引用本文：蒋菲. 面向中大型仓库的可见光通信系统目标定位[J]. 光通信技术，2025，49(3)：1-9.

面向中大型仓库的可见光通信系统目标定位

蒋 菲
（广西社会科学院 产业经济研究所，南宁 530022）

摘要：为解决中大型智能仓储系统的货物位置估计问题，提出一种基于对比学习和接收信号强度(RSS)的室内可见光定位模型，即对比学习变换器(CLTf)模型。首先，对数以百计的光功率值进行过滤，选择其中强度最高的若干个发光二极管(LED)光功率信号构造光功率向量；然后，利用 Transformer 模型捕捉长序列依赖关系，并结合对比学习技术挖掘锚点先验知识以优化特征表达。仿真结果表明：在 50 m×20 m×3 m 的中大型仓库空间内，CLTf 模型在货架第 1~5 层的平均定位误差分别为 0.292、0.344、0.375、1.133、2.471 cm，定位精度达到厘米级，显著优于传统方法。

关键词：智能仓储系统；中大型仓库；位置估计；室内定位；可见光通信
中图分类号：TN929.12；TP394.1　**文献标志码**：A　**文章编号**：1002-5561(2025)03-0001-09
DOI：10.13921/j.cnki.issn1002-5561.2025.03.001　**开放科学(资源服务)标识码(OSID)**：

Target localization of visible light communication system for medium and large warehouses

JIANG Fei
(Industrial Economics Research Institute, Guangxi Academy of Social Sciences, Nanning 530022, China)

Abstract: To address the issue of goods position estimation in medium and large intelligent warehouse systems, this paper proposes an indoor visible light positioning model based on contrastive learning and received signal strength(RSS), namely the contrastive learning transformer (CLTf) model. First, hundreds of optical power values are filtered to select the highest-intensity light-emitting diode(LED) signals for constructing the optical power vector. Then, the Transformer model is employed to capture long-sequence dependencies, while contrastive learning techniques are integrated to mine anchor point prior knowledge for feature representation optimization. The simulation results show that in a medium to large warehouse space of 50 m×20 m×3 m, the average positioning errors of the CLTf model on the 1st to 5th shelves are 0.292, 0.344, 0.375, 1.133, and 2.471 cm, respectively, with a positioning accuracy of centimeter level, significantly better than traditional methods.

Key words: intelligent warehouse system, medium and large warehouse, position estimation, indoor position, visible light communication

0 引言

随着网上购物和大型商超的普及，建设自动化、高效且低成本的智能仓储系统成为我国经济高质量发展的迫切需求。近年来，人工智能、大数据和物联网等技术不断融合，逐渐形成了一套较完整的智能仓储系统框架[1-2]。在该框架中，拣货机器人[3]从货架上取下货物是一项基本功能，而此功能的优劣主要取决于货物定位的性能。目前，实现货物定位主要有 2 种方式：数据库查询和室内目标定位。第一种方法在存放货物时记录货物位置信息，并在拣货时通过查询数据库来定位货物。然而，这种方法的缺点是在货物存储期间，货物的位置需始终保持不变。第二种方法则主要依靠物联网技术，例如超宽带(UWB)[4]和射频识别(RFID)[5]。

收稿日期：2024-12-20。
基金项目：国家社会科学基金项目(24CGJ108))资助；国家社会科学基金项目(20CFX047)资助；广西高校中青年教师科研基础能力提升项目(2022KY1252)资助。

作者简介：蒋菲(1985—)，女，广西桂林人，博士，副教授，CILT 三级、广西环境保护研究学会秘书长，主要研究方向为物流工程、智慧仓库、绿色供应链等。参与省部级、国家级项目 16 项，获批软著及专利 7 项。

理论上，此类基于电磁波的室内定位精度已达到厘米级[6]，但在中大型仓库（1 000 m² 以上），密集分布的货架会阻隔电磁波传播，导致基于电磁波通信技术的定位误差大幅增加，无法准确引导拣货机器人到达指定货物位置。

智能仓库过道上方安装大量的发光二极管（LED）用于照明。通过利用 LED 和室内可见光通信（VLC）技术，可以建立一个低成本的货物定位系统[7-8]。针对室内可见光定位问题，研究者们提出了多种有效方案，主要包括以下 3 种：三角定位法[9]、基于接收信号强度（RSS）的定位法[10]和融合多源信息的定位法[11]。三角定位法易于实现，但其定位精度相对较低；而融合多源信息的定位法虽然能提供更高的定位精度，但却需要额外的辅助设备支持[12]。综合来看，基于 RSS 的定位法在精度和成本之间取得了较好的平衡[13]，更适合智能仓储系统的货物定位问题。在现有基于 RSS 的室内可见光定位方法中，文献[14]提出了一种基于双向长短期记忆神经网络（Bi-LSTM）的单 LED 定位方法。文献[15]提出一种基于极限学习机神经网络和RSS 的室内可见光定位方法。文献[16]提出基于人工神经网络（ANN）和 RSS 的室内可见光定位方法。文献[17]提出基于级联残差神经网络（CRNN）和 RSS 的室内可见光定位方法。这些方法得益于神经网络技术的发展，在 5 m×5 m×3 m的室内场景下实现了厘米级的平均定位误差。尽管这种精度水平足以引导拣货机器人到达货物所在位置，但此类方法通常需要以 0.2~0.3 m 为间隔设立参考点，并需要使用少于 10 个 LED 来构建神经网络训练集，这在面积超过 1 000 m² 的中大型仓库中实施起来颇具挑战性。总体而言，在中大型仓库场景下，如文献[14-17]的方法面临两大主要挑战：首先，由于需要为每个 5 m×5 m×3 m 的子区域单独训练一个神经网络模型，对于中大型仓库来说，至少需准备 40 个独立的网络模型，这极大地增加了网络训练的复杂度；其次，为了采集 RSS 训练数据，必须以 0.2~0.3 m 的间隔设立参考点，这意味着每个接收面至少需要采集 10^5 个 RSS 样本，从而导致数据采集和训练过程极其困难。因此，如何有效地克服这些问题成为提升室内可见光定位系统在实际应用中的关键所在。

