

WiFi信号传播与无线AP定位研究

陈丽琴¹, 高凌宇¹, 杨晶晶², 黄 铭²

¹怒江州工业和信息化局, 云南 六库

²云南大学无线创新实验室, 云南 昆明

Email: huangming@ynu.edu.cn

收稿日期: 2021年4月2日; 录用日期: 2021年4月17日; 发布日期: 2021年4月28日

摘 要

由于泛在、传输速率高和免费等特点, WiFi是最受欢迎的上网方式, 无线AP也是目前世界上数量最多的在用无线电收发设备。因此, 研究WiFi信号传播与无线AP定位是掌握无线电监测技术基础的理想实现平台。本文基于电磁场仿真软件WinProp和人工智能AI算法, 结合一栋单层别墅的无线电传播特性研究了无线AP定位技术。仿真结果表明: 采用卷积神经网络CNN、K邻近算法KNN和支持向量机SVM, 监测接收点的数目为3时, 识别无线AP所在房间的准确率分别为76.7%、76.3%和70.0%; 数目为5时, 识别准确率分别为83.6%、91.4%和87.9%; 数目为7时, 准确率分别为92.6%、98.1%和96.8%。上述工作对理解非视通环境下的无线电监测定位原理有重要意义。

关键词

WiFi, 非视通定位, 无线电监测

Study of WiFi Signal Propagation and the Location of the Wireless AP

Liqin Chen¹, Lingyu Gao¹, Jingjing Yang², Ming Huang²

¹Nujiang Prefecture Bureau of Industry and Information Technology, Liuku Yunnan

²Wireless Innovation Lab of Yunnan University, Kunming Yunnan

Email: huangming@ynu.edu.cn

Received: Apr. 2nd, 2021; accepted: Apr. 17th, 2021; published: Apr. 28th, 2021

Abstract

Due to its ubiquity, high transmission rate, and free of charge, WiFi is the most popular way to

access the Internet, and wireless AP is currently the world's largest number of radio transceiver devices in use. Therefore, studying WiFi signal propagation and wireless AP positioning is an ideal platform for mastering the basics of radio monitoring technology. Based on the electromagnetic field simulation software WinProp and artificial intelligence algorithm, combined with the radio propagation characteristics of a one-story villa, the wireless AP positioning technology is studied in this paper. The simulation results show that: using convolutional neural network, K neighboring algorithm and support vector machine, when the number of monitoring receiving points is 3, the accuracy rates of identifying wireless AP are 76.7%, 76.3% and 70.0% respectively; when the number is 5, the recognition accuracy rates are 83.6%, 91.4%, and 87.9%, respectively; when the number is 7, the accuracy rates are 92.6%, 98.1%, and 96.8%, respectively. The above work is of great significance for understanding the principle of radio monitoring and positioning in a Non Line of Sight environment.

Keywords

WiFi, Non Line of Sight Location, Radio Monitoring

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

无线电频谱是连接陆、海、空、天和网络空间的媒介，是国家重要的战略性稀缺资源。由于无线电波传播的开放性，无线电业务容易受到干扰，因此世界各国都对无线电使用有着严格的规定，并指定由专门的职能部门进行监测和管理。伴随数字经济、新基建和工业互联网等国家战略的深入实施，以 5G、物联网等为代表的新一代信息通信技术与实体经济深度融合的步伐不断加快，无线电新技术、新业务、新应用大量涌现，成为各行业各领域加速数字化转型的重要引擎。与此同时，干扰事件逐年增多，电磁环境越来越复杂，无线电安全问题也受到越来越多的关注[1]。

