上海交通大学学位论文

基于毫米波雷达的无线感知智能化算法设计

姓　　名： **Daniel-ChenJH**

学　　号： **-**

导　　师： **-**

学　　院： **-**

学科/专业名称： **信息工程**

申请学位层次： **学士**

2023 年 05 月

A Dissertation Submitted to

Shanghai Jiao Tong University for Bachelor Degree

DESIGN OF INTELLIGENT WIRELESS SENSING ALGORITHM BASED ON MILLIMETER-WAVE RADAR

Author: **Daniel-ChenJH**

Supervisor: **-**

School of Electronic Information and Electrical Engineering

Shanghai Jiao Tong University

Shanghai, P.R.China

May 24th, 2023

上海交通大学

学位论文原创性声明

此处隐去。

上海交通大学

学位论文使用授权书

此处隐去。

摘 要

智能家居领域中基于毫米波雷达的无线感知智能化算法正在被广泛研究、应用，然而传统的智能家居感知算法无法对人际遮挡进行很好地处理，环境中移动的目标如窗帘、风扇等也都会对传统算法的性能产生较大影响；常见的人工智能算法在智能家居领域又存在需要大量实场数据训练、模型精度不高的问题。针对上述问题，本文提出了一种将模型的特征队列提取与任务输出需求相解耦的系统架构、设计了一套面向实场环境的毫米波雷达人体位置预测与动作分类算法，并通过将采集到的仿真、真实数据输入模型进行实验验证了算法架构的合理性与有效性。

本文的具体研究内容包括：

（1）设计了一种将模型的特征队列提取与任务输出需求相解耦的系统架构。该架构通过寻找多种任务需求中通用的中间信息，将特征提取和任务需求解耦，针对特定任务使用特定输出头，解决了传统端到端网络中可解释性、可操纵性差的问题，提高了模型的鲁棒性和可移植性，为无线感知智能化算法提出了新的架构。

（2）设计了一套可调节的、面向实场环境的毫米波雷达人体位置预测与动作分类算法。该算法充分考虑到毫米波雷达数据不具有平移不变性、包含大量位置信息的特点，通过对计算机视觉领域的 Detection Transformer 模型进行分析和针对毫米波雷达领域的创新性、适配性修改，解决了智能家居领域中基于毫米波雷达的无线感知智能化算法模型精度不高的问题。

（3）完成了本文系统架构与算法在不同类型数据输入情况下的可行性与实用性评估。通过将生成的仿真数据和实场采集、解析得到的真实数据输入模型进行分类预测，实验结果证明了本文系统架构的可行性与算法的有效性和实用性。

关键词：智能家居，无线感知，Detection Transformer，毫米波雷达，人体位置检测

ABSTRACT

Intelligent algorithms based on millimeter-wave radar for wireless perception in the field of smart homes are widely researched and applied. However, traditional smart home perception algorithms struggle to handle occlusions caused by human presence, and the movement of objects such as curtains and fans in the environment significantly impacts the performance of conventional algorithms. Common artificial intelligence algorithms in the smart home domain face challenges such as the need for extensive real-world data training and low model accuracy. To address these issues, this paper proposes a system architecture that decouples the feature queue extraction from task output requirements and designs a set of millimeter-wave radar-based algorithms for human position prediction and action classification in real-world environments. The validity and effectiveness of the algorithm architecture are experimentally verified by inputting simulated and real data into the model. The specific research contributions of this paper are as follows:

(1) Designing a system architecture that decouples the feature queue extraction from task output requirements. By identifying common intermediate information among multiple task requirements, the architecture decouples feature extraction from task demands. Specific output heads are used for specific tasks, addressing the issues of poor interpretability and controllability in traditional end-to-end networks. This improves the robustness and portability of the model and presents a novel architecture for intelligent algorithms in wireless perception.

(2) Designing an adjustable millimeter-wave radar-based algorithm for human position prediction and action classification in real-world environments. This algorithm considers the characteristics of millimeter-wave radar data, including the lack of translation invariance and the abundance of positional information. It analyzes the Detection Transformer model in computer vision and makes innovative and adaptive modifications for the millimeter-wave radar domain to overcome the issue of low model accuracy in intelligent algorithms for wireless perception in the smart home domain.

(3) Conducting feasibility and practicality evaluations of the proposed system architecture and algorithm under different types of data inputs. By inputting generated simula-

tion data and real data obtained from field collection and parsing into the model for classification prediction, the experimental results demonstrate the feasibility, effectiveness, and practicality of the system architecture and algorithm proposed in this paper.

Key words: smart home, wireless perception, Detection Transformer, millimeter-wave radar, human position detection

目 录

第一章 绪论

1.1 研究背景

在人们生活走入万物互联时代的过程中，样式繁多的应用场景被催生，其中非常重要的一个场景便是智能家居。在智能家居场景下，家居是宽泛的，可泛化为各种半开放/封闭的室内场景。在这些场景下，人们需要的不仅仅是温馨放松的环境，还有智能舒适的家居体验。在逐渐普及的众多智能家居中，各种传感器如家用摄像头、感应灯、语音助手等被用于记录、监测人们的生活环境与行为，由上游服务器对采集的数据处理后返回相应的指令，从而对人们的生活产生实际影响。

但这些传感器数据的收集与利用也带来了各种各样的隐私问题，比如在智能家居场景中十分常见的议题人体位置检测。目前已经落地应用的人体位置检测技术主要是基于摄像头完成。摄像头拍下的实时照片或视频经用户侧芯片传输到云端，在服务器进行解码、分析后完成指定任务，并将结果传输回用户侧。然而，摄像头数据难加密、易泄漏、对普通人没有理解门槛。摄像头的高清与无处不在带来了不可忽视的隐私问题。针孔摄像头的出现、偷拍产业链与用户数据泄漏问题都让人们提心吊胆。因此，针对目前项目的背景，我们必须提出一种全新的技术路线，替代传统以摄像头采集信息的方案。

在这种情况下，基于毫米波雷达的无线感知智能化算法设计被提出、考虑。

我们希望采用毫米波雷达替代传统摄像头，完成人体位置检测与动作分类任务。采用毫米波雷达代替摄像头主要是基于两点考虑：一方面，毫米波雷达精度远不及高清摄像头，其数据不涉及到偷窥等问题；另一方面，来自毫米波雷达的原始信息经过变换后不具有直接意义，普通人无法提取出有用信息。这两点共同实现了对用户隐私的保护，在一定程度上解决了传统摄像头带来的隐私问题。

1.2 国内外研究现状与发展动态

在毫米波雷达解算的传统算法中，雷达发生调频连续波，采用多发多收(MIMO)

天线完成角度解算。将采集到的信息经过距离维快速傅里叶变换（Fast Fourier Transform，下称 FFT）得到距离信息，再进行多普勒 FFT 得到速度信息。然而，传统算法在时域转换到频域的过程中都对接收端信息进行了取模操作，丢失了相位信息；同时，传统方法得到的都是物理信息，在白噪声下结果虽然较好，但无法处理一些特别的干扰信息，如家庭场景中的窗帘、鱼缸等干扰[1][2][3][4][5]。

目前毫米波雷达主要应用于车载场景下，因为毫米波雷达具有穿透能力强、探测距离远、不易受恶劣天气的影响的优点，因此毫米波雷达数据一般与摄像头数据同时被采集后输入神经网络进行特征融合、作为摄像头数据的辅助。

在智能家居领域中，单独使用毫米波雷达的智能产品非常少。

在业界，德州仪器公司有一个名为 3D People Counting Demo 的项目①，其获取点云的算法被称为 TI-mPoint。此项目走社区常规路线，对采集到的原始雷达数据首先执行 1D-FFT，然后通过移动目标检测（MTI）移除背景信息，再通过恒虚警检测器（CFAR）解析目标的距离信息，最后通过自适应波束成形算法（MVDR）和 CFAR 解算目标的平面角，仰角信息，从而得到目标点云信息，点云实际上包含了三个信息：距离、方位角、俯仰角。

在项目的实际效果方面，在 TI 公司的测试中②，虽然此项目成功检测到了移动的人，但房间里面摇头的风扇和被吹动的盆栽都对人的检测造成了影响；在我们实际测试时，场景中有两个人交错走过，但该项目解析得到的点云数量不足以描绘人体边框，且只识别到了一个人。点云数量较少、无法描述人体边框其实是传统路线的通病[6][7][8]。另外，此项目并没有对两个人的身份做出必要的区分。当甲从左向右、同时乙从右向左交错走过时，此项目识别的结果为：乙先从右向左走一半路程，再从左向右走一半路程。这样的识别效果说明目前业界的算法对目标识别的精度性能以及抗噪性能还需要提升。

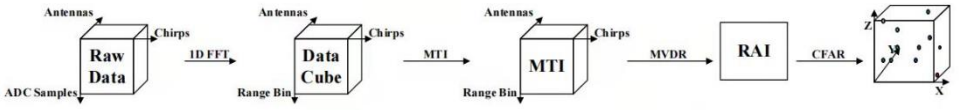

图 1-1 德州仪器公司从原始数据提取点云的 TI-mPoint 技术路线①

① People Counting Labs Directory, https://dev.ti.com/tirex/explore/node?node=AITDkxgrRIdA.6m43Ts-sg__VLyFKFf__LATEST

② TI Ifdm Intro, https://dev.ti.com/tirex/explore/content/mmwave_industrial_toolbox_4_10_1/labs/People_Counting/Sense_and_Direct_HVAC_Control/docs/images/ifdm_intro.mp4

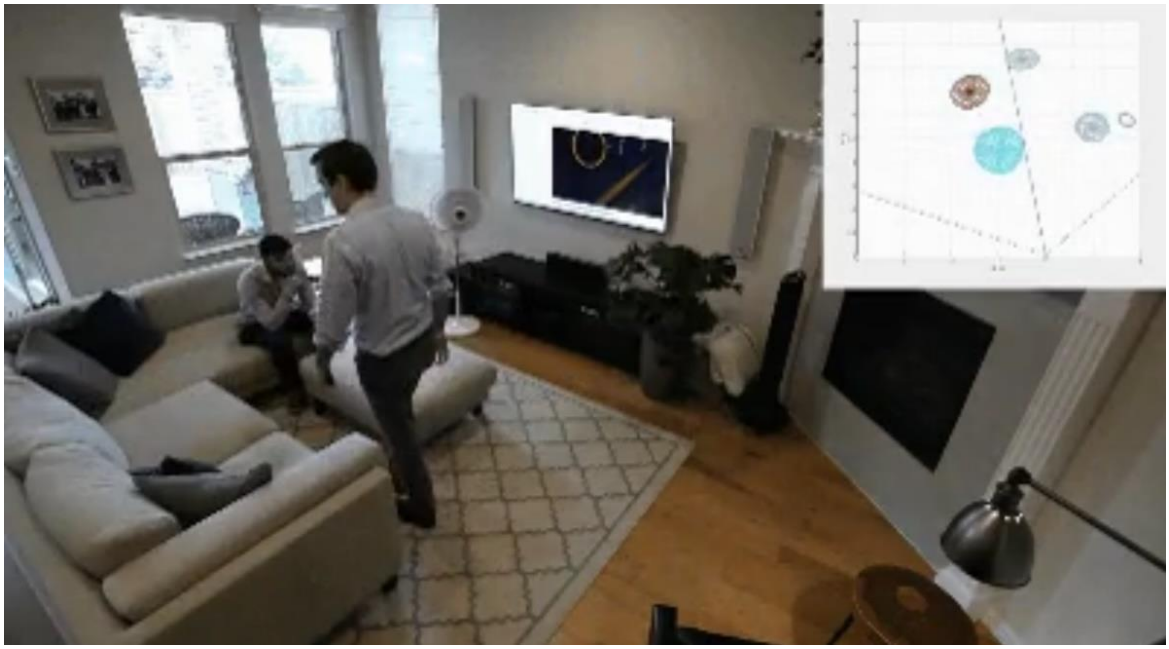

图 1-2 德州仪器公司 People Counting Demo 受风扇、盆栽干扰示意图①

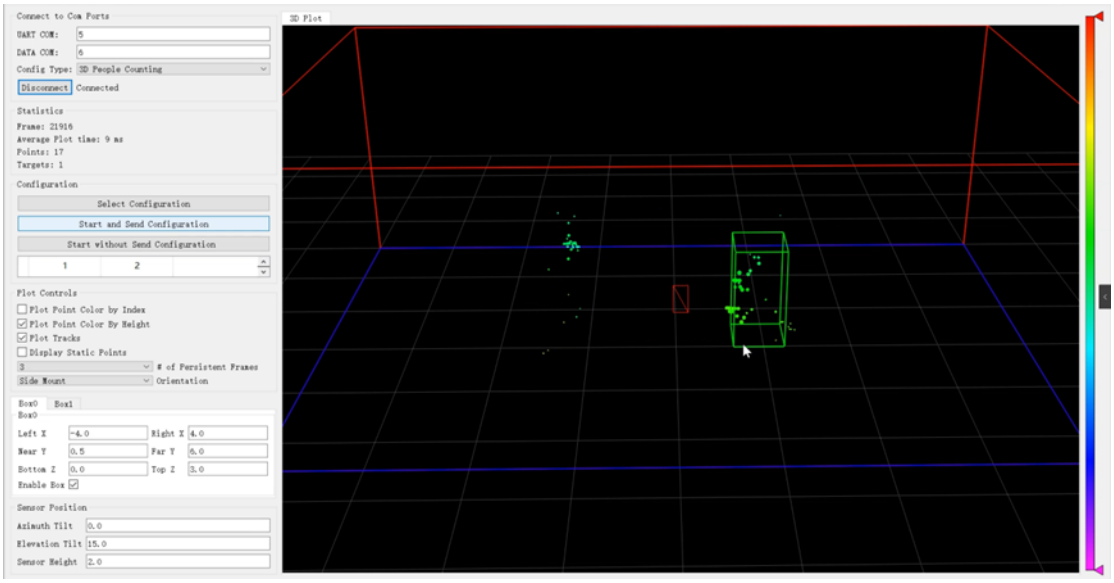

图 1-3 实测德州仪器公司 People Counting Demo 的截图

在学界，2018 年 CVPR 会议上，来自 MIT 的 CSAIL 团队提出单独使用毫米波雷达进行多目标跟踪任务[9]。他们采集数据时所采用的毫米波雷达具有 8 根接收天线，带宽为 3GHz，方位角和俯仰角分辨率为 15°。文中采用老师-学生网络，在模型训练

① TI Ifdm Intro, https://dev.ti.com/tirex/explore/content/mmwave_industrial_toolbox_4_10_1/labs/People_Counting/Sense_and_Direct_HVAC_Control/docs/images/ifdm_intro.mp4

时将同步采集到的视频数据输入视频神经网络，得到检测结果；再将原始雷达数据解析得到点云，点云向水平、垂直两个方向分别投影得到热力图，通过各自 encoder 网络后将输出级联通过 decoder 网络，最终得到基于毫米波雷达数据的输出。在模型推理中仅输入毫米波雷达数据，返回跟踪结果。

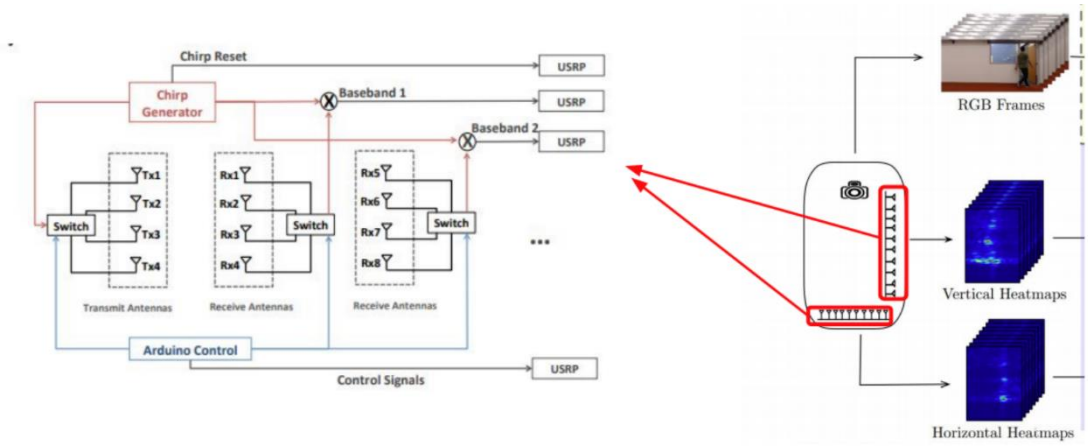

图 1-4 MIT 团队原始数据采集示意图[9]

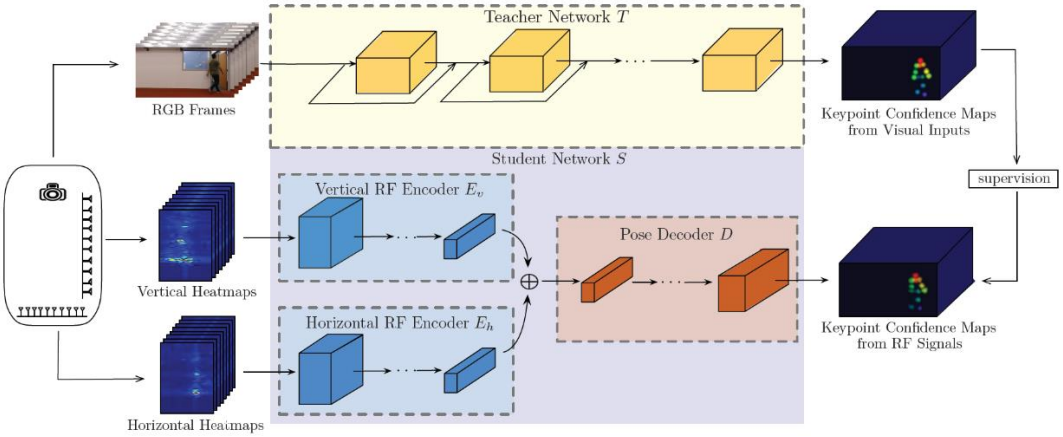

图 1-5 MIT 团队采用的神经网络结构示意图[9]

