克服这些非理想因素，我们提出了一种基于全卷积神经网络的新型图像增强技术，可提高二维位置估计的精度、后续跟踪精度以及多普勒频谱信噪比。与采用远场假设并使用合成生成数据进行训练的先前全卷积神经网络合成孔径雷达（SAR）技术相比 [20], ，我们提出的增强型全卷积神经网络方法直接处理近场图像，即使在小孔径条件下也能提高定位精度，并采用一种新颖的训练技术，使网络能够学习环境噪声和设备噪声、近场波束模式以及多基地效应。

为训练增强型全卷积神经网络，我们构建了一个包含真实人手数据和合成生成数据的数据集。真实人手数据通过采集用户将手放在相对于设备已知位置时的帧来获取，合成数据则用于补充训练集。每个合成样本都通过使用公式（6）模拟 MIMO 拍频信号生成，其中包含一个位于已知位置的理想点目标，以及从雷达采集的加性真实设备噪声。模拟位置经过随机化，以均匀覆盖感兴趣区域（ROI）。真实数据和合成数据均作为全卷积神经网络训练过程中的特征，使网络能够拟合非理想波束模式、真实多径和多基地效应、人手的经验反射特性、设备噪声和环境噪声，以及整个感兴趣区域内的手部位置。

为训练图像到图像回归全卷积神经网络，每个训练特征（真实或合成图像）都必须对应一个真实值标签。真实值标签图像由该模型合成生成

$$\mathcal{I}(y, z) = e^{-(y-y_0)^2/\sigma_y^2 - (z-z_0)^2/\sigma_z^2} \tag{15}$$

其中，位于 (y_0, z_0) 的预期目标宽度分别由 y 和 z 维度上的 σ_y 和 σ_z 决定；根据 3 dB 波束宽度定义 [20]，分辨率分别为 $1.18\sigma_y$ 和 $1.18\sigma_z$。每个标签均根据每幅特征图像中人手或目标位置的必要信息生成。在训练过程中，FCNN 学习失真、模糊的 RMA 图像与使用 (15) 生成的理想图像之间高度非线性的关系。我们提出的新型训练技术通过拟合非理想成像约束，训练出稳健且具有良好泛化能力的 FCNN，从而提高图像信噪比和定位能力。此外，训练后的网络能够实现精确定位

突破设备的物理限制，显著提升跟踪性能。全卷积神经网络的训练将在第 4.4 节中讨论，相关结果将在第 5.3 节中展示和讨论。

此外，通过分离对应于人手的峰值，可以减轻其他位置的杂波和相位噪声，从而提高多普勒频谱信噪比以及后续的速度估计。因此，FCNN增强了粒子滤波器之前从雷达拍频信号中提取的空间和时间特征。通过结合所提出的粒子滤波器和增强型全卷积神经网络，我们的新算法能够稳健地跟踪距离、横向距离振荡和速度，并将其映射到音乐交互界面控制项。

4 系统设计与实现

在本节中，我们介绍上一节讨论的经典跟踪技术以及我们提出的新型超分辨率特征提取与跟踪算法的系统实现。

4.1 硬件与软件实现

所提出系统采用的硬件包括一台德州仪器（TI）AWR1243 汽车雷达，以及一台 DCA1000EVM 实时数据采集适配器。该 TI 雷达是一款 MIMO-FMCW 毫米波雷达，工作带宽为 4 GHz，中心频率为 79 GHz。在本研究中，我们使用由 2 个发射天线（TX）阵元组成的线性 MIMO 阵列，阵元间距为 $2\lambda_c$，并使用由 4 个接收天线（RX）阵元组成的线性 MIMO 阵列，阵元间距为 $\lambda_c/2$。由此形成的虚拟阵列包含 8 个等间距的虚拟阵元，间距为 $\lambda_c/4$ [24]。我们采用 [27] 中讨论的校准方法，以减小距离偏差、固定相位误差和仪器延迟。在此过程中，从已知位置的角反射器采集数据，用于确定天线之间的距离偏差和相位偏移。与光学或红外校准不同，该过程不受照明、温度条件以及用户手掌大小等因素的影响。因此，只需进行一次校准，即可适用于多种环境和不同用户。

用于信号处理、可视化和机器学习的软件平台以 MATLAB 编写。尽管与其他语言相比，其计算效率较低，但采用 MATLAB 能为研究人员提供易于使用的平台，帮助他们开展这项工作，并利用我们的定制工具快速构建定制实时算法的原型。算法在 PC 上通过验证后，即可部署到此类嵌入式设备上，以优化特定应用场景中的使用。为了提供良好的音乐交互界面使用体验，必须考虑延迟和时序问题，相关讨论见第 6 节。

4.2 实时数据获取与交互式 MATLAB 用户界面

为了将设备中的数据流传输至 MATLAB，我们编写了定制的 UDP 接口软件。该程序以 C++ 高效实现，能够接收顺序传入的 UDP 数据包，将数据包整理成每个线性调频信号，并通过共享内存将数据提供给 MATLAB。

我们编写了一个定制的交互式 MATLAB 图形用户界面（GUI），如图 3 所示，作为整个框架的统一用户界面。MATLAB 图形用户界面与 TI mmWave Studio [32] 交互，以控制硬件设置并初始化 UDP 接口，从而无需用户在 GUI 之外进行设置。雷达通过完全集成的实现持续采集数据并将其流式传输至 MATLAB。根据我们的实验结果，尽管 MATLAB 的计算速度不足以支持实时系统实现，但它能够以约 250 Hz 的频率完成数据采集、信号处理、深度学习、可视化和信号输出。在早期原型开发阶段，我们认为这一吞吐量足以用于研究超分辨率跟踪算法以及简单音乐交互界面的性能。

借助我们提出的方法，软件能够提取用户的手的高分辨率时空特征，并通过内置音频输出工具或附带的 MIDI 输出将其映射为相应的输出。定制的 MATLAB 图形用户界面为研究和展示我们的方法提供了易于使用的选项，同时也提供了一个开源平台，以促进多媒体和雷达领域开展进一步的合作研究。

如前所述，所提出的音乐交互界面主要通过用户的手与设备之间的距离（z 位置）来控制输出。使用内置的音频输出工具和 MIDI 输出时，用户的手与设备之间的距离直接控制音符选择。与允许连续选择音符的特雷门琴不同，我们的界面会将用户输入量化为预先定义的子区域，每个子区域对应由用户定义的音符。要演奏所需的音符，用户必须将手垂直移动到与该音符对应的位置。用户可以在交互式 MATLAB 图形用户界面中设置子区域和可用音符。使用我们的用户界面可以轻松实现类似特雷门琴的连续控制。不过，与模拟特雷门琴一样，要实现真正的连续音高控制仍有一定限制，因为数字软件必须在某个环节进行离散化。我们选择了