众所周知，由于泛在、传输速率高和免费等特点，WiFi 是最受人们欢迎的上网方式，也是物联网中传感器信息传输的基础。由于公共 WiFi 环境缺少甚至毫无安全防护措施，WiFi 安全问题十分突出，主要的包括：WiFi 路由器 DNS 遭篡改导致上网时广告弹窗或在购物时进入钓鱼网站；在公共场所使用免费 WiFi 时进入黑客设置的陷阱或被黑客监听，从而使上网记录、账号密码等关键信息被窃取；用户设置 WiFi 密码过于简单，黑客可瞬间暴力破解，并进行蹭网或盗取信息等[2]。WiFi 无线网络在给企业的生产和工作都带来了极大便利的同时，也引发了许多安全方面的问题，为此文中提出了可能采取的措施[3]，例如，对 WiFi 无线网络的访问进行严格的控制；通过与 MAC 地址进行绑定，对接入 WiFi 网络中的非法者进行防范；提高无线设备的加密与校验技术；避免受到 DOS 的攻击对站点定期加以审查；对核心网络进行隔离；强化对链路的监测与保护，从而提高其安全性，确保用户的信息数据和通信传输安全。同时，由于 WiFi 无线网络无所不在，利用 WiFi 信道状态信息 CSI (Channel State Information) 的幅度和相位特征可进行人数统计，在室内环境下进行下我们实验室采用支持向量机 SVM (Support Vector Machine)、神经网络 NN (Neural Networks) 和卷积神经网络 CNN (Convolutional Neural Networks) 进行室内人数的统计分类，结果表明三种机器学习算法分别达到了 94%、96% 和 88% 的准确率[4]。WiFi 无线网络的大量使用，在方便人们的同时也可能暴露用户的隐私[5]。因此，本文以 WiFi 信号传播和无线 AP

(Wireless Access Point)定位为研究对象, 不仅有理论意义, 而且对无线电监管工作有应用价值。论文的主要工作包括仿真模型、WiFi 信号传播特性仿真和数据预处理、基于 AI (Artificial Intelligence)的发射源定位研究, 在发射源定位研究部分, 论文采用卷积神经网络 CNN、K 邻近算法 KNN (K-Nearest Neighbor) 和支持向量机 SVM 对 WiFi 信号发射源进行定位, 并讨论了监测接收点数量对定位准确率的影响。

2. 仿真模型

本文假设在一栋单层别墅的各个房间中分别放置无线 AP (WiFi 信号发射源, 发射功率为 10W, 天线为偶极子), 研究了无线 AP 的发射特性和定位方法。图 1 为单层别墅房外观示意图, 图中房子外围白色柱子为放置的监测接收点, 高度为 1.5 m, 接收天线为偶极子。图 2 为房间布局平面示意, 图中房间内 Site1、Site2、Site3 表示三个房间内放置的无线 AP 的典型位置, 外围红点为监测接收点, 与图 1 中的白色柱子对应。



Figure 1. Schematic graph of the appearance of a one-story villa

图 1. 单层别墅的外观示意图

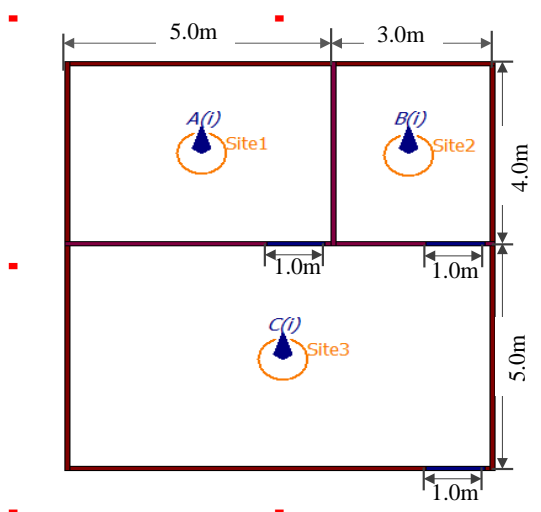


Figure 2. Floor plan of wireless AP and receiving point

图 2. 无线 AP 及接收点位置平面图

3. WiFi 信号传播特性仿真和数据预处理

WiFi 信号传播特性仿真采用 WinProp 软件[6], 仿真时假设无线 AP 在室内房间内按照从左到右, 从上到下的路径共移动 5001 次, 每次移动时进行场分布计算, 并在监测接收点接收和记录 WiFi 信号场强, 针对每个房间的仿真将获得 8×5001 组场强数据, 为进行发射源定位准备数据。图 3(a)-(c)分别是在 3 个房间发射 WiFi 信号时得到的仿真区域电场强度分布, 仿真时截取的平面高度为 1.5 m; 图 3(d)是在截取高度为 1.5 m、2 m 和 2.5 m 的平面上得到的场强分布仿真结果。可以看出, 信号最强区域位于无线 AP 所属房间, 靠近该房间的监测接收点测量到的场强次之; 当无线 AP 在房间 1 时, 位于左上角的监测接收点接收到的场强较强, 而距离其较远的右下角的监测接收点较弱; 同理, 当无线 AP 在房间 2 和 3 发射时, 位于右上侧和下方的监测接收点接收到的场强较强; 当无线 AP 在各个房间内移动时, 随着位置的变化各个监测接收点收到的信号将发生变化。因此, 基于接收场强对非视通环境下的无线 AP 进行定位是可能的。假设每个房间内的无线 AP 在 5001 个不同的位置处进行发射, 房子外围设置 8 个监测接收点进行接收时获得的接收功率数据矩阵为:

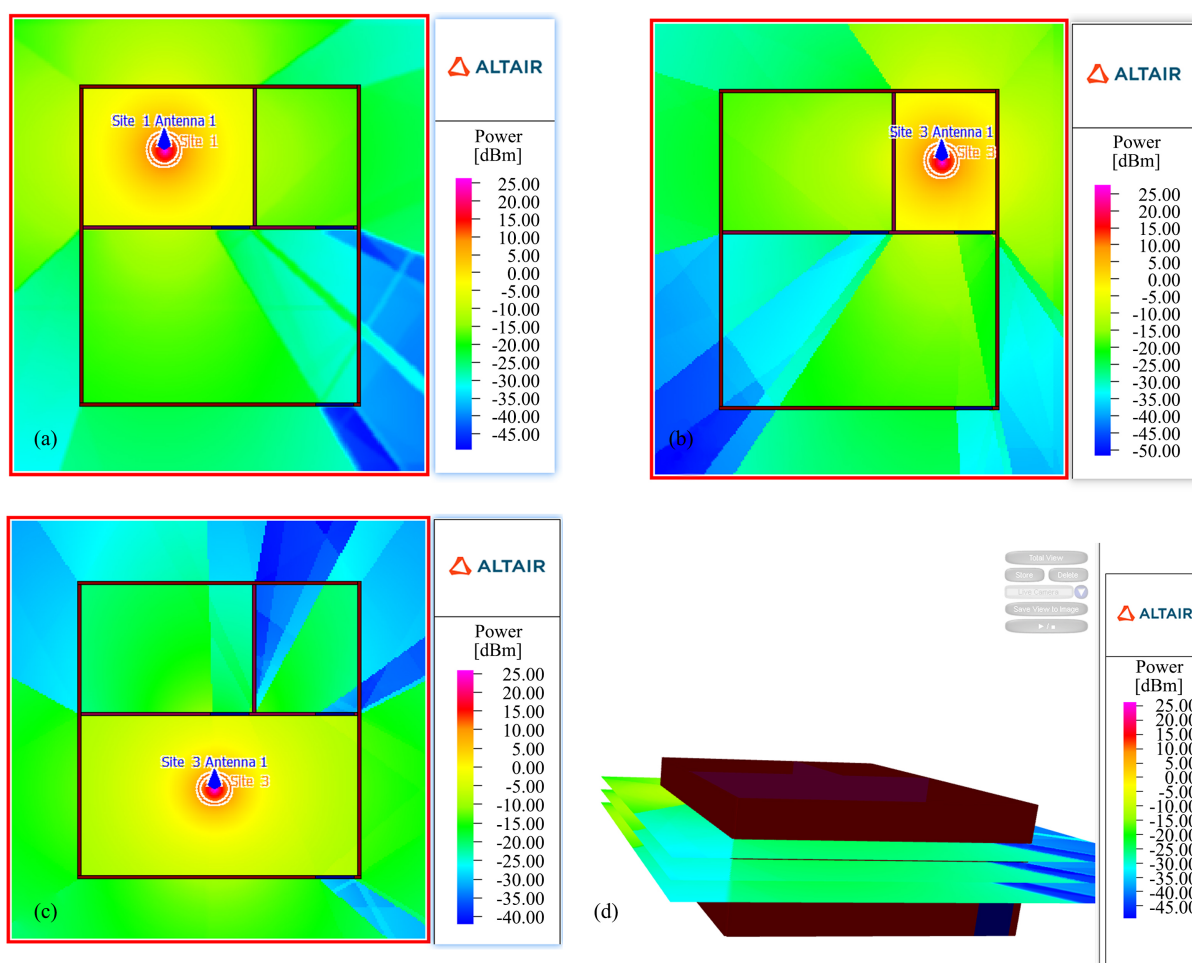


Figure 3. The power distribution inside and outside the villa and at different heights when WiFi signals are transmitted in 3 rooms respectively. (a) Wireless AP is located in room 1; (b) Wireless AP is located in room 2; (c) Wireless AP is located in room 3; (d) The distribution of received power inside and outside the villa under different receiving antenna heights