其中，他们采用由香港中文大学 Sijie Yan.团队中提出的时空卷积网络 ST-GCN[10]，根据人体结构，将每帧的人体节点对应连接成边，做三维的时空卷积，实现动态人体骨架模型的跟踪任务。Encoder 网络由时空卷积网络、批量归一化层、Relu 层构成；Decoder 网络由时空卷积网络、反卷积层、PRelu 层构成。

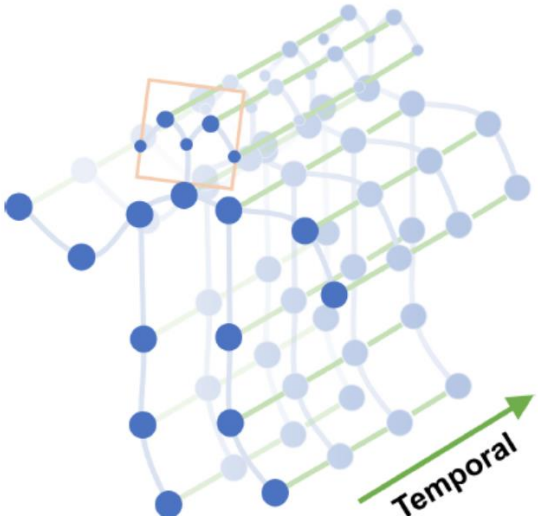

图 1-6 ST-GCN 时间轴节点对应连接[10]

然而，他们的工作中仍存在一些问题亟待解决：RF 信号在定位头部和躯干（颈部和臀部）方面非常准确，但在定位四肢方面不太准确。射频反射的数量取决于身体部位的大小与四肢相比，因为头部和躯干具有较大的反射区域和相对较慢的运动，所以更容易被 RF-Pose 捕捉。当人被阻挡射频信号的金属结构遮挡时，或者当人靠得太近而导致射频信号分辨率降低时，定位就会出现问题，动作也就无法很好地被识别。另外，人际遮挡对模型识别的精确度产生了很大限制。

令人遗憾的是，文章作者并没有将实验所用数据集与代码公开发表，网上也暂时没有找到相应的复现代码。我们曾尝试邮件联系作者获取相应数据集与代码，但并未得到回复。

虽然基于毫米波雷达的模式识别工作较少，但基于激光雷达的模式识别工作在学界较为丰富。本着跨界类比参考学习的想法，我们同样对基于激光雷达的模式识别工作进行了调研。此处主要介绍其中代表性较强的 PointNet、PointNet++与 CenterPoint 工作。

Qi C R, Su H, Mo K 等人创新性地提出了 PointNet 架构[11]，指出了利用点云进行模式识别的三大核心要点：利用点云的无序性、点云的相互作用性以及刚体旋转不变性。

在 PointNet 中，作者使用 Max Pooling 层来利用三维点云的无序性。Max Pooling 层的输出不受输入数据顺序的影响；另外，文中作者采用只包含乘法和加法运算的多层感知机结构进行模型的特征提取。通过这样的方式，PointNet 规避了输入点云顺序

对模型的影响，成功利用了三维点云的无序性。

点云的无序性直接说明了 RNN 和 LSTM 并不适合用于三维点云数据的处理。RNN 和 LSTM 被广泛应用于时间或空间序列数据的处理[8]，但点云具有无序性，即点云数据中点的输入顺序与模型的输出没有明确的规律，因此直接使用 RNN 和 LSTM 效果不佳。

在点云分类任务中，可以直接利用特征向量对 SVM 或 MLP 进行训练。然而，在以点为单位的点云分割或分块任务中，为了完成逐点的分类，需要将每个点的局部特征和全局特征结合起来，进行特征融合和处理。在 PointNet 中，作者将经过特征对齐后的 64 维特征视为点的局部特征，而将最后的 1024 维特征视为点的全局特征。通过简单地将局部和全局特征拼接在一起，并利用 MLP 进行特征融合，PointNet 最终可以训练其分类器实现逐点分类。

PointNet 引入了一个名为 T-Net 的子网络，用于预测特征空间变换矩阵。T-Net 学习生成与特征空间维度相同的变换矩阵，然后用这个变换矩阵作用于原始数据，完成对输入特征空间的变换，使每个点与输入数据中的所有点相关联。通过这种数据融合方式，PointNet 在保持刚体旋转不变性的同时实现了对原始点云数据特征的逐级抽象。

综上所述，PointNet 提出了一种处理三维问题的新方法，即将点云作为直接输入，通过在批处理维度上使用对称函数如 Max Pooling 层、MLP 层来处理数量上的不稳定性。它利用 T-Net 解决了刚体旋转不变性的问题，具有$O(N)$的空间和时间复杂度。然而，PointNet 在处理相互作用性质方面还存在不足，它对局部特征的处理不够精细。此外，它使用多层感知机提取特征的方法忽略了图的拓扑结构，使其在分析复杂场景时面临一定的困难，需要进一步优化。

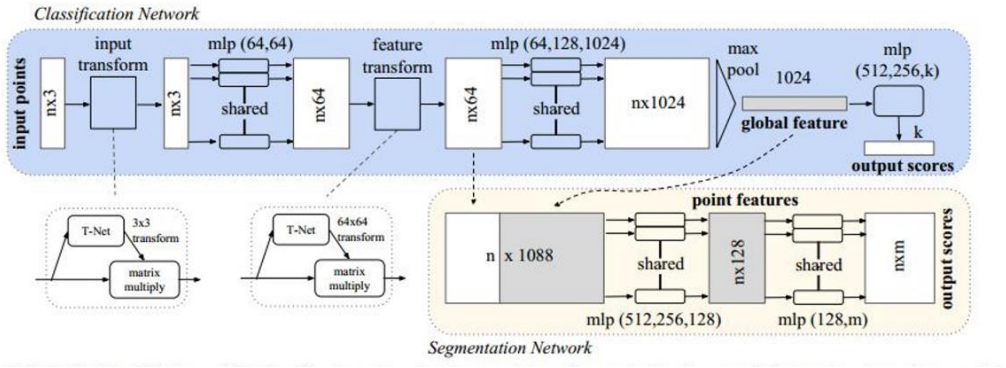

图 1-7 PointNet 网络模型示意图[11]

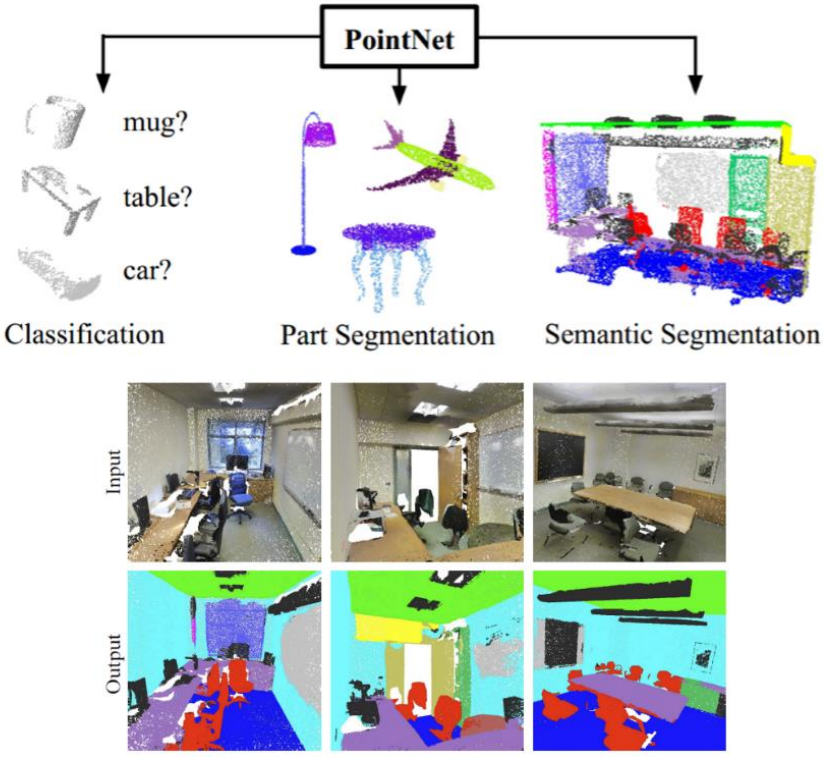

图 1-8 PointNet 模型工作示意图[11]

Qi C R, Yi L, Su H 等人使用了两个方法对 PointNet 的缺点进行了改进，改进后的架构被称为 PointNet++[12]。首先，他们利用空间距离，在局部区域上使用 PointNet 进行特征迭代提取，使其能够学习到不断增大的局部尺度的特征。第二，作者提出了一种全新的特征提取方法来应对点集分布的不均匀特性，这种特征提取方法具有针对密度的自适应性。通过以上两种方法，模型的鲁棒性得到提升，模型对特征的学习也更加高效。

众所周知，卷积神经网络可以通过提升卷积层的感受野进行特征提取。在 PointNet++中，作者利用空间距离度量将点集划分成多个相互重叠的局部区域。以 PointNet 作为基本的特征提取器，模型在每个局部区域内先提取局部的浅层特征，然后逐渐扩大范围，在局部特征的基础上提取更高层次的特征，直到获得整个点集的全局深层特征。

Set abstraction layers 模块主要由采样层 Sampling layer, 分组层 Grouping layer 和 PointNet 层 PointNet layer 三大部分组成。在采样层中，模型对输入点进行最远点采样，从中选则若干个中心点；分组层利用采样层得到的中心点，通过邻域球（ball query）

算法将点集划分为若干个区域；PointNet 层对得到的每个区域提取特征并编码，最终拼接得到各区域的特征向量。

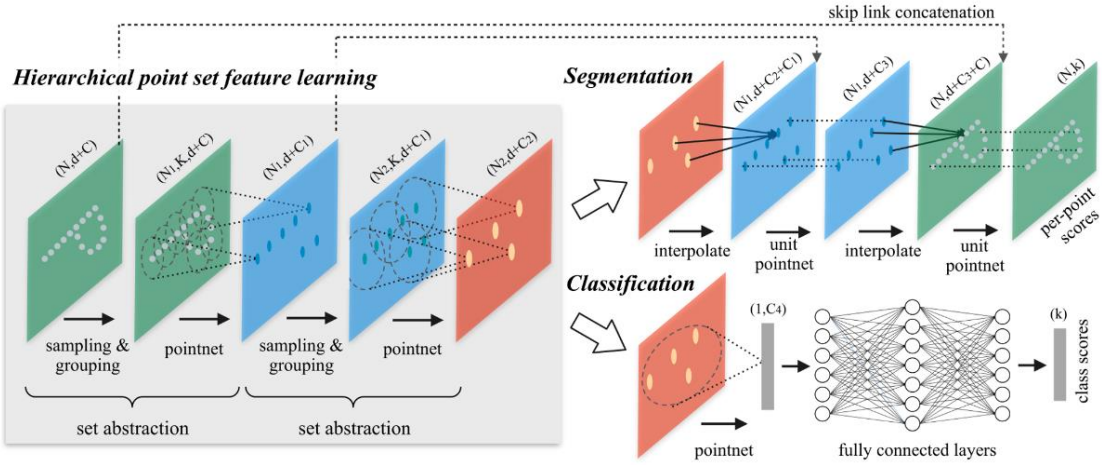

图 1-9 PointNet++网络结构图[12]

在 CVPR2021 会议上，Yin T, Zhou X 和 Krahenbuhl 提出了 CenterPoint 模型[13]，利用激光雷达数据完成了对三维空间物体的检测和跟踪任务。激光雷达可以得到密集的点云，因此他们采用密集体元表征物体，并主要采用鸟瞰图+体元作为模型的输入。他们把整个模型的锚定框（Bounding Box，下文简称$Bbox$）预测分为两个阶段，即中心点预测和修正。

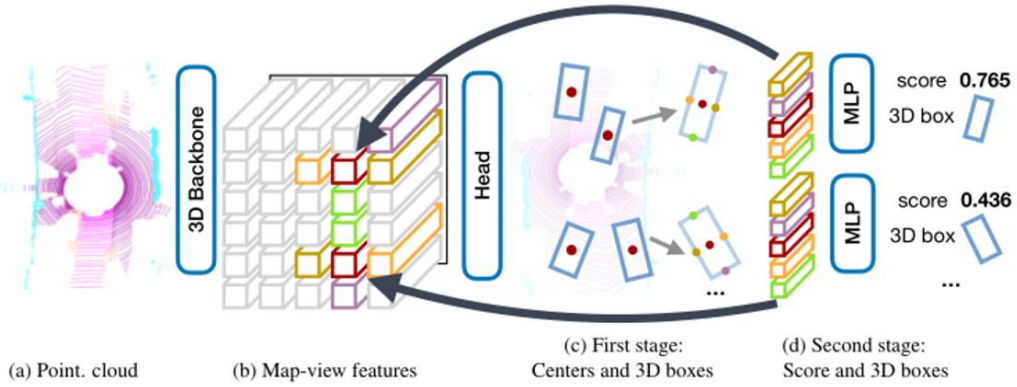

图 1-10 CenterPoint 网络结构图[13]

在中心点预测阶段，CenterPoint 预测各个类别的热力图、目标的面积分布、旋转速度及角度，并进行体元位置的细化，最终得到初步的三维检测锚定框。由于在训练过程中，目标物体中心在空间中的分布是非常稀疏的，不利于训练，所以文章以检测到的中心点为中心，为每个点生成了满足高斯分布的热力图，采用高斯分布的方式将

信息散布到关键点的附近去，即越靠近关键点的像素对应的值越接近 1，越远离关键点的像素对应的值越接近 0。高斯分布的期望值是 1，方差则采用 CornerNet[14]中半径的设置方法，即保证该点的高斯分布范围内，至少有 1 个点能够产生IOU>0.3 的$Bbox$。在得到多个中心点后，使用中心点的特征回归出$Bbox$的多个参数，包括体元化位置的优化值、高度值、在 3D 空间中目标的尺寸以及使用正弦、余弦表示的偏角。

在中心点修正阶段，CenterPoint 从除顶面和底面以外的其他四个面中提取每个面的中心点。然后，通过双线性插值的方法，从这四个点中提取相应的特征并拼接输入到一个多层感知器（MLP）中进行预测。预测结果包括每个边界框的置信度分数和相应的优化参数。中心点修正阶段通过这样的方式完成了对中心点预测阶段检测结果的优化。

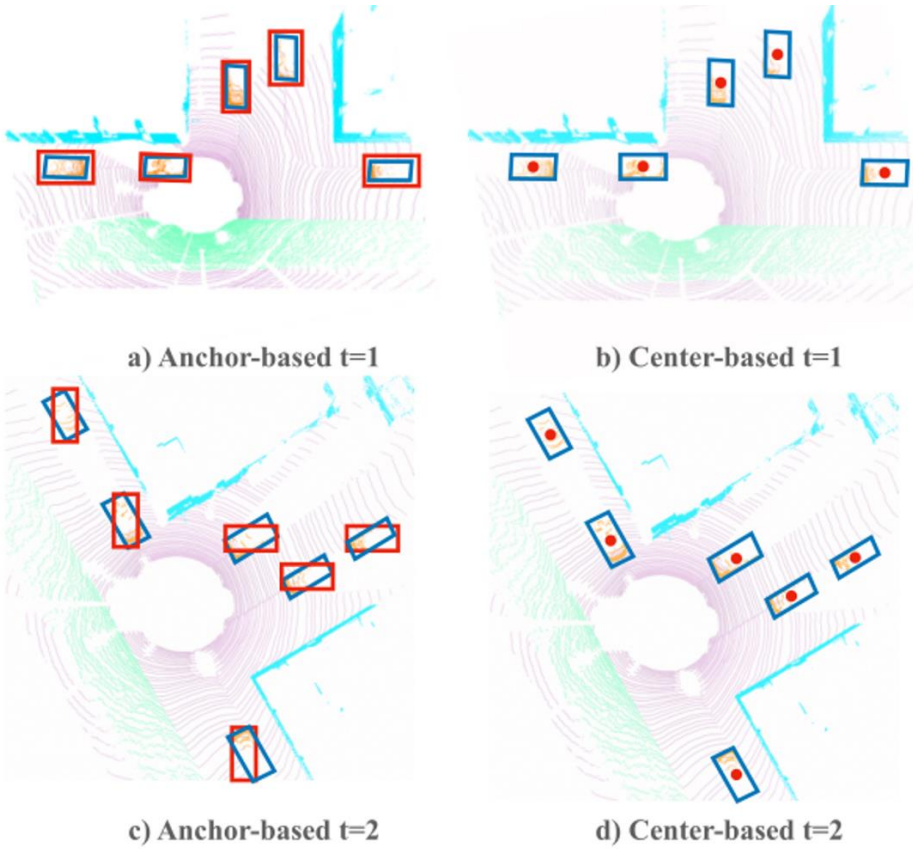

图 1-11 Anchor-Based Bbox 与 Center-Based Bbox 的区别[13]

在点云中，三维目标的朝向是任意的，这使得基于锚点（anchor-based）的检测器难以枚举所有可能的方向，更无法为旋转的目标适配平行于坐标轴的检测框。CenterPoint 的检测框基于中心点（Center-based）生成，因此它生成的检测框更加灵

活、更利于描绘物体的空间范围。

此外，由于点云中的点没有朝向性，因此旋转不变性得到了满足，同时搜索空间的复杂性得到了较大程度的降低。CenterPoint 简化了跟踪任务，将热力图损失和回归损失统一为同一个目标，简化并加速了之前计算成本高昂的基于 PointNet 的特征提取器。

然而，PointNet、PointNet++和 CenterPoint 模型都是基于激光雷达数据的，激光雷达可以得到目标密集的点云、用体元来构成目标，但毫米波雷达无法做到这一点，因此无法简单地在毫米波雷达领域迁移使用这些模型。后续有工作者用与摄像头数据进行特征融合的方式将 PointNet 和 PointNet ++模型用于车载毫米波雷达，但笔者没有发现有人将这些模型用于基于毫米波雷达的室内 3D 模式识别领域。

1.3 本课题的目的与意义

本项目基于毫米波雷达原始数据，通过借鉴计算机视觉（Computer Vision，下称 CV）领域的 Detection Transformer（DETR）模型并对数据处理算法进行多重优化，实现基于毫米波雷达的无线感知智能化算法设计。

