

# 改进AlexNet的细粒度汉字手势识别研究

王聪聪<sup>1</sup>, 左 洋<sup>1\*</sup>, 陈兰兰<sup>2</sup>, 葛宝泉<sup>1</sup>, 张亚军<sup>3</sup>

<sup>1</sup>新疆理工职业大学人工智能学院, 新疆 喀什

<sup>2</sup>新疆理工职业大学通识学院, 新疆 喀什

<sup>3</sup>新疆大学软件学院, 新疆 乌鲁木齐

收稿日期: 2025年12月6日; 录用日期: 2026年1月7日; 发布日期: 2026年1月15日

## 摘 要

射频识别在人机交互中的应用不断拓展, 但现有方法在细粒度手势区分、跨场景适应性和模型轻量化方面仍存在不足。为此, 提出一种改进AlexNet融合注意力机制的细粒度汉字手势识别方法。该方法通过标签矩阵与双天线融合增强信号采集完整性, 结合非视距信号基线修正与马尔可夫转移场映射强化时序特征表达, 并在轻量化卷积神经网络中嵌入注意力模块, 在降低计算复杂度的同时保持较高识别准确率。实验结果显示, 系统总体识别准确率为98.88%, 跨用户与跨场景平均准确率分别为97.49%和97.74%。研究表明, 该方法在细粒度汉字手势识别中具有较强鲁棒性与泛化能力, 对智慧课堂人机交互及教育信息化发展具有应用价值。

## 关键词

RFID, 细粒度手势识别, AlexNet, SE注意力模块, 轻量化模型, 人机交互

# Fine-Grained Chinese Character Gesture Recognition Using Improved AlexNet

Congcong Wang<sup>1</sup>, Yang Zuo<sup>1\*</sup>, Lanlan Chen<sup>2</sup>, Baoquan Ge<sup>1</sup>, Yajun Zhang<sup>3</sup>

<sup>1</sup>College of Artificial Intelligence, Xinjiang Vocational University of Technology, Kashgar Xinjiang

<sup>2</sup>College of General Education, Xinjiang Vocational University of Technology, Kashgar Xinjiang

<sup>3</sup>College of Software, Xinjiang University, Urumqi Xinjiang

Received: December 6, 2025; accepted: January 7, 2026; published: January 15, 2026

## Abstract

Radio Frequency Identification (RFID) has been increasingly applied in human-computer interaction

\*通讯作者。

文章引用: 王聪聪, 左洋, 陈兰兰, 葛宝泉, 张亚军. 改进 AlexNet 的细粒度汉字手势识别研究[J]. 计算机科学与应用, 2026, 16(1): 169-181. DOI: 10.12677/csa.2026.161014

(HCI), yet existing approaches still face challenges in fine-grained gesture differentiation, cross-scenario adaptability, and model lightweighting. To address these issues, this study proposes a fine-grained Chinese character gesture recognition method based on an improved AlexNet integrated with a Squeeze-and-Excitation (SE) attention mechanism. The method enhances signal acquisition integrity by combining label matrices with dual-antenna fusion, applies non-line-of-sight baseline correction and Markov Transition Field mapping to strengthen temporal feature representation, and embeds an attention module into a lightweight convolutional neural network to reduce computational complexity while maintaining high recognition accuracy. Experimental results demonstrate an overall recognition accuracy of 98.88%, with average accuracies of 97.49% and 97.74% under cross-user and cross-scenario conditions, respectively. These findings confirm that the proposed approach exhibits strong robustness and generalization in fine-grained Chinese character gesture recognition, and provides a reliable solution for smart classroom interaction and the advancement of educational informatization.

## Keywords

RFID, Fine-Grained Gesture Recognition, AlexNet, SE Attention Module, Lightweight Model, Human-Computer Interaction

Copyright © 2026 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

随着智慧教育与智能交互技术的发展, 课堂教学正逐渐从传统的板书和口头讲解模式, 转向基于人机交互的多模态信息交流方式[1]。手势识别作为其中的重要交互手段, 能够实现教师与教学设备之间的自然沟通, 为课堂互动与教育信息化提供新型支持。然而, 现有基于射频识别(Radio Frequency Identification, RFID) [2]的手势识别方法仍存在局限: 一是细粒度手势, 特别是汉字书写类手势, 相似汉字的信号波形高度相似, 传统方法难以实现有效区分; 二是跨场景识别性能不稳定, 受多径效应和环境噪声干扰较大; 三是部分模型计算复杂度高, 限制了其在实时性要求较高场景中的应用。

针对上述问题, 本研究提出一种改进 AlexNet 融合注意力机制的细粒度汉字手势识别方法, 并引入双天线与标签矩阵融合方案以增强信号采集完整性, 结合马尔可夫转移场(Markov Transition Field, MTF)映射强化时序特征表达[3], 同时在轻量化卷积神经网络中嵌入注意力模块[4], 以实现关键特征的高效建模。实验结果显示, 该方法在不同场景与用户条件下均保持较高识别准确率, 验证了其鲁棒性与泛化能力。本研究不仅在细粒度手势识别方法上具有创新价值, 更在智能人机交互应用中展现了实际意义[5]。通过实现对教师书写与指令类手势的精准识别, 可促进课堂人机交互效率的提升[6], 为智慧教育建设提供技术支持, 并为教育信息化与教学质量提升开辟新的路径。

## 2. 手势信号特征转换与增强

细粒度汉字手势的识别对特征提取能力提出了更高要求。卷积神经网络在图像特征建模方面具有显著优势, 但直接处理一维时序信号存在适配性不足的问题。为充分发挥深度学习模型在细粒度特征提取中的能力, 本研究采用马尔可夫转移场将采集到的 RFID 手势相位序列映射为二维特征图像[7]。该方法不仅将稀疏的时序信息转化为直观的空间特征, 还为卷积神经网络提供丰富的语义输入, 有助于提升模型学习效率与识别精度。