图 3. 分别在 3 个房间发射 WiFi 信号时别墅内外及其不同高度下的功率分布。(a) 无线 AP 位于房间 1; (b) 无线 AP 位于房间 2; (c) 无线 AP 位于房间 3; (d) 不同接收天线高度下别墅内外的接收功率分布

$$P_k = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,8} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,8} \\ \vdots & \vdots & \cdots & \vdots \\ p_{5001,1} & p_{5001,2} & \cdots & p_{5001,8} \end{bmatrix}, k=1,2,3 \quad (1)$$

上式中, $k=1,2,3$ 对应模型中的三个房间; 该矩阵的每个元素 $p_{i,j}$ 表示在第 i 个位置处发射 WiFi 信号时, 第 j 个监测点收到的功率值, $i=1,2,\dots,5001$, $j=1,2,\dots,8$ 。每个房间采集 5001 条数据, 三个房间将获得 15,003 条数据。在采用人工智能算法进行定位研究时, 必须先对数据进行预处理, 本文采用的方法是先取出 WiFi 信号接收功率原始数据集, 求出数据的最大值和最小值后, 采用 $X^* = (X - X_{\min}) / (X_{\max} - X_{\min})$ 将原始数据归一化到区间[0, 1]。

4. 基于 AI 的发射源定位研究

1) 人工智能 AI

人工智能 AI 是计算机科学的一个分支, 主要研究开发用于模拟、延伸和扩展人类智能的理论、方法、技术及应用系统, 研究内容包括机器学习和深度学习。机器学习算法包括支持向量机、K 邻近算法、随机森林、梯度提升和 k 均值等[7]; 深度学习算法包括卷积神经网络、循环神经网络、自编码、集成学习、多任务学习、迁移学习、终身学习、玻尔兹曼机、深度信念网络、深度生成模型、深度强化学习和序列生成模型等[8] [9]。下面结合发射源定位介绍支持向量机 SVM、K 邻近算法 KNN 和卷积神经网络 CNN。

SVM 是一种监督机器学习算法, 本文采用成对分类方法将非线性多分类问题转为 binary SVM 进行处理。设给定含 N 个样本的训练数据集 $X = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 对于第 i 类和第 j 类数据, 训练一个 SVM 分类模型即等效为求解二次规划问题:

$$\begin{aligned} \min_{\mathbf{w}^{ij}, b^{ij}, \xi^{ij}} & \frac{1}{2} (\mathbf{w}^{ij})^T \mathbf{w}^{ij} + C \sum_t \xi_t^{ij} \\ \text{s.t.} & (\mathbf{w}^{ij})^T \phi(\mathbf{x}_t) + b^{ij} \geq 1 - \xi_t^{ij}, \text{ if } y_t = i \\ & (\mathbf{w}^{ij})^T \phi(\mathbf{x}_t) + b^{ij} \leq -1 + \xi_t^{ij}, \text{ if } y_t = j \\ & \xi_t^{ij} \geq 0 \end{aligned} \quad (2)$$

其中, 上标表示是 i 类和 j 类之间的 binary SVM 的参数; 下标 t 表示 i 类和 j 类的并集中样本的索引; ϕ 表示输入空间到特征空间的非线性映射, 核函数为:

$$\phi(\mathbf{x}_i, \mathbf{y}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{y}_j\|^2}, \gamma > 0 \quad (3)$$

索引 $i, j \in \{1, \dots, M\}, i < j$, 一共需训练 $C_M^2 = \frac{1}{2} M(M-1)$ 个 binary SVMs。平均每个类包含 N/M 个样本, 所以平均每个对偶问题包含 $2N/M$ 个变量。第 i 类和第 j 类之间 binary SVM 的决策函数为:

$$y_{\text{new}}^{ij} = \text{sign} \left[(\mathbf{w}^{ij})^T \phi(\mathbf{x}_{\text{new}}) + b^{ij} \right] = \text{sign} \left[\sum_{\substack{\text{support} \\ \text{vectors}}} y_t \alpha_t^{ij} k(\mathbf{x}_t, \mathbf{x}_{\text{new}}) + b^{ij} \right] \quad (4)$$

(4)式用于判断数据是属于 i 类还是 j 类。

KNN 是数据挖掘分类技术中最简单的机器学习算法之一, 其指导思想是“近朱者赤, 近墨者黑”, 即由你的邻居来推断出你的类别。KNN 算法的核心思想是: 为了判断未知样本的类别, 以所有已知类别

