IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

出版日期 xxxx 00，0000年；当前版本日期 xxxx 00，0000年。数字对象
标识符 10.1109/ACCESS.2017.Doi Number

一种利用卷积神经网络和卷积自编码器的基于边缘的 WiFi 指纹室内定位方法

Amin Kargar-Barzi[1], Ebrahim Farahmand[2], Nooshin Taheri Chatrudi[3], Ali Mahani[2], Muhammad Shafique[4], （IEEE 高级会员）
[1]爱尔兰科克科克大学 Tyndall National Institute[2]伊朗克尔曼沙希德·巴霍纳尔大学电气工程系可靠与智能系统实验室[3]美国亚利桑那州立大学健康解决方案学院 EMIL 实验室[4]阿拉伯联合酋长国纽约大学阿布扎比分校（NYUAD）工程部 eBrain 实验室

通讯作者 通讯作者：Ebrahim Farahmand（电子邮箱：ebi271@ gmail.co m).

本研究部分得到了纽约大学阿布扎比分校城市交互网络中心（CITIES）的支持。该中心由 Tamkeen 通过 NYUAD研究院CG001奖项资助；同时，本研究也得到了人工智能与机器人中心（CAIR）的支持，该中心由 Tamkeen 通过 NYUAD研究院CG010奖项资助。此外，第一作者声明，本研究是在其担任伊朗克尔曼沙希德·巴霍纳尔大学可靠与智能系统实验室研究人员期间开展的。

摘要 随着室内位置-服务的持续发展，近年来，室内环境中用户的位置信息一直是一个具有挑战性的问题。由于 Wi-Fi网络的广泛使用，WiFi指纹定位已成为定位移动用户最实用的方法之一。除了定位精度之外，延迟和用户隐私等其他关键因素也应在室内定位系统中予以考虑。在本研究中，我们提出了一种面向边缘设备（例如智能手机）的轻量级卷积神经网络方法，通过消除定位系统对云/服务器的需求来解决上述问题。为验证其可扩展性，我们在三个不同的开放数据集（即 UJIIndoorLoc、坦佩雷和 UTSIndoorLoc）以及我们采集的名为 SBUK-D 的数据集上对所提出的方法进行了评估。我们还在一部安卓智能手机上评估了该定位方法的性能效率，以展示其在边缘设备上的适用性。对于 UJIIndoorLoc 数据集，我们的模型取得了约 99% 的建筑物准确率、超过 90% 的楼层准确率以及 9.5 米的定位平均误差；模型大小和推理时间分别为 0.5 MB 和 51 μs。这些结果表明，该方法不仅达到了当前最先进研究成果范围内的高准确率，而且适用于资源受限的边缘设备。

索引词 室内定位、深度学习、卷积神经网络、WiFi指纹定位、基于边缘的模型

I. 引言

如今，室内环境中与用户位置相关的信息在大多数应用程序中都受到了广泛关注，尤其是在室内位置服务（ILBSs）中 [1], 。在当今时代，ILBSs可应用于多个领域，例如室内导航与跟踪、基于位置的广告（购物广告）、基于位置的信息检索（博物馆中的游客导览服务、医疗保健领域的员工和患者跟踪）等 [1][2]。准确且低成本的定位系统是ILBSs的重要组成部分，已受到学术界和工业界的重视。通常，在室外环境中，这一问题已通过全球定位系统（GPS）技术得到解决，但由于卫星信号受到阻挡、衰减或反射，这种方法并不适用于室内场所 [1]。因此，如何找到准确且低成本的室内定位系统，已成为该领域持续面临的挑战。

近年来，摄像机 [3],可见光 [4], 蓝牙 [5], Wi-Fi [6], 超宽带 [7] 和射频识别 [8] 等不同技术已被用于室内定位。在这些技术中，Wi-Fi技术备受关注，因为大多数公共建筑物（如办公室和购物中心）都部署了Wi-Fi网络及其基础设施。此外，大多数用户都拥有支持Wi-Fi技术的智能手机。因此，位置信息

可以通过这项技术获取，无需额外的硬件和成本。

传统定位算法，如三边测量或三角测量，基于从某些参考节点测得的距离或角度等信息来估计位置 [9]。这些方法需要视距（LOS）通信来测量准确的距离或角度。因此，显然这些方法不适用于具有大量墙壁和其他类型障碍物的室内环境 [10]。在不同方法中，WiFi指纹定位方法无需距离或角度信息即可轻松克服上述问题，因此是一种适用于非视距（NLOS）环境的合适方法。实际上，在WiFi指纹定位方法中，仅使用来自各个接入点（AP）的WiFi信号的接收信号强度指示器（RSSI）来计算用户的位置。在这些方法中，假设某一点处多个WiFi信号的接收信号强度指示具有唯一性；因此，可以利用这种模式（即WiFi指纹）来估计位置 [11]。

通常，WiFi指纹定位主要分为两个阶段：离线阶段和在线阶段。在离线阶段，通过采集目标区域内若干已知位置处可接入接入点的 RSSI 值，构建 WiFi指纹数据集（也称为无线电地图），其中每种 RSSI 模式都标注了对应的位置。在在线阶段，用户利用采集的数据集估计自己的位置。在此阶段，用户在所在位置测量 RSSI 模式，并将数据发送至系统（服务器或云端），以便通过将该 RSSI 模式与数据集中的可用模式进行匹配来确定位置。匹配的目的是从数据集中找出与测得的 RSSI 模式最相似的模式。

WiFi指纹定位有多种用于匹配并估计位置的方法，包括概率方法、K近邻（KNN）和支持向量机（SVM）[12]。这些方法需要复杂的滤波和参数调整，耗时且计算量大。为了减少耗时和计算负担，近年来深度神经网络（DNN）已被用于定位 [13]–[15]。尽管许多研究尝试降低计算量和耗时，但仍有一个主要问题亟待解决：如何有效优化基于 DNN 的方法，使其适用于室内定位应用程序。为此，本文提出一种轻量级的基于 CNN 的模型，以解决上述问题。

此外，大多数最先进方法都基于云，会收集数据并将其发送到服务器、云端或其他设备，以分析并计算位置 [16]。这一过程存在以下缺点，会显著影响定位过程的效率。

• 将用户数据传输到云端或服务器等第三方平台，会引发隐私和安全问题。• 由于数据必须发送到服务器进行计算，随后服务器再将位置响应发送给用户，因此定位过程的延迟会增加。• 采用集中式方法需要服务器或云端进行计算，这会增加网络流量和系统成本。

近年来，基于边缘的系统主要用于各种应用程序，所有计算都在边缘完成，因此无需将数据传输到其他地方。为了解决上述缺点，我们采用了基于边缘的室内定位系统，以在隐私、延迟和成本方面取得更好的表现。然而，边缘设备无法运行复杂算法，例如近期提出的用于室内定位的复杂 CNN。为此，我们提出了一种轻量级卷积神经网络模型，可在资源有限的边缘设备上运行。该模型可部署在智能手机等边缘设备上，从而显著提升定位性能，并改善准确率、延迟和成本表现。图1展示了我们创新工作的概况，本文的主要贡献如下：