在业界已有的基于毫米波雷达的智能家居解决方案中，主要采用的即为点云投影热力图方案，但在实际使用场景中，人们发现这种路线一方面容易丢失信息，另一方面家居环境内诸如窗帘、盆栽等移动物体对识别效果有较大干扰；而学界中基于雷达的模式识别、分割工作也主要是基于激光雷达实现的，但激光雷达与毫米波雷达采集到的数据有较大区别，使得激光雷达领域使用的相关技术无法很好地简单移植到毫米波雷达领域中。

因此，我们希望提出一套全新的解决方案，能较好地解决上述这些问题，实现基于毫米波雷达的无线感知智能化算法设计。

1.4 本文的章节安排

本文旨在提出一种基于毫米波雷达的无线感知智能化算法设计，通过对毫米波雷达信号的处理和机器学习算法的应用，实现人体动作分类和位置预测等任务。本文主要包括五个章节。

本文第一章绪论部分介绍了研究的背景和动机，探讨了国内外毫米波雷达智能家

居领域无线感知技术的研究现状与发展动态、明确了本课题的目的与意义，并简要概述了整篇论文的章节安排。

第二章介绍了本课题中所用的基于毫米波雷达的信号处理方法，包括距离 FFT、多普勒FFT和MIMO虚拟天线技术等。这些技术后续会在仿真数据和真实数据采集、解析、处理过程中被使用。

第三章确定了本文的系统架构和技术路线，同时对 DETR 在毫米波雷达领域的使用进行了分析并完成了适配性修改。第三章首先提出了将模型的特征队列提取与任务输出需求相解耦的系统架构与技术路线，并阐述了在毫米波雷达智能家居领域中构建通用的中间特征信息的方式，包括人体骨架关键点或锚定框的使用，实现了不同任务的统一处理。然后对 DETR 用于毫米波雷达领域的可行性进行了讨论，提出 DETR 需要经过相应适配性修改才能在毫米波雷达数据集上使用，并在最后详细地介绍了对 DETR 模型的各个子模块进行适配性修改的方法，包括如何对数据传输方式、损失函数设计、骨干网络等模块进行调整，同时也完成了对采集得到真实数据集的锚定框数据分布的分析，解决了 DETR 在毫米波雷达领域使用时的适配性问题。

第四章介绍了本文实验过程中使用的仿真数据集的生成方法与真实数据集的采集、解析方法，给出了经过适配性修改后的 DETR 模型在不同输入情况下针对仿真、真实数据集的多组训练结果，并对每组实验结果进行了详细的分析，验证了本项目技术路线的可行性与鲁棒性以及本项目工作的合理性，实现了对整个工作流程的阐述。第四章首先介绍了针对单人仿真数据的实验，展示了算法在该场景下的性能表现；然后介绍了针对多人仿真数据的实验，验证了算法在复杂场景下的鲁棒性；最后详细描述了基于真实数据的锚定框实验，包括基础实验、加入动作分类实验以及加入动作分类和倾角预测实验的结果。通过对各实验结果的对比分析，评估了所设计算法的有效性和算法的性能。

第五章总结了本项目的研究成果与创新点，同时展望了后续值得深入研究的工作内容。

通过以上章节的展开，本文提出了一种将模型的特征队列提取与任务输出需求相解耦的系统架构，同时对 DETR 模型进行了毫米波雷达领域的创新性、适配性修改，实现了基于毫米波雷达的无线感知智能化算法设计，并在实验中得到了良好的性能表现。本研究为毫米波雷达智能家居领域的算法设计提供了一种新思路和方法。

1.5 本章小结

本章先介绍了本项目的研究背景，然后对国内外本领域的研究动态分业界、学界两个部分进行了介绍。基于目前领域内当前的研究动态，明确了本项目的目的与意义是基于毫米波雷达原始数据，通过借鉴计算机视觉领域的 Detection Transformer（DETR）模型并对数据处理算法进行多重优化，实现基于毫米波雷达的无线感知智能化算法设计。本章最后介绍了本文的章节安排。

第二章 基于毫米波雷达的信号处理方法

本章主要介绍了本课题中所用的基于毫米波雷达的信号处理方法，包括距离 FFT、多普勒 FFT 和 MIMO 虚拟天线技术等。

2.1 雷达信号处理

图 2-1 是调频连续波雷达简化框图。合成器生成频率随时间线性增加的线性调频脉冲，并送至 TX 天线进行发射。RX 天线会在一定时延后捕捉到物体对该线性调频脉冲的反射。RX 接收到的信号和 TX 发射的信号在混频器被合并而生成中频信号。其中，中频信号的相位和频率分别为 TX 信号和 RX 信号的相位差和频率差。

2.1.1 距离 FFT

接收脉冲是延时后的发射脉冲，两者之间频率关系如图 2-2 所示 。中频信号的频率是定值，对其做 FFT ，找到频率峰值 f，由此频率可以计算出收发信号的延迟 τ，从而计算出物体距离 d。 这样的 FFT 称为距离 FFT。

$$f = S\tau \tag{2-1}$$

$$\tau = \frac{2d}{c} \tag{2-2}$$

$$d = \frac{fc}{2S} \tag{2-3}$$

距离d与中频信号的频率f呈线性关系。根据傅里叶变换理论，长度为 T 的观测窗口可以区分间隔超过 1/T Hz 的频率分量。因此，我们可以由频率分辨率导出距离分辨率为：

$$d_{res} = \frac{c}{2ST_c} = \frac{c}{2B} \tag{2-4}$$

式（2-4）中Tc表示脉冲持续时间，S表示线性调频脉冲频率变化率，B表示线性调频脉冲扫频带宽。

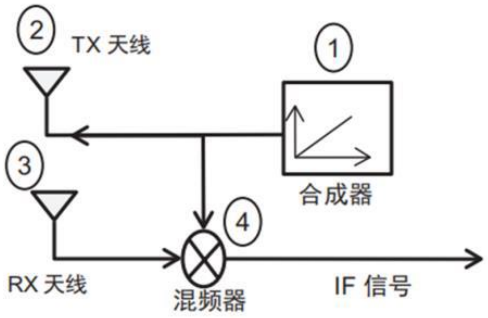

图 2-1 FMCW 雷达简化框图[①]

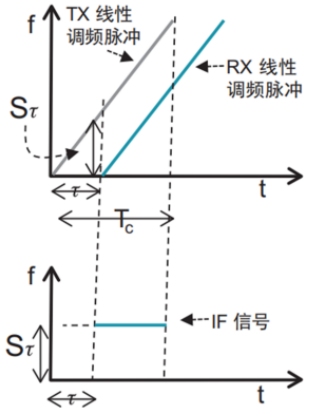

图 2-2 收发信号与中频信号的频率关系[①]

2.1.2 多普勒 FFT

测量速度，需要连续发射两个或以上调频连续波。每个反射信号通过 FFT 可以解算目标距离，但解算速度不能依靠两个反射信号 FFT 之后的频率差实现。这是因为物体运动对中频信号频率的影响很小，达不到距离检测的分辨率。

由于我们采用的雷达扫频带宽$B = 4GHz$，持续时长$Tc = 40\mu s$，则由式（2-4）可以计算出距离分辨率$d_{res} = 0.0375m$。如果能利用距离检测物体的速度，则该物体需要在Tc内移动0.0375m，换算为速度即$937.5m/s$，这个速度在家居场景下显然是无法达到的。因此，检测速度需要用到更灵敏的相位信息，即多普勒 FFT。中频信号的相位φ_0表示为：

$$\varphi_0 = 2\pi f_c \tau = \frac{4\pi d}{\lambda} \qquad (2-5)$$

连续发射的两个时间间隔Tc的调频连续波，由于物体移动造成的相位差为：

$$\Delta\varphi = \frac{4\pi\Delta d}{\lambda} = \frac{4\pi v T_c}{\lambda} \qquad (2-6)$$

其中v表示物体移动的速度。由于速度测量基于相位差，而相位差具有模糊性，要求$|\Delta\varphi| < \pi$，否则就无法判断物体是靠近雷达还是远离雷达。测量速度应该满足：

$$v < \frac{\lambda}{4T_c} \qquad (2-7)$$

如果有多个速度不同，但距离雷达相同的物体，则雷达系统必须发射两个或以上的调频连续波用于检测物体速度。这是由于这些物体与雷达的距离相同，因此中频信

[①] 毫米波雷达传感器基础知识, https://www.ti.com.cn/cn/lit/wp/zhcy075/zhcy075.pdf

号的频率完全相同，所以距离 FFT 会产生单个峰值，该峰值的相位是这些距离相同的目标信号求和得到的，因此无法直接利用相位计算速度。

若有两个距离雷达距离相同，速度不同的物体，则距离 FFT 处理反射回来的N个线性调频信号将产生N个位置完全相同但相位不同的峰值，其相位为这两个物体的相位和。

在脉冲持续的短间隔之内，如果认为物体速度基本不变，则距离 FFT 峰值处的相位是均匀变化的，由每个物体引起的相位也是均匀变化的。对距离 FFT 峰值的相量序列进行 FFT，这称为多普勒-FFT，多普勒 FFT 在 N 个相量上执行，其频率峰值ω对应的就是连续两个反射脉冲信号的相位差。物体的速度由式（2-8）可计算得出，速度分辨率与 FFT 的点数有关，由式（2-9）给出：

$$v = \frac{\lambda\omega}{4\pi T_c} \tag{2-8}$$

$$v_{res} = \frac{\lambda}{2NT_c} = \frac{\lambda}{2T_f} \tag{2-9}$$

式（2-9）中T_f为脉冲帧持续时间。

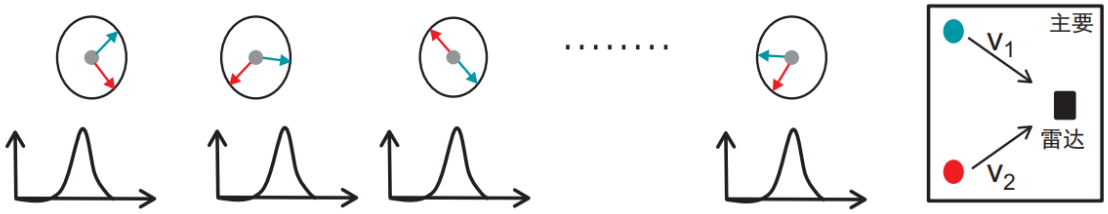

图 2-3 距离 FFT 峰值处的相位组成[①]

2.1.3 MIMO 虚拟天线

在时分复用的多发多收（TDM-MIMO）中，发射信号的正交性体现在时间上。图 2-4 的每个发射帧里有若干个发射块，每个发射块发射的脉冲数等于发射天线的数量，每个发射天线依次发射脉冲，时域上有序不重叠发射。

[①] 毫米波雷达传感器基础知识，https://www.ti.com.cn/cn/lit/wp/zhcy075/zhcy075.pdf

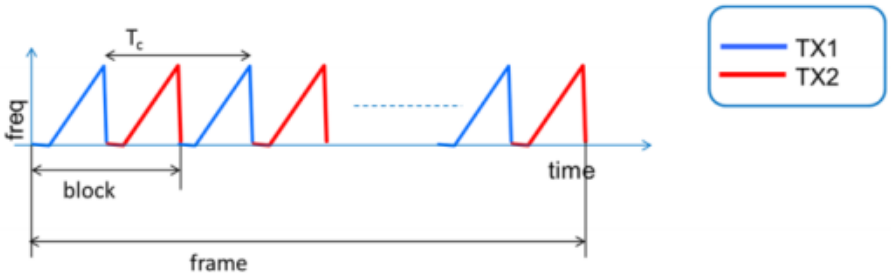

图 2-4 基于时分复用的 MIMO 技术（TDM-MIMO）[①]

图 2-5 中两个发射天线相距 $4d$，假设间隔为d引入的相位差是ω，那么蓝色发射天线与红色发射天线之间已经引入了 4ω的相位差。以第一个接收天线接收到蓝色天线的相位作为基准，第 2、3、4 根接收天线接收到蓝色天线发射信号的相位为ω，2ω，3ω，接收天线接收到红色发射天线发射的信号的相位在原来的基础上加上 4ω，因此相对相位是 4ω，5ω，6ω，7ω。那么总的来说，相当于有 8 根接收天线接收 1 根发射天线发射的信号，相邻接收天线的相位差为ω，因此 2-5 的等效虚拟天线阵列就是 2-6。

通常来说，N 发 M 收的阵列可以虚拟成 1 发 N*M 收的虚拟天线阵列。更多的虚拟天线意味着更高的分辨率。TDM-MIMO 的优点是简单，容易满足检测系统的实时性；缺点是发射天线依次发射，因此接收功率较低，天线的利用率不高。

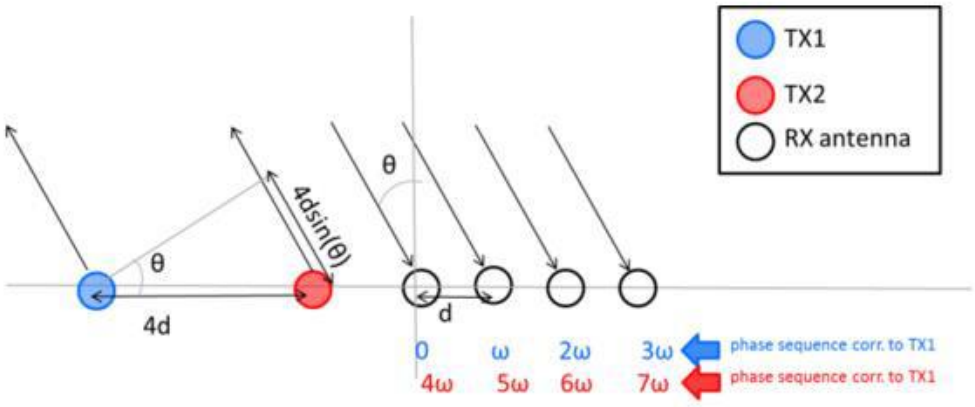

图 2-5 一种 2T4R 的天线阵列[①]

[①] MIMO Radar, https://blog.csdn.net/xiao_jie123/article/details/112093513

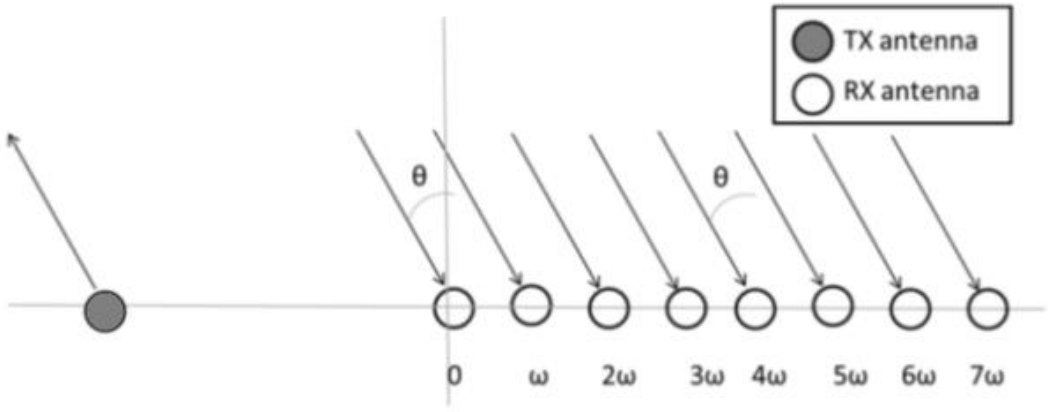

图 2-6 该 2T4R 天线对应的虚拟天线阵列[①]

2.2 本章小结

本章主要介绍了本课题中所用的基于毫米波雷达的信号处理方法，包括距离 FFT、多普勒FFT和MIMO虚拟天线技术等。这些技术后续会在仿真数据和真实数据采集、解析、处理过程中被使用。

[①] MIMO Radar, https://blog.csdn.net/xiao_jie123/article/details/112093513

第三章 系统架构、技术路线与 DETR 模型分析修改

本章首先确定了本项目的系统架构、明确了所采用的技术路线，然后 DETR 用在毫米波雷达领域的可行性进行了分析，最后介绍了对 DETR 模型的创新性、适配性修改，包括数据预处理、骨干网络、模型推断等方面的修改，同时也对采集得到真实数据集的锚定框数据分布进行了分析。修改好的 DETR 模型将会用于本文第四章进行模型训练并得到实验结果。

3.1 系统架构与技术路线

对于整个系统，我们的需求是输入雷达原始信号，根据实际任务得到各种功能性结果，如识别、追踪、姿态、动作等。如果直接使用一个端到端的网络，则该系统的可解释性与操控性是很差的[5][15]。

经过调研后我们发现，这些实际任务在输出最终结果的时候都依赖于对人体某些关键特征的提取与定位。因此，一个简单的思路是将人视为“火柴人”，即将复杂的人体结构简化为人体几十个关键点部位的连接，可以大大简化网络规模与计算量。骨架关键点是一个比较通用且关键的中间信息，得到了比较准确的骨架关键点，就可以方便的进行后续识别、追踪、姿态、动作等多样化的功能，Transformer 的骨干网络则能与后面的部分解耦。

因此，如果能构建通用的中间特征信息如人体骨架关键点或锚定框，即可将模型的特征队列提取与任务输出需求解耦，针对不同的任务需求将特征队列输入不同的输出头、得到相应输出。

我们考虑将整个系统分成以下三个部分：

第一部分是对雷达 RDM 数据通过一些简单方式进行预处理，尽可能多的保留数据的幅值信息和相位差信息。

第二部分是以预处理好的雷达信号作为输入，使用 Transformer 作为主体网络进行处理，最终期望得到人体的人体锚定框信息。

第三部分是通过得到的人体锚定框信息，根据实际任务需求进行后续功能的实现。

本项目的技术路线如图 3-1 所示。我们首先将雷达和 Kinect 采集到的数据进行解

析、匹配，得到完整的、可用于训练的数据。针对毫米波雷达信号，我们进行 Range-Doppler FFT，得到距离-多普勒热力图并输入到经过毫米波雷达适配性修改后的 DETR 主体网络 Transformer 中，由 Transformer 网络将输入数据升维到高维可分可预测空间中，并输出由各数据的嵌入空间向量构成的特征队列。这个特征队列是非常重要的中间数据，它包含了使原始输入数据高维可分可预测的所需全部信息，可根据实际使用场景及需求将该特征队列输入到对应的后处理算法、分类器中，如卷积神经网络、人体检测器、循环神经网络等，完成姿态检测[16][17]、人体存在性检测、人体位置预测以及多人追踪等不同任务。