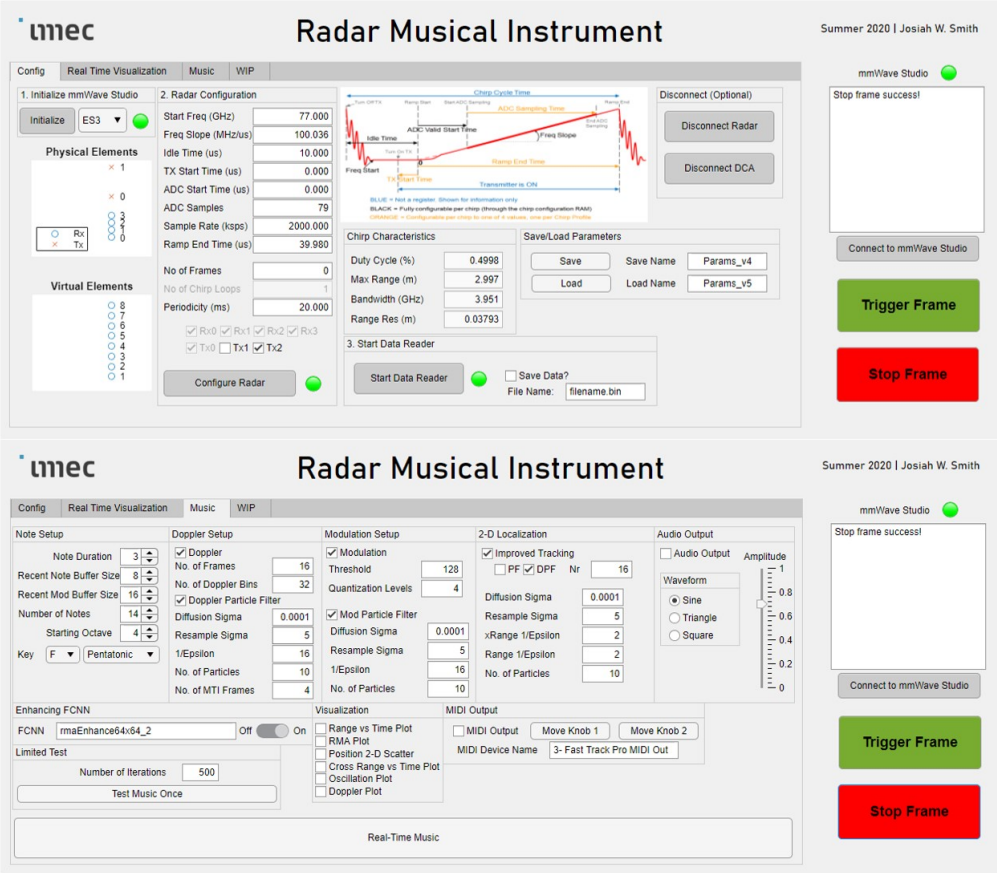

图 3：交互式 MATLAB 图形用户界面：用于设置雷达设备、指定跟踪和音乐生成参数，以及开始和停止设备采集的单一界面。

离散音高控制作为我们框架的默认模式，以便简化音符选择并促进其在众多领域的可访问性。特雷门琴通常被认为是最难控制的乐器之一，因为它需要连续选择音符；即使只是演奏简单的乐句并保持正确的音准，也需要丰富的经验。我们的方案允许经验丰富的音乐家启用连续音高控制，同时为经验较少的用户提供简单易用的界面。因此，我们的平台适用于众多用户，既能带来愉悦的体验，也能让用户无需精湛的音乐技艺便轻松进行演示。同样，用户可以通过让手沿 y 方向来回摆动，或以较快或较慢的速度移向下一个音符，来调节次要参数——横向距离振荡和速度。内置的音频输出工具利用横向距离振荡来控制颤音效果（对音频信号进行低频调制）。因此，使用该工具时，用户可以通过改变手与设备之间的距离来选择所需音符，并通过以所需速率摆动手来演奏相应速率的颤音。或者，MIDI 输出工具会将横向距离振荡和速度作为 MIDI 参数提供，供用户在连接到我们音乐交互界面 MIDI 输出端口的虚拟乐器环境中进行设置。因此，我们提出的算法旨在实现类似 MIDI 键盘的操作方式：手与设备之间的距离控制音符选择，而横向距离振荡和速度则作为 MIDI 参数供用户分配。

4.3 简单特征提取与跟踪算法信号处理链

简单特征提取和跟踪方法的信号处理链如图 4 所示。将拍频信号载入 MATLAB，并在其中执行上一节所述的预处理（距离迁移算法和峰值检测）；通过使用 (10) 和 (12) 提取时空特征，将用户输入（二维位置和速度）转换为音频或 MIDI 输出。本文使用用户的手的位置和速度构建音乐手势界面；不过，只要有足够的二维阵列，我们的新算法便可轻松应用于多种不同的人机交互应用，甚至用于三维定位。重建后的 RMA 图像和提取的原始特征

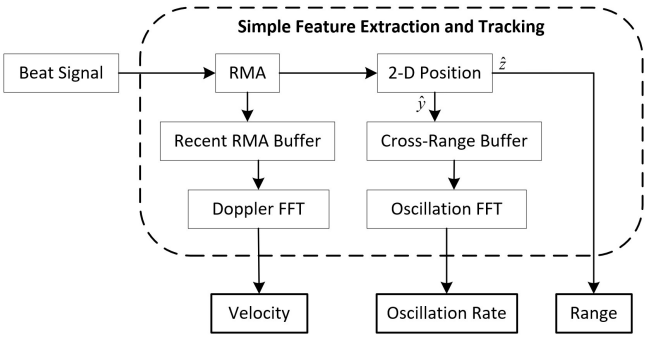

图 4：简单的跟踪信号处理链。对拍频信号执行 RMA 后，直接从原始 RMA 图像中提取特征。

经典技术所获得的信息可供粒子滤波算法和超分辨率 FCNN 利用，从而提高跟踪性能。

4.4 超　分辨率框架——训练 FCNN 并实现粒子滤波算法

为实现我们的超分辨率特征提取与跟踪框架，必须先训练超分辨率 FCNN。增强型全卷积神经网络使用来自人手的真实数据，以及受加性真实雷达噪声污染的仿真数据进行训练。FCNN 以 65536 幅仿真 RMA 图像和 23040 幅真实人手 RMA 图像作为输入和输出图像进行训练，$\sigma_y = \sigma_z = 1$ mm 对应的横向距离和距离分辨率为 1.18 mm。每个仿真样本均在 ROI $y \in [-0.1, 0.1], z \in [0.1, 0.5]$ 内的随机位置生成。合成数据覆盖整个 ROI，使网络在学习成像方案非理想特性的同时，能够很好地泛化至不同位置。如图 5 所示，在 ROI 内的 45 个位置上，每个位置均采集 512 个真实手部样本。对于合成样本和真实人手样本，均使用公式 (15) 生成相应的真实值图像，并将其用作训练标签。因此，训练集由真实数据和仿真数据构成的特征，以及表示各已知位置理想预期响应的标签组成。

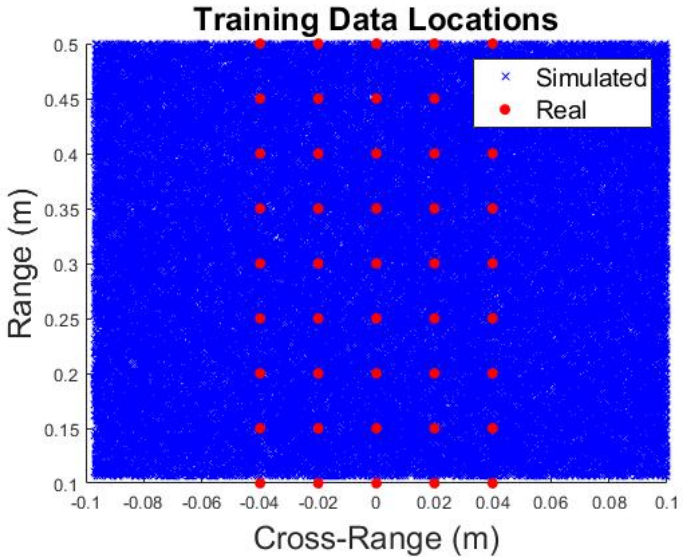

图 5：用于训练增强型全卷积神经网络的训练数据位置。真实数据（红色）通过将手静止保持在已知位置采集。仿真数据（蓝色）通过从连续感兴趣区域（ROI）中随机选择位置生成。

所提出的增强型全卷积神经网络的架构如图 6 所示。该网络由四个卷积层组成，卷积核尺寸依次递减，每个卷积层后均连接一个非线性修正线性单元（ReLU）层。每个卷积层均采用零填充，使输出尺寸与输入尺寸相同。对该网络进行 100 个训练周期的训练需要