本文面向中大型仓库，提出一种基于对比学习和 RSS 的室内可见光通信系统目标定位模型，即对比学习变换器（CLTf）模型，以期提高中大型仓库内货物定位的准确性。

1 中大型仓库货物定位

1.1 中大型仓库模型

智能仓库的模型图如图 1 所示。仓库内分布着成排的货架，相邻两排货架间设有过道供叉车或机器人移动，过道正上方的天花板装有用于照明的 LED。拣货机器人接收拣货命令，沿过道到达目标货物所在的货架，使用机械臂取下目标货物。

图 1 智能仓库模型图

仓库的二维平面分布图如图 2 所示。图中，L 和 W 分别表示仓库的长度和宽度。

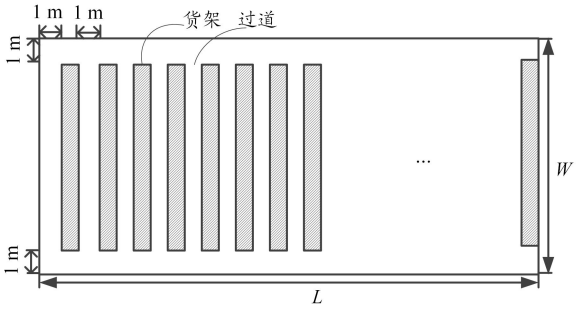


图 2 仓库的二维平面分布图

1.2 VLC 系统定位模型

VLC 系统定位模型如图 3 所示。仓库过道的顶部设有用于照明的 LED，货架的载物面、天花板和地面三者之间平行，第 i 个 LED 发射光线与法线夹角为 α_i，

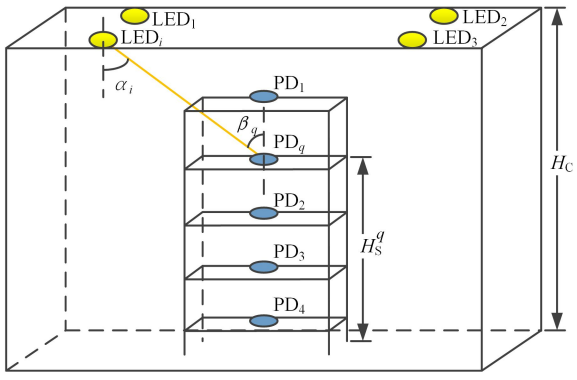


图 3 VLC 系统目标定位模型

β_q 为第 q 个光电探测器(PD)入射光线与法线夹角，H_S^q 为第 q 个 PD 高度，H_C 为天花板高度。

假设 LED 为 Lambertian 光源，第 i 个 LED 发射可见光经过自由空间信道后到达第 q 个 PD，接收光功率可表示为

$$P_R(i,q)=\begin{cases}\dfrac{(m_i+1)A_qP_b(i,q)}{2\pi}\times\\ \dfrac{\cos^{m_i}(\alpha_i)\cos(\beta_q)T_s(\beta_q)g(\beta_q)}{d_{i,q}^2},0\leqslant\phi_q\leqslant F_q\\ 0,\phi_q>F_q\end{cases}\tag{1}$$

其中，A_q 为第 q 个 PD 有效区域，$g(\beta_q)$ 为第 q 个 PD 前的聚光器增益，$T_s(\beta_q)$ 为第 q 个 PD 前的光滤波器增益，$P_b(i,q)$ 表示从第 i 个 LED 到第 q 个 PD 的发射光功率，$d_{i,q}$ 为第 i 个 LED 与第 q 个 PD 间的直线距离，ϕ_q 为第 q 个 PD 的入射光角度，F_q 为第 q 个 PD 的视场角(FOV)。m_i 为第 i 个 LED 的 Lambertian 阶数，其计算方法为

$$m_i=-\frac{\ln 2}{\ln\left(\cos\alpha_i^{1/2}\right)}\tag{2}$$

其中，$\alpha_i^{1/2}$ 为第 i 个 LED 的半功率角度。

为简化分析，假设 PD 法线垂直于地面，即 PD 无倾斜，此时 LED 发射角等于 PD 入射角($\alpha=\beta$)，可将式(1)改写为

$$P_R(i,q)=\frac{(m_i+1)A_qP_b(i,q)}{2\pi d_{i,q}^2}\cos^{m_i+1}(\theta)T_s(\theta)g(\theta)\tag{3}$$

其中，θ 为入射角或辐射角，此处 $\theta=\alpha_i=\beta_q$。

室内 VLC 的信道噪声主要由散粒噪声和热噪声构成[19]，可将信道噪声表示为

$$N_{Gaus}=\sigma_{TH}^2+\sigma_{SH}^2\tag{4}$$

其中，σ_{TH}^2 为接收端热噪声方差，σ_{SH}^2 为入射光功率散粒噪声方差，它们的计算方法可参考文献[20]。

第 i 个 LED 传输可见光信号的信噪比可表示为

$$S_i=10\lg\frac{P_R(i)}{N_{Gaus}}\tag{5}$$

2 CLTf 模型的定位方法

本文提出的 CLTf 采用 Transformer 模型[18]处理中大型仓库的可见光定位问题，其核心优势在于能够有效学习 RSS 向量分量间的依赖关系，并在深层网络结构中保持稳定的训练效果。相较于传统序列模型，CLTf 通过以下 2 个创新设计显著提升定位性能：嵌入层和对比学习联合训练。1)嵌入层：将 RSS 分量与对应 LED 标识(ID)进行联合编码，构建信息增强的 RSS 表征向量。针对 PD 采集的稀疏 RSS 数据，采用滑动窗口机制提取 Top-M 显著分量，减少 RSS 向量的维度；同时，通过窗口滑动生成多样化训练样本，有效缓解参考点稀疏问题。2)对比学习联合训练：基于锚点位置先验知识，对 Transformer 编码器提取的特征进行对比优化，增强特征空间中对位置差异的敏感性，提升定位判别性。