的样本作为参照，计算未知样本与所有已知样本之间的欧式距离，从中选取与未知样本距离最近的 K 个已知样本，根据少数服从多数的投票法则，将未知样本与 K 个最邻近样本中所属类别占比较多的归为一类。假设给定两个样本 $X = (x_1, x_2, \dots, x_n)$ 和 $Y = (y_1, y_2, \dots, y_n)$ ，其中 n 表示特征数， X 和 Y 两个向量间的欧氏距离表示为：

$$\text{dist}_{ed}(X, Y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} \quad (5)$$

KNN 算法的分类原理如图 4 所示，图中当选择的 K 值取 3 时，待测样本的预测标签为类别 1；当 K 取 5 时，待测样本预测标签为类别 2。因此 K 值的选取十分重要， K 越小越容易过拟合，当 $K = 1$ 时只根据单个近邻进行预测，如果离目标点最近的一个点是噪声就会出错，此时模型复杂度高，稳健性低；但是如果 K 取的过大，这时与目标点较远的样本点也会对预测起作用，就会导致欠拟合，此时模型变得简单；如果 $K = N$ 的时候，那么就是取全部的样本点，这样预测新点时最终结果都是取所有样本点中某分类下最多的点，分类模型就完全失效了。

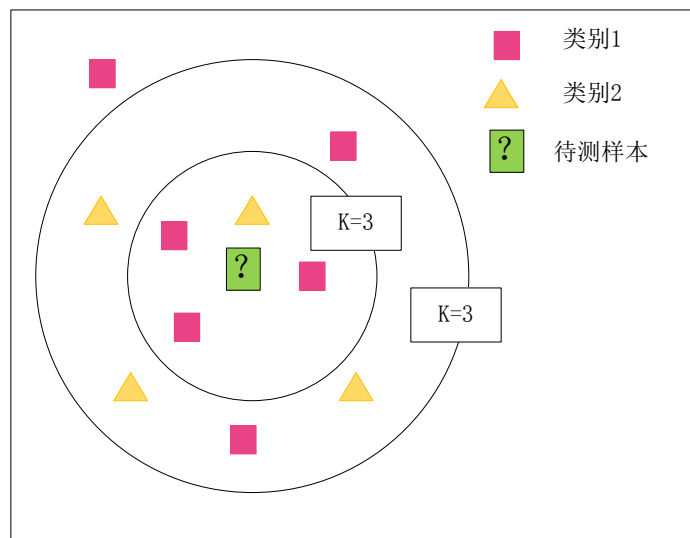


Figure 4. Schematic diagram of KNN classification principle
图 4. KNN 分类原理示意图

CNN 是一种前馈型的神经网络，相比于其他神经网络结构，卷积神经网络需要的参数相对较少，应用广泛。其核心思想是用卷积核 $\mathbf{W}^{p,1}, \mathbf{W}^{p,2}, \dots, \mathbf{W}^{p,D}$ 分别对输入特征映射 $\mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^D$ 进行卷积，然后将卷积结果相加，并加上一个标量偏置 b 得到卷积层的净输入 \mathbf{Z}^p ，再经过非线性激活函数后得到输出特征映射 \mathbf{Y}^p ，最后通过汇聚层、全连接层和激活函数完成输入特征分类。

$$\mathbf{Z}^p = \mathbf{W}^p \otimes \mathbf{X} + b^p = \sum_{d=1}^D \mathbf{W}^{p,d} \otimes \mathbf{X}^d + b^p \quad (6)$$

$$\mathbf{Y}^p = f(\mathbf{Z}^p) \quad (7)$$

其中， $\mathbf{W}^p \in \mathbb{R}^{U \times V \times D}$ 为三维卷积核， $f(\cdot)$ 为非线性激活函数。本文设计的 CNN 结构和参数如图 5 所示，采用 8 个卷积层对特征进行提取，通过一层 Flatten 将卷积层的多维输出一维化，最后通过全连接层和带有 3 个神经元和 softmax 的激活函数实现对 3 个房间内 WiFi 发射源位置的识别。