在配备单块 NVIDIA GTX1080TI 显卡的机器上耗时 5 小时。我们还研究了其他网络架构和训练时长，但这一组合在实现高性能的同时，也具备实时效率。

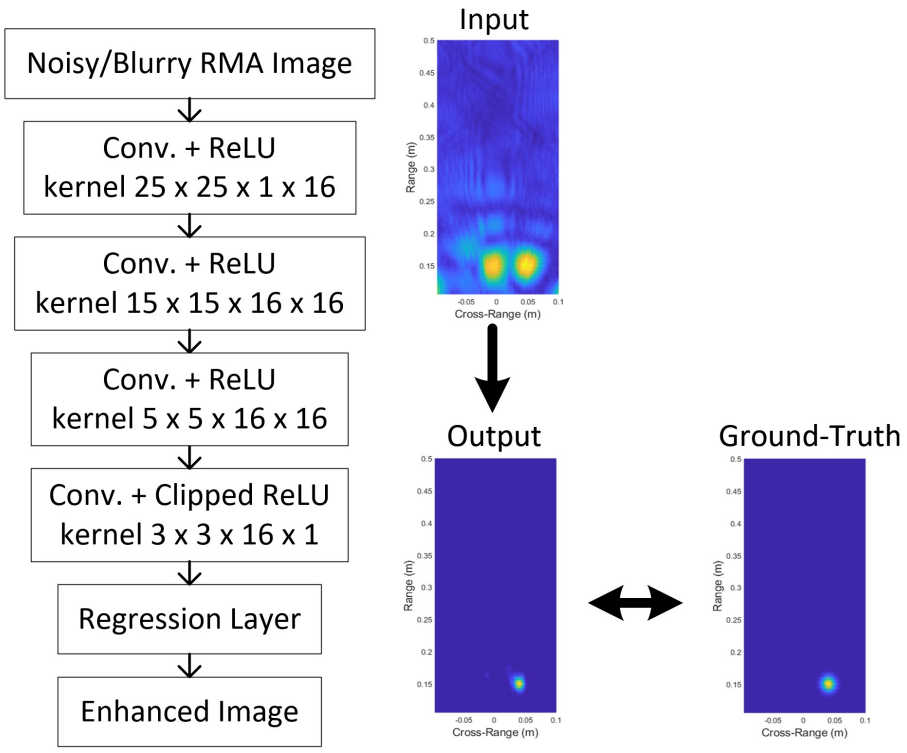

图 6：超分辨率增强FCNN的网络拓扑。所选卷积核和层尺寸能够充分学习失真的RMA图像的非理想形状，同时保持较高的计算效率，以实现实时部署。

采用所提出的技术训练超分辨率 FCNN 后，我们的新型跟踪算法即可利用前文讨论的粒子滤波器实现。该多普勒佐证粒子滤波器用于跟踪手在 y-z 平面上的位置，另外两个粒子滤波器则用于跟踪多普勒速度和横向距离振荡。增强特征提取与跟踪方法的完整信号处理链如图 7 所示。算法输出时空特征，可用于多种跟踪应用。此外，如果需要获取手在距离和横向距离振荡率方面的二维位置，则可轻松调整算法，以输出所需的空间特征。

5 时空特征提取与跟踪结果

在本节中，我们将概述所提出的新型跟踪和特征增强算法的结果，从简单的经典技术开始，并将其性能与我们提出的方法进行比较。与传统方法相比，我们改进后的跟踪方案显著提升了性能，并可在小型雷达平台上实现现有方法无法企及的稳健超分辨率跟踪。

5.1 真实值 - 理想运动轨迹

为验证特征估计技术，采用虚拟原型方法。根据公式（2），模拟一个点目标运动，其位置和速度如图 8 所示，位置和速度为 y-z。利用这一理想运动剖面，将我们提出的方法的跟踪性能与传统技术进行比较。将雷达在空场景下采集的真实噪声添加到每个合成拍频信号中，作为.....

$$\tilde{s}(y_T, y_R, k) = \frac{p}{R_T R_R} e^{jk(R_T+R_R)} + \alpha\tilde{\omega}(y_T, y_R, k), \tag{16}$$

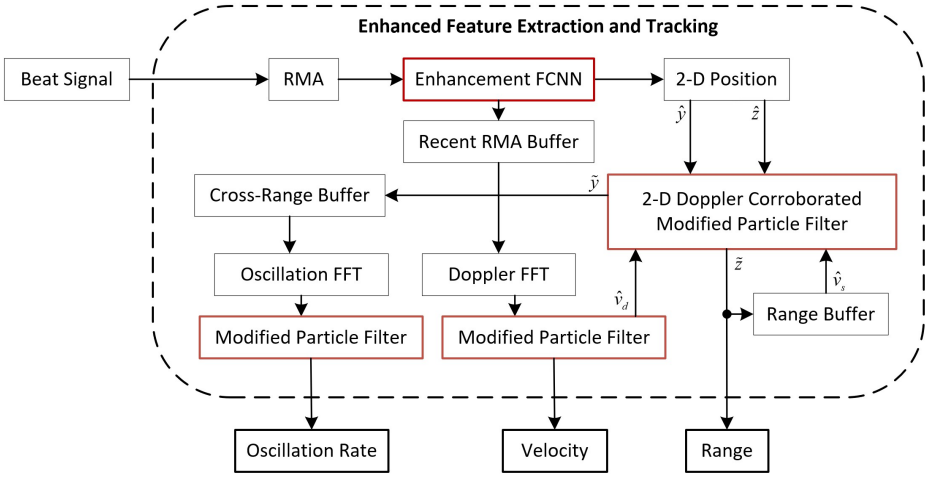

图 7：增强后的跟 踪信号处理链。增强方法的关键要素以红色突出显示。

其中， $\tilde{\omega}$ 是会使理想模拟拍频信号的幅度和相位失真的复值噪声样本， α 控制信噪比。

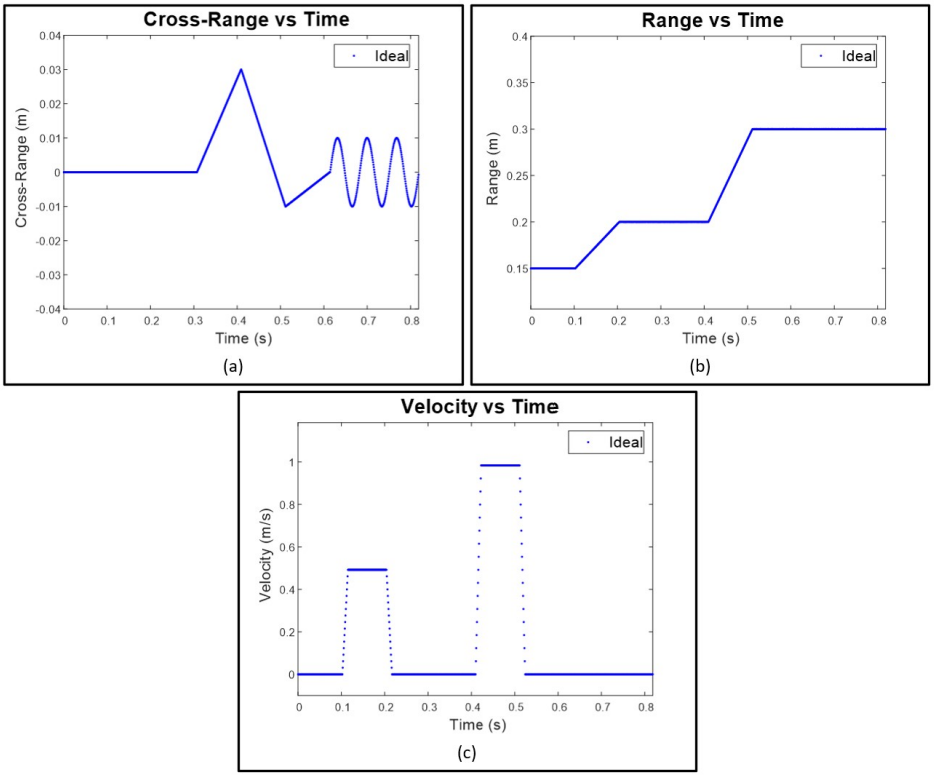

图 8：目标在 (a) 横向距离和 (b) 距离方向上的理想运动剖面，以及随时间变化的 (c) 距离速度轮廓。

图 8 所示的运动轮廓展示了目标的理想距离（z）、横向距离（y）和速度（v）。该运动轮廓除了正弦横距振荡外，还包括距离域和横向距离域中的独立运动和联合运动。对于我们的仿真，我们使用 $p \in [0.5, 1]$ 生成 4096 个时间采样，以模拟根据先前手部数据观测到的手部实测雷达截面积（雷达散射截面）的方差，并使用 $\alpha \in [1, 3]$ 改变信噪比

在各样本之间。对于每个时间样本， p 和 α 的值会在指定区间内随机选取，从而为仿真数据提供一定程度的随机真实性。

5.2 经典时空成像结果

首先，实现第 3.1 节中讨论的简单跟踪方法，以提供基线性能指标。执行图 4 所示的信号处理链，提取时空特征。在每次迭代中，特征均直接从原始 RMA 图像中提取，因此容易出现不稳定行为。