在模型选型方面，传统 RNN 因梯度问题难以处理长序列依赖，LSTM 虽通过门控机制有所改善，但在处理含数百个 LED 的 RSS 数据时仍面临效率瓶颈。相比之下，Transformer 的多头自注意力(MHSA)机制具有三大优势：精准捕获长距离序列依赖，支持变长输入的高效并行处理，以及通过层级编码有效提取 RSS-位置映射特征。该设计使 CLTf 能够从复杂的光信号环境中提取关键定位特征，为稀疏参考点下的高精度定位提供可靠解决方案。

CLTf 模型的运行流程图如图 4 所示。由于 LED 的 ID 是 RSS 向量的关键特征，因此本文的 RSS 向量每个分量由 RSS 值与其对应的 LED 索引构成，可表示为 $G=\{(g_1,l_1),(g_2,l_2),\cdots,(g_t,l_t),\cdots,(g_T,l_t)\}$，其中 T 为仓库内安装的 LED 总数，l_t 为第 t 个 LED，g_t 为 PD 从 LED 接收的光功率值。由于货架遮挡会导致每个锚

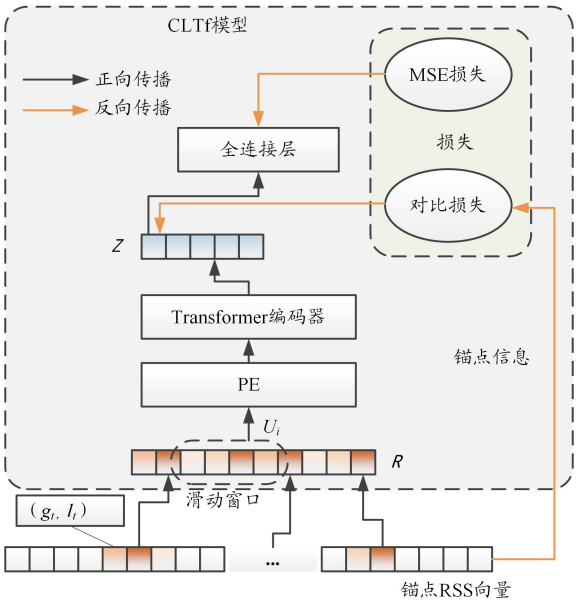


图 4 CLTf 模型的运行流程图

点处 PD 采集的 RSS 向量十分稀疏，因此从 RSS 向量中选取Top-M的 RSS 分量构成新 RSS 向量 R，R的分量顺序保持与原始 RSS 向量一致。采用一个尺寸为 w 的滑动窗口从 R向量中提取 RSS 子序列 U_i。U_i经过位置编码(PE)和Transformer 编码器后得到对应的特征向量 Z，将 Z输入全连接层和输出层得到其坐标，再计算预测坐标的均方误差(MSE)损失。另外，将同一个 batch 中的特征向量 Z之间进行对比学习，利用锚点的先验信息对 CLTf 模型进行微调，进一步提高提取特征向量 Z关于锚点位置的显著性。

2.1 PE

Transformer 模型的嵌入层主要由输入嵌入(IE)和 PE 这 2 个关键步骤组成。IE 的作用是将原始输入数据转换为模型可以处理的向量形式；PE 则通过在这些嵌入向量中添加表示各个元素在其序列中的位置信息，以解决 Transformer 无法直接处理序列顺序的问题。在本文的应用场景中，RSS 向量可以直接作为输入嵌入向量使用，因此只需对已有的 RSS 向量实施位置编码，即可保留其序列特性，同时使其适应 Transformer 模型的要求。假设 PE 向量为 P，其表达式为

$$\begin{cases} P_{(p,2i)}=\sin\left(\dfrac{p}{10\ 000^{2i/d}}\right) \\ P_{(p,2i+1)}=\cos\left(\dfrac{p}{10\ 000^{2i/d}}\right) \end{cases} \tag{6}$$

其中，p为 RSS 向量中的分量位置，即 RSS 分量的索引，d为 P的维度，i为 LED 的索引。

2.2 Transformer 编码器

将嵌入层产生的序列 X^{E}传入 Transformer 编码器，编码器将 x压缩成一个包含上下文信息的特征向量 Z。Transformer 编码器的核心模块为 MHSA，其结构和处理流程如图 5 所示。

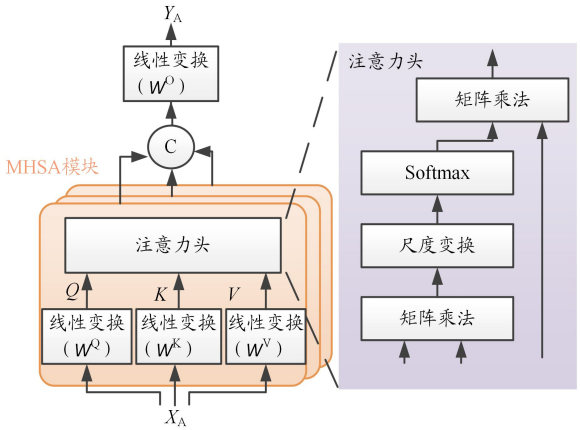


图 5 MHSA 处理流程

假设 MHSA 的输入为 X_A，MHSA 的输出 Y_A表示为

$$Y_A=M(Q,K,V)=C(h_1,h_2,\ldots,h_{N_A})W^O \tag{7}$$

其中，Q、K和 V均由 X_A线性变换得到的不同信息，$M(\cdot)$表示 MHSA 运算，N_A为注意力头数，h_i为第 i个注意力头输出的矩阵，$C(\cdot)$表示特征拼接操作，W^O为输出层的权重矩阵。第 i个注意力头的输出的矩阵可表示为