MTF 映射能够在二维图像中保留原始信号的时序结构，同时捕捉不同信号点之间的动态转移模式。在生成转移概率矩阵的过程中，随机噪声得到平滑处理，从而增强了模型在不同场景和环境条件下的鲁棒性。该特征增强策略为后续轻量化卷积神经网络的高精度识别提供了稳定的输入基础。在实验中，选取了 6 个常见汉字作为细粒度手势实验对象。采集的手势相位信号经 MTF 映射后生成二维特征图像，图像清晰呈现了不同手势之间的时序模式差异，验证了所提方法在复杂手势特征表达与可视化方面的有效性。通过一维信号向二维图像的映射，系统能够更准确地捕捉手势动作细节，实现对复杂汉字手势的高精度识别，为智慧课堂的人机交互应用提供可靠的技术支持。

图 1 展示了经 MTF 转换后的 6 个汉字手势特征图，分别对应“大”“人”“不”“口”“日”和“中”。其图像生成过程包括以下步骤：

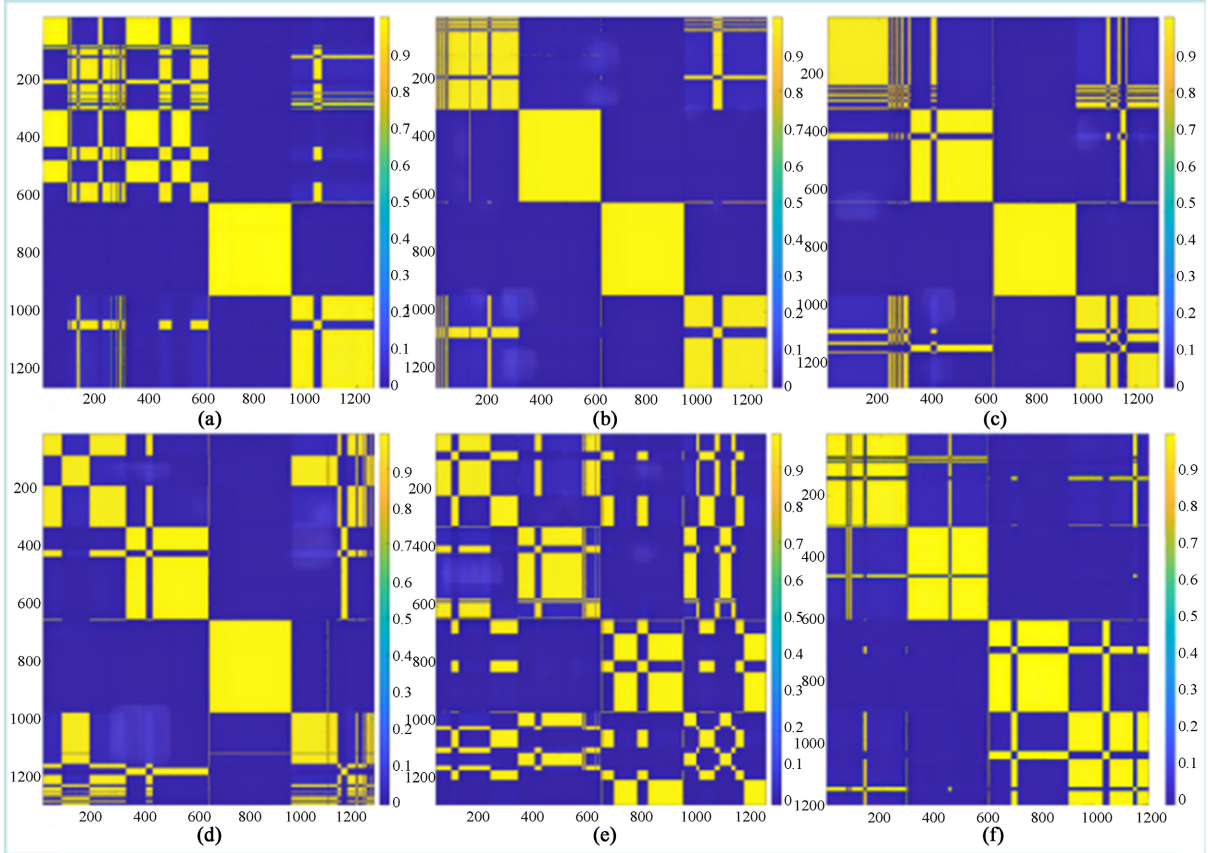


Figure 1. Feature images generated by the Markov Transition Field  
图 1. 马尔可夫转移场生成的特征图像

- (1) 时间序列分箱  
设原始相位序列为

$$X = [x_1, x_2, \dots, x_T] \quad (1)$$

将其归一化并划分为  $K$  个分位箱，记为  $(q_1, q_2, \dots, q_K)$ ，保证每个分位箱中包含的数据点数量大致相同。

- (2) 构造转移矩阵

基于相邻点的转移关系，得到一阶马尔可夫转移矩阵  $w \in R^{K \times K}$ ，如式(2)所示：

$$w_{i,j} = \frac{C(x_t \in q_i, x_{t+1} \in q_j)}{\sum_{j=1}^K C(x_t \in q_i, x_{t+1} \in q_j)} \quad (2)$$

(3) 生成马尔可夫转移场

结合时序信息，将转移概率映射到二维矩阵  $M \in R^{H \times H}$ ，如式(3)所示：

$$M_{i,j} = w_{q(x_i), q(x_j)}, i, j \in [1, H] \quad (3)$$

(4) 图像化与输入网络

最终得到的矩阵  $M$  编码为特征图，作为卷积神经网络的输入数据，用于后续训练与分类。

### 3. 轻量型 AlexNet 识别模型设计

#### 3.1. SE 注意力模块

在对模型进行轻量化改进以降低计算复杂度的过程中，往往会导致识别精度下降。为在保持轻量化的同时提升模型的判别能力，引入 Squeeze-and-Excitation (SE) 注意力机制[8]，如图 2 所示。该机制通过模拟人类视觉的选择性关注过程，在特征提取阶段动态调整通道权重，从而抑制冗余信息并突出关键手势特征，实现对细粒度特征的高效建模。