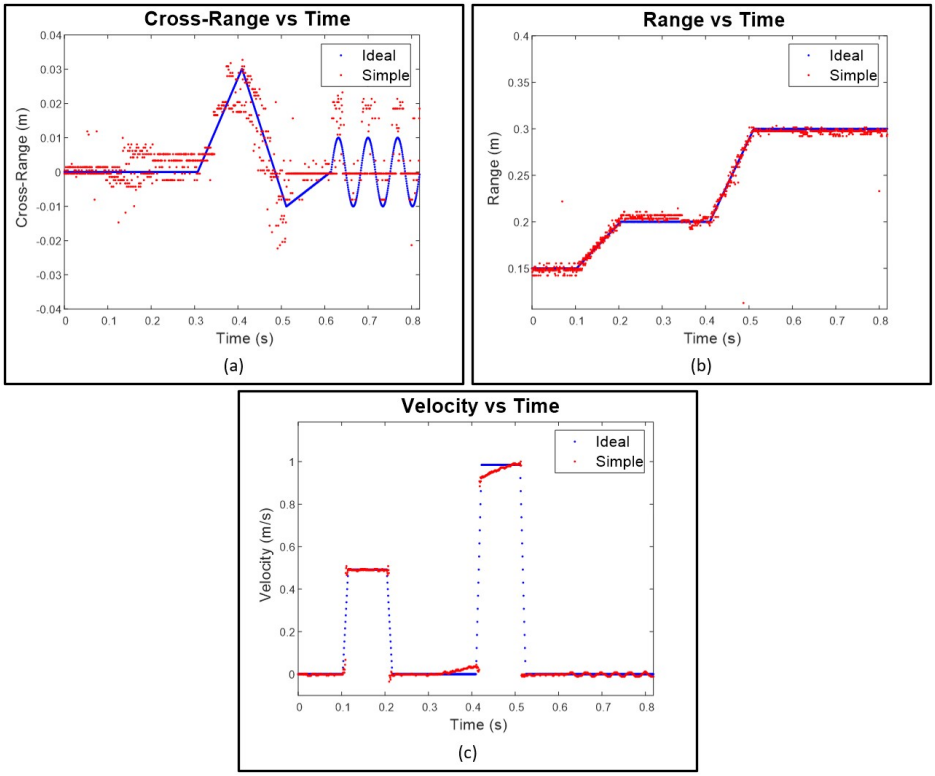

图 9：使用简单特征提取技术，在每个时间步（红色）上获得的运动轮廓，并与理想的运动和速度轮廓（蓝色）进行比较。(a) 横向距离和 (b) 距离直接从每帧 RMA 图像的峰值测得，(c) 速度则使用式（11）和（12）通过原始 RMA 图像的多普勒快速傅里叶变换测得。

图 9 展示了使用简单方法从式（16）生成的数据中估计出的特征。真实雷达噪声和变化的反射率会导致目标估计位置和速度中出现离群值与误差，尤其是在横向距离域中。若没有更稳健的特征提取和跟踪技术，性能仍远未达到理想水平。在以下各节中，将对简单跟踪方法与增强型跟踪方法的性能进行定量比较，并讨论设计考量。

5.3 基于全卷积神经网络的超分辨率跟踪结果

假设图 8 中的运动轮廓，我们采用所提出的粒子滤波算法，以更稳健地跟踪目标随时间变化的二维位置和多普勒速度，从而显著提升用户对界面的控制。

首先，使用图 9 中的数据作为含噪测量向量 $\mathbf{r}$ 的元素，实现不含多普勒佐证的粒子滤波算法（粒子滤波器）。如图 10 所示，粒子滤波器减弱了噪声对位置估计的影响，并提升了时空跟踪性能。与传统方法相比，横向距离位置跟踪的改善最为显著。接下来，将多普勒佐证粒子滤波器（DPF）应用于同一组数据，进一步改善距离估计。图 10b 中的离群值经图 10d 中的 DPF 缓解，因为这些离群样本产生的样本速度 $\hat{}v_s$ 与多普勒速度 $\hat{}v_d$ 相矛盾，因此被赋予权重，如♀♀♀♀♀♀

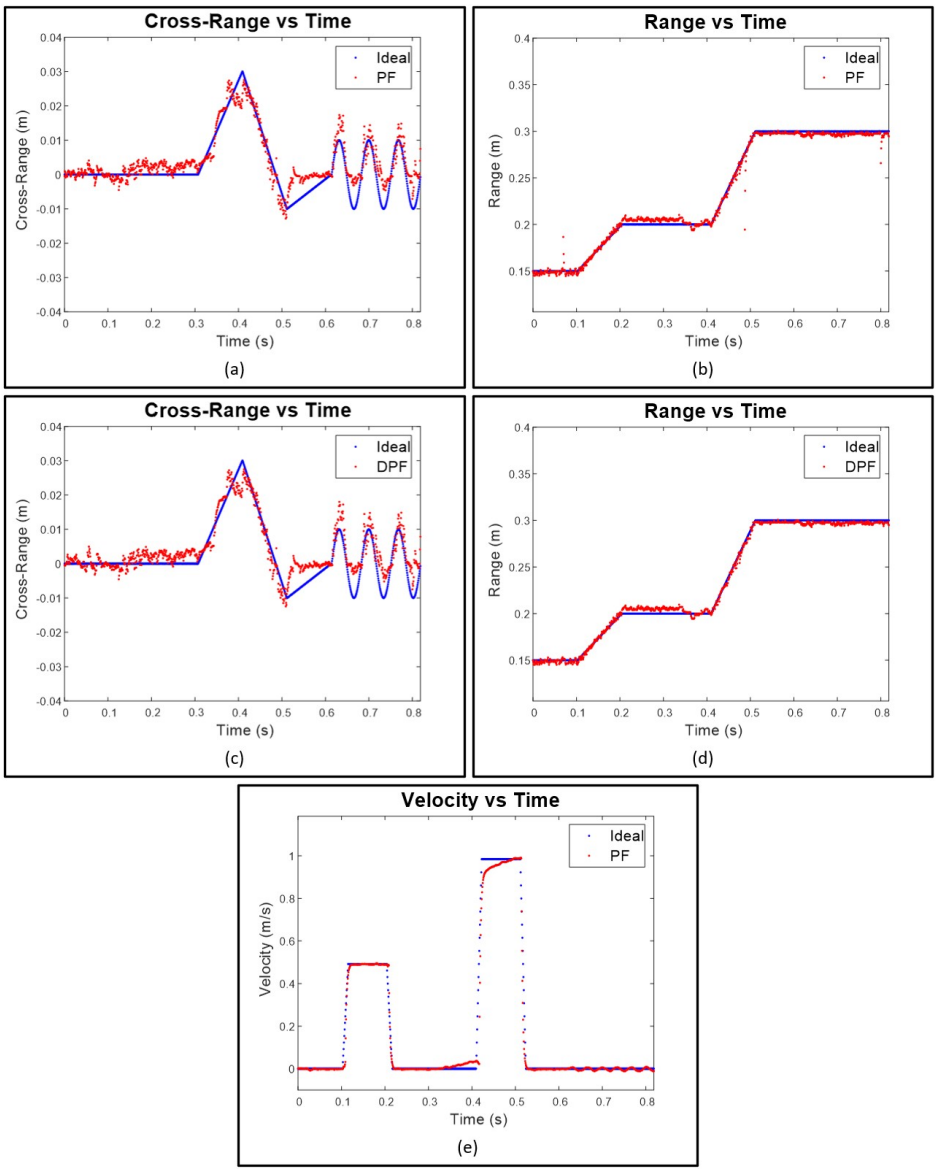

图10：粒子滤波器（PF）和多普勒验证粒子滤波器（DPF）的跟踪。使用 $N_z = 16$ 的 PF/DPF，改善了（a）/（c）横向距离位置和（b）/（d）距离位置随时间的跟踪效果；（e）采用 PF 方法得到的多普勒速度随时间的变化。

在重采样过程中并不重要。DPF 算法通过提供一种稳健、一致的跟踪算法，平滑地估计用户的手的二维位置和时空特征，从而改善了我们界面的用户体验。然而，通过实施所提出的增强型全卷积神经网络，还可以进一步改进 PF 和 DPF。