简要描述如下。

• 一种可在处理器和内存等资源有限的边缘设备（如智能手机）上运行的轻量级卷积神经网络模型，适用于室内定位应用。为此，采用卷积自编码器进行特征提取、去噪和降维。• 在预处理方面，采用区域网格化方法，将定位任务从回归转化为分类任务，并通过提高定位精度、减小网络大小和降低复杂度来改善网络性能。• 在安卓智能手机上评估所提出的网络模型，并通过在三个不同的公共数据集以及我们采集的名为 SBUK-D 的数据集上进行评估，验证该模型的可扩展性。

论文结构：本文其余部分组织如下：第二节概述若干与室内定位方法相关的研究。随后，第三节分三个主要小节介绍我们提出的模型及其结构。之后，模型

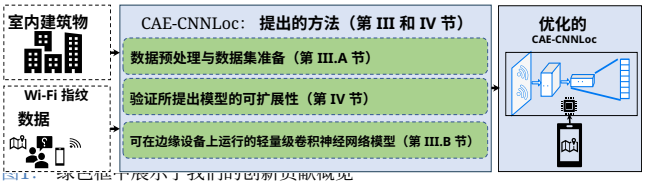


图1：绿色框中展示了我们的创新贡献概览

评估和实验结果将在第四节和第五节报告。最后，第六节总结全文。

II. 相关工作

下面简要介绍一些近期的室内定位研究及其特点和问题。室内定位方法总体上分为两大类：无设备定位方法和基于设备的方法。顾名思义，在无设备定位方法中，无需用户携带任何特殊设备即可估算其位置，例如利用闭路电视检测并定位室内环境中的人员。相反，基于设备的方法需要用户佩戴某些特定设备来计算其位置[2], ，例如利用智能手机的 Wi-Fi 或其他传感器进行用户定位。

大多数无设备定位（DFL）技术都需要借助一些基础设施来估算位置。一些 DFL 研究（例如[17][18], ）利用无线传感器网络（WSN）进行定位，其主要思路是在目标区域周围布置多个无线传感器，并使它们彼此通信。如果有人出现在该区域或在其中移动，这些无线链路就会受到影响。因此，可以根据无线链路的变化估算位置。例如，[19] 中的作者采用了这种方法，并将 DFL 问题表述为图像分类问题。随后，他们提出了一种三层卷积自编码神经网络，用于提取特征，并根据与不同位置相关的多种模式，从原始数据中计算位置。此外，[20], 中的作者提出了一种 DFL 方法：在两面相邻墙上安装正交的 RFID 标签，并使用一些 RFID 读写器测量相位信息和接收信号强度指示。采集到的信息被输入基于粒子群优化的算法，以估算物体在二维空间中的位置。

研究发现，无设备方法需要额外的基础设施，从而增加定位成本。此外，这些方法不适用于大型建筑物；在大多数情况下，它们需要视距通信，因此只能用于单层建筑物。

相反，基于设备的定位方法试图通过用户携带的设备估算位置。这些便携式设备利用各种类型的模块（例如蓝牙或惯性传感器）来估算用户在室内环境中的位置。如今，大多数人都使用配备了不同类型设备（例如 Wi-Fi、蓝牙和惯性传感器）的智能手机；因此，研究人员最近开始尝试利用用户的智能手机进行定位。由于在大多数建筑物中都能广泛使用，基于 Wi-Fi 的方法在众多技术中引起了人们对室内定位的极大兴趣 [2]。在基于 Wi-Fi 的方法中，一些研究（如 [21] ）采用到达时间（TOA）技术估算位置，而 TOA 的主要缺点是所有收发器之间需要同步。此外，还有一些研究通过估算位置基于到达角（AOA）的方法进行定位，该方法需要配备多根天线的接入点，这也是其缺点之一。此外，一些研究采用 WiFi指纹定位方法确定用户位置。指纹定位的主要思路是，将采集到的周围接入点的接收信号强度指示集合（称为指纹）与预先构建的 WiFi指纹数据集进行匹配，从而估算位置 [22]。

在 [23], 中，作者提出了用于 WiFi 指纹定位的 DeepFi，这是一种基于深度学习的方法。该方法使用来自所有天线及其所有子载波的信道状态信息（CSI），并通过包含四个隐藏层的深度网络进行分析。通常，基于 CSI 的方法比基于 RSS 的方法更准确，因为它们使用信号的幅度和相位。然而，需要注意的是，现代智能手机无法提取 CSI；因此，基于 CSI 的方法似乎不适用于 ILBS。

与基于 CSI 的方法不同，如今的智能手机可以轻松计算 RSS；因此，大多数先前研究都聚焦于基于 RSS 的 WiFi 指纹定位。在这方面， [24] 中的作者利用基于 RSSI 相似度和空间位置的加权 K 近邻（WKNN）提高了 WiFi 指纹定位的准确率；而其他基于 K 近邻算法的方法，例如 [25], ，则基于欧几里得距离。近年来，深度学习方法也被用于 WiFi 指纹定位。在 [23] 中，有研究提出了一种用于 WiFi 指纹定位的深度学习模型。此外，一些研究将自编码器（AE）与其 DNN 模型结合，以提高定位精度。例如，在 [26], 中，作者提出了一种用于建筑物和楼层分类的 DNN 系统，并采用堆叠自编码器（SAE）来缩减特征空间、提高准确率。此外， [27] 中提出了另一种结合 SAE 的 DNN 架构，用于对建筑物ID、楼层ID和位置进行多标签分类。通常，DNN 模型会随着隐藏层数量的增加而达到更高的准确率。然而，更深的 DNN 模型会增加计算复杂度和计算时间。为解决上述问题， [28] 中提出了一种基于卷积神经网络（CNN）的结构，用于建筑物和楼层分类，该结构可降低复杂度，并减少模型对信号变化的敏感性。此外， [13], 中还提出了用于 WiFi 指纹定位的 CNNLoc 方法，这是一种包含 SAE 和一维卷积神经网络的基于 CNN 的模型。这些方法均基于云/服务器，所有数据处理都在第三方平台上完成。因此，用户与云之间的数据传输会带来隐私和安全问题，并造成显著的时间开销。相关工作的总结见表I。

总而言之，大多数研究的重点是实现尽可能高的准确率，这需要较高的计算能力和内存，而此类研究大多基于云/服务器平台。因此，这些基于云/服务器的模型不适合在资源受限的边缘设备上直接运行，也缺乏效率

表I 相关研究概述

研究	类别	方法	技术	属性和局限性
15,16,17	无设备	无线传感器网络（WSN）	物体出现时的信号变化位于节点之间	• 需要额外的基础设施 • 不适用于大型建筑物 • 需要视距通信 • 适用于单楼层建筑物。
18	无设备	RFID标签	将相位和接收信号强度指示信息输入粒子群优化算法以估计位置	
19	基于设备	WiFi信号	利用WiFi信号的TOA来估算位置	需要同步所有收发器之间
21	基于设备	基于CSI的Wi-Fi指纹	深度学习（DL）	• 基于CSI的方法比……更准确基于RSS • 现代智能手机和设备无法提取 CSI
22	基于设备	基于 RSSI 的 WiFi 指纹	加权 K 近邻(WKNN)	• 现代智能手机和设备支持接收信号强度 • 非边缘方法会导致隐私和安全问题以及时间开销
23	基于设备	基于 RSSI 的 WiFi 指纹	欧几里得距离	
13,24,25,26	基于设备	基于 RSSI 的 WiFi 指纹	深度学习（DL）	