SE 模块由“Squeeze”和“Excitation”两个阶段构成：前者通过全局平均池化提取每个通道的全局信息，其计算过程如式(4)所示：

$$s_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (4)$$

其中： $s_c$  为通道  $c$  的全局平均池化值， $H$  和  $W$  分别代表特征图的高度和宽度， $x_c(i, j)$  为通道  $c$  在位置  $(i, j)$  的像素值。在“Excitation”阶段，SE 模块通过两层全连接网络自适应生成通道权重如式(5)所示：

$$z = \sigma(W_2 \cdot \delta(W_1 \cdot s)) \quad (5)$$

其中： $s$  为所有通道的全局池化向量， $W_1$  与  $W_2$  分别为降维与升维的权重矩阵， $\delta(\cdot)$  表示 ReLU 激活函数， $\sigma(\cdot)$  表示 Sigmoid 函数， $z$  为最终得到的通道权重向量。最后对特征图进行加权重标定如式(6)所示：

$$X_c = z_c \cdot x_c \quad (6)$$

其中： $X_c$  为重标定后的通道特征， $x_c$  为原始通道特征， $z_c$  为对应的权重。将 SE 注意力模块嵌入至轻量化后的模型中，不仅缓解了轻量化带来的精度损失，还显著增强了模型在跨场景手势识别中的鲁棒性与泛化能力，为轻量级细粒度汉字手势识别提供了有效解决方案。

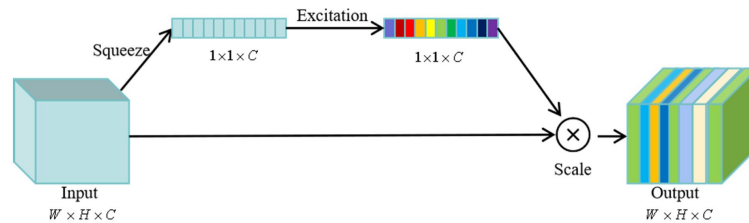


Figure 2. Structure of the SE attention mechanism  
图 2. SE 注意力机制结构图



3.2. AlexNet 网络模型的优化

传统 AlexNet 网络由 8 层组成，包括 5 个卷积层与 3 个全连接层[9]。该模型依托 ReLU 激活函数与并行计算架构，在大规模图像识别中表现突出。然而，该结构的 3 层全连接层引入了大量参数，容易导致模型过拟合[10]，并显著增加存储与计算负担；加之较深的层次结构，需要消耗更多内存来存储权重和激活值，从而限制了其在实时交互场景中的应用。为克服这些不足并提升模型适配性，本文在原始 AlexNet 基础上设计了轻量化改进模型 C-AlexNet。

C-AlexNet 的主要优化体现在 4 个方面，如图 3 所示。首先，采用  $1 \times 1$  卷积与全局平均池化替代原有全连接层，其中  $1 \times 1$  卷积通过加权整合通道间的特征，既提升了特征表达能力，又显著降低参数规模和计算复杂度。其次，在每个卷积层后引入 SE 注意力模块，使模型能够聚焦于关键手势特征，从而提升识别稳定性与鲁棒性。再次，通过减少每层卷积核数量并引入 Batch Normalization 层，有效缓解过拟合并稳定训练过程。最后，删除原始架构中的第 5 个卷积层，并重新配置卷积通道数，以进一步压缩模型规模。实验结果表明，改进后的 C-AlexNet 在保持较高识别精度的同时，训练时间缩短、收敛速度加快，整体实时性显著增强。

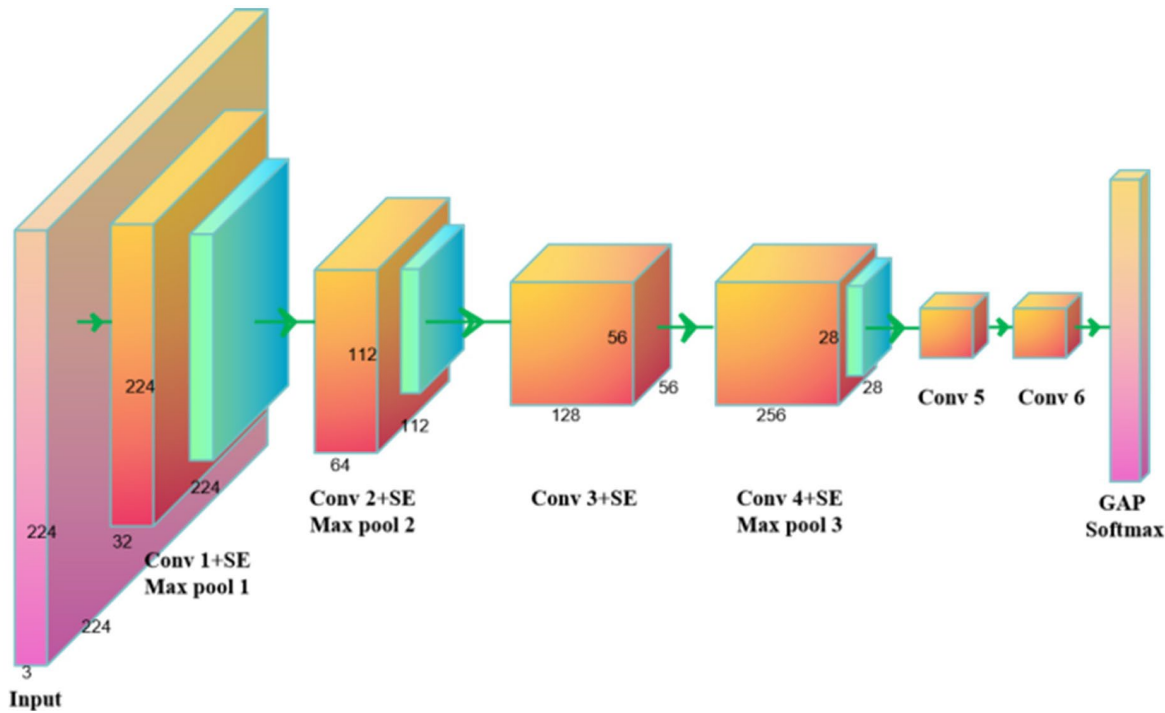


Figure 3. The C-AlexNet network model  
图 3. C-AlexNet 网络模型

3.3. 模型优化后的网络结构

基于上述改进策略，C-AlexNet 的整体网络结构由卷积层、池化层、注意力模块和分类层构成，形成轻量化且高效的识别框架。优化后的网络结构如图 4 所示，网络前端以多层卷积与池化操作实现对输入手势图像的特征提取，同时在卷积层后嵌入注意力机制以强化关键特征表达，末端通过全局平均池化与 Softmax 分类器完成最终识别。该结构在保持特征表达能力的同时显著降低了参数规模与运算量，使模型兼具准确性与实时性，适用于资源受限的交互式应用场景。