按照第 4.4 节所述方法训练超分辨率 FCNN 后，我们采集了一个与训练集大小相同的验证数据集。图 11 展示了经增强型全卷积神经网络处理的图像，证明了该网络的鲁棒性。图 11a 和 11b 展示了经 FCNN 增强的模拟点目标图像，实现了定位超分辨率。图 11c 是根据在横向距离域中部附近采集的真实手部数据重建的 RMA 图像。与理想图像相比，手部的二维位置得到了精确定位。同样，图 11d–11f 展示了该网络增强图像的能力，这些图像因手部 RCS 较小而相对于噪声受到干扰，并受到非理想波束方向图造成的鬼影、环境噪声、设备噪声及其他非理想因素的影响。所提出的增强型全卷积神经网络既能实现定位超分辨率，又能克服设备和环境问题。因此，与 FCNN 处理前的原始 RMA 图像相比，从增强图像中提取的特征大幅改善，进而带来更优异的跟踪性能。

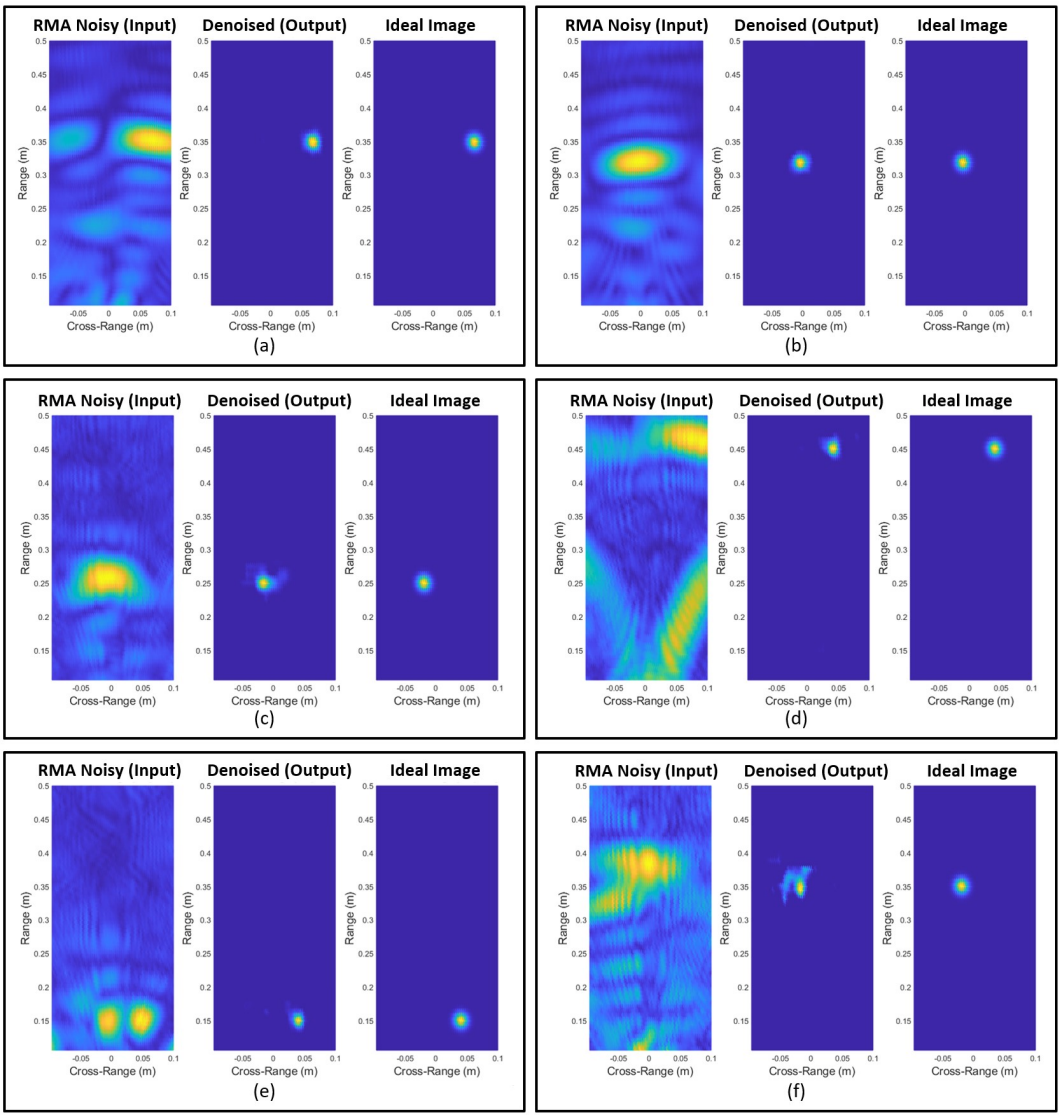

图 11：将增强型全卷积神经网络应用于模拟场景（a、b）和真实手部（c–f）的 RMA 图像，以实现图像增强和定位精度提升。

表 1：简单定位与增强定位的均方根误差

	y (m)	z (米)
简单	0.0154	0.023
增强	0.0085	0.0083

为定量比较增强型全卷积神经网络相较于简单方法在定位方面的改进，我们使用两种技术在验证数据集上计算了距离和横向距离位置的均方根误差，并将结果列于表 1。对于模拟数据和真实数据，增强型全卷积神经网络均提升了 RMA 图像的分辨率和定位精度。

图12展示了全卷积神经网络和DPF（FCNN-DPF）在与前述跟踪示例相同数据上的跟踪性能，并显示出相较于单独使用DPF的明显改进。应用FCNN-DPF后，目标的距离和横向距离跟踪结果几乎与理想运动剖面完全一致，且速度估计得到了改进。使用图9a中导致横向距离位置估计不佳的相同稀疏数据时，FCNN-DPF得到的估计结果几乎与理想运动剖面完全一致。同样，图10a和图10c中的横向距离估计在

图12a中被FCNN-DPF超越。与经典技术以及单独使用PF/DPF相比，FCNN-DPF的定位性能明显更优。

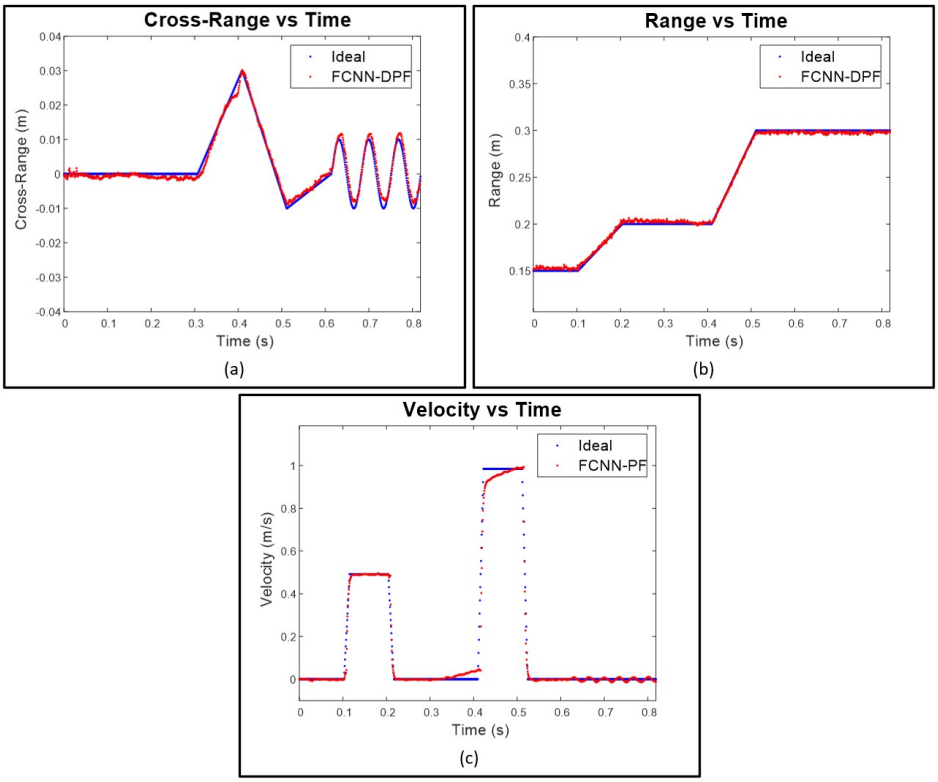

图 12：时空temp　采用全卷积神经网络增强的多普勒佐证改进粒子滤波算法进行口部跟踪　　.