，因为这些设备的功率、内存和计算资源有限。此外，室内定位还涉及其他一些关键因素，例如成本、延迟以及隐私/安全问题，这些因素在系统设计过程中都必须加以考虑。因此，本研究提出了一种新的基于边缘的WiFi指纹定位方法，以解决这些问题，同时满足多建筑物、多楼层场所中ILBS对尽可能高准确率的需求。

III. 所提方法

本节介绍一种新的 WiFi 指纹定位系统，用于估计人员在室内环境中的位置。所提出的架构基于卷积神经网络分类器，可识别个人在室内场所中的位置。所提方法的系统示意图如图2所示。如图所示，该定位系统由三个主要阶段组成，包括预处理、网络训练和训练后优化。以下将详细介绍各部分。

A. 数据预处理与数据集准备

The first step is to prepare the input data, i.e. the fingerprints (数据以便将其输入所提出的网络，并且只输入相关数据。为我们的网络准备数据主要分为以下三个阶段：

1）区域网格划分

通常，在大多数室内定位与导航服务应用中，并不需要用户的精确位置，只需知道用户所在的区域或范围即可 [1]。

这一事实促使我们从不同的角度审视室内定位问题。利用这一事实，我们旨在提升室内定位的性能，不仅关注准确率，也关注延迟、隐私以及前文提到的其他重要指标。

因此，为了提升定位性能，我们采用网格化技术[22]，将定位区域划分为若干边长为 L 的方形单元格。更重要的是，通过区域网格化，我们将目标区域划分为多个区域，从而将定位问题从回归问题（精确位置）转变为分类问题

问题（区域）。

这显著简化了定位过程，因此与精确定位相比，可以使用更简单的硬件来实现。

2）数据集与输入准备

通常，WiFi指纹数据集是基于精确位置采集的。如本研究所述，我们采用了区域网格化来提升定位性能。因此，需要将精确位置转换为基于区域的形式。为此，我们首先

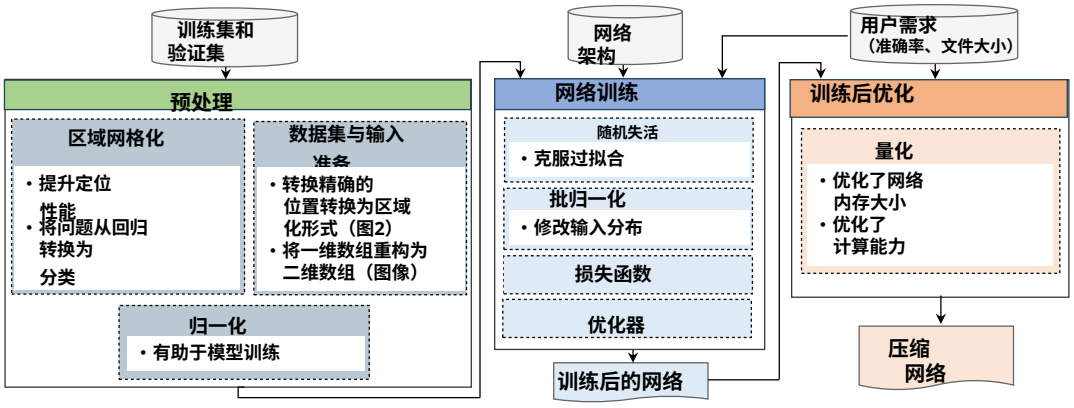


图2: 我们的定位系统示意图

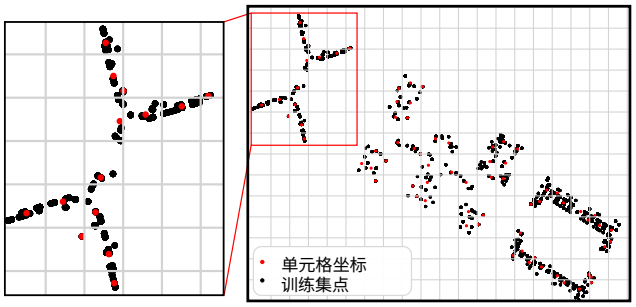


图 3： UJIIndoorLoc 数据集（训练集）的区域网格化，按 L 划分=20

将数据集区域划分为若干长度为 L 的区域，并为每个网格单元分配一个唯一坐标，以表示给定区域中的各个点位置。

UJIIndoorLoc 数据集的这一处理流程如图 3 所示。从图中可以明显看出，每个单元格包含多个样本，并假设这些样本的位置就是该单元格的坐标。对于每个单元格，通过对其中所有样本的位置取平均值，计算出该单元格唯一的坐标。图 3 中的黑点表示精确位置，红点表示每个单元格计算得到的位置。因此，每个区域中的所有点都具有相同的坐标，即红点所在的位置。

需要注意的是，在结果部分，我们计算每个点的精确位置（黑点）与预测位置（即对应单元格的位置，红点）之间的距离误差。

此外，本研究提出了一种基于 CNN 的模型用于 WiFi指纹定位，因此模型输入应为图像。在 WiFi指纹定位问题中，输入是一个 RSSI 值数组，因此我们需要将这个一维数组重构为二维数组（图像），以适配所提出的基于 CNN 的模型。因此，对于 UJIIndoorLoc 数据集，我们根据输入创建一个二维数组，输入是一个包含 520 个元素的向量。为此，我们首先在每个输入的末尾添加 9 个零元素，使其成为包含 529 个元素的向量，然后将输入数组重塑为 23×23 的二维数组，如图 4 所示。

3) 归一化

归一化技术旨在缩小输入分布，同时不丢失信息，并使模型训练更加容易。本文的输入数据是相邻接入点的接收信号强度指示器（RSSI）值，经归一化后，根据以下公式映射到 [0,1]：

$$RSSI_{inew} = \begin{cases} 0, & RSS_i > 0\ (no\ signal) \\ \dfrac{RSSI_i - RSSI_{min}}{-RSSI_{min}}, & otherwise \end{cases} \quad (1)$$

其中， RSS_{min} 是数据集中的最小值。例如，在 UJIIndoorLoc 数据集中，RSSI 值介于 $-104\ dbm$ 和 $0\ dbm$ 之间，因此 $RSS_{min} = -104$；故对于该数据集。

B. 网络架构

本文采用卷积神经网络（CNN）作为网络的基础，并部署卷积自编码器（CAE）以提升网络性能。下文将详细介绍所提出的网络，如图6所示。

1）卷积神经网络

如上所述，本文提出了一种轻量级网络，用于 WiFi指纹定位，力求实现尽可能高的准确率，并可在用户设备上运行。我们利用卷积神经网络（CNN）缩减输入规模，同时使数据更易于处理，且不会丢失重要特征。此外，与支持向量机（SVM）和K近邻算法（KNN）等其他机器学习方法相比，CNN网络对输入数据变化的敏感性较低，因此更加稳健 [28]。这是 WiFi指纹定位的一项关键特性，因为在室内环境中，信号强度很容易受到多种因素影响而发生变化，因素包括多径效应（主要因素）、电磁干扰，以及阻挡 Wi-Fi信号的临时障碍物 [29][30]。

如 III-A-2 小节所述，我们将输入数据从向量重塑为二维形式，因此可将输入视为灰度图像，即以图像形式表示的无线电地图，其中每个像素都等于来自不同接入点的接收信号强度指示。因此，所提出的基于卷积神经网络的网络能够从接收信号强度指示值（像素值）以及周围接入点的无线电地图（模式）中学习 [28]。