本研究将经特征映射得到的手势图像输入卷积神经网络，以充分利用其在细粒度特征自动提取方面的优势，实现汉字手势识别。卷积层的核心计算过程可表示为

$$X = \text{POOL}(\text{ReLU}(WX_0 + B)) \quad (7)$$

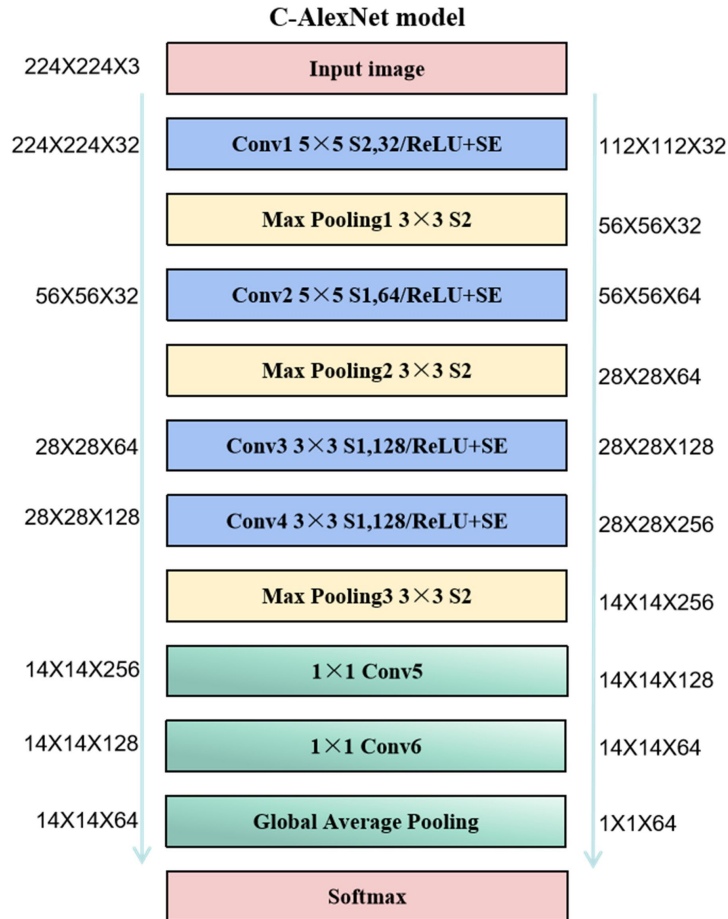
其中： $W$  为卷积核， $X_0$  为输入特征图， $B$  为偏置项。卷积操作能够从局部区域中提取关键特征，ReLU 激活函数进一步增强了网络对非线性关系的建模能力，同时有效缓解梯度消失，加速收敛。经过池化操作后得到的特征图输入后续卷积层，逐层提取细粒度特征。卷积层提取到的汉字手势特征可表示为

$$PF = CK(X, \theta_{fe}) \quad (8)$$

其中： $CK(\cdot)$  表示卷积核提取运算， $X$  为预处理后的手势相位图像， $\theta_{fe}$  为卷积层参数集合。在完成特征提取后，利用 Softmax 分类器将输出向量映射为各类别的概率分布，如式(9)所示：

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^C e^{x_j}} \quad (9)$$

其中： $x_i$  是输入向量的第  $i$  个分量， $C$  为手势类别总数， $y_i$  表示样本属于第  $i$  个类别的概率。通过该归一化函数，模型能够完成多类别手势的判别与分类。



**Figure 4.** Parameters of the C-AlexNet network  
**图 4.** C-AlexNet 网络参数

### 3.4. 模型总体架构设计

本研究构建的汉字手势识别系统基于轻量级卷积神经网络,整体流程如图5所示,主要由信号采集、数据预处理、特征转换与手势识别4个模块构成。首先,在信号采集阶段,系统采用 $2 \times 2$ 标签矩阵与双天线组合布局,有效扩大了手势信号的覆盖范围,并缓解了标签间的电磁干扰问题。双天线的引入为相似手势提供补充特征信息,从而增强了数据的区分度。随后,在数据预处理阶段,依次进行静态反射消除、相位解缠绕、Savitzky-Golay滤波与归一化处理,以降低环境噪声、校正信号连续性并提升跨场景的稳定性,同时利用数据增强策略扩充不足类别,保证模型的均衡学习能力。

在特征信息转换阶段,采用马尔可夫转移场将预处理后的相位序列映射为二维特征图像。该方法通过刻画数据点之间的转移概率,保留了时序依赖与空间分布信息,从而避免一维信号直接建模带来的特征稀疏问题,并强化了细粒度特征表达。最后,在手势识别阶段,引入改进的C-AlexNet轻量级卷积神经网络对特征图像进行分类。该模型在降低参数规模与训练开销的同时提升了识别的鲁棒性和实时性。