此外，全卷积神经网络提高了多普勒估计鲁棒性。如图13所示，与对原始 RMA 图像进行多普勒处理相比，对增强后的 RMA 图像进行多普勒处理可提高多普勒频谱信噪比。因此，增强网络提高了多普勒速度估计的可靠性，有助于时空跟踪。

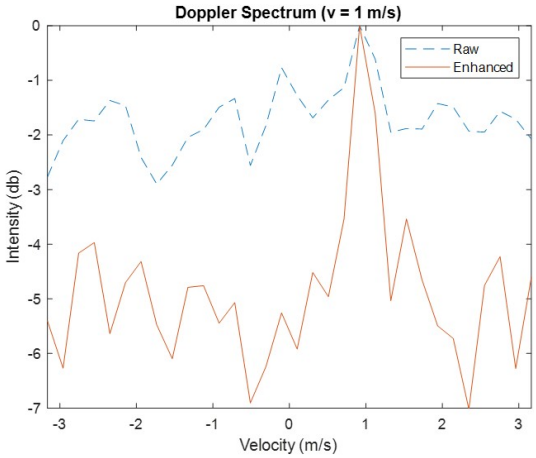

图 13：对原始 RMA 图像与增强后的 RMA 图像执行多普勒快速傅里叶变换和视频脉冲积分步骤时的多普勒速度频谱对比。仿真数据包含 128 帧，每次采集使用 $\alpha = 3$，以模拟低信噪比场景。

6 讨论与未来工作

为了定量比较所提出的各种方法的跟踪性能，我们生成了 4096 种独特的运动轨迹，并计算横向距离、距离和速度的相应跟踪均方根误差。如表2所示，本文提出的新算法降低了横向距离（y）、距离（z）和速度（v）的均方根误差。

如预期，基线简单方法在全部三个特征上的误差均最大。比较 PF 和 DPF 可知，两种技术的横向距离和速度均方根误差相同，但由于采用了动态加权技术，距离均方根误差得到了改善。单独使用全卷积神经网络的性能优于简单方法，但在图像增强后加入 PF 和 DPF 还可以进一步提升性能。最后，如预期，FCNN-PF 和 FCNN-DPF 在横向距离和速度均方根误差方面的结果相同，但距离误差得到了显著改善。表 2 中的结果表明，与简单跟踪方法相比，增强型跟踪方法，即 FCNN-DPF，具有明显优越的跟踪性能。通过实现超分辨率全卷积神经网络所获得的性能提升，证明了该网络能够在训练阶段利用真实数据和合成数据学习系统噪声和模糊性。

每种方法的平均延迟 $\bar{\tau}$ 定义为采集新样本到完成对该样本的估计过程之间的时间。生成的估计结果会通过 MIDI 端口流式传输，或发送至内置音频信号生成工具。后续合成引擎带来的额外延迟高度取决于所使用的软件和受测设备，因此不计入我们方法产生的延迟。

表 2：跟踪方法的平均均方根误差

	y (毫米)	z (毫米)	v (毫米/秒)	$\bar{\tau}$ (毫秒)
简单	7.86	22.0	72.4	2.29
PF	5.27	13.6	52.9	2.36
DPF	5.27	6.85	52.9	2.41
FCNN	7.74	12.3	58.4	2.67
FCNN-PF	3.70	7.44	44.5	3.92
FCNN-DPF	3.70	3.07	44.5	3.96

增强型跟踪方法在定位分辨率、多普勒频谱信噪比和跟踪精度方面优于现有技术；不过，这种性能提升需要付出一些必要的代价。新颖的超分辨率 FCNN 带来了显著的分辨率提升，超越了理论界限。在理想条件下，我们系统的横向距离分辨率和距离分辨率分别受限于 $\delta_y = 7.5$ cm 和 $\delta_z = 3.75$ cm。根据表 1 中真实手部数据与合成数据的组合，我们通过实证计算得出各方向的空间分辨率分别为 $\delta_y = 2.3$ mm 和 $\delta_z = 1.96$ mm。另一方面，增强型全卷积神经网络的效果受到训练集的限制。由于 FCNN 仅使用预期区域内的图像进行训练，将 ROI 扩展到训练区域之外会导致性能下降。如果 ROI 发生变化，就应相应地重新训练 FCNN。相比之下，简单方法具有很高的灵活性，但性能无法与增强技术相媲美。不过，我们研究了特定 TI 毫米波雷达设备的局限性，发现如果将手放在上一节所定义的 ROI 之外，设备将无法检测到手。受设备信噪比和波束宽度的影响，对于大多数手部尺寸，反射回雷达的信号都不足以被检测到。此外，我们在较小的 ROI 中测试了所提出的 FCNN，发现无需重新训练也能获得类似结果。对于其他阵列拓扑结构，所提出的方法也可轻松应用，不过需要相应地训练 FCNN。

虽然多普勒佐证粒子滤波器提高了跟踪的鲁棒性，但要使其正常运行，需要高吞吐量的框架。由于 DPF 依赖准确的多普勒速度估计，因此脉冲重复间隔 PRI，T_{PRI}，必须足够短，才能使较高的手部速度处于可分辨范围内。脉冲重复间隔受信号处理链每次迭代所需时间的影响最大。鉴于我们的框架目前以 MATLAB 程序的形式处于原型开发阶段，其延迟性能不及在效率更高的嵌入式设备上的实时实现。因此，典型的吞吐时间将脉冲重复间隔限制在约 4 ms。使用 77 GHz 设备时，在 $T_{PRI} = 4$ ms 下，可分辨的最大速度为 0.24 m/s。受此限制，某些高速快速运动可能导致多普勒频谱混叠。

尽管本文提出的软件包旨在作为展示和原型验证所提出的跟踪与超分辨率算法的框架，但信号处理步骤固有的延迟是人机交互中的关键问题，必须加以解决。在我们的研究中，所提出系统中延迟的最大来源，是雷达设备与 MATLAB 之间通过 UDP 和共享内存进行的数据交接，平均耗时为 1.93 毫秒。与其将数据流传输到 MATLAB，不如在 TI 雷达设备内置的数字信号处理器上实现实时解决方案，从而提供更高的

吞吐效率，因为数字信号处理器可以直接访问采集到的样本。此外，采用嵌入式解决方案还能提高信号处理链中多个步骤的效率。采用较小的窗口尺寸， $N_z = 16$且 FFT 空间点数为 64 时，与效率相对较低的 MATLAB 实现相比，DPF 和 FFT 的计算时间还可以进一步缩短。我们还想指出，通过采用图形处理器加速编码优化实现，延迟显著降低。嵌入式解决方案也可采用类似方法，利用信号处理链中许多步骤（FFT、卷积神经网络、高斯分布计算）高度并行化的特性。比较各算法的计算效率可知，即使采用 MATLAB 实现，更稳健算法所付出的延迟代价相对于性能提升也微不足道。在延迟测试中，使用 FCNN-DPF 时，从用户输入到 MIDI 信号发出的平均响应时间为 3.96 毫秒。虽然大多数 MIDI 接口的性能优于这一指标，但我们认为，本框架展示了相较于现有技术具有竞争力的吞吐周期时间，并且通过更高效的实现还可进一步改进。

与其他传感方式相比，使用毫米波雷达进行手部跟踪既有优势，也有不足。本文使用单个雷达，为毫米波设备开发并展示了鲁棒跟踪算法。虽然采用传感器融合技术很可能取得最佳性能，但如果使用光学摄像头会引发隐私顾虑，或必须考虑遮挡和光照条件等问题，那么基于雷达的实现方案可能是最佳选择。与光学和 RGB+D 方案相比，毫米波用途更广、可靠性更高，在遮挡情况下以及任何温度或光照环境中都能良好工作，还能提供整个场景的精确深度信息。对于音乐交互界面而言，这些优势或许无法经常得到充分发挥；不过，本文提出的新型跟踪方法适用于许多人机交互应用。另一方面，由于孔径尺寸有限，毫米波传感器的横向分辨率无法达到光学方案的水平，因此多目标跟踪和手指跟踪的难度要大得多。因此，许多人机交互、计算机视觉、自动驾驶等领域的应用会采用雷达、激光雷达和光学成像设备，并结合传感器融合算法，以更高的成本换取进一步提升的性能。对于这些应用，我们提出的算法可显著提升雷达传感器的性能贡献，从而助力传感器融合。