由于得到中间数据特征队列后，根据目标任务的需要进行一定的后处理与输出头设计，即可很好地完成对应任务，因此在本文中我们主要讨论人体位置预测（回归）与动作识别（分类）两种任务，其余任务可以视为这两种任务的简单类比。

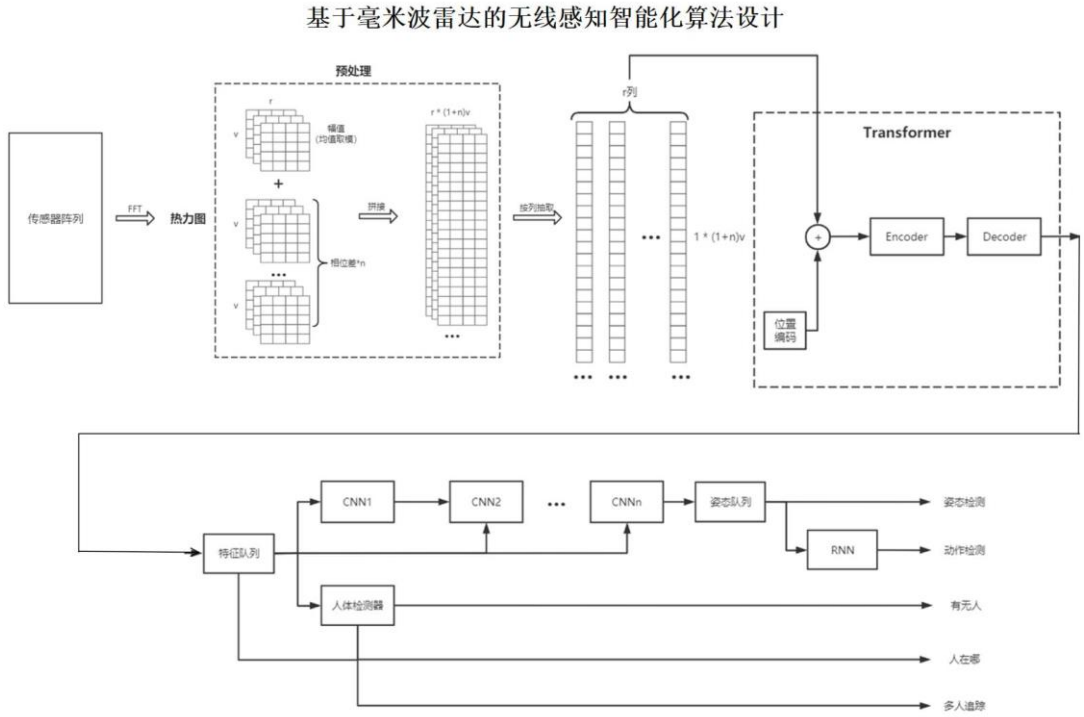

图 3-1 基于毫米波雷达的无线感知智能化算法设计技术路线图

3.2 DETR 用在毫米波雷达领域中的可行性分析

Transformer 是自然语言处理（Natural Language Processing, NLP）领域的杰作之

一[18]，由谷歌团队提出。Transformer 中创新性地提出了加入位置编码、编解码器等结构，在对每个单词进行编码后，通过将词嵌入与三个权重矩阵相乘，创造了查询向量 Q、键向量 K 和值向量 V。通过这三个向量，模型根据式（3-1）得到表征每个单词对当前编码位置的贡献的$softmax$分数，即体现了模型的注意力机制。这种注意力与“贡献”思想能对模型的训练提供指导。

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (3-1)$$

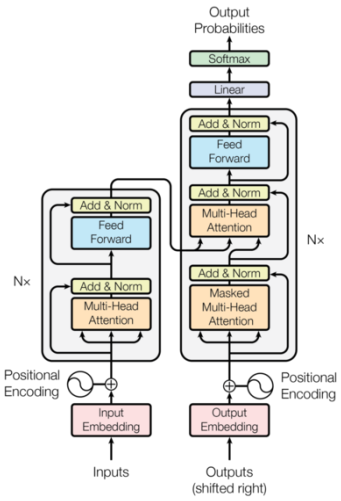

图 3-2 Transformer 模型架构图[18]

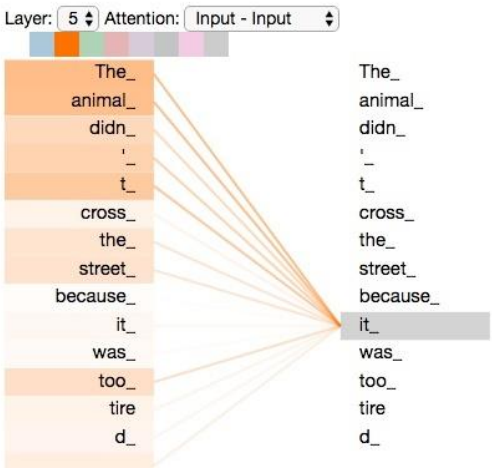

图 3-3 Transformer 注意力机制示意图[18]

DETR（Detection Transformer）[19]为目前计算机视觉领域最先进的多目标识别算法之一，由 Facebook 团队于 2020 年提出，是一种基于 Transformer 的端到端目标检测方法，是 Transformer 技术在图像目标检测领域的推广应用。DETR 把 Transformer 中的位置编码从词与词之间转变为了像素与像素之间、各包含框之间，把词语与句子之间的特征转变为了包含框与整个图像之间的相对特征。DETR 没有非极大值抑制 NMS 后处理步骤、没有先验知识和约束，对目标检测的项目流程进行了较大程度上的简化。DETR 网络的输入为图像，输出为检测的结果，检测结果包含类别、置信度与包含框位置。

相比于其他目标识别算法，DETR 是使用 CV 领域图像处理手段相对较少的一种方案，因此我们认为用 DETR 对雷达信号进行处理，所要遇到的问题会相对比采用别的 CV 模型少一些。DETR 具有位置编码，同时在修改前置处理网络后对输入数据没有平移不变性的要求，这对我们的毫米波雷达数据非常友好。

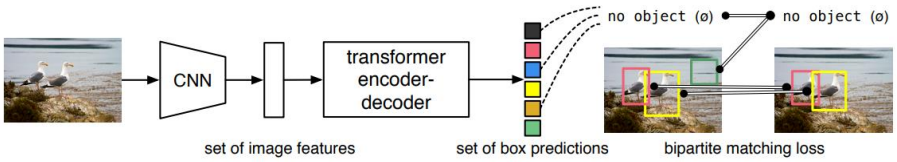

图 3-4 DETR 网络结构示意图[19]

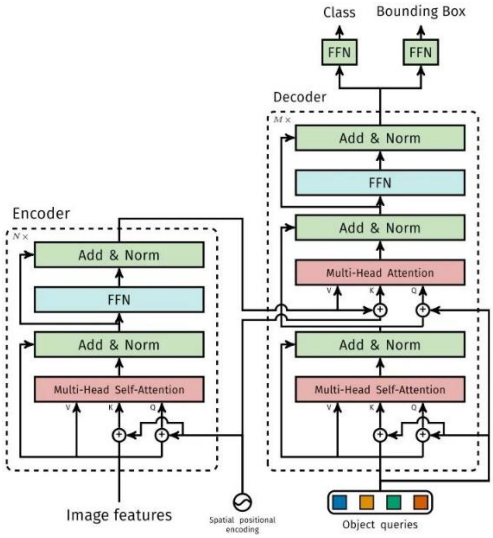

图 3-5 DETR 网络结构示意图[19]

在我们修改后的 DETR 模型中，我们的输入信息不再是 RGB 图像信息，而是将原始信号进行 Range-Doppler FFT 处理得到的 RDM。在计算机视觉领域里，RGB 图像信号中无论一个目标出现在图片的哪个位置，都是同一个目标种类，具有平移不变性；RDM 数据仅在距离 Range 维度有一定的平移不变性，整体上不具有平移不变性。因此对于雷达数据，我们在 DETR 前置网络上不能采用卷积神经网络，转而采用简单的多层感知机实现特征提取。

虽然图像数据和 RDM 数据在平移不变性上存在区别，但对于网络来说，两种数据在数据特征提取上的底层逻辑是一样的。当网络接收到输入数据后，将输入数据进行升维、编码，使得看似杂乱无章、无法区分的低维原始数据在高维空间中可分或可预测，再将编码后的数据调整到统一的维度作为数据特征输出，即可通过高维特征空间的构建实现对数据的区分或预测。

因此，对于 RDM 数据，我们通过 MLP 网络抽取特征，并进行位置编码，然后将所得到的编码特征信号送入 Transformer 中进行训练，这样的流程在理论上是可行的。

在我们的实验中，DETR 接受经过解析的毫米波雷达数据输入，通过提取数据特征，最终输出雷达前方各人物的锚定框位置，同时返回预测类别与相应置信度。

3.3 DETR 模型适配性修改

为了将用于 CV 的原始 DETR 模型修改至在 RDM 数据上使用，我们需要对 DETR 模型中的一些部分进行适配性修改。

3.3.1 数据传输方式

由于 DETR 源码使用的数据集是 coco 数据集，因此不管是在数据加载过程中还是在精度评价中，都包含了大量 coco API 的使用。由于 coco API 高度集成化、不易跨领域使用的特点，对于我们的雷达数据，我们需要采用新的数据传输方式。

我们新的数据加载方式将数据集分为训练集和验证集，根据不同数据集提取各自的雷达数据文件和对应的真值数据到列表中。在训练或者验证时，从序列最开始依次随机抽取对应的数据进行训练或验证。相比于调用 pytorch 库中 dataloader 类的传统方法，我们新的数据加载方式具有训练速度快、数据加载快的特点，但由于提前载入全部的数据信息，我们需要在训练开始前花一些时间载入全部数据集，并且将这些数据一直存放在内存中。

采用新的数据输入方式是因为我们的雷达数据以距离多普勒热力图来呈现，由于距离多普勒热力图是复数信息，而 DETR 是实数神经网络，因此需要将数据信息进行合理的转换，使之能够适配实数神经网络。在这里，我们选择幅值、相位差拼接的方法。由于 12 个等效 MIMO 天线的距离多普勒热力图幅度十分接近，因此我们将幅值求均值作为幅值输入。由于目标的俯仰角、方位角与这些天线的相位差有明显的线性关系，因此对于相位差信息，我们根据等效出的 12 个虚拟接收天线的排布，选择等效位置的天线相位做差并进行解绕。最后我们将幅值与相位差拼接后输入神经网络。

3.3.2 数据真值调整

在 DETR 源码中，实验数据的真值是图中目标的类别及其位置。在我们的实验中，我们希望网络能预测目标中人体的位置，因此模型的输出为 1∗7 的列表，包括人体骨架中心点（脊柱底部）的三维坐标、人体$Bbox$框的三维尺寸以及人体与地面的夹角（下称倾角，范围$[0,\pi]$）。

我们利用采集到的 17 个骨架关键点生成当前人体的$Bbox$框，方法为取脊柱底部作为人体中心点；在$X/Y/Z$三个维度上分别取各骨架关键点与脊柱底部的距离在该维度上的投影中的最大值的两倍作为$Bbox$框在该维度的尺寸；取脊柱顶部与脊柱底部的连线与X轴正方向的夹角作为倾角。

3.3.3 锚定框真值分布

真实数据集的真值$Bbox$数据分布小提琴图如图 3-6 所示。由于 Kinect 采集到的 25 个骨架关键点我们只取了 17 个，且 Kinect 无法采集到头顶以及脚底位置，因此$Bbox$的Y_size会小于人的实际身高。对于全部 24018 组数据，$Bbox$的平均尺寸大小为$0.64m * 1.44m * 0.64m$。考虑雷达的距离分辨率为 0.0375m，则在后续实验中，我们模型的平均预测IOU理论最大值为 0.8631。

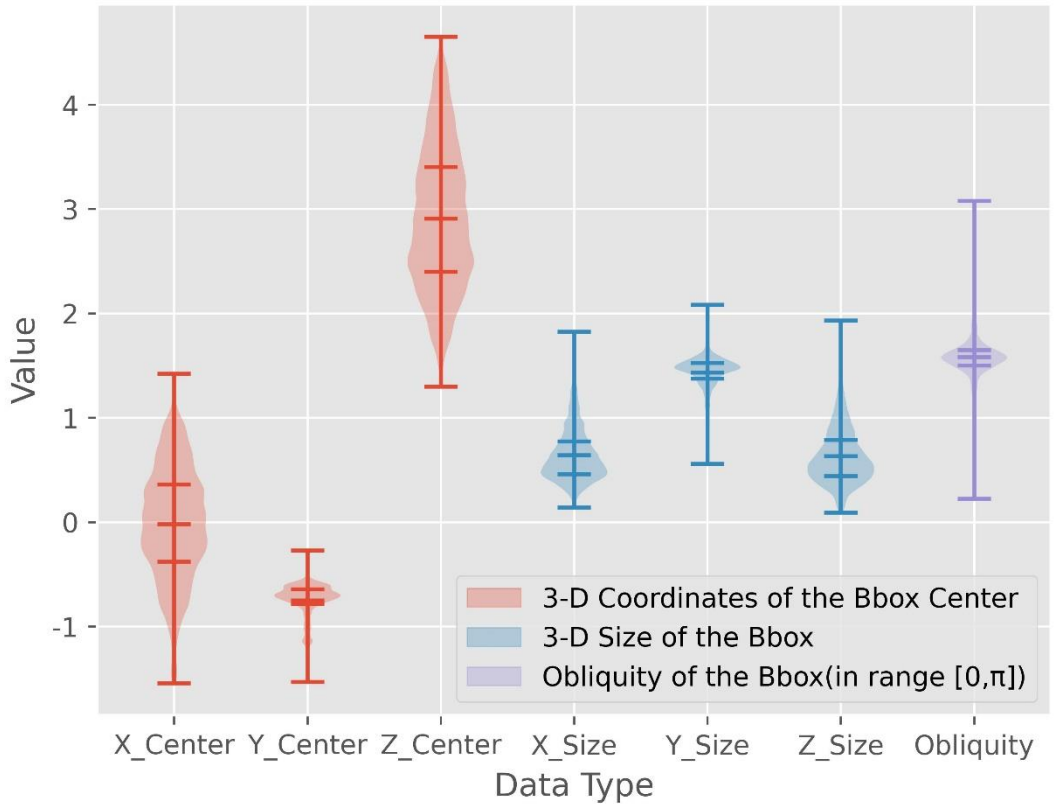

图 3-6 真实数据集真值$Bbox$数据分布小提琴图

3.3.4 骨干网络

在原始 DETR 中，骨干网络是 resnet50 模型，用于提取图像特征并为序列进行位

置编码。由于我们的输入 RDM 并不具有平移不变性，CNN 无法保留输入数据特征，因此我们选择采用多层感知机（MLP）提取信息，其中 MLP 包含三个线性层和 2 个激活函数层（$Relu$），并将 MLP 的输出进行三角函数位置编码。由于输入数据是由幅度和相位差拼接而来，同一个数组包含两个不同维度的数据，因此将数组拆分，分别将幅度信息和相位差信息输入到各自的 MLP 网络，再将两个 MLP 的输出拼接，最后将结果进行位置编码。

3.3.5 输出头调整

在 Transformer 解码输出结果后，需要将结果经过一层线性层和一层 MLP 层，分别得到类别置信度和位置预测边界框。在我们的模型中，针对类别，我们只需要修改线性层参数，使之适配于我们数据集的类别。针对关键点的坐标输出，为了契合雷达的探测原理，我们选择采用球坐标系的表示方法，输出关键点相对于雷达的距离R、方位角φ、俯仰角θ。由于在最后的坐标输出中，距离、俯仰角、方位角、$Bbox$尺寸、倾角并不是一个维度的参数，因此我们将 Transformer 的输出分别经过它们对应的 MLP 层，分别输出所需信息，从而提高我们模型的正确率。

3.3.6 后处理调整

在网络最后的 MLP 层得到输出后，后处理对数据进行格式转换，使输出直观易理解，并方便后续将输出输入到精度评价函数测算精度。由于图片尺寸不同，网络一般会在输入时将图片归一化，最后在后处理还原图片尺度。而对于我们的模型，输入是雷达数据，并不会存在尺度不同的问题，因此模型输出的坐标就是真实空间的尺度，不需要进行后处理。

3.3.7 损失函数设计调整

在这部分中，我们参考 DETR 将损失函数设计为标签损失$loss_labels$、中心点损失$loss_kpts$、锚定框损失$loss_bbox$、交并比损失$loss_iou$、倾角损失$loss_obliquity$共五个损失的加权和。

由于标签预测可以视为简单的分类问题，因此标签损失$loss_labels$主体采用交叉熵损失函数。

中心点损失$loss_kpts$为真值中心点与预测框中心点的 L2 范式距离。

锚定框损失$loss_bbox$为真值$Bbox$尺寸与预测框$Bbox$尺寸的 L2 范式距离。

交并比损失$loss_iou$为真值$Bbox$尺寸与预测框$Bbox$的$DIOU$交并比值。

倾角损失$loss_obliquity$为倾角（弧度表示）的真值与预测值差值的绝对值。

3.3.8 匈牙利匹配算法

匈牙利匹配算法的修改与损失函数修改类似，不过只保留标签损失、中心点损失、锚定框损失三个部分做加权和，最终输出能使匈牙利匹配代价最小的索引列表。

3.3.9 精度评价

在模型预测出输入数据的$Bbox$位置信息后，我们计算真值与预测值$Bbox$在三维空间中的IOU值，对三维空间中$IOU > 0.5$的预测框视为成功检出。函数返回验证集预测的平均IOU值与成功检出的真值数量。