2）卷积自编码器

卷积自编码器（CAE）兼具卷积神经网络（CNN）和自编码器（AE）的特征。自编码器是一种无监督学习方法，通常通过在输出中重构输入数据来实现去噪、降维和特征提取。

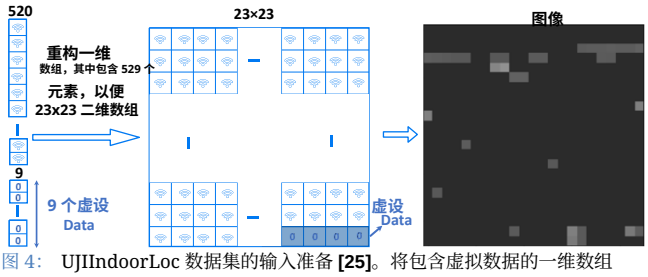


图 4： UJIIndoorLoc 数据集的输入准备 **[25]**。将包含虚拟数据的一维数组重构为二维数组。

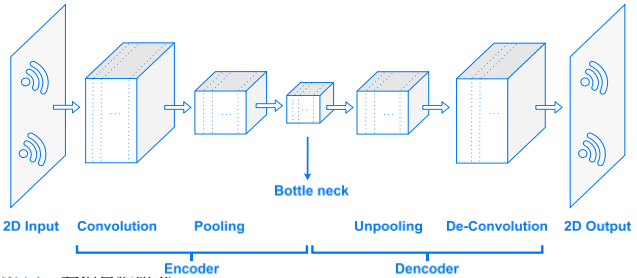
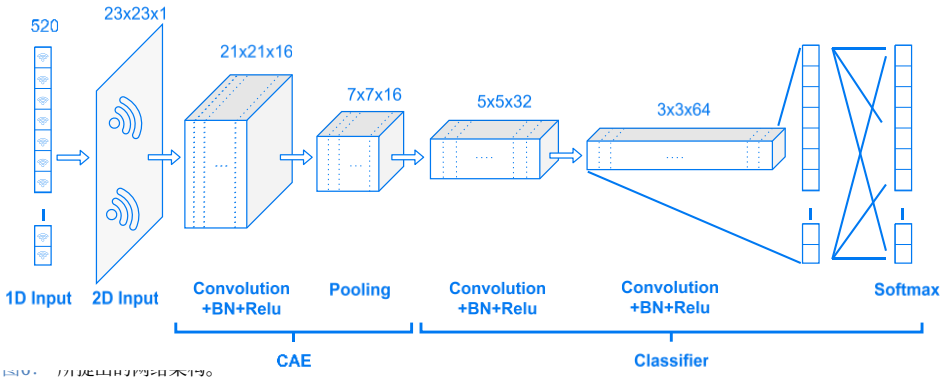


图 5： 卷积自编码器



因此，在 WiFi指纹定位中，这是一种合适的方法，可减小室内环境中由不同噪声源（如多径效应及前文提到的其他来源）引起的 RSSI值波动。

与包含若干全连接层的通用自编码器不同，卷积自编码器使用若干卷积层，从而减少网络参数，进而缩小网络大小。如图5所示，卷积自编码器包含两个主要部分：由若干卷积层和池化层组成的卷积编码器，以及与之配套的反卷积解码器，后者包含若干反卷积（转置卷积）层和反池化（上采样）层。

3）网络设计

所提出的网络架构如图6所示。该网络由输入、卷积自编码器和分类器三部分组成。输入是送入卷积自编码器层的灰度无线电地图图像；该层实际上是卷积自编码器的编码部分，用于压缩输入并提取其主要特征。随后，分类器用于识别用户所在的建筑物、楼层和位置。为提升该网络的性能，我们使用以下层：

- 丢弃层：随机失活通常用于解决过拟合问题，其主要思路是在每次训练迭代中随机舍弃一些单元；因此，网络不会针对训练集进行过度训练，从而避免过拟合。在我们提出的网络中，我们在最后一层之前使用这一技术，以避免过拟合。

- 批量归一化（BN）：该层通过调整每个小批量中各层的输入分布来解决内部协变量偏移问题；因此，使用批量归一化可以提高网络的收敛速度 [31]。在所提出的方法中，如图6所示，每个卷积层之后都使用 BN。

C. 训练后优化I

n this pe l ,

该模型经过所提出的 d optimi ation 因为它需要部署在内存和计算能力有限的智能手机或其他嵌入式设备上。较小的模型不仅占用手机更少的存储空间，运行时也会使用更少的随机存取存储器。

因此，其他应用程序可以使用更多内存，从而提升性能和稳定性。此外，延迟较低的模型运行速度更快，也会直接影响功耗。需要注意的是，训练后优化通常会降低模型准确率；因此，在设计过程中必须权衡准确率与模型大小或延迟。

训练后优化采用量化技术 [32]。量化是降低模型中数值（权重）精度，以减小模型大小并缩短计算时间。默认的数值类型为 float32。本文采用 float16 和 int8 量化，并评估它们对模型性能的影响，评估指标为模型大小和延迟。

IV. 模型评估与实验结果

本节将所提出的 CAE-CNNLoc 与最先进的方法进行比较，以评估其优越性。为考察所提方法的性能，我们将 CAE-CNNLoc 应用于 UJIIndoorLoc 数据集 [25]。此外，我们还将所提模型用于本系的室内定位，并通过采集名为 SBUK-D 的 WiFi 指纹数据集进行案例研究。另外，CAE-CNNLoc 网络模型采用 Tensorflow 2.4.0 框架，在配备 Tesla 4 GPU 的 Google Collaboratory Cloud 上实现，并在安卓智能手机上测试其性能。网络参数如下所示（表 II）。

本研究使用了三个不同的公共数据集，其详情如下所述。

UJIIndoorLoc：这是最常用于 WiFi指纹定位的数据集，包含 3 栋建筑物，每栋有 4 或 5 层，覆盖大学内 108,703 m^2 的区域，位于

表 II 一般网络参数

参数	数值	参数	数值
激活函数	Relu	优化器	Nadam
CNN损失函数	MSE	输出层激活函数	Sigmoid
	稀疏分类交叉熵		

西班牙 Jaume I 大学。该数据集包含训练集和测试集，分别有 19,938 个和 1,111 个样本，每个样本有 529 个特征。前 520 个特征表示来自不同接入点的接收信号强度指示值，范围为 -104 分贝毫瓦至 0 分贝毫瓦；空值以 100 表示，代表无法访问的接入点。每个样本的位置信息由建筑物ID、楼层ID以及以米为单位的经度和纬度值组成 [25]。

坦佩雷：该数据集采集自芬兰坦佩雷的一栋大学建筑。该建筑占地面积约为 22,750 m^2，共有五层，但仅使用其中四层创建此数据集。该建筑内共有 991 个接入点；因此，每个位置都会记录这些接入点的 RSSI 值。坦佩雷数据集中的位置信息包含 X、Y 和 Z，其中 Z 表示楼层 [33]。

UTSIndoorLoc：该数据集采集自悉尼科技大学（UTS）的 FEIT大楼。这栋建筑占地近 44,000 m^2，共有 18 层，其中 16 层用于创建 UTSIndoorLoc 数据集。位置信息包含 X、Y 和楼层 ID，并且每个样本还包含来自不同接入点的 590 个 RSSI 值 [13]。