毫米波雷达感知有几种替代方案，即可穿戴、手持式和光学设备。可穿戴和手持式传感解决方案能够提供精度很高的时空特征，但与非接触式传感器相比，通常不是首选方案 [33,34]。从成本来看，毫米波雷达设备与广受欢迎的 Kinect 和 Leap Motion 光学传感器处于同一价格区间，约为 $100 – $200。采用多个 RGB 摄像头的尝试 [35,36] 展现出良好的前景；然而，考虑到多摄像头系统使用不便，人们更倾向于使用单台设备。已有研究提出了采用生成式姿态跟踪 [37,38] 和基于学习的生成式姿态跟踪 [39,40]的单台 RGB+D 解决方案。然而，这些方法在受到遮挡或场景杂乱影响时，性能都会大幅下降，而毫米波雷达能够克服这两方面的问题。一些面向深度学习的解决方案已展现出相当不错的效果 [41,42], ，但要构建足以开展有效监督训练的数据集，仍然是一项挑战。

我们提出的界面通过跟踪用户的手的二维位置和速度，来控制音符选择和两个用户选择的参数；与先前采用 Google Soli 毫米波雷达、仅跟踪一维距离以进行参数控制的工作相比，这是显著的改进 [17]。不过，光学方案能够跟踪双手 [3,8,23,37–42]或手和手指位置 [7,11] ，从而实现更精细的音乐控制，但也存在一些特定场景下的缺点。随着雷达技术不断进步、较大孔径日益普及，跟踪单个手指将变得越来越可行；由于深度分辨率更高，这种方法有望取得与光学方案相当或更优的效果。

与此前使用毫米波设备进行手部跟踪的研究相比，我们提出的方法取得了具有竞争力的结果。以往使用雷达设备的研究在人手定位方面所取得的最佳平均距离跟踪误差为 2 cm [18]。我们的增强型跟踪技术将平均距离跟踪误差降至 1.89 mm，使跟踪精度提高了十倍以上。在 [43], 中，一种带宽为 4 GHz 的毫米波传感器在跟踪拇指时，当距离小于 10 cm，可实现 1.16 mm 的二维位置均方根误差（RMSE）。相比之下，我们的增强型跟踪技术能够在更远的距离范围内跟踪人手，同时仍能实现具有竞争力的 3.4 mm 二维位置均方根误差（RMSE）。截至本文撰写时，我们尚未发现其他使用毫米波设备进行手部跟踪的相关先前研究。据我们所知，本文提出的系统使用单个毫米波传感器，实现了前所未有的手势跟踪性能。

与我们的框架最直接的音乐交互界面对比对象是特雷门琴，因为二者都是通过手与传感器之间的接近程度进行控制的。特雷门琴的音高由手部的垂直位置连续控制，而我们的交互界面则以数字方式跟踪手部距离，并从用户定义音阶中选择音符。特雷门琴使用两根天线，一根用于音量控制，另一根用于音高控制，共提供两个自由度；而我们的框架提供三个自由度（距离、横向距离和速度），因此能够提供三个可控参数。如前所述，本文提出的音乐交互界面支持用于音符选择和参数控制的特雷门琴式手势。然而，利用我们的高保真跟踪算法，也可以实现高速打击式手势，但存在一定局限性。在粒子滤波算法中，加权向量$\mathbf{a}$的值较小时，系统可能会出现过度平滑和阻尼过大的情况，从而限制系统跟踪突发变化的能力。

动作。根据所需的应用，精细调节此参数对于实现恰当的手势控制至关重要。我们提出的界面是改进版的特雷门琴，采用现代毫米波传感器，可在二维空间中实现精确跟踪（配备适当硬件后，可轻松扩展至三维）。与特雷门琴相比，我们的音乐交互界面在音符选择上省力得多，能够以简单直观的方式纳入额外的参数控制，并提升用户群体的可访问性。作者之一是一位技艺娴熟的吉他和小提琴演奏者，同时拥有电子音乐制作背景。从经验丰富的音乐家的角度来看，所提出的方法为音乐家提供了充分且稳定的控制能力，并带来一种优雅的新型音乐交互界面，能够生成独特的乐句；以往，这类乐句只能通过手动将乐器数字接口音符和控制参数转录到数字音频工作站（DAW）或其他实时数字合成平台中才能实现。

未来工作仍有若干值得探索的方向。首先，可将我们提出的方法部署到实时嵌入式平台上，进一步加以开发。此外，利用多个 MIMO 雷达或更大的 MIMO 阵列，可以研究支持多手和单指跟踪的界面，从而进一步拓展我们稳健跟踪方法的应用范围。最后，我们新颖的超分辨率跟踪算法易于调整，可为人机交互领域中众多棘手的手部跟踪问题提供优雅、高效的解决方案，甚至可用于传感器融合系统。

7 结论

我们基于FCNN的超分辨率框架成功证明了使用毫米波传感器进行HCI急速手部跟踪的可行性。我们以非接触式音乐交互界面的形式验证并实现了时空信号处理算法和鲁棒跟踪算法；不过，本文也旨在展示毫米波技术在多种近场急速手部跟踪应用中的广泛效用。首先，我们介绍了简单的特征提取和跟踪方法，随后提出一种增强方法，利用多普勒佐证粒子滤波器算法和增强型全卷积神经网络，在非理想成像场景中实现鲁棒跟踪和超分辨率。通过比较这些方法，我们发现，与经典技术相比，FCNN-DPF带来了显著改善。此外，与文献中针对毫米波和光学实现的既有方法相比，我们的方法在跟踪估计和定位性能方面也具有竞争力。我们的完整软件实现和实时雷达接口平台可应要求免费获取。本文提出的新型基于FCNN的超分辨率和跟踪算法，为许多非接触式HCI问题提供了优雅的解决方案。

致谢

Josiah Smith 的工作得到了 imec USA 暑期实习项目的支持。作者感谢 Gonzalo Vaca Castano 博士在开发粒子滤波算法和计算机视觉方法方面提供的见解。Murat Torlak 在 NSF 任职期间的工作得到了 NSF 的支持。本材料中表达的任何观点、研究发现、结论或建议均代表作者的观点，并不一定反映 NSF 的立场。

参考文献

[1] T. Winkler，《让动作成为音乐：交互式计算机音乐的手势映射策略》，载于国际计算机音乐会议论文集，加拿大艾伯塔省班夫，1995年9月，第261–264页。[2] K. D. Skeldon、L. M. Reid、V. McInally、B. Dougan 和 C. Fulton，《特雷门琴的物理学》，美国物理学杂志，第66卷，第11期，第945–955页，1998年12月。[3] R. Polfreman，《多模态乐器：构建比较控制器评估平台》，载于国际计算机音乐会议论文集，英国哈德斯菲尔德，2011年7月，第147–150页。[4] S. Trail、M. Dean、G. Odowichuk、T. F. Tavares、P. F. Driessen、W. A. Schloss 和 G. Tzanetakis，《使用 Kinect 为定音打击乐超乐器进行非侵入式感知与手势控制》，载于*NIME*论文集，美国密歇根州安娜堡，2012年5月。[5] S. Sentürk、S. W. Lee、A. Sastry、A. Daruwalla 和 G. Weinberg，《Crossole：使用 Kinect 进行作曲、即兴创作与表演的手势界面》，载于*NIME*论文集，美国密歇根州安娜堡，2012年5月，第449–502页。[6] R. Schramm、C. R. Jung 和 E. R. Miranda，《用于评估音乐指挥手势的动态时间规整》，*IEEE* 多媒体汇刊，第17卷，第2期，第243–255页，2015年12月。