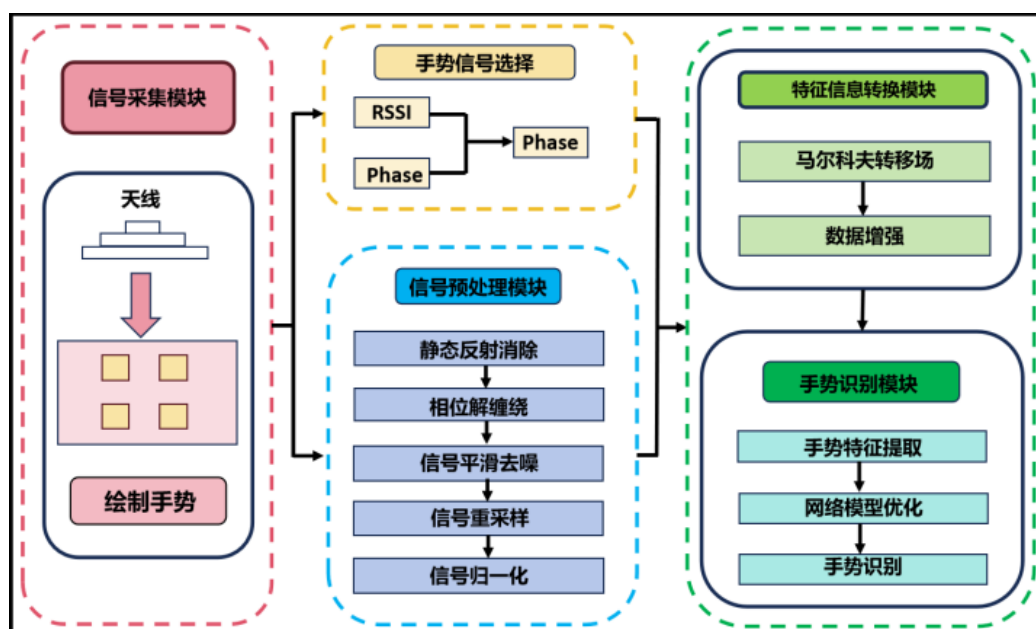


Figure 5. Design of the system network model

图5. 系统网络模型设计

综上,本研究通过多模块协同设计,在数据采集、预处理、特征转换和模型优化等方面实现了有机融合。实验结果表明,该系统在多场景及新用户条件下均表现出优异的识别精度与泛化性能,为基于RFID的细粒度汉字手势识别提供了一种高效且可扩展的解决方案。

## 4. 实验验证

### 4.1. 实验部署环境

本实验所采用的硬件设施主要由4部分组成,如图6所示。具体包括:工作频率为920.875 MHz的Impinj Speedway R420型RFID读写器、一副圆极化UHF天线、4个尺寸为 $5\text{ cm} \times 5\text{ cm}$ 的UHF无源标签,以及一台搭载Intel i5-1135G7处理器的笔记本电脑。上述设备共同构建了实验所需的射频识别采集平台。

实验软件运行在一台配备 2.4 GHz CPU 和 32 GB 内存的计算机上, 主要用于数据采集与后续处理。计算机通过以太网与 RFID 读写器连接, 通信协议采用低级阅读器协议(Low-Level Reader Protocol, LLRP)。在实现层面, RFID 数据的采集与预处理模块由 C 语言编写, 而卷积神经网络模型的构建与训练则基于 Python 实现。

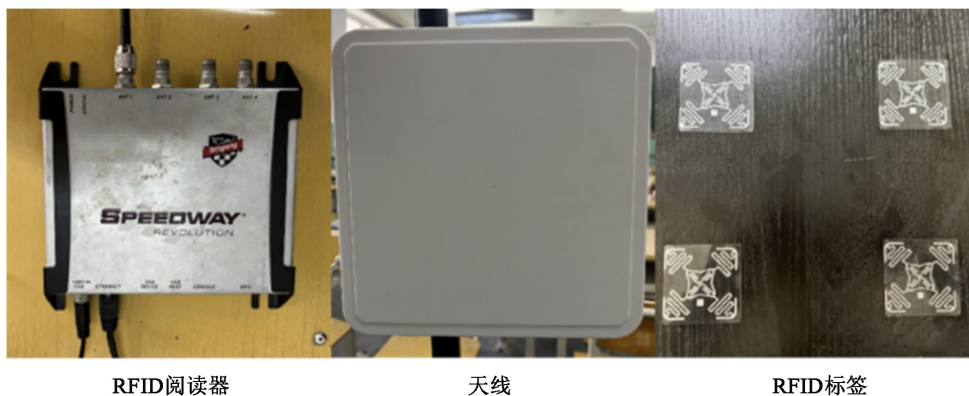


Figure 6. Experimental equipment  
图 6. 实验器材

## 4.2. 实验数据集与场景

为构建覆盖性较强的数据集并验证所提模型在多场景条件下的适应性, 本研究共招募 20 名志愿者参与实验。每位志愿者在 3 个典型场景中分别完成 6 种汉字手势的书写动作, 每个手势重复 55 次。为进一步提升实验的真实性与鲁棒性, 采集过程中额外引入人为噪声以模拟复杂环境干扰。经统计, 3 个场景共获得 59,400 组有效样本, 其中 70% (41,580 组) 作为训练集, 其余 30% (17,820 组) 用于测试集。在实验环境布置上, 均采用  $2 \times 2$  标签矩阵与双天线组合, 以保证信号覆盖完整性与采集一致性。具体而言, 场景 1 设于约  $10 \text{ m} \times 8 \text{ m}$  的会议室, 室内存在金属柜与圆形会议桌, 具备较强多径反射特性; 场景 2 位于  $4 \text{ m} \times 18 \text{ m}$  的走廊, 环境空旷无遮挡, 主要用于评估低干扰条件下的系统性能; 场景 3 则布置在  $8 \text{ m} \times 12 \text{ m}$  的教室, 室内分布有桌椅及杂物, 可有效模拟真实教学环境的复杂干扰。在所有场景中, 受试者均站立于天线与标签的布设区域中央完成指定汉字手势, 所得数据被输入至后续模型训练与验证流程。

## 4.3. 评价指标

为全面评估所提模型在细粒度汉字手势识别中的性能, 本研究采用混淆矩阵、准确率(Accuracy, ACC)与 F1-score 作为主要评价指标。这些指标在分类任务中被广泛应用, 能够从整体表现与局部特性两个层面刻画模型的识别能力。混淆矩阵通过矩阵形式直观展示预测结果与真实类别的对应关系, 不仅能揭示模型整体分类正确性, 还能反映具体的误分类模式。准确率用于衡量模型在所有类别上的综合识别水平, 是评价整体性能的核心指标。而 F1-score 则在精确率(Precision)与召回率(Recall)之间取得平衡, 尤其适用于样本分布存在不均衡的场景, 能够更准确地反映模型对不同类别的区分能力。评价指标的数学定义如下:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$F1 - score = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (11)$$