$$h_i=S(QW_i^Q,KW_i^K,VW_i^V) \tag{8}$$

其中，W_i^Q、W_i^K和 W_i^V为注意力头 h_i的可学习权重矩阵，$S(\cdot)$表示单头注意力运算。单头注意力运算可表示为

$$S(Q,K,V)=\sigma\left(\frac{QK^T}{\sqrt{d_K}}\right)V \tag{9}$$

其中，d_K为 K的维度，$\sigma(\cdot)$为归一化指数函数(softmax 函数)。

Transformer 编码器的结构图如图 6 所示。该编码器由 N_B个结构相同的编码块级联组成，每个编码块由 MHSA、添加与规范化(Add&Norm)、前馈层和 Add&Norm 级联组成。MHSA 通过并行运行多个独立的注意力头，能够捕捉输入序列中复杂的依赖关系和丰富的模式信息。MHSA 的输出经过残差连接和层标准化(LN)处理，可使网络的训练过程更稳定并防止过拟合。在第一个编码块中，PE 被加入到 IE 向量中，以捕获序列的位置信息和上下文信息。随着数据流经后续的编码块，每一层基于前一个编码块的输出进一步提取序列中更复杂、更深层次的模式和依赖性。因此，整个编码器能够有效地学习输入序列的多层次表示。综上所述，使用 Transformer 编码器处理 RSS 向量能有效提取 RSS 向量中的模式特征及其分量间的依赖关系。

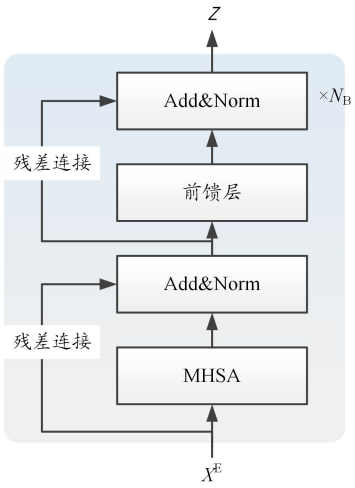


图 6 Transformer 编码器的结构图

在每个编码块中，序列 X^{E}经过 MHSA 和 Add&Norm 处理的结果可表示为

$$X'^{\mathrm{E}}=N(M(X^{\mathrm{E}})+X^{\mathrm{E}}) \tag{10}$$

其中，$N(\cdot)$表示层标准化处理。

中间编码序列经过前馈层和 Add&Norm 处理的结果可表示为

$$Z=N(F(X'^{E})+X'^{E}) \tag{11}$$

其中，$F(\cdot)$表示前馈网络的等效函数。

2.3 全连接层与损失函数

全连接层与输出层的网络结构图如图 7 所示。将 Transformer 编码器输出的特征向量 Z 输入全连接层，全连接层可将不同维度的特征向量 Z 线性地转换成固定维度的向量，输出层基于当前 RSS 向量输出其坐标点$(x',\ y')$。

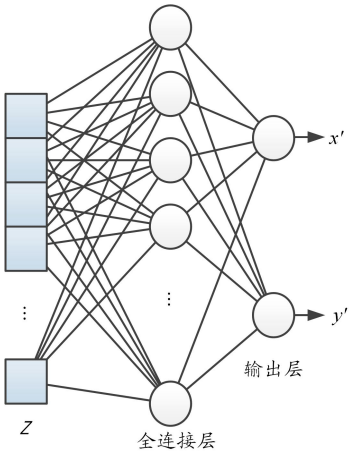


图 7 全连接层与输出层的结构

目标定位可视为一个回归问题，将预测坐标和真实坐标(x,y)之间的 MSE 作为定位损失，表示为

$$l_{\mathrm{MSE}}=\frac{1}{N}\sum_{i=1}^{N}\left|x_i-x_i'\right|+\left|y_i-y_i'\right|^2 \tag{12}$$

其中，N为样本数，(x_i,y_i)、(x_i',y_i')分别为第 i 个滑动窗口对应的真实坐标、预测坐标。

为了充分利用锚点的先验知识，通过对比学习对 Transformer 编码器提取的特征进行微调，从而进一步提高所提取特征关于锚点位置的表达力。对比损失的目标是放大不同锚点间特征的距离，缩小同一锚点间特征的距离，由此可增强同一锚点特征关于其位置的表达力。将对比损失定义为

$$l_{\mathrm{ctra}}=-\lg\frac{\sum_{i'\in I(i)}\exp(s(Z_i,Z_{i'}))}{\sum_{i'\in I(i)}\exp(s(Z_i,Z_{i'}))+\sum_{j\in J(i)}\exp(s(Z_i,Z_j))} \tag{13}$$

其中，Z_i和 $Z_{i'}$为第 i 个锚点不同滑动窗口经过 Transformer 编码器输出的特征向量，Z_i和 Z_j分别为第 i 个和第 j 个锚点滑动窗口经过 Transformer 编码器输出的特征向量，$s(\cdot)$为余弦相似性，$I(i)$表示第 i 个锚点不同滑动窗向量集，$J(i)$表示非第 i 个锚点的滑动窗口向量集。