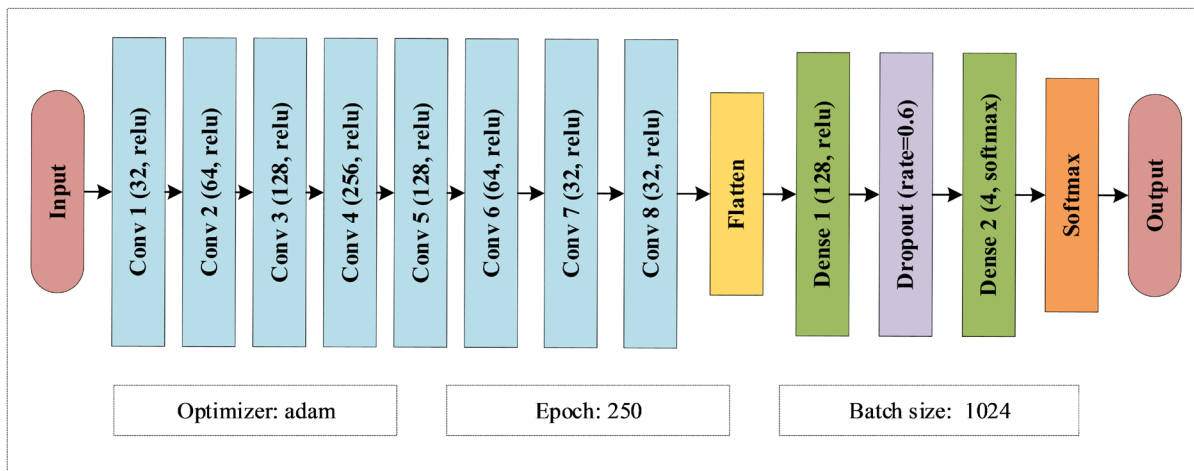


Figure 5. Structure CNN model and parameters
 图 5. CNN 深度学习模型结构和参数

2) 实验结果

首先，在房子外围监测接收点的数量为 8 的情况下，用采集到是数据对 CNN 模型进行训练，训练次数为 50 次，每次训练的样本数为 1024。图 6 为监测接收点位置编号示意图和通过模型训练和测试得到的混淆矩阵。结果表明采用 8 个监测接收点时，识别准确率达到 98.5%，此时采用 KNN 和 SVM 的准确率分别为 98.8%和 97.7。

其次，将监测接收点的数量从 8 个分别减少至 7 个、6 个、5 个、……和 2 个，采用 CNN，KNN 和 SVM 对定位研究并对精度进行分析和比较。图 7 所示为混淆矩阵计算结果，表 1 为采用不同算法进行位置识别得到的精度比较。结果表明，当监测接收点数量仅为 2 个时，三种算法定位准确率均较低；当监测接收点数量为 3 个时，三种算法的定位准确率均达到 70%；当监测接收点数量为 4 个时，KNN 的定位准确率提高最明显，达到 81.7%；当监测接收点的数量为 5 个时，KNN 的定位准确率达到 91.4%，CNN 和 SVM 分别为 83.6%和 87.9%。因此要达到较高的定位精度，监测接收点的数量至少要有 5 个。

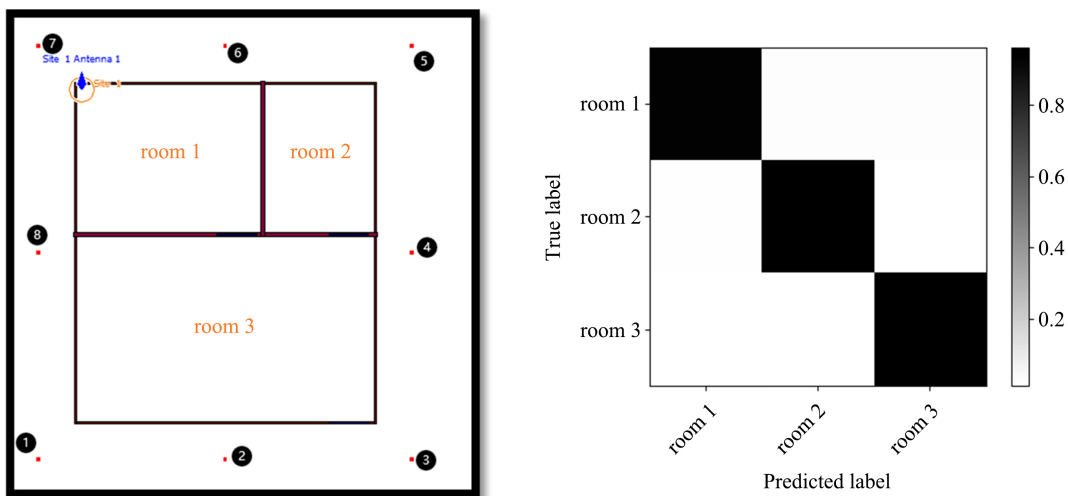


Figure 6. Antenna numbering diagram and confusion matrix (8 monitoring nodes)
 图 6. 天线编号示意图和混淆矩阵(8 个监测节点)