A. 卷积自编码器和卷积神经网络优化

首先，我们研究所提出模型的各种结构，以找出最适合我们应用的模型。为此，我们测试了具有不同参数的卷积层和池化层等不同层，以获得最佳结构。需要注意的是，我们只研究较小的结构，因为我们的目标是构建一个轻量级网络。建议的结构如表 III 所示。该模型在卷积自编码器部分仅包含一个具有 16 个通道的卷积层，后接一个最大池化层；分类器部分则由两个卷积层组成，分别具有 32 和 64 个通道。

B. 区域网格化对定位性能的影响

本小节考察区域网格化对定位性能的影响。区域网格化的参数 L 表示定位区域中每个正方形的边长。表 IV 列出了不同 L 值下 CAE-CNNLoc 的结果，并对比了定位精度和模型大小。根据给出的 Data，L 为 = 7 的模型在位置预测中的误差最低，同时模型参数数量和大小

表III 建议的网络模型结构

	层 层特征		输出尺寸
	输入	-	23x23x1
CAE	二维卷积	f:16, k:3	21x21x16
	最大池化	p: 2	7x7x16
分类器	二维卷积	f:32, k:3	5x5x32
	二维卷积	f:64, k:3	3x3x64
	展平	-	576
	稠密的	-	类别数

F：滤波器尺寸，k：卷积核尺寸，p：池化尺寸

表IV 区域网格化对定位性能的影响

L	欧几里得 **Mean** 误差	建筑 命中率	楼层 命中率	**Num** 类别	**File Size (MB)**
1	10.369	0.994	0.892	1,834	1.04
3	10.024	0.994	0.892	1,496	0.88
5	10.364	0.995	0.901	1,217	0.72
7	9.524	0.994	0.905	823	0.50
10	12.285	0.995	0.900	766	0.46
20	10.774	0.995	0.915	285	0.19
30	11.703	0.996	0.917	259	0.18
50	15.511	0.996	0.919	107	0.09

也显著减少。这是合理的，因为增大 L 会减少类别数量，从而减少模型参数，进而缩小模型大小。此外，较大的 L 会形成覆盖更多点的较大正方形，而这些点的位置实际上都被视为该正方形的唯一坐标，因此会增大定位误差。因此，定位精度与模型大小之间存在权衡，我们可以根据应用场景调整 L，以实现最佳性能。

C. 与现有方法的比较

现在，为了评估 CAE-CNNLoc 方法的性能效率，我们将其与近期的一些研究进行了比较。一些现有研究仅关注建筑物准确率和楼层准确率，而本研究除建筑物准确率和楼层准确率外，还对定位平均误差进行了研究。表V报告了 CAE-CNNLoc 与一些相关方法的定位精度。可以观察到，所提出的方法达到了与其他方法相当的准确率。在这方面，CAE-CNNLoc 的楼层命中率为 0.90，定位平均误差为 9.52，最后，与其他一些方法一样，建筑物准确率约为 100%。

D. CAE-CNNLoc 的抗噪性

室内环境中的WiFi信号极易受到噪声影响；因此，RSSI值会根据III-B-1小节所述的不同条件轻易发生变化。因此，对于室内定位方法而言，抵御噪声至关重要，

表 V CAE-CNNLoc 与其他方法的比较

	欧几里得 平均误差	建筑物 命中率	楼层 命中率	模型
CAE-CNNLoc	9.52	0.994	0.905	卷积自编码器+2维卷积神经网络
CNNLoc [13]	11.78	1	0.96	SAE+1维卷积神经网络
可扩展深度神经网络 [27]	9.27	1	0.93	SAE+DNN
二维卷积神经网络 [28]	-	0.958*		二维卷积神经网络
SAE+分类器 [26]	-	0.911*		SAE+DNN
基线 [25]	7.9	0.899*		KNN
RTLS@UM [34]	6.2	1	0.94	滤波 +KNN

*建筑物和楼层类别合并

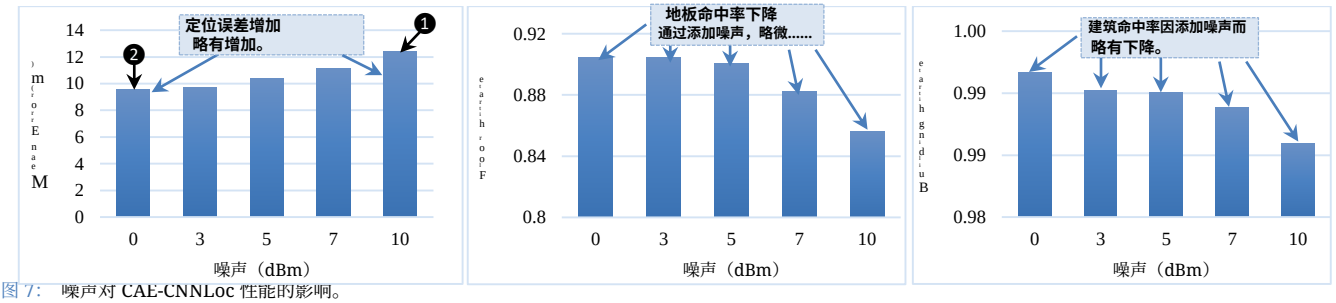


图 7： 噪声对 CAE-CNNLoc 性能的影响。

在本小节中，我们通过向数据添加噪声来检验模型的性能。为此，我们随机向测试数据添加 3、5、7 和 10 分贝毫瓦的噪声，并评估模型的准确率。如图 7 所示，向信号添加噪声会降低定位精度，这是合理的；见 ❶ 和 ❷。但必须指出，在最差情况下，即添加 10 分贝毫瓦噪声时，CAE-CNNLoc模型的误差为 12.4 米，与 0 分贝毫瓦噪声相比，误差增加了约 3 米。因此，这些结果充分验证了CAE-CNNLoc模型中卷积自编码器部分的去噪能力。

E. CAE-CNNLoc 在安卓智能手机上的性能如前所述，本文的主要目的是提出一种可在用户设备上运行的设备端室内定位模型。因此，我们在真实世界中使用安卓智能手机考察了 CAE-CNNLoc 模型的性能。为此，我们在 Redmi Note 8 上测试了所提出的 CAE-CNNLoc 模型，并报告了测试结果。作为设备端实现中的两个关键因素，我们考虑了推理时间（延迟）和模型大小，并报告了不同技术所产生的影响。Redmi Note 8 搭载高通骁龙 665芯片组，配备主频最高可达 2.0 GHz 的八核中央处理器，并配有 4GB 随机存取存储器。

区域网格化：首先，我们考察了区域网格化方法的效果；为此，图8展示了所提方法在不同 L 值下的结果。从 IV-B 小节可以看出，增大 L 会减少类别数量，从而显著减少网络参数，进而降低延迟和网络大小，如图8所示。例如，设置 L=10 时，与 L=1 相比，延迟可降低约 2 倍（50 μs），网络大小可缩减至原来的二分之一以下（460 KB），而定位精度仅下降约 2 m。因此，区域网格化方法适用于设备端 WiFi指纹定位。