综上，CLTf 训练的总损失可表示为

$$l_{\mathrm{Total}}=l_{\mathrm{MAE}}+l_{\mathrm{ctra}} \tag{14}$$

2.4 训练方法

CLTf 模型的训练流程如算法 1 所示。首先，在所有锚点采集的 RSS 向量构造 RSS 矩阵，并确定批大小、编码器注意力头数和编码块数。一批 RSS 向量传入 CLTf 模型，依次进行 PE、多头注意力处理、Add&Norm、前馈层和 Add&Norm，得到包含上下文信息的特征序列向量 Z，Z经过全连接层和输出层产生预测坐标，将预测坐标与真实坐标进行比较，得到损失 l_{MSE}；基于 Z 计算损失 l_{ctra}。通过优化总损失 l_{Total} 来更新 CLTf 模型的超参数 Θ_{CLTf}。

算法 1 CLTf 模型的训练流程

输入：RSS 矩阵 D，批大小 n，注意力头数 N_{A}，Transformer 编码块数 N_{B}，滑动窗口尺寸 S，筛选的 RSS 向量长度 M

输出：CLTf 模型超参数 Θ_{CLTf}，预测坐标$(x',\ y')$

1)FOREACH batch IN D;　//分批训练

2)　FOREACH j FROM 0 TO n−1 DO　//输入每个 RSS 向量

3)　　FOREACH i FROM 0 TO M−S+1 DO

4)　　$U_i=SW(D_{\mathrm{batch}\times n+i})$;//滑动窗口产生子 RSS 向量

5)　　$X^{E}=PE(U_i)$;//位置编码//

6)　　FOREACH k FROM 1 TO N_{B} DO

7)　　式(10);　//多头注意力机制

8)　　式(11);　//前馈层和 Add&Norm

9)　　ENDFOR

10)　　$(x',\ y')=FC(Z)$;//全连接层和输出层

11)　ENDFOR

12)　ENDFOR

13)　通过优化 l_{Total} 来更新 Θ_{CLTf};　//反向传播

14)ENDFOR

3 仿真与结果

3.1 仿真设计

本文基于 Matlab 平台和第 1.2 节的可见光定位模型建立仿真。仓库模型的二维平面图如图 8 所示。仓库的长、宽、高为 50 m×20 m×3 m，过道宽度和货架宽度均为 1 m，货架长度为 18 m，每个过道天花板上等间距地分布 9 个 LED，相邻 LED 间的距离为 2 m，过道1的 LED 坐标为 (0.5 m,2 m)，(0.5 m,4 m)，(0.5 m,6 m)，…，(0.5 m,18 m)；过道 2 的 LED 坐标为(2.5 m,2 m)，(2.5 m,4 m)，(2.5 m,6 m)，…，(2.5 m,18 m)；以此类推，整个仓库的 LED 数量为 9×25=225。

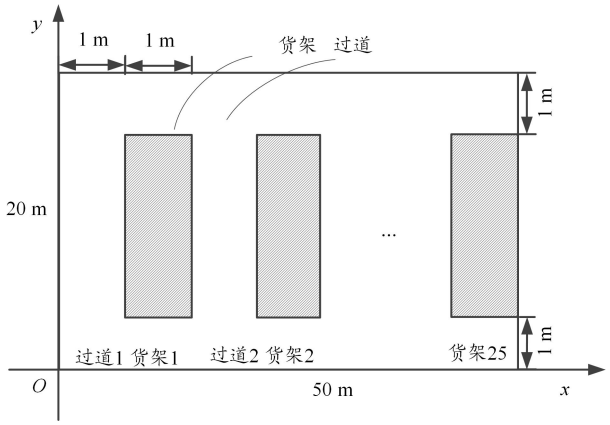


图 8 仿真的仓库平面模型

可见光定位系统的仿真参数如表 1 所示。系统采用视距(LOS)信道模型，忽略多径效应和非直射分量。在仿真环境中，货架各层载物面上以 1 m 间隔均匀布置锚点，其中第一层锚点坐标示例为(1.5 m，1 m，0.2 m)、(1.5 m，2 m，0.2 m)、(1.5 m，3 m，0.2 m)等，其它层锚点的 z 坐标值随货架层高相应调整。每个锚点处部署 PD 用于接收可见光信号，每个 PD 检测来自不同 LED 的 225 个光功率值，这些测量值与对应的 LED ID 共同组成 RSS 向量。所有锚点采集的 RSS 向量构成完整数据集，并按 9:1 的比例随机划分为训练集和验证集，其中 90%的数据用于神经网络训练，剩余 10%用于模型验证和性能评估。

表 1 可见光定位系统的仿真参数

仿真参数	值
FOV/(°)	90
PD 有效区域/cm²	1
光滤波器增益	1
光聚合器增益	1
LED 半功率角/(°)	60
LED 发射功率/W	4

3.2 性能评价指标

本文使用平均定位误差(MPE)和均方根误差(RMSE)评估可见光定位系统的定位性能，MPE 和 RMSE 的计算方法分别为

$$E_{\mathrm{MPE}}=\frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i-x_i')^2+(y_i-y_i')^2} \tag{15}$$

$$E_{\mathrm{RMSE}}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-x_i')^2+(y_i-y_i')^2} \tag{16}$$

3.3 参数分析

仿真环境为一台服务器，其处理器型号为 Intel Xeon Silver 4216，显卡型号为 NVIDIA GeForce RTX 4090，内存大小为 64 GB。操作系统为 Ubuntu 20.04，编程环境为 Python 3.9.17。使用 PyTorch 2.1.0 构造神经网络，基于 Hugging Face 库实现 CLTf 模型。CLTf 模型相关的超参数如表 2 所示。

表 2 CLTf 模型的超参数

超参数	值
注意力头数	5
编码器的编码块数	1,2,3,4,5
Batch 尺寸/个	32
学习率	0.001
滑动窗口尺寸	4,5,6,7,8,9,10
Epoch 数	100
筛选 RSS 分量数 O	14