3.4 本章小结

本章首先确定了本项目的系统架构和技术路线，即通过构建通用的中间特征信息如人体骨架关键点或锚定框，将模型的特征队列提取与任务输出需求解耦，针对不同的任务需求将特征队列输入不同的输出头、得到相应输出。然后本章分析了 DETR 用在毫米波雷达领域的可行性，最后介绍了对 DETR 模型中各模块的创新性、适配性修改，包括数据预处理、骨干网络、模型推断等方面的修改，同时也完成了对采集得到真实数据集的锚定框数据分布的分析。修改好的 DETR 模型将会用于本文第四章进行模型训练并得到实验结果。

第四章 数据采集与实验结果

本章介绍了本文实验过程中使用的仿真数据集的生成方法与真实数据集的采集、解析方法，并给出了经过适配性修改后的 DETR 模型在不同输入情况下针对仿真、真实数据集的多组训练结果，最后给出了各组实验结果对应的结论。

4.1 单人仿真数据实验

4.1.1 单人仿真数据采集

实验数据的规范性和正确性是对实验结果的重要保证。在进行深入的实验之前，我们花费了充足的精力用于生成仿真数据、采集真实数据并做预处理，同时验证这些数据的合理性。

仿真数据利用 MATLAB 中的雷达仿真工具模拟雷达板卡的天线分布生成[①]。通过不断在仿真场景范围内随机产生若干个目标，保存得到的雷达信号和目标位置，作为实验的数据集。

在仿真中，可以设置不同参数改变生成数据集的数目、每张图片目标个数，并通过随机函数，随机生成在一定范围内的不同位置、不同速度、不同雷达截面积的物体。然后我们将仿真得到的雷达信息分别按行、列进行两次 FFT（Range-Doppler FFT）得到距离多普勒热力图，并存放在 csv 文件中。仿真数据的真值信息会按照标签、距离、方位角、俯仰角的顺序存放到 txt 文件中。由于仿真数据无法生成完整的人体各关键点位置数据，因此在仿真数据的实验部分中，我们的模型仅预测人体的中心点三维坐标，即人体脊柱底端的三维坐标。

我们共收集 5000 帧单点仿真数据，目标的距离范围为 0-12.8m，方位角范围为±60°，俯仰角范围为±40°。

4.1.2 实验结果

将单点仿真数据集中的 3500 份作为训练集，1500 份作为验证集，设置模型仅预测人体的中心点目标，训练 100 代后，模型的训练结果如表 4-1 所示。验证集欧氏距

[①] FMCW-2T4R-SIM-MUSIC, https://github.com/liynjy/FMCW-2T4R-SIM-MUSIC

离误差为 0.42m，径向距离误差为 0.07m，方位角误差 2.8°，俯仰角误差 5°。

表 4-1 单目标仿真数据训练效果表

	欧氏距离误差	径向距离误差	方位角误差	俯仰角误差
训练集	<0.15m	/	/	/
验证集	0.42m	0.07m	2.8°	5.0°

单人仿真数据实验验证集的精度结果径向距离误差较小、欧式距离误差较大，有以下三个方面的原因：

1. 仿真数据在 0-12.8m 的探测范围内进行了 256 次采样，则 RDFFT 的径向距离分辨率为 0.05m，因此存在 0.07m 的径向误差完全可以接受；

2. 因为目标在 0-12.8m 范围内均匀分布，但是网络的输出结果在角度层面上，俯仰角和水平角都存在 2°-5°的误差，这导致径向距离越远的目标，所产生的误差越大。在实际的家居场景中，径向范围一般在 0-5m，角度误差导致的欧式距离误差会小很多；

3. 此时的 DETR 还比较接近于原始的结构，并没有针对我们的雷达数据进行结构和训练上的优化，因此本次实验中的精度结果还有较大提升的空间。

同时，我们对单人仿真数据集也使用传统雷达算法 CFAR 进行处理作为对照组。在 CFAR 得到的结果中，径向距离误差约为 0.1m，俯仰角和水平角的误差大约在 2°，少部分时候会到达 3°- 4°。CFAR 的结果比我们网络训练的结果略好，但从整体上说明了我们算法的正确性，也说明我们单点仿真数据的模型还有进一步的调优、提升空间。

4.2 多人仿真数据实验

4.2.1 多人仿真数据采集

在多人仿真数据的制作中，我们采用类似单人仿真数据制作的方法，利用 MATLAB 雷达仿真工具在 0-12.8m 的范围内随机生成 1-4 个目标点，对同样的场景进行仿真，得到 10000 份仿真数据，其中含 1 个真值、2 个真值、3 个真值、4 个真值的数据各 2500 份。多目标仿真数据中训练集与验证集数量比例为 4：1。

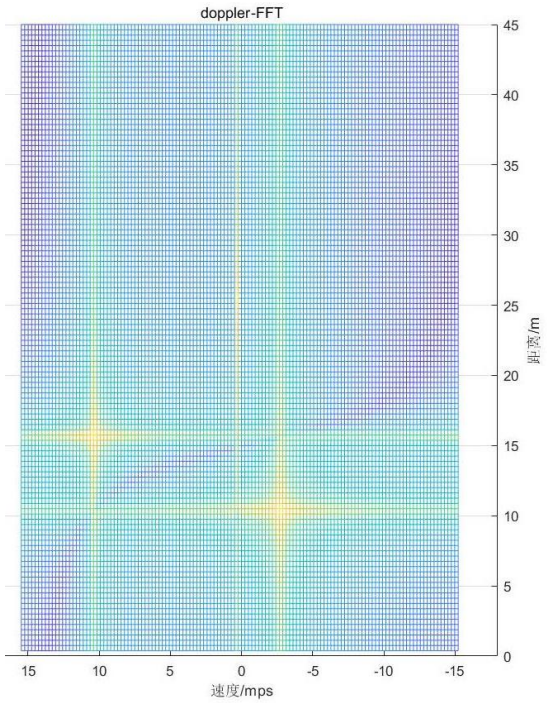

图 4-1 仿真数据 RDM 示意图

4.2.2 实验结果

对 10000 份随机包含 1-4 个真值点的多点仿真数据，取 8000 份作为训练集，2000 份作为验证集。设定网络的查询个数 NUM_QUERIES 为 4，训练 100 代，模型验证结果如图 4-2、表 4.2 所示：

表 4-2 多目标仿真数据训练效果表

	欧氏距离误差	标签正确率
训练集	0.15m	/
验证集	0.51m	84%

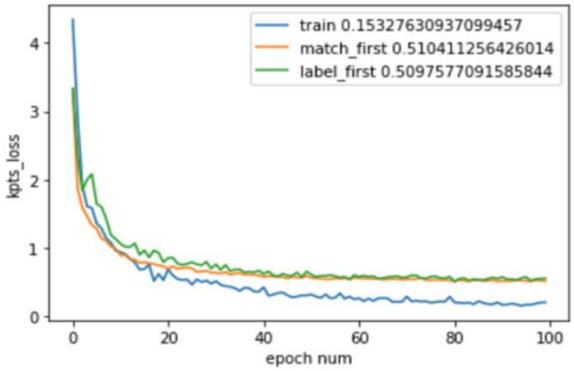

图 4-2 多目标仿真数据训练效果图

多目标的训练的结果在欧式距离的误差上比单目标有所减小，原因是多目标使得在相同范围内点的数量增多，网络的输出点位置分布和实际点位置分布都更广，因此匹配点的位置关系更加接近。同时，从单点仿真数据到多点仿真数据的数据集调整过程中，我们对 DETR 进行了结构和训练的优化，因此精度也有所提升。

4.3 真实数据锚定框实验

用单点来描述一个人的位置显然是不够的，我们希望能用一个完整包含人体的锚定框准确地框出目标人体的位置。真实数据具有由 Kinect 同步采集的人体骨架关键点三维坐标生成的人体锚定框作为真值，因此我们的模型针对真实数据的预测内容为人体中心点坐标、$Bbox$尺寸、人体倾角及人体姿态。我们选择用五个指标来评价$Bbox$的准确性，分别是预测目标框与人体实际位置中心点的误差、倾角的误差、输出框和实际框的交并比、验证集中的检出率和动作分类正确率。

从仿真数据到真实数据，通过模型的输出从单点改成锚定框，模型一方面可以延续仿真数据实验中对中心点学习的效果，另一方面又可以学习新的数据类型而不至于完全无经验，可以较好地完成模型学习上的过渡。

考虑到我们采用雷达的距离分辨率为$0.0375m$，$Bbox$平均尺寸为$0.64m * 1.44m * 0.64m$，则我们模型在三维空间中预测锚定框交并比均值的理论极限为0.8631。此处我们设置人体检出的交并比门限为0.5，即对$IOU > 0.5$的$Bbox$预测框视为成功预测。

4.3.1 真实数据采集

我们完成了真实数据采集环境的搭建，在不同场景下采集了各种人体姿态下的雷

达数据 24018 组，并且利用 Kinect 识别了人体关键点，与采集的雷达数据成功匹配，为后续神经网络提供了带标签的数据集。

我们采用的毫米波雷达为德州仪器（TI）公司的 IWR6843AOP，频段为 60-64GHz，为三发四收（3T4R）天线，天线的俯仰角方位角视野为 120°，最小距离分辨率为 0.0375m。雷达采集数据的真值由 Kinect 提供。Kinect 是一种 3D 体感摄影机，由微软公司发布，可用于提取摄像头目标范围内人体的骨架关键点位置。我们的计算实验平台为一张显存为 24G 的 RTX3090 显卡。数据采集模块功能说明见表 4-3。

在实际采集时，Kinect 与雷达正对人体、高度为 1.7m、离人体距离约为 2-4m。在坐标轴朝向方面，$X/Y/Z$轴遵循右手螺旋定则。如图 4-5 所示，站在雷达位置向人体看去时，左侧为X轴正方向，上侧为Y轴正方向，前侧为Z轴正方向。

表 4-3 数据采集模块功能说明表

模块	功能
IWR6843AOP	FMCW 雷达主射频部件，工作频段 60-64GHz
DCA1000EVM	用于雷达感应应用的实时数据捕捉
MMWAVEICBOOST	毫米波传感器承载卡平台，提供高级软件开发、追踪、调试、采集模式设置等功能
Kinect	用于识别人体关键点提供实验数据的真值，需要与雷达板卡同步采集数据

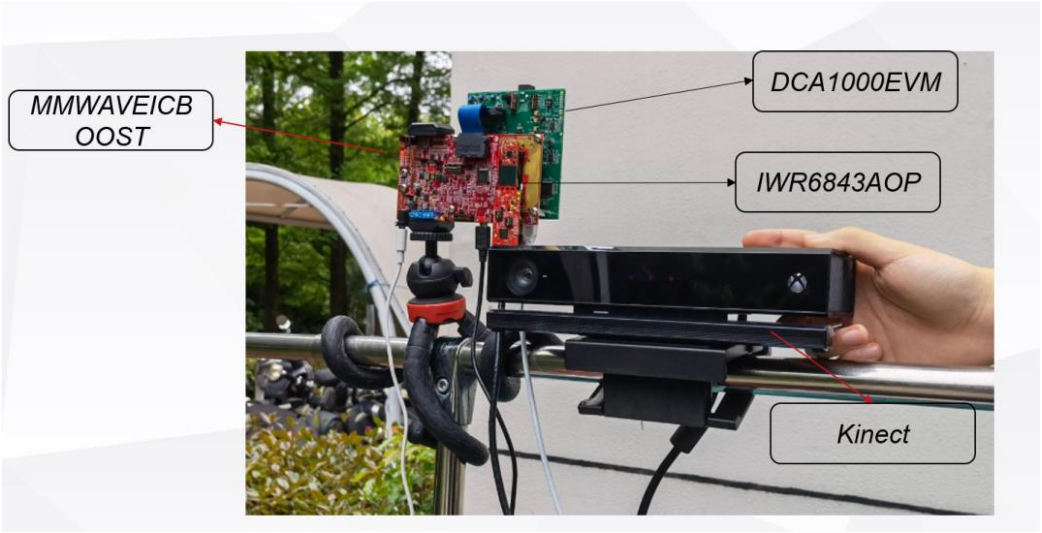

图 4-3 真实数据采集环境搭建

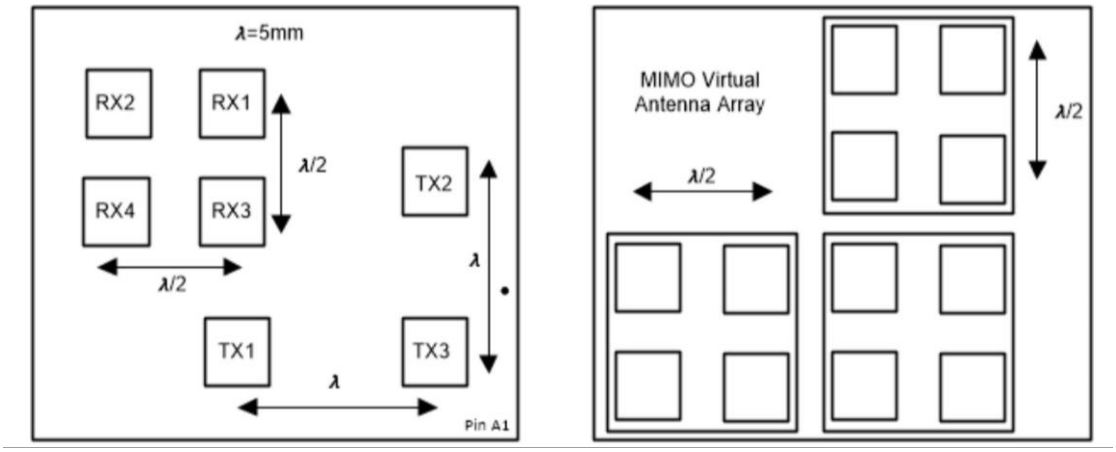

图 4-4 IWR6843AOP 天线阵列及虚拟天线阵列 ①

真实数据采集示意图与采集效果如图 4-5、图 4-6 所示。

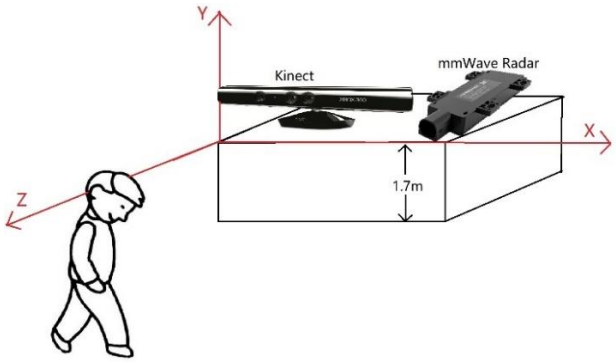

图 4-5 真实数据集采集示意图

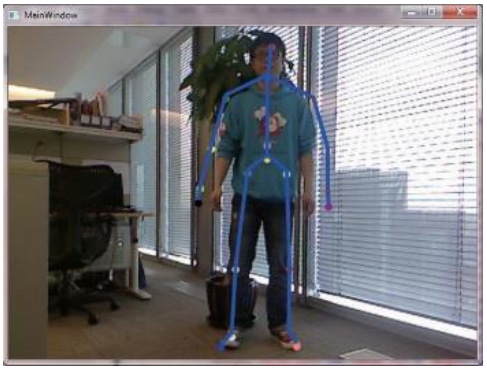

图 4-6 真实数据采集效果②

① IWR6843AOP 单芯片 60GHz 至 64GHz 毫米波传感器封装天线 (AOP), https://www.ti.com.cn/document-viewer/cn/IWR6843AOP/datasheet/GUID-1976E361-FC6E-435E-BE72-9103B7A5177F#TITLE-SWRS188X8752

② Kinect 骨骼追踪数据的处理办法, https://blog.csdn.net/qq_33835307/article/details/81269372

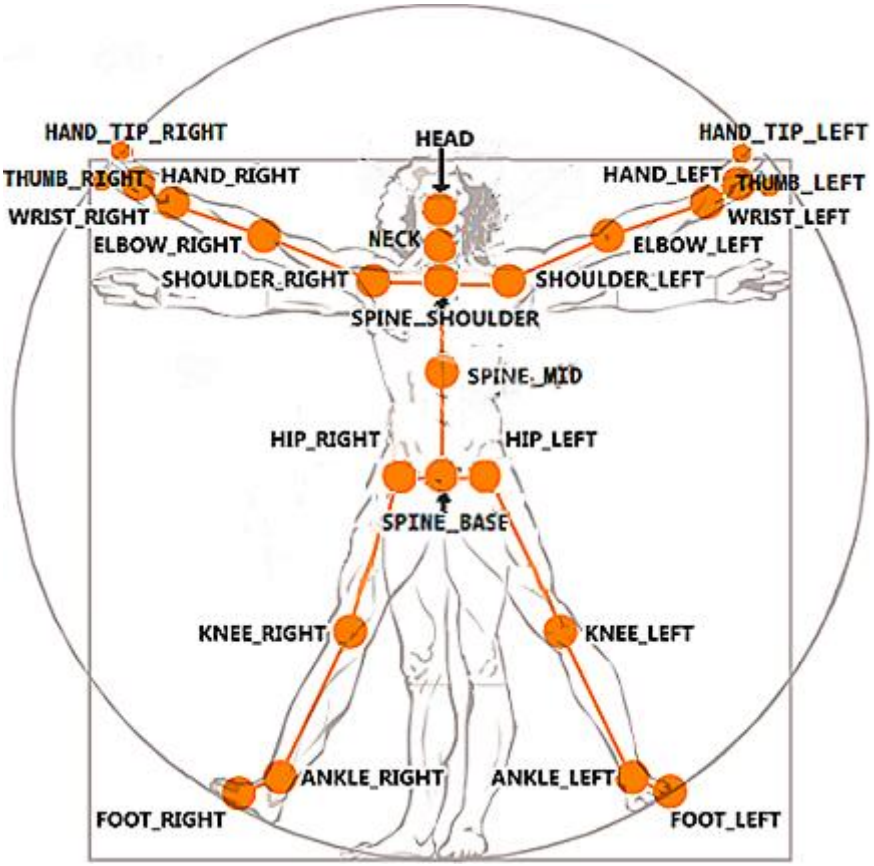

图 4-7 Kinect 采集骨架关键点坐标图 [①]