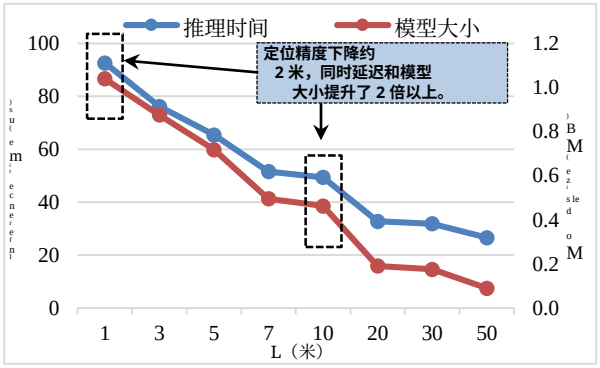


图8： 区域网格化对定位推理时间和模型大小的影响。

量化：本文采用了 float16 和 int8 量化，结果如图 9 所示。首先，使用 float16 量化后，权重类型由 float32 变为 float16。显然，从

如图 9 所示，对于 float16 量化，尽管网络大小缩小了约 2 倍，但定位精度和推理时间均未受到影响。此外，int8量化能够将网络大小缩小超过 3.9 倍，并使延迟提升近 2 倍，同时不会显著降低定位精度。

F. CAE-CNNLoc 可扩展性

在本小节中，我们将 CAE-CNNLoc 模型应用于另外两个公开的 WiFi 指纹定位数据集，以评估所提模型的可扩展性。表 VI 展示了

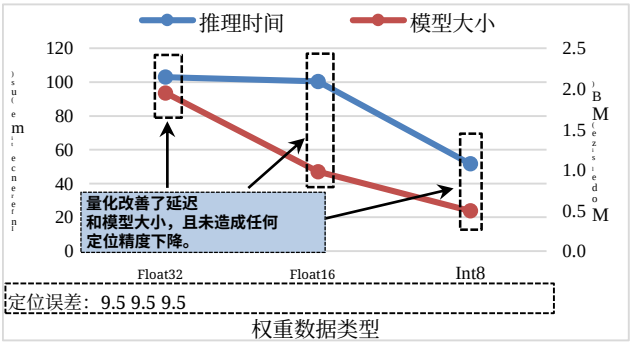


图 9： 量化对定位误差、延迟和模型大小的影响。

表 VI CAE-CNNLoc模型的可扩展性

	CAE-CNNLoc的最佳结果			其他基于DNN的研究中的最佳结果		
数据集	**Mean** 错误	建筑物命中率	楼层命中率	**Mean** 误差	建筑物命中率	楼层命中率
UJIInd 室内定位	9.5	0.994	0.905	9.27	1	0.93
坦佩**re**	10.24	-	0.889	10.88	-	0.94
UTSIn 室内定位**c**	7.7	-	0.92	7.60	-	0.946

*坦佩雷的 UJIIndoorLoc 和 [13] 的 UTSIndoorLoc 中的最佳结果 [27],

与近期研究相比，CAE-CNNLoc 在坦佩雷和 UTSIndoorLoc 数据集上的性能。从该表可以清楚地看出，我们的模型性能良好，在上述数据集上的表现接近其他基于DNN的研究中的最佳水平，证明了 CAE-CNNLoc 的可扩展性。

V. 讨论

如前所述，在本研究中，我们将定位问题从回归问题转化为分类任务，并利用区域网格化来定义类别。这意味着每个单元格（由网格化技术生成）都被视为一个类别，且该单元格中的所有点都具有相同的坐标，因为大多数基于位置的服务并不真正需要精确位置。

为此，我们对训练集进行了区域网格化。然而，我们的分析发现，测试集中有许多点在区域网格化时未落入已定义的网格单元。这意味着这些点属于模型在训练期间未见过的某些类别，而这也是我们提出的基于分类的方法中定位误差增大的原因之一。如图10所示，测试集中的许多点（红点）位于独立的网格单元内，而训练集中的点（蓝点/黑点）并未覆盖这些单元；这意味着在训练阶段，模型不会将这些网格单元视为独立类别。由于本研究将定位问题从回归转化为分类任务，因此需要一个符合分类问题标准的数据集。为此，我们对数据集进行了一些修改，使其更符合分类任务的要求。在数据集与输入准备阶段（第 III-A 小节），我们首先合并训练集和测试集；随后将合并后的数据随机划分为训练集、验证集和测试集，再使用网格化方法定义网格单元。通过

表VII 建议的网络模型结构

	Mean 误差	建筑物命中率	楼层命中率
UJIIndoorLoc	2.36	0.9992	0.993
坦佩雷	6.30	-	0.961
UTSIndoorLoc	1.72	-	0.996
SBUK-D	1.73	-	0.985

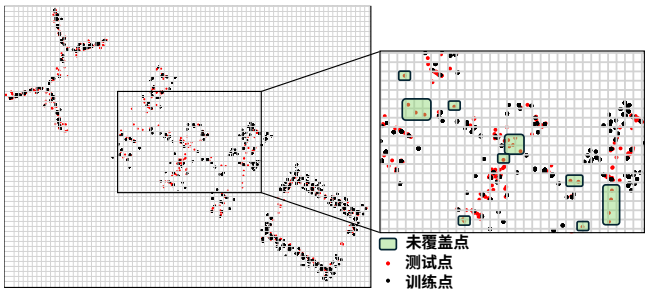


图10： 测试集中有若干孤立点（红点），训练集中的点（黑点）未覆盖这些位置。

这样做，我们可以确保模型在训练期间见到所有可能的类别（网格单元），并为其提供候选样本，从而显著降低定位误差，表 VII 报告了这一结果。显而易见，对于 UJIIndoorLoc，楼层命中率提高至 99%，建筑命中率接近 100%，更重要的是，欧几里得误差降至 2.36 米。

A. 在我们的数据集上进行实验（SBUK-D数据集）

在本小节中，我们评估所提出的网络模型在本系进行定位的性能。为此，我们生成了名为 SBUK-D 的 WiFi指纹定位数据集。该数据集的详情如下：

SBUK-D数据集：作者在克尔曼沙希德·巴霍纳尔大学（SBUK）工程系采集了该数据集。该建筑物共有3层，覆盖面积近11,500 m^2 。位置信息包括楼层ID、X坐标和Y坐标；建筑物内有198个不同的接入点，用于记录各位置的RSSI值。记录的RSSI范围为-100分贝毫瓦至0分贝毫瓦，与其他数据集一样，无法访问的接入点数值设为100。该数据集包含使用4部Android智能手机在70个不同位置采集的2292个样本。

表VII还报告了CAE-CNNLoc在SBUK-D数据集上的结果，并与三个不同的数据集进行了比较。从表中可以清楚地看出，CAE-CNNLoc在预测SBUK-D数据集的楼层时，误差仅为1.73米，准确率超过98%。此外，针对我们的数据集，模型大小约为103 KB，在Android手机上的推理时间为198 µs。

VI. 结论

本文提出了一种名为CAE-CNNLoc的轻量级WiFi指纹定位方法，用于估计用户在室内环境中的位置。CAE-CNNLoc由卷积神经网络（CNN）与卷积自编码器（CAE）结合而成，可显著降低输入维度以及模型对

输入波动。所提出的模型可以轻松部署在边缘设备（例如智能手机）上，从而在延迟和用户隐私方面显著提升定位性能。实验结果表明，考虑到该模型……，其性能优于其他研究。

是一种基于边缘的方法。对于 UJIIndoorLoc 数据集，采用 int8量化的模型大小仅为 0.5 MB，推理时间约为 51 µs，定位误差为 9.5 米。尽管其定位性能与其他研究相当，但这是一个基于边缘的系统，因此能够在定位过程中消除云端/服务器缺陷。此外，所提出的模型在我们名为 SBUK-D 的新数据集上表现出显著性能：定位误差为 1.73 米，楼层检测准确率为 98%，推理时间为 198 µs，模型大小为 103 KB。