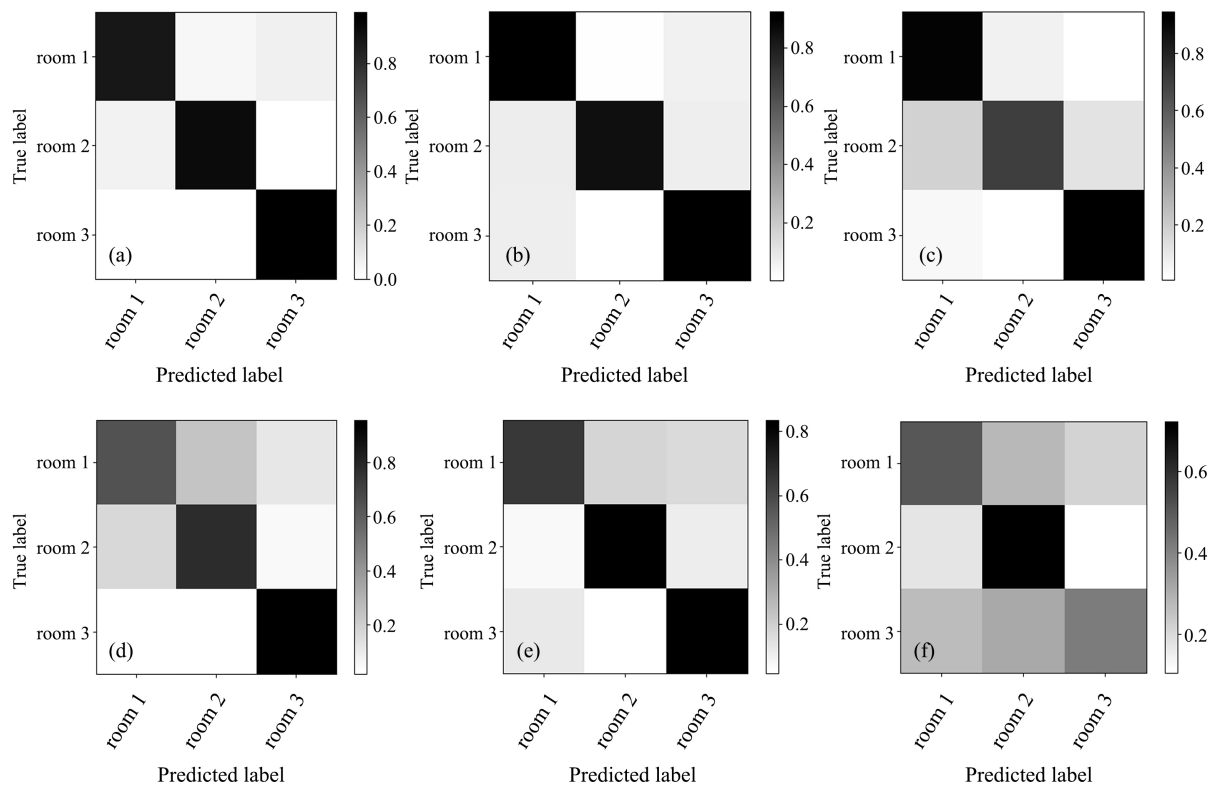


Figure 7. The confusion matrix for different number of receiving points. (a) $N = 7$; (b) $N = 6$; (c) $N = 5$; (d) $N = 4$; (e) $N = 3$; (f) $N = 2$

图 7. 监测接收点数量 N 变化时的混淆矩阵。(a) $N = 7$; (b) $N = 6$; (c) $N = 5$; (d) $N = 4$; (e) $N = 3$; (f) $N = 2$

Table 1. Comparison of location recognition accuracy of CNN, KNN and SVM

表 1. CNN, KNN 和 SVM 位置识别精度比较

接收点数	方法	接收点位置编号	CNN (%)	KNN (%)	SVM (%)
8		1, 2, 3, 4, 5, 6, 7, 8	95.2	98.8	97.7
7		1, 2, 3, 4, 5, 6, 7	92.6	98.1	96.8
6		3, 4, 5, 6, 7, 8	88.3	96.8	93.6
5		1, 3, 4, 5, 6	83.6	91.4	87.9
4		1, 3, 5, 7	76.4	81.7	69.9
3		1, 4, 8	76.7	76.3	70.0
2		1, 5	54.9	55.5	46.4