我们分别在三个小房间内进行数据采集，其中两个房间根据不同的视角各采集了两次，因此共在五个不同场景下进行了采集。在这些场景中，雷达和 Kinect 均从房间的一端朝向另一端，两端中间基本没有多余的物体干扰。在数据采集过程中，由一个人在房间两端中间的地上做各种静止或运动的动作。

我们在五个场景下都进行时长 500 秒、10 FPS 的数据采集，然后进行数据的筛选，排除同步不佳或者硬件出错的数据。在采集得到数据后，我们发现某些数据相似度过高、帧间IOU过大，在经过筛选评估后，仅剩 24018 组数据。最终我们得到了包含静止、走路、小跑、坐下、站起、随机动作共六种不同运动状态的数据集，各动作的数据量如图 4-8 所示。

在 24018 组数据中，我们随机取其中 4000 组数据集作为验证集，即训练集与验证集数据量比例为 5:1。在实验中，我们只设计了训练集和验证集，并没有设计测试集。考虑到测试集和验证集同样都没有梯度传播、没有参与网络的训练，我们在此将

① Kinect for Windows 资料下载, http://www.k4w.cn/news/1.html

测试集和验证集视为等效，统一用验证集代替。

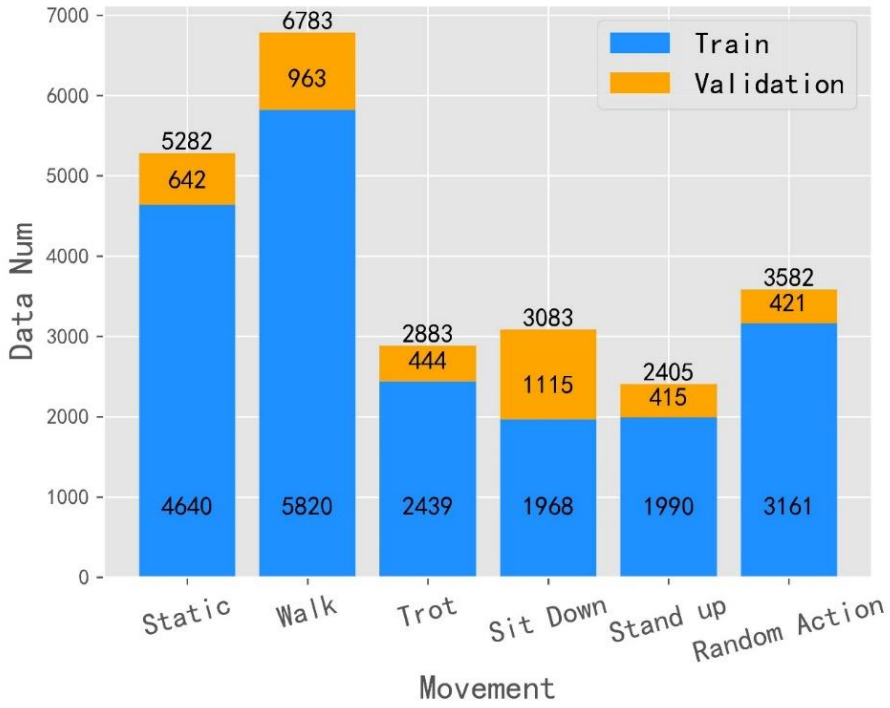

图 4-8 采集真实数据动作分布图

我们采集到的数据需要经过解析才能被输入网络使用。采集得到的每帧原始雷达数据为一个三维的二进制文件，三个维度分别为天线维、慢时间维（脉冲维）和快时间维（采样维）。我们首先在快时间维进行距离 FFT 得到物体与雷达之间的距离，然后在慢时间维进行多普勒 FFT 得到物体的速度信息。在对 12 个虚拟天线的数据各自进行两次 FFT 之后，我们就得到了某帧数据的距离-多普勒热力图（Range-Doppler Map，下称 RDM），该数据由天线维、距离维、速度维三个维度构成，我们将解析后的数据保存为 json 文件，等待后续输入网络使用。

采集到的每帧 Kinect 数据保存了图 4-7 中所示的人体 25 个骨架关键点在空间中的$X/Y/Z$三维坐标信息。考虑到部分骨架关键点较密、区分性不大，因此我们只取其中 17 个骨架关键点生成的人体锚定框作为真值，17 个关键点即头、颈、脊柱顶端、脊柱中端、脊柱底端、两侧肩膀、两侧手肘、两侧手腕、两侧臀部、两侧膝盖、两侧脚踝。

如图 4-5 所示，通过将 Kinect 和雷达放置在非常接近的位置，我们可以近似认为待测物体与 Kinect 的相对位置即待测物体与雷达的相对位置。

在实际训练时，我们以解析后的雷达数据作为实验数据，以 Kinect 得到的关键点为真值，并通过采集数据的时间戳将实验数据与其真值匹配起来输入网络。

4.3.2 基础实验

对 24018 份真实数据，我们取 20018 份作为训练集，4000 份作为验证集。设定网络的查询个数 NUM_QUERIES 为 1。

在不进行动作分类、倾角预测的基础设定下，模型训练 300 代后结果如图 4-9、表 4-4 所示。本次实验的验证集中心点误差为 0.0547m，$Bbox$交并比为 0.7748，4000 组验证集中检出率为 92.53%。相比于仿真数据，真实数据的精度无论是在训练集还是验证集上，都提升了很多，达到了理想的效果，原因可以归结于以下两方面：第一，真实数据集的雷达参数在距离分辨率上达到了 0.0375m，相比于仿真数据提升较大；第二，在从仿真数据到真实数据这个过程中，我们在 DETR 上从数据传输方式、网络结构、损失函数设计等方面都做了许多优化。

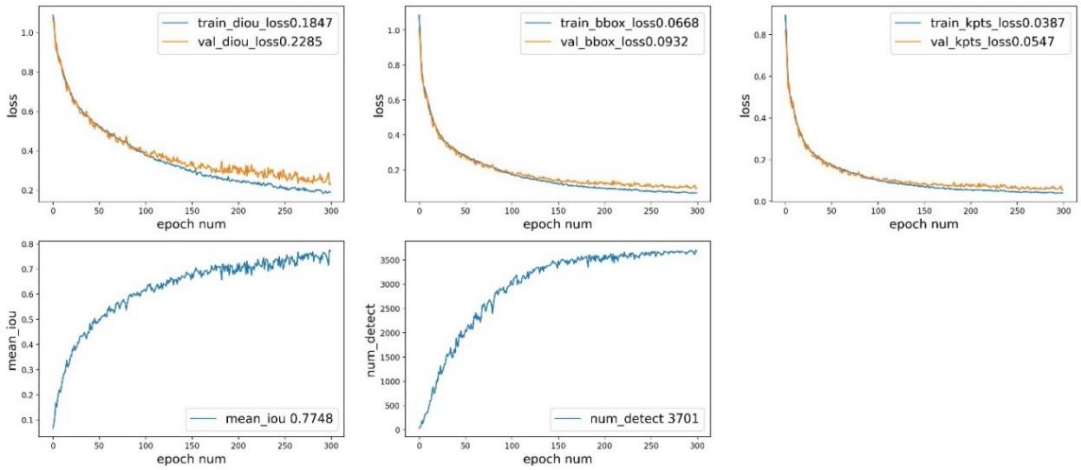

图 4-9 真实数据锚定框基础实验结果

表 4-4 真实数据人体 Bbox 定位实验结果表

实验类型	验证集中心点误差	验证集倾角误差	Bbox 交并比	验证集检出率	验证集动作分类正确率
基础实验	0.0547m	/	0.7748	92.53%	/
加入动作分类	0.0517m	/	0.7854	93.30%	99.12%
加入动作分类与倾角预测	0.0486m	1.39°	0.7913	93.73%	99.22%

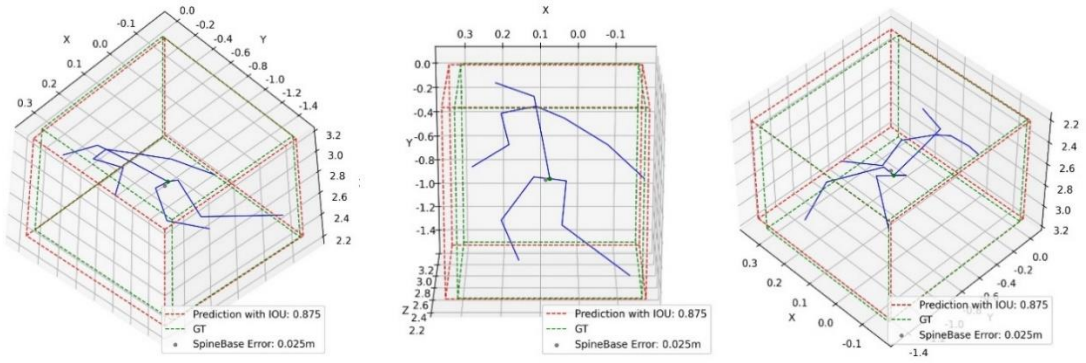

图 4-10 真实数据锚定框训练效果图

图 4-10 展示了模型对其中一组雷达数据的预测结果示意图。图中蓝色线条为人体真实姿态构成，绿色框为由人体姿态生成并作为真值输入模型的锚定框，红色框为模型预测得到的锚点框，灰色圆点为模型预测的人体脊柱底端位置。在图 4-10 中，真值框与预测框的交并比值为 0.875，对框中心点（即人体脊柱底部）的预测误差为 0.025m。

4.3.3 加入动作分类实验

受到 CenterPoint[13]的启发，在上一节的基础实验基础上，我们尝试让模型在预测人体$Bbox$的同时对样本中人的动作姿态做出分类。调整的具体方法为在网络的最后一层输出层处并联一个用于动作分类的$MOVEMENT_MLP$网络。当特征数据输出网络时，从$BBOX_MLP$输出人体的$Bbox$预测信息，从$MOVEMENT_MLP$输出人体的动作姿态分类信息。该实验很好地验证了本文技术路线的合理性，即不直接使用端到端的网络，而是将模型的特征队列提取与任务输出需求解耦开来，针对不同的任务需求我们只需要将同样的特征队列输入不同的输出头即可。本小节中，我们的模型在一次训练任务中很好地完成了机器学习领域中预测（回归）与分类这两种看似完全不相关的任务，这充分地验证了我们技术路线的可行性与鲁棒性。

如表 4-4、图 4-11 所示，在加入动作分类的实验中，验证集中心点误差为 0.0517m，$Bbox$交并比为 0.7854，4000 组验证集中检出率为 93.30%，验证集动作分类的准确率为 99.12%。同时在图 4-11 中可见，模型对人体动作的分类准确率在训练过程中快速提升。在模型训练到第 95 代时，验证集动作分类的准确率就已经达到了 98%以上。

这样的结果比上一节中基础实验的结果有所提升，我们认为这是因为当动作分类的损失加入损失函数并进行梯度反向传播后，动作分类的损失可以辅助网络学习到不同动作对应的$Bbox$分布和雷达数据的相对关系，因此模型的整体性能得到了提升。

为了说明在模型输出层处并联*MOVEMENT_MLP*层用于动作分类的可行性，我们取出*MOVEMENT_MLP*层的最后一层输入进行 UMAP（统一流形逼近与投影，Uniform Manifold Approximation and Projection）降维展示。UMAP 是基于流形学习技术和拓扑数据分析思想的一种非线性降维算法，常用于对高维数据降维可视化。在 UMAP 降维可视化图中，各数据点簇之间分的越开、混淆点越少，就说明 UMAP 的输入数据低维可分性越强，即说明我们的模型越成功地学习到了各类数据之间的差异。

我们取*MOVEMENT_MLP*中最后一层的 32 维输入作为模型的动作特征嵌入向量，将 4000 组验证集的嵌入向量输入 UMAP 并降维至二维，得到模型训练到不同代数（第 10/30/50/70/95/best 代）时 UMAP 降维图的结果图 4-12。从图 4-12 中可见，模型的动作分类准确率随着训练逐步提升。在训练到第 10 代时，模型动作分类准确率只有 52.65%，代表各类动作特征向量的降维散点胡乱地分散在图中各处、类与类之间交错混淆；在训练到第 30 代时，模型动作分类准确率为 88.92%，各动作类的特征向量开始逐渐成簇，其中静止动作散点簇与其余动作散点出现界限；在训练到第 50 代时，模型动作分类准确率达到 94.43%，各动作类进一步成簇、分界；在训练到第 70 代时，分类准确率为 96.73%，除了行走与小跑、坐下与站起两对动作区分不够明显以外，其余动作之间都有了明显的界限；在训练到第 95 代时，模型的分类准确率达到了 98.10%；在模型训练最佳时，模型分类准确率为 99.12%，六种不同动作散点各自成簇、每个动作与其余动作都有非常明显的界限。UMAP 降维图中各动作散点的成簇情况随模型训练的结果变化，说明了我们模型动作分类训练的有效性，也进一步验证了通过提取数据中的通用特征，我们的模型可以很好地同时完成分类与预测（回归）两种完全不同的任务。

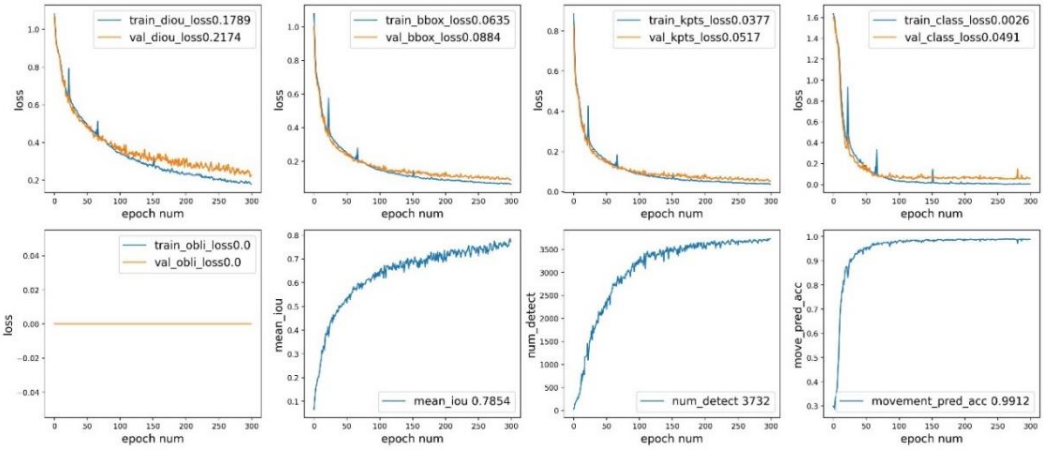

图 4-11 真实数据锚定框实验加入动作分类结果

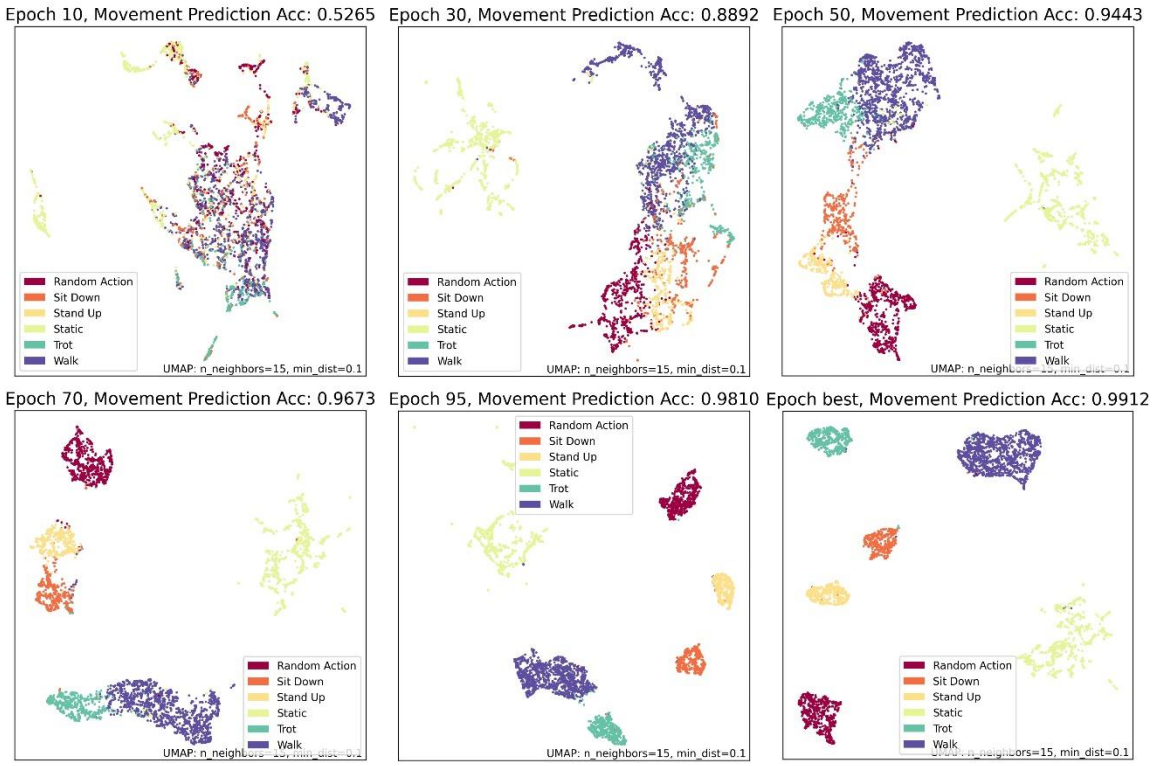

图 4-12 真实数据锚定框实验加入动作分类 UMAP 降维图

4.3.4 加入动作分类与倾角预测实验

受到 Centerpoint 模型的启发，我们开始思考$Bbox$框的倾角或人体的倾角是否对网络训练能起到帮助作用。倾斜的人体相比姿态未知的人体可以提供更多信息，在$Bbox$框的预测上可以起到一定的辅助作用。因此在上一节的实验基础上，我们给模型增加输入了每组雷达数据的人体倾角，并试图让模型在验证集上预测人体$Bbox$框的同时输出人体此时的倾角与动作分类结果。具体实现方法与上一节类似，只需在网络的输出层上再额外并联一个用于输出倾角的$OBLIQUITY_MLP$层。