其中： $TP$  (True Positive)表示模型正确预测为正类的样本数量， $TN$  (True Negative)表示正确预测为负类的样本数量， $FP$  (False Positive)表示将负类错误判定为正类的样本数量， $FN$  (False Negative)则表示将正类误判为负类的样本数量。通过结合上述指标，可以对识别模型在整体精度与类别区分能力两个方面进行全面、客观的评价。

4.4. 实验结果

为全面评估所提出模型的识别性能，本研究将 3 个实验场景下采集的数据按 70%训练、30%测试进行划分，并采用 10 折交叉验证策略提升泛化能力。具体操作为将训练集均分为 10 个子集，每次选取 1 个子集作为验证集，其余 9 个子集用于训练，循环 10 次并取平均结果。实验结果如图 7 所示，系统整体识别准确率达到 98.88%，表现出优异的稳定性。在细粒度汉字手势识别中，即便“口”和“日”笔画相似，系统识别精度仍保持在 98%以上，进一步验证了模型在复杂手势区分上的鲁棒性与应用潜力。

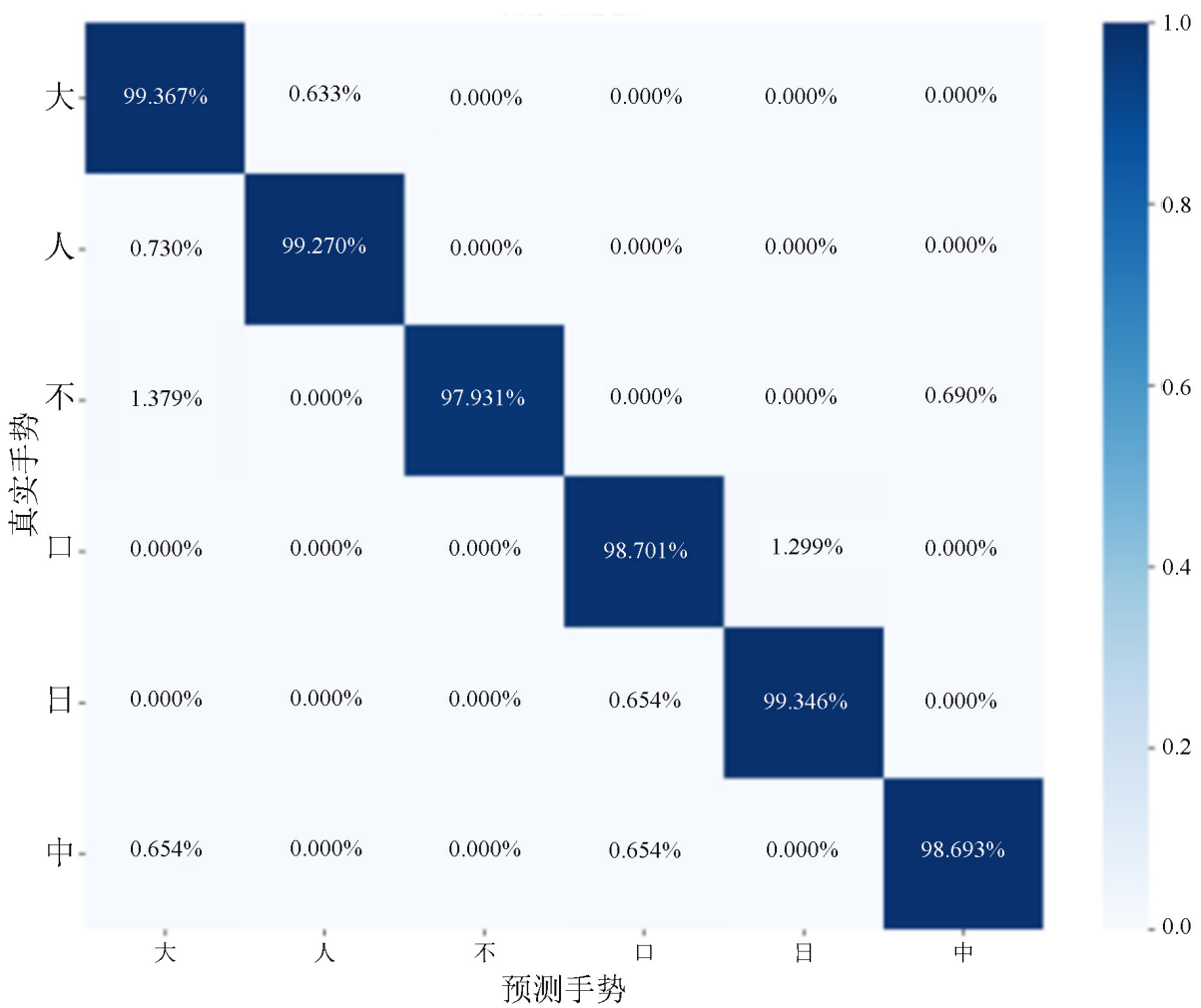


Figure 7. Confusion matrix of overall system performance  
图 7. 系统整体性能混淆矩阵

在空中绘制手势缺乏坐标参考，不同用户在速度和尺寸上存在差异，因此系统需要具备较强鲁棒性。实验邀请 6 名志愿者完成 6 个汉字手势动作，每个手势重复 55 次。



**Table 1.** Recognition performance of different users  
**表 1.** 不同用户的识别性能

用户	测试	
	<i>ACC/%</i>	<i>F1-Score/%</i>
用户 1	99.02	97.17
用户 2	97.17	97.33
用户 3	96.56	96.33
用户 4	98.15	97.67
用户 5	96.88	97.00
用户 6	97.18	97.16