为简化分析，假设仓库内的过道宽度和货架宽度均为 1 m，仓库内 LED 的高度为 3 m，货架总高为 2.7 m，货架离地间隙 0.2 m；货架共 5 层，每层高度为 0.5 m，货架长度为 2 m。

本文通过仿真实验分析了滑动窗口尺寸和编码器块数对可见光定位系统性能的影响。实验模拟了 PD 在 5 层货架载物面上的工作场景，重点研究不同高度条件下这 2 个关键参数对系统性能的影响规律。在固定编码器块数为 3 的情况下，不同滑动窗口尺寸对各层载物面 RMSE 的影响如图 9 所示。可以看出，随着滑动窗口尺寸的增大，训练集和验证集的 RMSE 均呈现逐渐降低的趋势；当窗口尺寸达到 8 时，定位误差趋于稳定。虽然继续增大窗口尺寸可以带来微小的性能提升，但会显著增加计算成本。通过综合分析模型的泛化能力和计算效率，本文最终确定将滑动窗口尺寸固定为 8，这一取值在所有货架层高下都能实现性能与效率的最佳平衡。

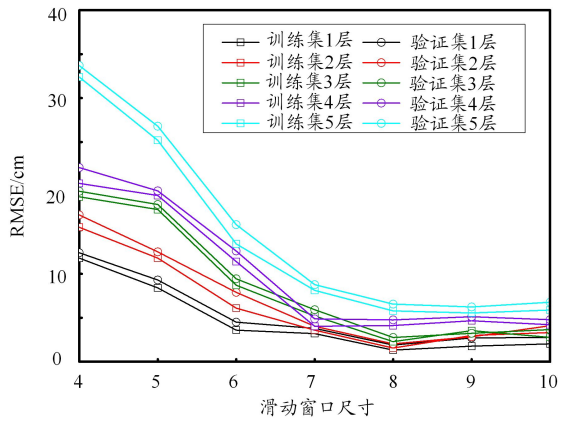


图 9 不同滑动窗口尺寸下收敛的 RMSE

不同编码块数下每层载物面上收敛的 RMSE 如图 10 所示。可以看出，随着 Transformer 编码块数的增加，模型对 RSS 向量的特征提取能力显著增强，这体现在训练集和验证集的 RMSE 均呈现持续下降趋势。当编码块数达到 3 时，系统在所有货架层高下均能获得较优的泛化性能；继续增加编码块数虽然能带来微弱的性能提升，但同时会导致神经网络层数增加、模型复杂度显著升高。基于模型性能与计算效率的综合考量，本文最终将 Transformer 编码器的块数确定为 3。

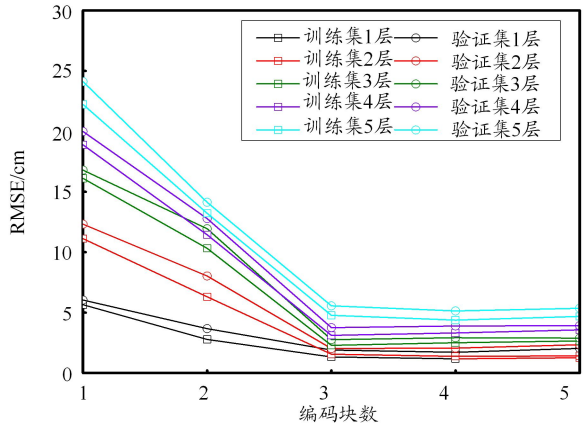
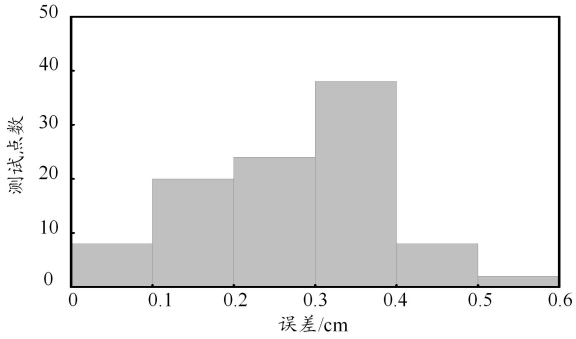