[7] J. Han 和 N. Gold，《探索 Leap Motion 传感器以进行基于手势的乐器设计：经验总结》，载于 NIME 论文集。英国伦敦：伦敦大学金史密斯学院，2014年6月，第371–374页。[8] L. Hantrakul 和 K. Kaczmarek，《Leap Motion 在声音合成、效果调制和辅助表演工具中的实现》，载于国际计算机音乐会议论文集，希腊雅典，2014年9月，第648–653页。[9] D. Brown、N. Renney、A. Stark、C. Nash 和 T. Mitchell，《Leimu：使用腕戴式 Leap Motion 实现无手套音乐交互》，载于 NIME 论文集，澳大利亚布里斯班，2016年7月，第300–304页。[10] A. Tindale、A. Kapur 和 G. Tzanetakis，《在音乐手势采集系统中训练替代传感器》，IEEE 多媒体汇刊，第13卷，第1期，第50–59页，2011年10月。[11] O. Nieto 和 D. Shasha，《移动设备中的手势识别：提升音乐体验》，音乐表征与检索国际会议论文集，第10卷，2013年10月。[12] M. Akbari 和 H. Cheng，《基于计算机视觉的实时钢琴音乐转录》，IEEE 多媒体汇刊，第17卷，第12期，第2113–2121页，2015年8月。[13] S. Z. Gurbuz、A. C. Gurbuz、E. A. Malaia、D. J. Griffin、C. S. Crawford、M. M. Rahman、E. Kurtoglu、R. Aksu、T. Macks 和 R. Mdrafi，《使用射频感知识别美国手语》，IEEE Sensors J.，第21卷，第3期，第3763–3775页，2020年9月。[14] Y. Sang、L. Shi 和 Y. Liu，《基于超声主动感知的微手势识别系统》，IEEE开放期刊，第6卷，第49 339–49 347页，2018年9月。[15] Y. Kim 和 B. Toomajian，《使用卷积神经网络和微多普勒特征进行手势识别》，IEEE开放期刊，第4卷，第7125–7130页，2016年10月。[16] J. W. Smith、S. Thiagarajan、R. Willis、Y. Makris 和 M. Torlak，《利用新型无菌训练技术改进深度卷积神经网络上的静态手势分类》，IEEE开放期刊，第9卷，第10 893–10 902页，2021年1月。[17] F. Bernardo、N. Arner 和 P. Batchelor，《O soli mio：探索毫米波雷达在音乐交互中的应用》，载于 NIME 论文集，丹麦奥尔堡，2017年5月，第283–286页。[18] K. Joshi、D. Bharadia、M. Kotaru 和 S. Katti，《WiDeo：利用射频反向散射进行精细的无设备运动追踪》，载于 USENIX网络系统设计与实现研讨会论文集，美国加利福尼亚州奥克兰，2015年5月，第189–204页。[19] S. Maragliulo、P. F. A. Lopes、L. B. Osório、A. T. De Almeida 和 M. Tavakoli，《通过双通道可穿戴式肌电图系统识别足部手势》，IEEE Sensors J.，第19卷，第22期，第10 187–10 197页，2019年。[20] J. Gao、B. Deng、Y. Qin、H. Wang 和 X. Li，《使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。[21] H. Kim、S. You、B. J. Jeong 和 W. Byun，《利用神经网络提高方位角分辨率的技术》，载于国际信息与通信技术融合会议论文集，韩国济州，2020年10月，第1384–1387页。[22] Y. Sun、X. Liang、H. Fan、M. Imran 和 H. Heidari，《使用二维匹配滤波器对深度图像进行视觉手部跟踪》，载于大学工程与技术会议论文集，英国格拉斯哥，2019年8月，第1–4页。[23] A. R. Jensenius，《Kinectofon：使用平面图形进行表演》，载于 NIME 论文集，韩国大田，2013年5月，第196–197页。[24] S. Rao，《毫米波感知简介：调频连续波雷达》，2017年6月26日。访问日期：2021年4月20日。 [在线]。网址：https://training.ti.com/node/1139153[25] J. W. Smith、M. E. Yanik 和 M. Torlak，《近场多输入多输出逆合成孔径雷达毫米波成像》，载于 IEEE RadarConf 论文集，意大利佛罗伦萨，2020年9月，第1–6页。[26] M. E. Yanik 和 M. Torlak，《基于稀疏采样孔径数据的近场多输入多输出合成孔径雷达毫米波成像》，IEEE开放期刊，第7卷，第31 801–31 819页，2019年3月。[27] 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开放期刊，第8卷，第126 019–126 038页，2020年7月。[28] V. Winkler，《汽车调频连续波雷达的距离多普勒检测》，载于欧洲雷达会议论文集，德国慕尼黑，2007年10月，第166–169页。[29] J. Kim、J. Chun 和 S. Song，《调频连续波多输入多输出雷达的距离与角度联合估计及其应用》，arXiv预印本 arXiv:1811.06715，2018年。[30] J. García、A. Gardel、I. Bravo、J. L. Lázaro 和 M. Martínez，《基于扩展凝聚算法的人体运动跟踪》，IEEE Trans. Cybern.，第43卷，第3期，第606–618页，2013年。

[31] F. Le Gland、N. Oudjane 等，“使用希尔伯特度量的非线性滤波器稳定性与一致逼近及其在粒子滤波器中的应用”，应用概率年刊，第14卷，第1期，第144–187页，2004年2月。[32]“德州仪器 mmWave Studio”，2016年。访问日期：2021年4月20日。 [在线]。网址：https://www.ti.com/tool/MMWAVE-STUDIO[33] L. Pardue 和 W. Sebastian，“用于触觉控制与运动跟踪的手控器”，发表于*NIME* 会议论文集，韩国大田，2013年5月，第90–93页。[34] P. Neto、J. N. Pires 和 A. P. Moreira，“工业机器人技术的高级编程与控制：使用手持式加速度计输入设备进行手势与姿态识别”，工业机器人，第37卷，第2期，第137–147页，2010年3月。[35] L. Ballan、A. Taneja、J. Gall、L. Van Gool 和 M. Pollefeys，“利用判别性显著点捕捉动作中的手部运动”，发表于欧洲计算机视觉会议论文集。意大利佛罗伦萨：施普林格，2012年10月，第640–653页。[36] S. Sridhar、A. Oulasvirta 和 C. Theobalt，“利用 RGB 和深度数据进行交互式无标记关节式手部运动跟踪”，发表于国际计算机视觉大会论文集，澳大利亚新南威尔士州悉尼，2013年12月，第2456–2463页。[37] I. Oikonomidis、N. Kyriazis 和 A. A. Argyros，“利用 Kinect 高效地进行基于模型的手部关节运动三维跟踪”，发表于*IEEE*英国机器视觉会议论文集，第1卷，第2期，英国邓迪，2011年8月，第101.1–101.11页。[38] D. Tang、J. Taylor、P. Kohli、C. Keskin、T.-K. Kim 和 J. Shotton，“揭开黑箱：用于估计人手姿态的分层采样优化”，发表于*IEEE*国际计算机视觉大会论文集，智利圣地亚哥，2015年12月，第3325–3333页。[39] S. Sridhar、F. Mueller、A. Oulasvirta 和 C. Theobalt，“使用检测引导优化实现快速稳健的手部跟踪”，发表于*IEEE*计算机视觉与模式识别会议论文集，美国马萨诸塞州波士顿，2015年6月，第3213–3221页。[40] J. Taylor、L. Bordeaux、T. Cashman、B. Corish、C. Keskin、T. Sharp、E. Soto、D. Sweeney、J. Valentin、B. Luff 等，“通过对姿态和对应关系进行联合连续优化，实现高效精准的交互式手部跟踪”，*ACM* 图形学汇刊，第35卷，第4期，第1–12页，2016年7月。[41] J. Tompson、M. Stein、Y. Lecun 和 K. Perlin，“使用卷积网络实时连续恢复人手姿态”，*ACM* 图形学汇刊，第33卷，第5期，第1–10页，2014年8月。[42] Q. Ye、S. Yuan 和 T.-K. Kim，“结合部分粒子群优化的空间注意力深度网络，用于分层混合手部姿态估计”，发表于*IEEE*欧洲计算机视觉会议论文集，荷兰阿姆斯特丹，2016年10月，第346–361页。[43] Z. Li、Z. Lei、A. Yan、E. Solovey 和 K. Pahlavan，“ThuMouse：一种基于毫米波雷达交互的微手势光标输入方式”，发表于*IEEE*消费电子国际会议论文集，美国内华达州拉斯维加斯，2020年1月，第1–9页。