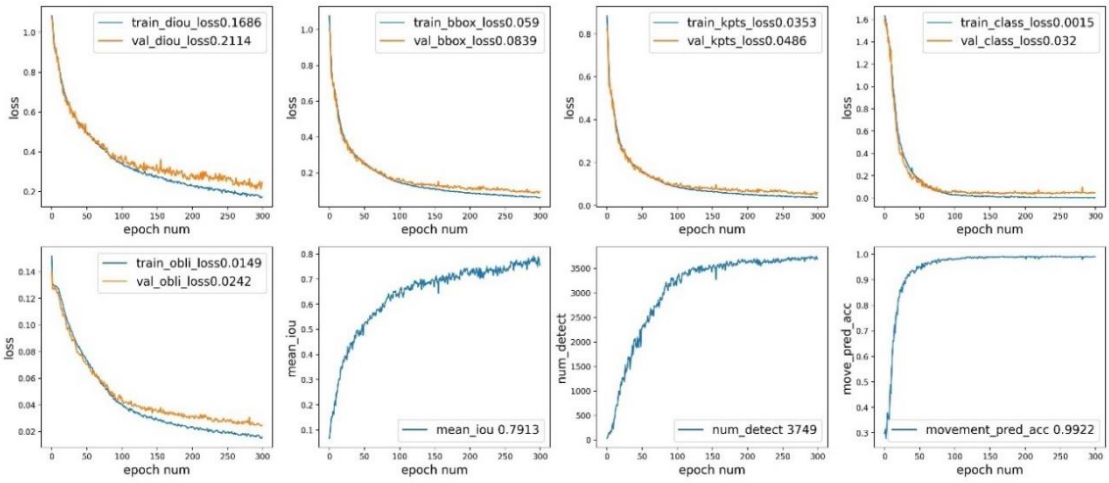

图 4-13 真实数据锚定框实验加入动作分类与倾角预测结果

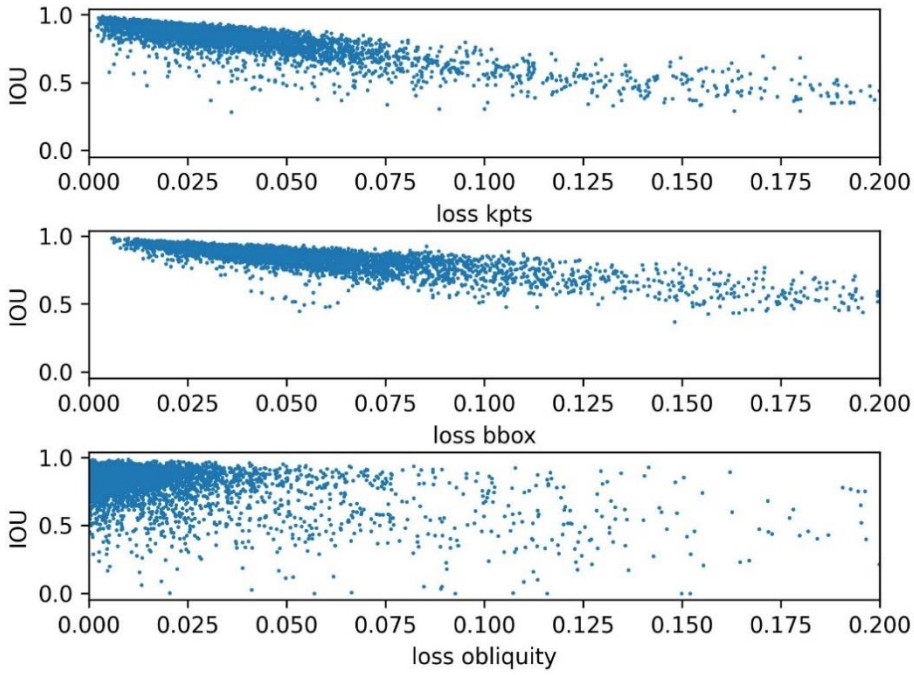

图 4-14 模式采用的三种 Loss 与 IOU 值的关系散点图

如表 4-4、图 4-13 所示，在加入动作分类与倾角预测的实验中，验证集中心点误差为 0.0486m，$Bbox$交并比为 0.7913，4000 组验证集中检出率为 93.73%，验证集动作分类的准确率为 99.22%。这样的结果比基础实验的结果有明显提升、同时也比加入动作分类的结果有所提升，我们认为这是因为倾斜的人体相比姿态未知的人体可以提供更多信息，在$Bbox$框的预测上可以起到一定的辅助作用；同时，倾角与$Bbox$尺

寸和动作有较强的相关性。在人体做不同类型的动作时，这些动作的倾角呈现出不同的分布，同时人的锚定框尺寸分布也发生相应的变化，因此加入动作分类、加入倾角预测都对模型的性能起到了增益作用。验证集散点图 4-14 展示了本实验中所采用的各种$loss$与IOU值之间的关系，图中可见$loss_kpts$、$loss_bbox$与IOU值呈现明显的线性关系，而$loss_obliquity$与IOU之间则没有这种线性关系。因此在这三种不同类型的损失中，可以认为$loss_kpts$、$loss_bbox$主要负责IOU值提升，而$loss_obliquity$主要负责人体倾角预测与协助$loss_labels$进行动作分类。图 4-14 很好地说明了本实验损失函数设计的合理性与有效性。

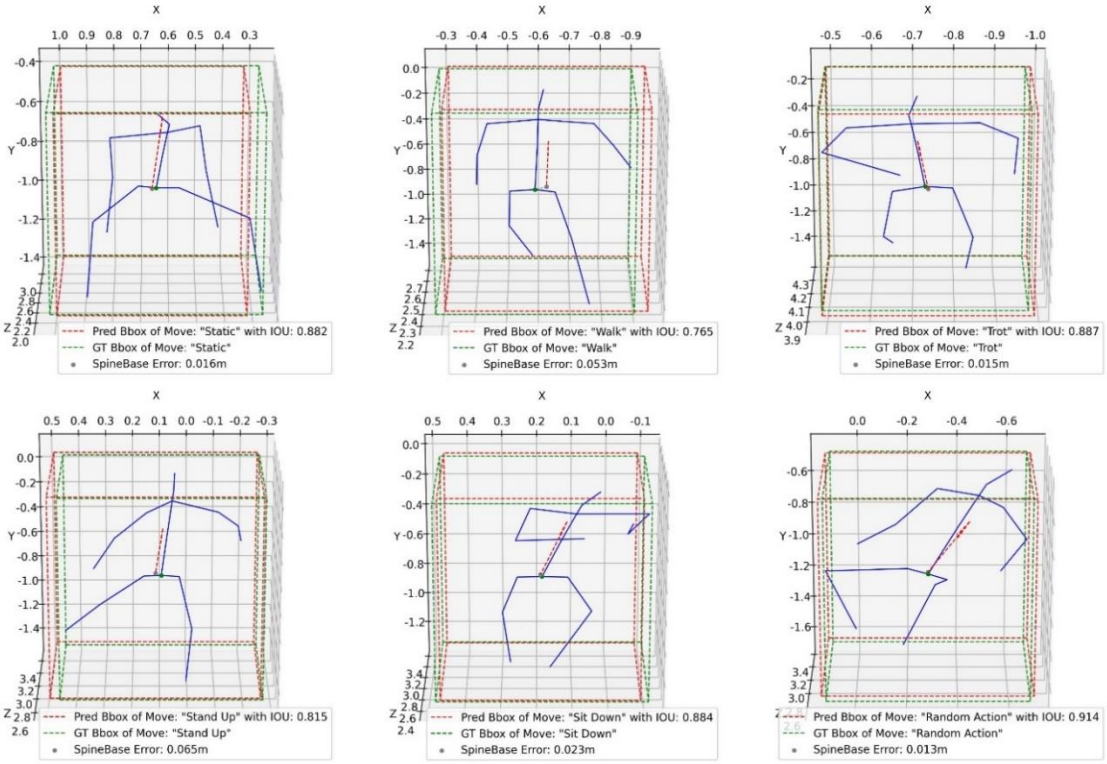

图 4-15 真实数据锚定框加入动作分类与倾角预测训练效果图

图 4-15 展示了模型对六组不同动作雷达数据的预测结果示意图。图中蓝色线条为人体真实姿态构成，绿色框为由人体姿态生成并作为真值输入模型的$Bbox$，红色框为模型预测得到的$Bbox$；灰色圆点为模型预测得到的人体脊柱骨架底端位置，灰色圆点处指出的红色箭头为模型预测的人体倾角朝向（$obliquity$）。图 4-15 中，我们的模型对六种不同动作姿态都给出了正确的姿态分类，同时$Bbox$交并比都较高、脊柱底端预测误差都较低。

图 4-16 给出了本模型在训练到不同代数（0/40/80/120/160/200/240/280/best 代）时对同一组数据的预测输出对比，很好地说明了本模型训练对正确预测人体位置与动作分类的有效性。从图 4-16 中可以看出，模型在前 80 代内错误地将本样本的动作分类为“行走”，直到第 80 代时才正确分类此样本动作为“小跑”；同时，该样本的预测IOU值随训练代数的增加逐渐提高、脊柱底端预测误差随训练代数的增加逐渐降低。在训练初期，模型对该样本预测的$Bbox$与真值$Bbox$完全不重合、脊柱底端估计误差高达 1.23m；在训练到 200 代时，脊柱底端估计误差首次收敛到 0.1m 内，同时IOU值首次达到了 0.7 以上。随着模型的进一步训练，最终在模型验证集均方误差最小时该样本的预测输出达到了 0.887 的交并比与 0.015m 的脊柱底端预测误差。

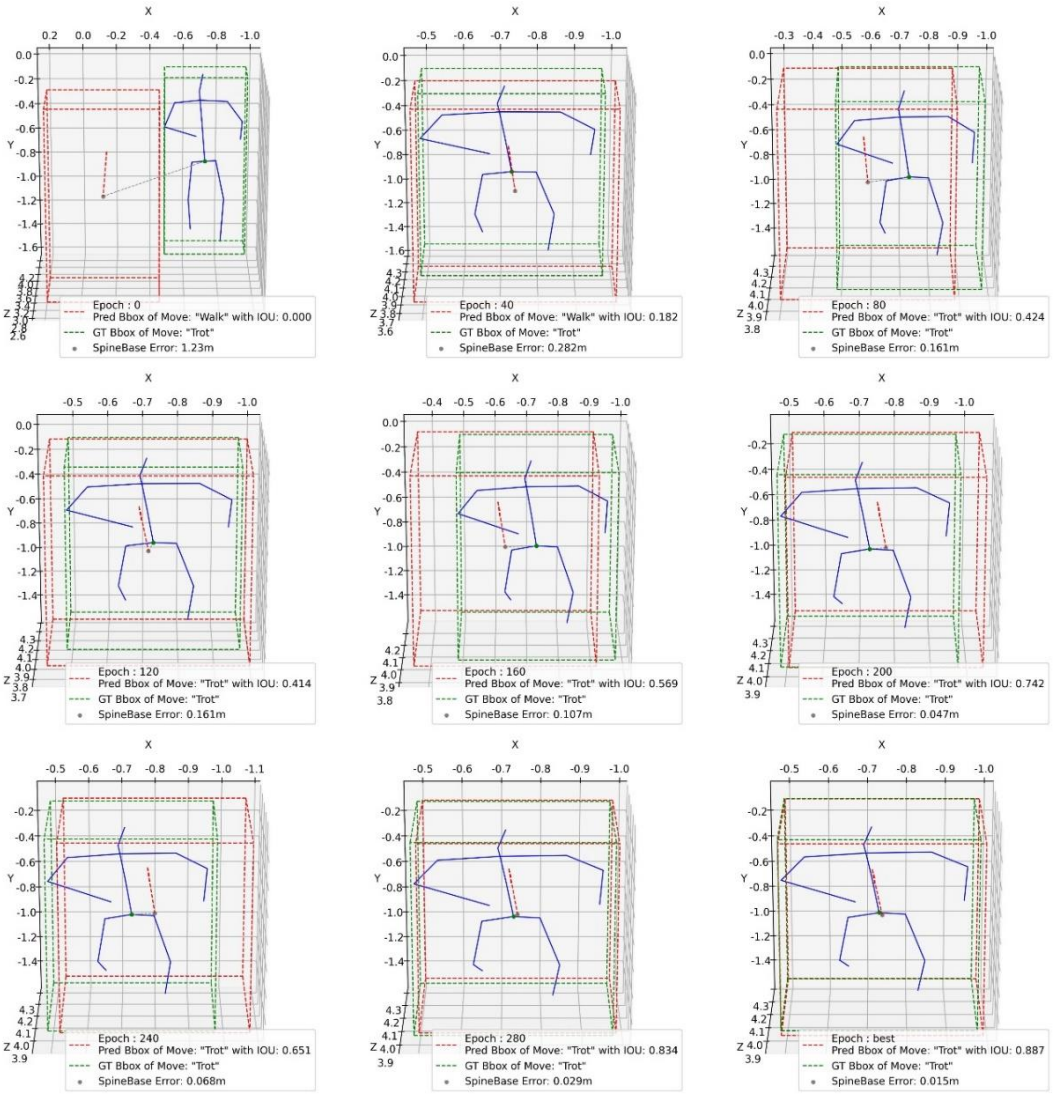

图 4-16 真实数据锚定框加入动作分类与倾角预测训练效果图

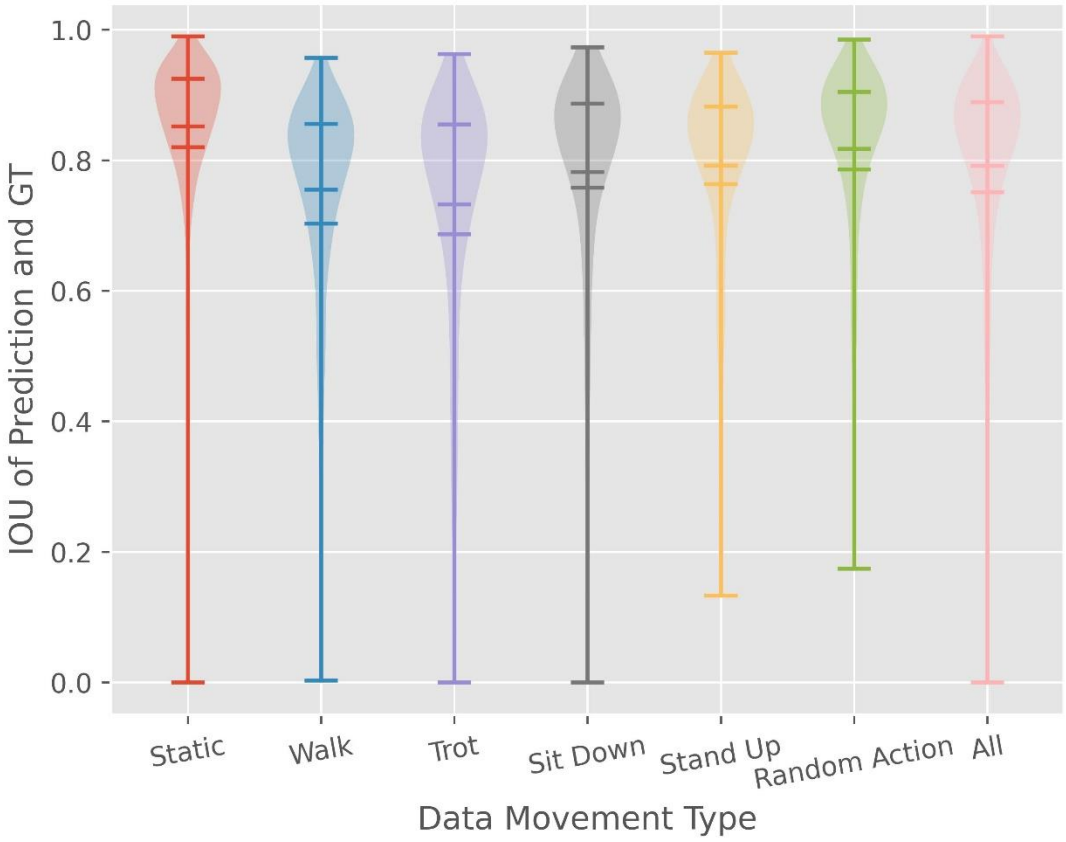

图 4-17 不同动作预测锚定框的 IOU 分布小提琴图

在三维空间中，任何一个维度的微小误差都会导致总体的交并比下降明显。图 4-17 展示了本模型对不同动作的锚定框预测IOU值分布，图中可见不同动作的平均预测IOU值都在 0.73 以上，其中静止动作的平均IOU值最高，可以达到 0.85 的水平；行走、小跑的平均IOU值相对较低一些，约为 0.74；所有动作的平均IOU值为 0.79。图 4-18 展示了行走、小跑动作中两组预测IOU相对较低的样本。关于行走、小跑的平均IOU值相对较低的现象，我们认为由于行走、小跑这两个动作都是相似的快速移动动作，因此此处主要考虑是雷达与 Kinect 数据标签同步时差导致的问题。由于雷达数据与 Kinect 数据在时间戳上没法完全对齐，因此在采集完数据进行数据匹配时，我们将时间戳相差 0.2 秒以内的雷达与 Kinect 数据进行匹配。但由于人离雷达的距离较近、且行走和小跑都为移动速度较快的动作，因此在 0.2 秒的时间间隔内，人体与雷达的相对位置关系（包括距离和方位角）可能发生了较大变化，从而影响了模型对这两个动作姿态下的人体进行锚定框预测的性能，所以这两个动作$Bbox$的交并比较低。