图 8 为三种算法准确率比较曲线, 可以看出, 在监测节点数量多($N \geq 4$)和少($N < 4$)的情况下 KNN 均具有较高的准确率; 在监测节点数量较多的情况下, SVM 的准确率高出 CNN, 而监测接收节点较少的情况下, CNN 和 KNN 的准确率优于 SVM。

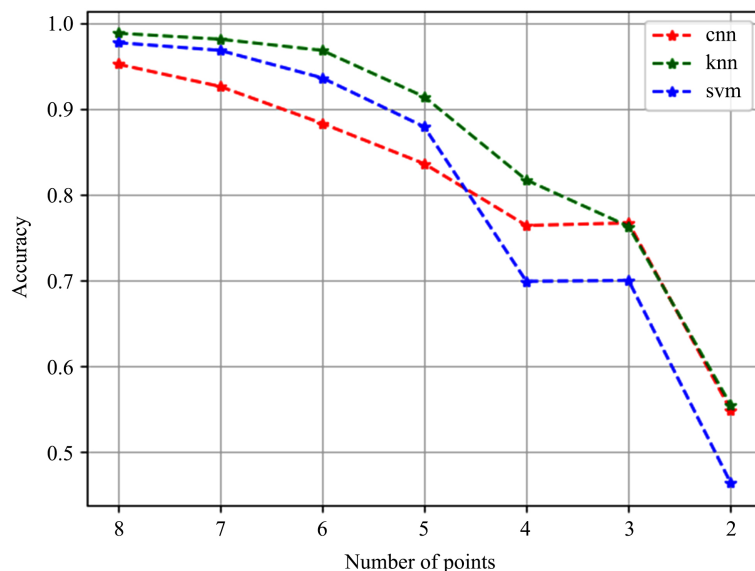


Figure 8. Comparison of location recognition accuracy of three algorithms

图 8. 三种算法准确率比较曲线

5. 结论

本文基于电磁场仿真软件 WinProp 建立了 WiFi 发射和接收功率仿真模型。通过模拟在一栋单层别墅内不同房间进行 WiFi 发射，在别墅外围设置监测接收点进行接收，研究了 WiFi 信号的传输特性；建立了 WiFi 信号接收功率数据集，分别采用 CNN、KNN 和 SVM 三种算法对房间内无线 AP 进行定位，并比较了监测接收节点数量变化时的定位精度。结果表明，在相同的节点数量下，KNN 具有较高的定位精度；当监测接收节点数量为 5 个时，CNN、KNN 和 SVM 算法的定位精度分别为 83.6%、91.4% 和 87.9%；当监测接收节点的数量大于等于 7 个时，三种算法均可以达到 90% 以上的准确率。本文的研究工作对解决非视通环境下的无线电监测定位问题有重要意义。

致 谢

本文得到了云南大学无线创新实验室的支持，对王晓燕博士和付英贤同学的支持表示衷心感谢。

基金项目

国家自然科学基金(61863035, 41461052, 11564044)。

参考文献

- [1] 边疆无线电编委会. 边境区域无线电管理技术设施建设及投资需求研究[M]. 云南: 云南大学出版社, 2015.
- [2] 中央网信办. 无线网络安全面临六大威胁[EB/OL]. http://www.cac.gov.cn/2014-03/19/c_126397441.htm
- [3] 林兴峰. 企业无线网络技术安全性问题探究[J]. 电子产品可靠性与环境试验, 2020, 38(S2): 84-88.
- [4] 张六, 曾正, 陈俊昌, 杨晶晶, 黄铭. 基于信道状态信息的室内人数统计方法[J]. 计算机工程与设计, 2020, 41(4): 1125-1132.
- [5] Belal, K., Hong, C. and Yasamin, M. (2021) Multiple People Identification Through Walls, Using Off-The-Shelf WiFi. *IEEE Internet of Things Journal*, **8**, 6963-6974.
- [6] 海凜, 王正斌, 程勇, 刘芫健. 基于 WinProp 仿真软件的《无线通信网络规划与优化》课程改革研究[J]. 教育现代化, 2019(74): 111-113.

- [7] Python 机器学习库[DB/OL]. <https://scikit-learn.org/stable/>
- [8] 邱锡鹏. 神经网络与深度学习[M]. 北京: 机械工业出版社, 2020.
- [9] Keras 深度学习[DB/OL]. <https://keras.io/about/>.