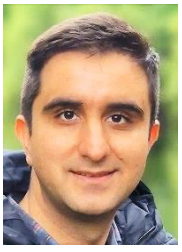
致谢

本研究部分得到了纽约大学阿布扎比分校城市交互网络中心（CITIES）的支持。该中心由 Tamkeen 通过纽约大学阿布扎比分校研究院CG001奖项资助；本研究还得到了人工智能与机器人中心（CAIR）的支持，该中心由 Tamkeen 通过纽约大学阿布扎比分校研究院CG010奖项资助。此外，第一作者谨此说明，本研究是在其任职于伊朗克尔曼沙希德·巴霍纳尔大学可靠与智能系统实验室期间开展的。

参考文献

[1] Basiri 等，“室内位置服务：挑战、需求与现有解决方案的可用性”，Comput. Sci. Rev.，第24卷，第1–12页，2017年5月，doi: 10.1016/j.cosrev.2017.03.002。[2] J. Xiao、Z. Zhou、Y. Yi 和 L. M. Ni，“从设备视角审视无线室内定位”，ACM Computing Surveys，第49卷，第2期。美国计算机协会，第1–31页，2016年6月。doi: 10.1145/2933232。[3] Q. Niu、M. Li、S. He、C. Gao、S. H. Gary Chan 和 X. Luo，“资源高效的自动化基于图像的室内定位”，ACM Trans. Sens. Networks，第15卷，第2期，第1–31页，2019年2月，doi: 10.1145/3284555。[4] M. Saadi、Z. Saeed、T. Ahmad、M. K. Saleem 和 L. Wuttisittikulkij，“采用 k 均值聚类和线性回归的基于可见光的室内定位”，Trans. Emerg. Telecommun. Technol.，第30卷，第2期，文章编号 e3480，2019年2月，doi: 10.1002/ett.3480。[5] L. Ciabattoni 等，“融合智能手机模型化行人航位推算与 BLE 信标的实时室内定位”，J. Ambient Intell. Humaniz. Comput.，第10卷，第1期，第1–12页，2019年1月，doi: 10.1007/s12652-017-0579-0。[6] H. Zhang、K. Liu、F. Jin、L. Feng、V. Lee 和 J. Ng，“Wi-Fi环境中基于距离拟合和指纹映射的可扩展室内定位算法”，Neural Comput. Appl.，第32卷，第9期，第5131–5145页，2020年5月，doi: 10.1007/s00521-018-3961-8。[7] J. J. Pérez-Solano、S. Ezpeleta 和 J. M. Claver，“利用UWB信号和未同步设备通过到达时间差进行室内定位”，Ad Hoc Networks，第99卷，文章编号 102067，2020年3月，doi: 10.1016/J.ADHOC.2019.102067。[8] A. A. Nazari Shirehjini 和 S. Shirmohammadi，“采用卡尔曼滤波和Tukey平滑提升基于HF-RFID的室内定位的准确率与鲁棒性”，IEEE Trans. Instrum. Meas.，第69卷，第11期，第9190–9202页，2020年11月，doi: 10.1109/TIM.2020.2995281。[9] A. Kargar-Barzi 和 A. Mahani，“抗障碍混合定位算法”，IET Wirel. Sens. Syst.，第10卷，第5期，第242–252页，2020年10月，doi: 10.1049/IET-WSS.2020.0052。[10] S. He 和 S. H. G. Chan，“基于Wi-Fi指纹的室内定位：最新进展与比较”，IEEE Commun. Surv. Tutorials，第18卷，第1期，第466–490页，2016年1月，doi: 10.1109/COMST.2015.2464084.[11] S. Khandker、J. Torres-Sospedra 和 T. Ristaniemi，《指纹定位中接收信号强度量化分析》，Sensors，第20卷，第11期，第3203页，2020年6月，doi: 10.3390/s20113203。[12] F. Zafari、A. Gkelias 和 K. K. Leung，《室内定位系统与技术综述》，IEEE Commun. Surv. Tutorials，第21卷，第3期，第2568–2599页，2019年，doi: 10.1109/COMST.2019.2911558。[13] X. Song 等，《一种基于Wi-Fi指纹的全新卷积神经网络室内定位框架》，IEEE Access，第7卷，第110698–110709页，2019年，doi: 10.1109/ACCESS.2019.2933921。[14] D. J. Suroso、P. Cherntanomwong 和 P. Sooraksa，《基于深度学习的指纹室内定位》，美国计算机协会国际会议论文集，第146–152页，2023年3月，doi: 10.1145/3592307.3592330。[15] Q. Wang 等，《DarLoc：基于深度学习和数据特征增强的鲁棒磁场室内定位》，Expert Syst. Appl.，第244卷，第122921页，2024年6月，doi: 10.1016/J.ESWA.2023.122921。[16] F. Alhomayani 和 M. H. Mahoor，《基于指纹的室内定位深度学习方法：综述》，Journal of Location Based Services，第14卷，第3期。Taylor and Francis Ltd.，第129–200页，2020年7月。doi: 10.1080/17489725.2020.1817582。[17] D. Ciuonzo 和 P. Salvo Rossi，《通过广义局部最优方法对非合作目标进行分布式检测》，Inf. Fusion，第36卷，第261–274页，2017年7月，doi: 10.1016/j.inffus.2016.12.006。[18] Y. Guo、K. Huang、N. Jiang、X. Guo、Y. Li 和 G. Wang，《一种用于基于RSS的无设备定位与跟踪的指数-瑞利模型》，IEEE Trans. Mob. Comput.，第14卷，第3期，第484–494页，2015年3月，doi: 10.1109/TMC.2014.2329007。[19] L. Zhao、H. Huang、X. Li、S. Ding、H. Zhao 和 Z. Han，《一种采用卷积自编码器的精确鲁棒无设备定位方法》，IEEE Internet Things J.，第6卷，第3期，第5825–5840页，2019年6月，doi: 10.1109/JIOT.2019.2907580。[20] L. Ma、M. Liu、H. Wang、Y. Yang、N. Wang 和 Y. Zhang，《WallSense：利用安装于墙面的UHF RFID标签进行无设备室内定位》，Sensors，第19卷，第1期，第68页，2018年12月，doi: 10.3390/s19010068。[21] H. Zhang、S. Y. Tan 和 C. K. Seow，《通过MRMSC-PHD滤波器，利用不准确的楼层平面图进行基于TOA的室内定位与跟踪》，IEEE Sens. J.，第19卷，第21期，第9869–9882页，2019年11月，doi: 10.1109/JSEN.2019.2926433。[22] F. Gu 等，《利用空间上下文改进室内定位——综述》，ACM Comput. Surv.，第52卷，第3期，第1–35页，2019年7月，doi: 10.1145/3322241。[23] X. Wang、L. Gao、S. Mao 和 S. Pandey，《用于室内定位的基于CSI的指纹定位：一种深度学习方法》，载于IEEE Transactions on Vehicular Technology，电气与电子工程师协会，2017年1月，第763–776页。doi: 10.1109/TVT.2016.2545523。[24] B. Wang 等，《一种基于RSS相似度和位置距离的全新加权KNN算法，用于Wi-Fi指纹定位》，IEEE Access，第8卷，第30591–30602页，2020年，doi: 10.1109/ACCESS.2020.2973212。[25] J. Torres-Sospedra 等，《UJIIndoorLoc：用于基于WLAN指纹的室内定位问题的全新多建筑多楼层数据库》，载于IPIN 2014——2014年国际室内定位与室内导航会议，电气与电子工程师协会，2014年，第261–270页。doi: 10.1109/IPIN.2014.7275492。[26] M. Nowicki 和 J. Wietrzykowski，《利用深度学习和Wi-Fi指纹实现低成本地点识别》，载于Advances in Intelligent Systems and Computing，Springer Verlag，2017年，第575–584页。doi: 10.1007/978-3-319-54042- 9 57。[27] K. S. Kim、S. Lee 和 K. Huang，《一种基于Wi-Fi指纹的可扩展深度神经网络架构，用于多建筑多楼层室内定位》，arXiv，第3卷，第