测试结果见表 1，系统整体识别精度为 97.49%，6 名志愿者的精度分别为 99.02%、97.17%、96.56%、98.15%、96.88%和 97.18%，尽管存在个体差异，系统整体仍保持较高识别水平。*F1-Score* 测试进一步表明，6 位用户的平均 *F1-Score* 达到 97.11%，验证了系统在多用户条件下的稳定性和高性能，为智慧课堂的人机交互应用提供可靠支持。

为验证系统在不同环境下的跨域识别能力，选取会议室、走廊和教室 3 个场景进行测试，其中走廊多径效应最小，教室多径效应最高。5 名志愿者在 3 个场景中完成 6 个汉字手势，每个手势书写 80 次，并将采集数据用于训练与测试。实验结果见表 2，系统在不同环境下的平均识别准确率达到 97.74%，会议室、走廊和教室的平均准确率分别为 98.17%、98.96%和 96.08%。尽管教室环境下多径和噪声干扰最大，识别率仍保持在 96%以上，充分表明系统在跨场景手势识别中具有良好的鲁棒性和泛化能力。

**Table 2.** Recognition performance in different scenarios  
**表 2.** 不同场景的识别性能

场景	测试	
	<i>ACC/%</i>	<i>F1-Score/%</i>
会议室	98.17	97.17
走廊	98.96	98.16
教室	96.08	96.00

**Table 3.** Recognition performance under different interferences  
**表 3.** 不同干扰下的识别性能

噪声情景	测试	
	<i>ACC/%</i>	<i>F1-Score/%</i>
基线环境	98.00	97.88
静态干扰	97.17	96.93
动态干扰	96.16	96.04

环境噪声可能导致手势相位信号的衰减与失真，从而影响系统识别性能。为评估系统的抗干扰能力，本实验设计了 3 种噪声情景模拟真实环境条件：情景 1 为基线实验，用户在正常环境下完成手势动作；

情景 2 在手指与标签之间放置金属障碍物以引入静态干扰；情景 3 邀请 4 名志愿者在用户周围随机走动产生动态噪声。每位用户在各情景下完成 6 个汉字手势，每个汉字 60 次，共 360 次。实验结果见表 3，系统在基线情景中识别率最高达到 98%，在动态噪声情景下仍能保持在 96% 以上，表明系统在复杂噪声条件下仍能稳定识别，具备良好的抗干扰性能。

为评估系统在复杂汉字手势识别中的性能，本研究在原有 6 个汉字的基础上，新增 12 个笔画复杂、结构多样的汉字，以模拟实际应用中可能出现的复杂手势场景。数据集按照 70% 训练集、30% 测试集划分，并采用 10 折交叉验证以防止过拟合，保证实验结果可靠性。实验结果如图 8 所示，系统在新增复杂汉字上的平均识别精度仍达 98.31%，表明改进的轻量化模型在处理笔画繁复、交叉封闭特征丰富的手势任务时仍保持稳定、准确的识别能力。

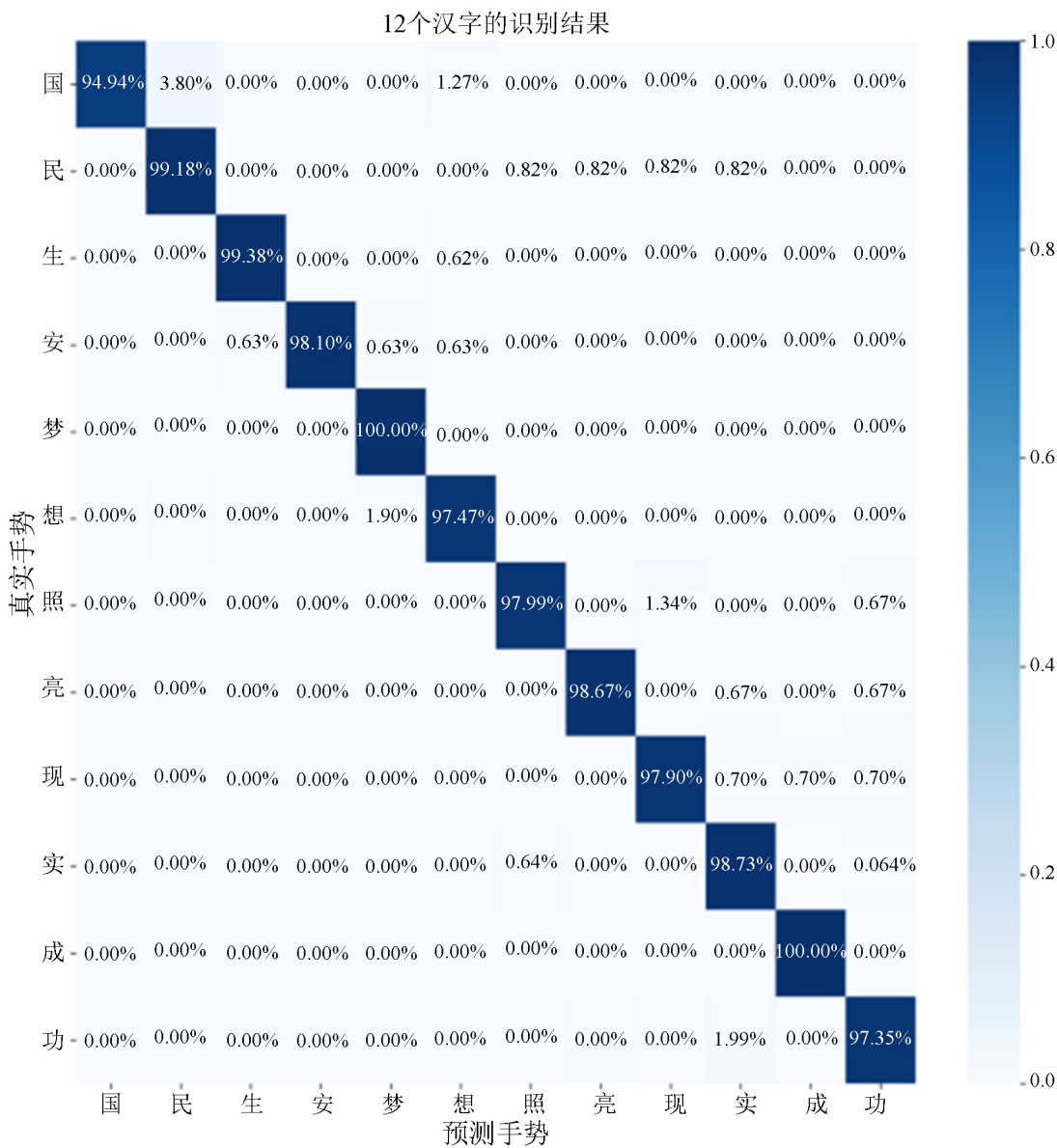


Figure 8. Confusion matrix of common complex Chinese characters  
图 8. 常见复杂汉字的混淆矩阵