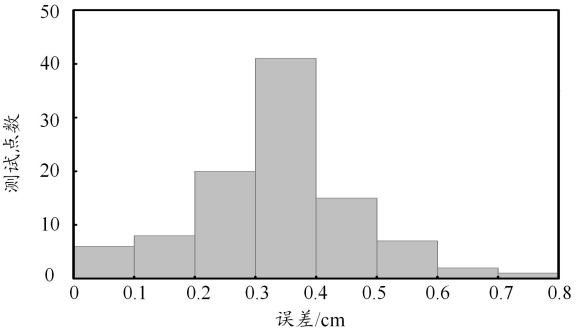
图 10 不同编码块数下收敛的 RMSE

3.4 定位结果与分析

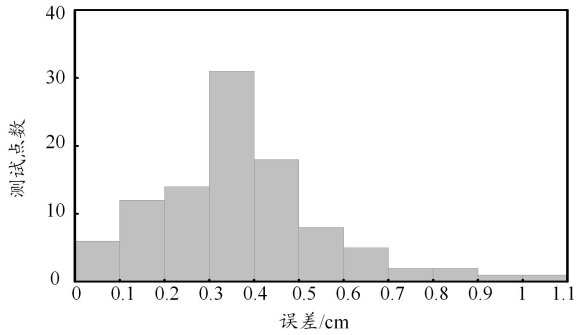
为评估所提方法在多层货架环境中的定位性能，本文构建了系统化的测试方案。在各层载物面上以 5 m 为间隔均匀选取 4×25 个测试点构成测试集，具体坐标示例如下：第一货架第 1 层测试点为(1.5 m,1.5 m,0.2 m)、(1.5 m,6.5 m,0.2 m)、(1.5 m,11.5 m,0.2 m)、(1.5 m,16.5 m,0.2 m)；第二货架第 1 层为(3.5 m,1.5 m,0.2 m)、(3.5 m,6.5 m,0.2 m)、(3.5 m,11.5 m,0.2 m)、(3.5 m,16.5 m,0.2 m)，其余测试点依此类推。计算货架各层测试点的平均定位误差，结果如图 11 所示。可以看出，在货架第 1 层上，90%测试点的定位误差小于 0.4 cm，最大定位误差小于 0.6 cm；在货架第 2 层上，90%测试点的定位误差小于 0.5 cm，最大定位误差小于 0.8 cm；在货架第 3 层上，90%测试点的定位误差小于 0.7 cm，最大定位误差小于 1.1 cm；在货架第 4 层上，85%测试点的定位误差小于 3.5 cm，最大定位误差小于 7 cm；在货架第 5 层上，80%测试点的定位误差小于 5 cm，最大定位误差小于 9.5 cm。由此可知，所提方法的定位误差达到厘米级，在货架第 1~3 层载物面上的定位精度较高。根据式(1)可知，随着 PD 高度增加，LED 辐射角和PD 入射角增大，导致可见光功率出



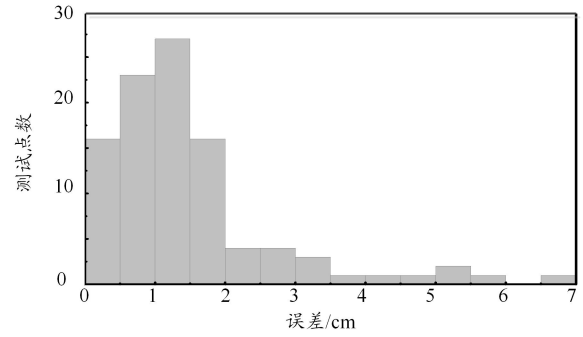
(a) 货架第 1 层



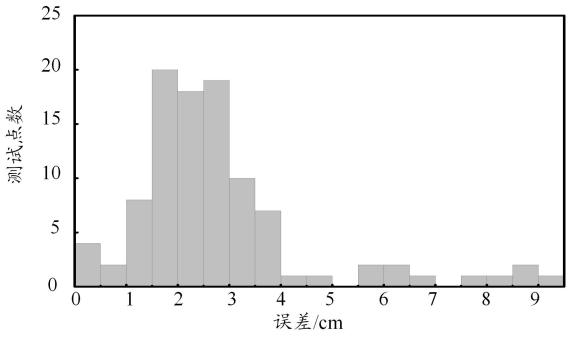
(b) 货架第 2 层



(c) 货架第 3 层



(d) 货架第 4 层



(e) 货架第5层

图11 货架各层的定位误差分布情况

现显著的非线性衰减。这种衰减使得基于接收光强的位置估计难度增加，特别是在第4~5层高度时，信号质量的下降直接影响了定位精度。

本文选择了3种基于RSS的可见光定位方法作为对比方法，分别是随机森林(RF)[21]、人工神经网络(ANN)[16]以及卷积循环神经网络(CRNN)[17]。其中，RF[21]的决策树数量设为200；ANN[16]的输入层神经元数为250，隐藏层数为5。3种模型均使用本文的训练集进行训练和验证，并在测试集上评估性能，其平均定位误差结果如表3所示。可以看出，RF在中大型仓库场景下的定位误差较高。但是，在以1 m为间隔的参考点集上训练的效果不佳，未能达到理想的泛化性能。与RF方法相比，ANN和CRNN的平均定位误差有所下降，ANN和CRNN能更好地捕捉RSS指纹关于锚点位置的特征。但由于锚点之间的间隔较远，ANN和CRNN的泛化能力受到明显的影响，它们在货架第5层的定位误差依然较高。CLTf在货架低层载物面上的平均定位误差较低，在高层载物面上的平均定位误差也保持了较低的水平。这是因为CLTf利用Transformer模型同时考虑RSS向量中的所有分量信息，能更好地捕捉分量间的依赖关系，提取的RSS向量特征关于锚点位置的表达力更强。此外，CLTf通过对比学习对所提取特征进行微调，进一步提高了特征表达力，进而提高了定位精度。

表3 不同算法的平均定位误差对比结果(单位：cm)

货架层号	不同算法的平均定位误差			
	RF	ANN	CRNN	CLTf
1	5.234	1.932	1.273	0.292
2	5.890	2.507	1.752	0.344
3	6.331	2.782	2.323	0.375
4	9.778	4.671	4.529	1.133
5	12.341	9.442	8.890	2.471

3.5 消融实验

本文通过消融实验验证对比学习机制对定位性能的提升效果。实验设置了2种模型对比：仅使用MSE损失函数的变形体(Tf)模型和引入对比学习的CLTf模型，两者网络结构和超参数保持完全一致，结果如图12所示。在低层货架(1~3层)区域，2种模型的定位误差差异较小(<0.5 cm)；而在高层区域(第5层)，CLTf模型展现出显著优势，平均定位误差比Tf模型降低了2.23 cm。这一现象主要源于高层区域LED辐射角和PD入射角增大导致的光功率非线性衰减特性，传统MSE损失函数难以有效拟合这种复杂非线性关系，而CLTf模型通过对比学习机制，充分利用锚点先验知识构建特征空间，在训练过程中实现网络参数的智能微调，显著提升了对非线性光功率特征的表征能力，从而有效提高了高层区域的定位精度。实验结果表明，对比学习机制在处理复杂光传播环境下的非线性特征提取问题上具有独特优势，为多层仓储环境下的高精度定位提供了可靠解决方案。