1. arXiv，第1–17页，2017年12月。doi: 10.1186/s41044-018-0031-2。[28] J. W. Jang 和 S. N. Hong，《利用卷积神经网络和WiFi指纹进行室内定位》，载于国际泛在与未来网络会议，ICUFN，IEEE计算机学会，2018年8月，第753–758页。doi: 10.1109/ICUFN.2018.8436598。[29] G. Zhang、P. Wang、H. Chen 和 L. Zhang，《利用卷积神经网络和高斯过程回归进行无线室内定位》，Sensors，第19卷，第11期，第2508页，2019年5月，doi: 10.3390/s19112508。[30] Z. Liu、B. Dai、X. Wan 和 X. Li，《一种基于卷积神经网络的混合无线指纹室内定位方法》，Sensors，第19卷，第20期，第4597页，2019年10月，doi: 10.3390/s19204597。[31] S. Ioffe 和 C. Szegedy，《批归一化：通过减少内部协变量偏移加速深度网络训练》，载于第32届国际机器学习会议论文集，F. Bach 和 D. Blei 编，机器学习研究论文集，第37卷。法国里尔：PMLR，2015年，第448–456页。[32] S. Han、H. Mao 和 W. J. Dally，《深度压缩：通过剪枝、训练量化和霍夫曼编码压缩深度神经网络》，载于第4届国际学习表征会议（ICLR 2016）会议论文集，国际学习表征会议（ICLR），2016年10月。[33] E. Lohan、J. Torres-Sospedra、H. Leppäkoski、P. Richter、Z. Peng 和 J. Huerta，《用于室内定位的Wi-Fi众包指纹数据集》，Data，第2卷，第4期，第32页，2017年10月，doi: 10.3390/data2040032。[34] A. Moreira、M. J. Nicolau、F. Meneses 和 A. Costa，《真实世界中的Wi-Fi指纹定位——RTLS@UM参加EvAAL竞赛》，载于2015年国际室内定位与室内导航会议（IPIN 2015），电气与电子工程师协会，2015年12月。doi: 10.1109/IPIN.2015.7346967。



Amin Kargar Barzi 是爱尔兰科克大学 Tyndall National Institute 的博士生。此前，他曾在伊朗克尔曼沙希德·巴霍纳尔大学（SBUK）的可靠与智能系统（RSS）实验室担任研究人员。他在 SBUK 获得电气工程理学学士学位和电子工程理学硕士学位。他的研究兴趣包括机器学习、深度学习、微型机器学习、边缘人工智能、机器视觉和无线智能系统。



Ebrahim Farahmand（S'20）是伊朗克尔曼沙希德·巴霍纳尔大学（SBUK）电气工程系的副研究员。他于 2012 年获得电气工程（通信系统）理学学士学位，并于 2016 年获得电气工程（电子学）理学硕士学位，均毕业于克尔曼沙希德·巴霍纳尔大学（SBUK）。自 2016 年至今，他一直在可靠与智能系统（RSS）实验室，SBUK，伊朗。他的研究兴趣包括类脑计算、深度学习、微型机器学习、近似计算、机器学习加速器、容错设计和网络系统。



Nooshin Taheri Chatrudi 是亚利桑那州立大学（ASU）健康解决方案学院的一名博士生。目前，我在亚利桑那州立大学（ASU）的嵌入式机器智能实验室（EMIL）攻读博士学位。我获得了通信工程理学硕士学位，研究方向为机器学习算法，毕业于克尔曼沙希德·巴霍纳尔大学（SBUK）。我的研究兴趣包括机器学习、临床信息学和健康监测系统开发。



Ali Mahani 于2001年获得伊朗克尔曼沙希德·巴霍纳尔大学的电子工程理学学士学位，随后于2003年和2009年分别获得伊朗德黑兰伊朗科技大学（IUST）的电子工程硕士学位和博士学位。此后，他一直在克尔曼沙希德·巴霍纳尔大学电气工程系任职，目前是副教授。他的研究兴趣主要集中在容错设计、基于FPGA的加速器、近似数字电路、随机计算和网络化系统。



Muhammad Shafique（M'11 - SM'16）于2011年获得德国卡尔斯鲁厄理工学院（KIT）的计算机科学博士学位。此后，他在KIT创立并领导了一个备受认可的研究团队数年，并在全球范围内开展了具有重要影响力的合作研发活动。2016年10月，他加入维也纳工业大学（TU Wien）信息学院的计算机工程研究所，位于维也纳，奥地利，担任计算机体系结构与稳健、节能技术领域的正教授。自2020年9月起，Shafique博士任职于纽约大学（NYU），目前担任正教授，并在阿联酋的纽约大学阿布扎比分校担任 eBrain 实验室主任，同时担任美国纽约大学纽约市校区坦登工程学院的全球网络教授。他还作为共同首席研究员/研究员参与多个 NYUAD中心，包括人工智能与机器人中心（CAIR）、网络安全中心（CCS）、交互式城市网络中心（CITIES）以及量子与拓扑系统中心（CQTS）。

他的研究兴趣包括人工智能与机器学习硬件及系统级设计、类脑计算、量子机器学习、认知自主系统、可穿戴医疗、节能系统、稳健计算、硬件安全、新兴技术、现场可编程门阵列、多核片上系统和嵌入式系统。他的研究特别关注计算与存储系统的跨层分析、建模、设计和优化。相关技术与工具已应用于物联网（IoT）、智能网络物理系统（CPS）以及发展信息与通信技术（ICT4D）领域的应用场景。Shafique博士曾在多个顶级会议上作主旨演讲、特邀报告和教程，并在这些会议上组织了许多专题会议。他曾担任多个知名 IEEE/ACM 会议的程序委员会主席、大会主席、分会主席和程序委员会委员。Shafique博士持有一项美国专利，并（共同）撰写了 6 本图书、 10+ 个书籍章节、 300+ 篇发表于顶级期刊和会议的论文，以及 50+ 篇档案文章。他曾获2015年 ACM/SIGDA 杰出新进教师奖、2020年和2022年 AI 2000芯片技术最具影响力学者奖、2021年 ASPIRE AARE卓越研究奖、六枚金牌，以及多个知名会议的最佳论文奖和提名。他是 IEEE 和 IEEE信号处理学会（SPS）的高级会员，也是 ACM、ACM计算机体系结构特别兴趣小组、ACM设计自动化特别兴趣小组、ACM嵌入式系统特别兴趣小组和欧洲高性能嵌入式架构与编译器网络的成员。