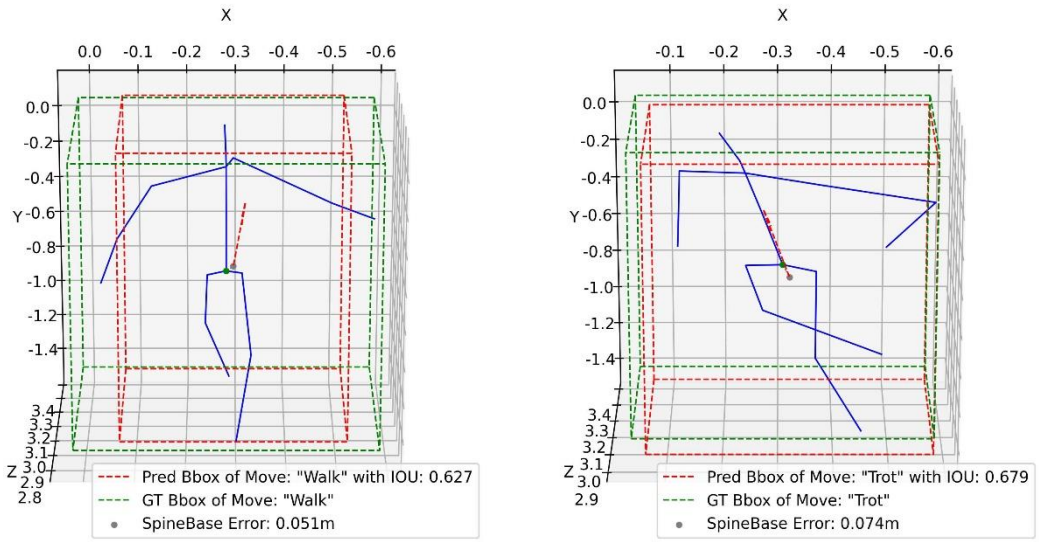

图 4-18 行走、小跑动作中两组预测 IOU 相对较低的样本

4.4 本章小结

本章介绍了本文实验过程中使用的仿真数据集的生成方法与真实数据集的采集、解析方法，并给出了经过本文第三章适配性修改后的 DETR 模型在不同输入情况下针对仿真、真实数据集的多组训练结果，最后给出了各组实验结果对应的结论。

仿真数据利用 MATLAB 中的雷达仿真工具模拟雷达板卡的天线分布生成。修改后的 DETR 模型在单人仿真数据集上的脊柱底部预测误差为 0.42m、方位角预测误差为 2.8°、俯仰角预测误差为 5.0°；在多人仿真数据集上模型的标签预测正确率为 90%，脊柱底部预测误差为 0.51m。

我们完成了真实数据采集环境的搭建、在不同场景下采集了各种人体姿态的雷达真实数据 24018 组，并且利用 XBOX360 体感周边外设 Kinect 识别了人体关键点，与采集的雷达数据成功匹配，提供了带标签的真实数据集。

修改后的 DETR 模型在单人真实数据集上完成了三个实验。其中，基础实验的脊柱底部预测误差为 0.0547m、IOU为 0.7748、验证集检出率为 92.53%；加入动作分类实验的脊柱底部预测误差为 0.0517m、IOU为 0.7854、验证集检出率为 93.30%、动作预测的准确率为 99.12%；加入动作分类与倾角预测实验的脊柱底部预测误差为 0.0486m、IOU为 0.7913、验证集检出率为 93.73%、动作预测的准确率为 99.22%、验证集倾角误差为 1.39°。

在仿真数据与真实数据的实验结果对比中，由于真实数据的雷达分辨率有所提升，且 DETR 在各方面进行了针对性优化，因此真实数据的实验结果比仿真数据更优。

在真实数据三组锚定框实验中，相比于基础实验，加入动作分类的实验中动作分类的损失可以辅助网络学习到不同动作对应的$Bbox$和雷达数据的相对关系，因此在预测类任务上比基础实验结果更好。同时加入动作分类的实验也很好地同时完成了预测与分类这两个不同领域的任务，这充分地验证了我们技术路线的可行性与鲁棒性。我们通过构建通用的中间特征信息如人体骨架关键点或锚定框，将模型的特征队列提取与任务输出需求解耦，针对不同的任务需求将特征队列输入不同的输出头、得到相应输出。

相比于真实数据的基础实验和加入动作分类实验，加入动作分类与倾角预测实验在各方面的指标都达到了三个实验中的最优。这是因为倾斜的人体相比姿态未知的人体可以提供更多信息，在$Bbox$框的预测上可以起到一定的辅助作用；同时，倾角与$Bbox$尺寸和动作有较强的相关性。在人体做不同类型的动作时，这些动作的倾角呈现出不同的分布，同时人的锚定框尺寸分布也发生相应的变化，因此加入动作分类、加入倾角预测都提升了模型的性能。上述三个真实数据实验说明了本项目系统架构与技术路线设计的合理性。

第五章 全文总结

5.1 主要结论与创新点

本文提出了一种将模型的特征队列提取与任务输出需求相解耦的系统架构，并在这个架构下进行了实验数据采集、训练，解决了智能家居领域中基于毫米波雷达的无线感知智能化算法模型精度不高的问题，完成了基于毫米波雷达的无线感知智能化算法设计。

本文提出在毫米波雷达智能家居领域，可以通过构建通用的中间特征信息如人体骨架关键点或锚定框的方式把不同的任务统一起来，这种实现方式相比端到端的网络有更好的鲁棒性、适用性和可移植性。在本文的技术路线下，我们针对计算机视觉领域的 Detection Transformer 模型进行了分析和针对毫米波雷达领域的创新性、适配性修改，并进行了仿真数据生成和真实数据采集与解析。

我们对修改后的 DETR 模型进行了实验评估，结果表明，在人体位置检测任务上，真实数据基础实验的人体检出率为 92.53%，锚定框交并比值为 0.7748，脊柱底端预测误差达到 0.0547m。在加入动作分类的实验中，人体检出率为 93.30%，交并比值为 0.7854，脊柱底端预测误差达到 0.0517m，动作分类准确率达到 99.12%。在加入动作分类与倾角预测实验中，人体检出率为 93.73%，交并比值为 0.7913，脊柱底端预测误差达到 0.0486m，倾角预测误差为 1.39°，动作分类准确率达到 99.22%。

对比真实数据锚定框系列实验的三个实验结果，我们认为模型很好地同时完成了预测与分类这两种不同的任务，且加入动作分类、加入倾角预测都对模型的性能有所提升，这证明了本项目系统架构与技术路线的合理性。

最终，我们的模型在人体动作分类任务上达到了 99.22%的准确率，在人体位置检测任务上达到了 93.73%的检出率和 0.7913 的锚定框交并比值。

5.2 研究展望

由于时间关系，本文的实验研究主要专注于将 DETR 从计算机视觉领域移植到毫米波雷达领域，因此本文的研究工作还有进一步完善的空间。在本文研究成果的基础上，还有值得进一步深入展开的研究工作，具体来说这些工作包括但不限于：

（1）减小网络结构与参数数量，提升网络的可部署性与实时性；

（2）通过数值对比实验进行超参数的进一步调优，提升实验结果；

（3）在本项目的系统架构下完成针对其余智能感知任务的输出头实现。

参 考 文 献

[1] Zhang G, Geng X, Lin Y J. Comprehensive mpoint: A method for 3d point cloud generation of human bodies utilizing fmcw mimo mm-wave radar[J]. Sensors, 2021, 21(19): 6455.

[2] Alizadeh M, Shaker G, De Almeida J C M, et al. Remote monitoring of human vital signs using mm-wave FMCW radar[J]. IEEE Access, 2019, 7: 54958-54968.

[3] Pegoraro J, Rossi M. Real-time people tracking and identification from sparse mm-wave radar point-clouds[J]. IEEE Access, 2021, 9: 78504-78520.

[4] Xu Z, Shi C, Zhang T, et al. Simultaneous monitoring of multiple people's vital sign leveraging a single phased-MIMO radar[J]. IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology, 2022, 6(3): 311-320.

[5] Zhao P, Lu C X, Wang B, et al. CubeLearn: End-to-end learning for human motion recognition from raw mmWave radar signals[J]. IEEE Internet of Things Journal, 2023.

[6] Singh A D, Sandha S S, Garcia L, et al. Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar[C]//Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems. 2019: 51-56.

[7] Zhao P, Lu C X, Wang J, et al. mid: Tracking and identifying people with millimeter wave radar[C]//2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS). IEEE, 2019: 33-40.

[8] Schumann O, Wöhler C, Hahn M, et al. Comparison of random forest and long short-term memory network performances in classification tasks using radar[C]//2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF). IEEE, 2017: 1-6.

[9] Zhao M, Li T, Abu Alsheikh M, et al. Through-wall human pose estimation using radio signals[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7356-7365.

[10] Yan S, Xiong Y, Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]//Proceedings of the AAAI conference on artificial intelligence. 2018, 32(1).

[11] Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 652-660.

[12] Qi C R, Yi L, Su H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in neural information processing systems, 2017, 30.

[13] Yin T, Zhou X, Krahenbuhl P. Center-based 3d object detection and tracking[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 11784-11793.

[14] Law H, Deng J. Cornernet: Detecting objects as paired keypoints[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 734-750.

[15] Li G, Zhang Z, Yang H, et al. Capturing human pose using mmWave radar[C]//2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). IEEE, 2020: 1-6.

[16] Gong P, Wang C, Zhang L. Mmpoint-gnn: Graph neural network with dynamic edges for human activity recognition through a millimeter-wave radar[C]//2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021: 1-7.

[17] Jin F, Zhang R, Sengupta A, et al. Multiple patients behavior detection in real-time using mmWave radar and deep CNNs[C]//2019 IEEE Radar Conference (RadarConf). IEEE, 2019: 1-6.

[18] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[19] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16. Springer International Publishing, 2020: 213-229.

符号与标记（附录 1）

符号内容	中文全称	英文全称
FFT	快速傅里叶变换	Fast Fourier Transform
Bbox	锚定框	Bounding Box
DETR	/	Detection Transformer
CV	计算机视觉	Computer Vision
RDM	距离多普勒热力图	Range-Doppler Map
NLP	自然语言处理	Natural Language Processing
CFAR	恒虚警率	Constant False-Alarm Rate
MTI	移动目标指示	Moving Target Indication
MIMO	多输入多输出	Multiple-Input Multiple-Output
MLP	多层感知机	Multilayer Perceptron
FOV	视场角	Field of View
GT	真值	Ground Truth
CNN	卷积神经网络	Convolutional Neural Network
RNN	循环神经网络	Recurrent Neural Network
GNN	图神经网络	Graph Neural Network
LSTM	长短期记忆	Long Short-term Memory
NMS	非极大值抑制	Non-Maximum Suppression
FFN	前馈神经网络	Feed Forward Network
MVDR	最小方差无畸变响应	Minimum Variance Distortionless Response
RDFFT	距离多普勒快速傅里叶变换	Range-Doppler Fast Fourier Transform
API	应用程序编程接口	Application Programming Interface
IOU	交并比	Intersection Over Union
DIOU	距离交并比	Distance Intersection Over Union
UMAP	统一流形逼近与投影	Uniform Manifold Approximation and Projection
NUM_QUERIES	查询个数	Number of Queries

致　谢

此处隐去。

DESIGN OF INTELLIGENT WIRELESS SENSING ALGORITHM BASED ON MILLIMETER-WAVE RADAR

As people enter the era of ubiquitous connectivity, various application scenarios have emerged, with one significant scene being smart homes. In the context of smart homes, the concept encompasses a wide range of semi-open/closed indoor environments. In these scenarios, people desire not only a warm and relaxing atmosphere but also an intelligent and comfortable home experience. Among the growing number of smart homes, various sensors such as home cameras, motion sensors, and voice assistants are used to record and monitor people's living environment and behaviors. The collected data is processed by upstream servers, which then provide corresponding adjustments that have practical impacts on people's lives.

However, the utilization of sensor data collection also brings about various privacy issues, such as the commonly encountered topic of object tracking in smart homes. The current deployed object tracking technologies primarily rely on real-time photos or videos captured by cameras, which are transmitted to the cloud for decoding and analysis. The results are then sent back to the user side. Nevertheless, camera data is difficult to encrypt, prone to leakage, and easily comprehensible to the public. The prevalence of high-definition cameras and their omnipresence have raised significant privacy concerns. The emergence of pinhole cameras, the illicit industry of covert filming, and user data breaches have made people increasingly anxious. Therefore, given the background of the current project, it is necessary to propose a new technological approach to replace the traditional method of information collection using cameras.

In this regard, the design of intelligent wireless sensing algorithms based on millimeter-wave radar is proposed and considered. We aim to employ millimeter-wave radar instead of traditional cameras to accomplish multi-object tracking tasks. The adoption of millimeter-wave radar over cameras is driven by two primary considerations. Firstly, mil-

limeter-wave radar has lower precision compared to high-definition cameras, and its data does not raise concerns about privacy infringement. Secondly, the raw information from millimeter-wave radar, after undergoing transformations, does not carry direct meaning and cannot be easily interpreted by the general public. These two factors collectively protect user privacy and to some extent address the privacy issues associated with traditional cameras.

This project is based on the raw data from millimeter-wave radar and incorporates multiple optimizations in the data processing algorithm, drawing inspiration from the Detection Transformer (DETR) model in the field of Computer Vision (CV). The goal is to achieve intelligent wireless sensing algorithm design based on millimeter-wave radar.

In existing solutions for smart homes based on millimeter-wave radar, the prevalent approach is the point cloud projection heatmap method. However, in practical usage scenarios, it has been observed that this approach tends to lose information and is susceptible to interference from moving objects within the home environment, such as curtains and fish tanks. On the other hand, the pattern recognition and segmentation work based on radar in the academic community has primarily relied on laser radar. However, the data collected by laser radar differs significantly from that of millimeter-wave radar, making it challenging to directly transfer the relevant techniques from the laser radar domain to the millimeter-wave radar domain.

Therefore, we aim to propose a completely new solution that can effectively address the aforementioned issues and realize the design of intelligent wireless sensing algorithms based on millimeter-wave radar.

This paper proposes a system architecture that decouples the feature extraction and task output requirements of models for intelligent sensing algorithms based on millimeter-wave radar in the smart home domain. We conducted experiments on simulated and real data, and successfully solved the problem of low accuracy of wireless sensing intelligent algorithms based on millimeter-wave radar. We also designed an intelligent sensing algorithm based on millimeter-wave radar.

In the millimeter-wave radar smart home domain, we proposed a way to unify different tasks by building universal intermediate feature information, such as human body skeleton keypoints or anchor boxes. This approach has better robustness, adaptability, and

portability than end-to-end networks. We generated simulated data and collected real data and analyzed the Detection Transformer (DETR) model in the computer vision field and made innovative and adaptive modifications for millimeter-wave radar field.

For the modified DETR model, the basic experiment on real data achieved a detection rate of 92.53% and an intersection-over-union (IOU) of 0.7748 on the human body position detection task, with a spine base prediction error of 0.0547m. In the experiment with the addition of action classification, the detection rate reached 93.30% with an IOU of 0.7854 on the human body position detection task, and the spine base prediction error was 0.0517m, with an action classification accuracy of 99.12%. In the experiment with the addition of action classification and obliquity prediction, the detection rate reached 93.73% with an IOU of 0.7913 on the human body position detection task, the spine base prediction error was 0.0486m, the obliquity prediction error was 1.39°, and the action classification accuracy was 99.22%.

For the three experiments on real data with anchor boxes, the addition of action classification and obliquity prediction improved the model's performance. The results of the three experiments on real data confirm the rationality of our technical route and loss function design.

Finally, our intelligent sensing algorithm based on millimeter-wave radar achieved an accuracy of 99.22% on the human body action classification task and a detection rate of 93.73% with an IOU of 0.7913 on the human body position detection task, demonstrating the feasibility of our proposed model architecture.

Future research could include optimizing the network structure and parameter number to improve the model's deployability and real-time performance, further tuning the neural network hyperparameters to enhance experimental results and implementing output heads for other smart sensing tasks based on our feature extraction in future work.

Here is a brief overview of the main content of each chapter:

Chapter 1 of this paper introduces the background and motivation of the research, discusses the current status and development trends of wireless sensing technology in the field of millimeter-wave radar-based smart homes, clarifies the purpose and significance of this study, and provides a brief overview of the chapter arrangement in the paper.

Chapter 2 introduces the signal processing methods based on millimeter-wave radar

employed in this study, including distance FFT, Doppler FFT, and MIMO virtual antenna technology. These techniques will be utilized in the subsequent processes of simulation data generation, real data acquisition, parsing, and processing.

Chapter 3 establishes the system architecture and technical roadmap of this study and analyzes the use of DETR in the millimeter-wave radar domain, including necessary adaptions. The chapter first proposes a system architecture and technical roadmap that decouples feature extraction from task-specific requirements and explains the approach of constructing universal intermediate feature information, such as human skeletal keypoints or anchor boxes, to unify different tasks in the field of millimeter-wave radar-based smart homes. The feasibility of using DETR in the millimeter-wave radar domain is then discussed, highlighting the need for adaptions to the model to work with millimeter-wave radar datasets. The chapter provides detailed methods for adapting the sub-modules of the DETR model, including adjustments to data transmission, loss function design, backbone network, etc. Additionally, an analysis of the distribution of anchor box data from the collected real dataset is presented to address the adaptability issues when applying DETR in the millimeter-wave radar domain.

Chapter 4 outlines the generation method for the simulated dataset used in the experimental process and describes the data acquisition and parsing methods for the real dataset. It presents multiple sets of training results obtained from the adaptability-modified DETR model under different input scenarios, using both the simulated and real datasets. Detailed analyses of each experimental result are provided to validate the feasibility, robustness, and rationality of the technical roadmap proposed in this project, thereby elucidating the entire workflow. In the beginning of Chapter 4, experiments conducted on single-person simulated data are introduced to showcase the algorithm's performance in that particular scenario. Subsequently, experiments conducted on multi-person simulated data are discussed to verify the algorithm's robustness in complex scenes. Finally, a comprehensive description of anchor box experiments based on real data is provided, including basic experiments as well as results incorporating action classification and tilt angle prediction. Through thorough analysis and discussions of the experimental results, the effectiveness and performance of the designed algorithm are evaluated.

Chapter 5 summarizes the research findings and innovations of this project and pro-

vides an outlook on future research directions that warrant further investigation.

Through the exploration of the chapters mentioned above, this paper proposes a system architecture that decouples feature extraction from task-specific outputs and introduces innovative and adaptive modifications to the DETR model in the millimeter-wave radar domain. This enables the design of intelligent wireless sensing algorithms based on millimeter-wave radar, which exhibit reliable performance in the conducted experiments. This research provides an innovative approach and methodology for algorithm design in the field of millimeter-wave radar-based smart homes.