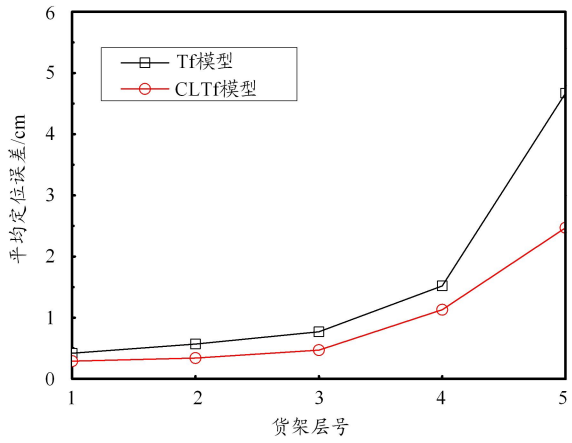


图12 2种模型的消融实验对比

4 结束语

本文提出了一种基于对比学习Transformer的中大型仓库可见光定位方法。该方法通过将RSS信号与LED ID联合编码构建增强特征，并采用滑动窗口机制提取Top-M显著分量，既降低数据维度又实现样本增广；利用锚点位置先验知识优化Transformer编码特征，增强位置判别性，有效提高了中大型仓库的货物位置估计精度。实验结果表明：所提方法的定位误差达到厘米级，在货架第1~5层载物面上的定位误差分别为0.292、0.344、0.375、1.133、2.471 cm，并且在第5

层载物面上的平均定位误差较原始模型降低 2.23 cm。

本文定位方法通过嵌入层的数据增强和对比学习的特征优化，有效解决了样本稀疏导致的泛化不足问题，为大型仓储环境提供了一种经济可靠的定位方案。未来工作将重点解决 NLOS 传播和复杂光环境干扰等实际挑战。

参考文献：

[1] 雷旭，陈静夷，陈潇阳. 改进哈里斯鹰算法的仓储机器人路径规划研究[J]. 系统仿真学报，2024，36(5)：1081–1092.

[2] 尹昊，董益民，冯卓，等. 有轨制导车堆垛的动态稳定性研究[J]. 计算机集成制造系统，2024，30(1)：316–328.

[3] 徐翔斌，马中强. 基于移动机器人的拣货系统研究进展[J].自动化学报，2022，48(1)：1–20.

[4] XIA B, XIE N, ZHANG L. Novel whale cooperative positioning algorithm for uwb sensor networks[J]. Wireless Personal Communications, 2024, 135(4): 2421–2438.

[5] ABUDALFA S, BOUCHARD K. Two-stage RFID approach for localizing objects in smart homes based on gradient boosted decision trees with under- and over-sampling[J]. Journal of Reliable Intelligent Environments, 2023, 1(10): 45–54.

[6] 邓诗凡，蒋伟，杨俊杰，等. 蓝牙到达角室内定位抗干扰优化研究[J]. 导航定位学报，2022，10(6)：75–80.

[7] DU X P, ZHANG Y, YE Z, et al. LED lighting area recognition for visible light positioning based on convolutional neural network in the industrial internet of things[J]. Optics Express, 2023, 31(8): 12778–12788.

[8] LONG Q, ZHANG J, CAO L W. Indoor visible light positioning system based on point classification using artificial intelligence algorithms [J]. Sensors, 2023, 23(11): 1–19.

[9] JUANDA A, CAHYADI W A, RUSDINAR A, et al. Indoor positioning system infrastructure based on triangulation method through visible light communication[J]. Ultima Computing, 2022, 25(7): 31–37.

[10] 朱亚丽. 基于深度神经网络的可见光通信室内定位[J]. 光学技术，2023，49(4)：452–458.

[11] ALBRAHEEM L, ALAWAD S. A hybrid indoor positioning system based on visible light communication and bluetooth RSS trilateration [J]. Sensors, 2023, 23(16): 1–32.

[12] 杨彦兵，胡超，鲁邦彦，等. 基于可见光和射频融合的通信定位一体化系统[J]. 通信学报，2023，44(12)：146–157.

[13] SALAMA W M, ALY M H, AMER E S. VLC localization: deep learning models by Kalman filter algorithm combined with RSS[J]. Optical and Quantum Electronics, 2022, 54(9): 1–18.

[14] 王乐乐，秦岭，胡晓莉，等. 基于 Bi-LSTM 神经网络的室内可见光定位方法[J]. 光通信技术，2024，48(2)：36–41.

[15] 赵霞，张君毅，龙倩倩. 基于 Circle 混沌映射的 ISSA-ELM 神经网络室内可见光定位方法[J]. 光学学报，2023，43(2)：25–34.

[16] NASCIMENTO M R F D, COUTINHO O G G, OLIVI L R, et al. A visible light positioning technique based on artificial neural network [J]. Journal of Control Automation and Electrical Systems, 2024, 35(4): 677–687.

[17] CHEN Q, CHEN H, WEN H, et al. Robustness enhancement of indoor visible light positioning system based on cascaded residual neural network [J]. Optics Communications, 2023, 546(4): 10–16.

[18] 李健，熊琦，胡雅婷，等. 基于 Transformer 和隐马尔科夫模型的中文命名实体识别方法[J]. 吉林大学学报(工学版)，2023，53(5)：1427–1434.

[19] GUAN W P, WEN S S, HU H X, et al. Research on visible light communication system based on hybrid modulation technique [J]. Journal of Optoelectronics·Laser, 2015, 26(11): 2125–2132.

[20] KOMINE T, NAKAGAWA M. Fundamental analysis for visible-light communication system using LED lights [J]. IEEE Transactions on Consumer Electronics, 2004, 50(1): 100–107.

[21] 曲佳，王旭东，吴楠，等. 基于随机森林算法的室内可见光指纹定位方法[J]. 光通信技术，2023，47(1)：1–7.