为进一步验证所提系统在跨域识别中的优势，本研究将其与 3 种最新手势识别方法进行对比，分别为 Wi-AM、ANN-to-SNN 和 Wi-Learner，对比结果见表 4。Wi-AM 基于元学习框架，在域内和跨域场景的平均识别准确率分别为 90.40%和 86.76%；ANN-to-SNN 利用脉冲神经网络(SNN)，在域内和跨域的平均识别率分别为 97.67%和 93.63%；Wi-Learner 通过新型元学习器实现少样本快速适应，域内与跨域准确率分别为 93.20%和 94.90%。相比之下，本系统在域内和跨域场景的平均识别率分别达到 98.88%和 97.74%，显著优于其他模型，充分体现了其在细粒度汉字手势识别中的鲁棒性与跨场景泛化能力。

**Table 4.** Comparison with other models  
**表 4.** 与其他模型的比较

模型	域内准确率/%	跨域准确率/%
Wi-AM	90.40	86.76
ANN-to-SNN	97.67	93.63
Wi-Learner	93.20	94.90
本模型	98.88	97.74

5. 结论

本文提出了一种改进 AlexNet 并融合 SE 注意力机制的轻量化细粒度汉字手势识别方法，通过标签矩阵与双天线增强信号采集完整性，并结合马尔可夫转移场映射实现时序特征强化。实验结果显示，系统在多场景、多用户条件下均具有优异性能：整体识别准确率达到 98.88%，跨用户与跨场景平均准确率分别为 97.49%和 97.74%；即便在笔画相似或复杂汉字手势上，识别精度仍保持在 98%以上。在环境噪声、金属干扰及动态干扰条件下，识别率仍超 96%，表明模型在保证高精度识别的同时具有良好的鲁棒性和跨场景泛化能力。该方法实现了从信号采集、特征增强到轻量化卷积神经网络训练的全流程优化，为智慧课堂的手势交互提供了高效可靠的解决方案。未来，可进一步探索多模态传感融合及自适应学习策略，以增强系统对复杂交互场景和新型手势的适应性。

基金项目

本项目受新疆维吾尔自治区自然科学基金(2022D01C54)、新疆大学博士科研启动基金(202212120001)的项目资助。

参考文献

[1] 程时伟, 沈泽鹏, 张章伟. 基于眼动跟踪的移动智慧课堂可视交互系统[J]. 浙江工业大学学报, 2023, 51(4): 362-371.

[2] 陈威宇. 基于 RFID 的人体动作识别研究[D]: [硕士学位论文]. 南京: 南京信息工程大学, 2024.

[3] Lin, M., Wu, J., Zou, J., Meng, J., Wang, W. and Wu, J. (2024) Retired Battery Screening Based on Markov Transition Field and Swin Transformer. *IEEE Transactions on Transportation Electrification*, **10**, 4217-4227. <https://doi.org/10.1109/tte.2023.3306393>

[4] Fang, T., Zhang, J., Qi, D. and Gao, M. (2024) BLSENet: A Novel Lightweight Bilinear Convolutional Neural Network Based on Attention Mechanism and Feature Fusion Strategy for Apple Leaf Disease Classification. *Journal of Food Quality*, **2024**, 1-11. <https://doi.org/10.1155/2024/5561625>

[5] Yuan, Y. and Lian, Y.H. (2025) Design Analysis of Human-Computer Interaction and Information Communication in Artificial Intelligence Environments. *International Journal of E-Collaboration*, **21**, 1-13. <https://doi.org/10.4018/ijec.371626>

[6] Qiu, Q., Wang, T., Chen, F. and Wang, C. (2024) LD-Recognition: Classroom Action Recognition Based on Passive

- 
- RFID. *IEEE Transactions on Computational Social Systems*, **11**, 1182-1191. <https://doi.org/10.1109/tcss.2023.3234423>
- [7] Pu, Y., Chen, J., Sheng, X., Wang, Z., Li, J. and Tang, H. (2025) Data-Driven Fault Diagnosis with Markov Transition Field and Deep Learning for Water Management of Automotive PEMFC Power System. *International Journal of Hydrogen Energy*, **148**, Article 150022. <https://doi.org/10.1016/j.ijhydene.2025.06.212>
- [8] Huang, M., Wan, N. and Zhu, H. (2025) Reconstruction of Structural Acceleration Response Based on CNN-BiGRU with Squeeze-and-Excitation under Environmental Temperature Effects. *Journal of Civil Structural Health Monitoring*, **15**, 985-1003. <https://doi.org/10.1007/s13349-024-00859-w>
- [9] Qiu, J., Lu, X., Wang, X., Chen, C., Chen, Y. and Yang, Y. (2024) Research on Image Recognition of Tomato Leaf Diseases Based on Improved Alexnet Model. *Heliyon*, **10**, e33555. <https://doi.org/10.1016/j.heliyon.2024.e33555>.
- [10] Aliferis, C. and Simon, G. (2024) Overfitting, Underfitting and General Model Overconfidence and Under-Performance Pitfalls and Best Practices in Machine Learning and AI. In: *Health Informatics*, Springer, 477-524. [https://doi.org/10.1007/978-3-031-39355-6\\_10](https://doi.org/10.1007/978-3-031-39355-6_10)