

# 基于 yolov11 手写汉字的识别

## 一、主要思路

本项目旨在实现基于深度神经网络的手写汉字识别功能，采用 YOLOv11 目标检测模型作为基础框架，构建完整的识别系统。项目的整体流程如下：首先，将任务形式定义为多类别目标检测问题：每张图像中包含一个手写汉字，模型需要识别出其具体类别。项目使用课程提供的汉字手写图像数据集，共包含 501 汉字，每类约 50 张图像，并使用 `char_dict.json` 文件建立类别索引。在数据处理阶段，使用自定义脚本 `get_dataset.py` 对原始数据进行清洗、预处理，并自动生成 YOLO 格式的数据集，完成训练集、验证集和测试集的划分，同时构建了数据配置文件 `dataset.yaml`。模型训练部分基于轻量级的 YOLOv11 框架，在已有预训练模型的基础上进行微调。通过 `train.py` 脚本，我们设置了合理的超参数，进行多轮训练，并保留最优性能模型 `best_model.pt`。训练完成后，使用 `test.py` 脚本对模型性能进行评估，主要指标包括 `mAP@0.5`、`mAP@0.5:0.95`，并生成学习曲线、混淆矩阵等图表，直观展示模型识别能力。最后，我们利用 `predict.py` 脚本对任意输入图像进行预测，实现模型的部署验证。模型能够准确识别图像中的手写汉字，并输出检测框、类别标签及置信度信息，完成从数据处理到模型预测的全流程闭环。

## 二、数据集处理

本项目所使用的数据集由课程提供，包含 500 个文件夹，每个文件夹中约有 50 张手写体汉字图片。每个文件夹的名称是该汉字对应的 Unicode 编码，例如“00699”表示汉字“的”。此外，配套的 `char_dict.json` 文件用于将这些 Unicode 编码映射为实际汉字字符，形成类别标签索引。为了适配 YOLOv11 目标检测模型的训练需求，本项目编写了 `get_dataset.py` 脚本对原始数据进行结构化处理。主要流程如下：

### 映射构建与清洗：

首先加载 `char_dict.json` 文件，并筛选出与实际图像文件夹匹配的有效

Unicode 编码。剔除空字符、无效类别后，构建连续编号的类别 ID (`class_id`)，确保 YOLO 模型的标签是从 0 开始的整数类型。

### 图像扫描与标签提取：

脚本支持常见图像格式（如 .jpg、.png 等），遍历每个文件夹下的图片，并

根据其上级文件夹名确定所属类别。每张图像被标注为一个完整覆盖图像区域的边界框（观察汉字在图像中居中且唯一），以满足 YOLO 所需的标注格式：1 <class\_id> 0.5 0.5 1.0 1.0 即类别编号 + 归一化的中心点 + 全图范围的宽高。

### 数据集划分：

使用 `sklearn.model_selection.train_test_split` 对图像及其标签进行两次分层抽样划分，确保各类别在训练集、验证集和测试集中分布均匀：测试集比例：15%；验证集比例：15%；训练集比例：70% 同时为每张图像生成对应的 .txt 标注文件，文件名由其原始路径中的父文件夹名和图像名组合而成，确保唯一性，避免重名冲突。最后生成 `dataset.yaml` 配置文件，记录数据集根路径、图像目录、类别数量和类别名称等信息，供 YOLOv11 模型加载使用。为保证训练过程对中文字符的兼容性，该文件采用 UTF8 编码保存，类名列表也通过 `json.dumps(..., ensure_ascii=False)` 写入，确保汉字不被转义。

### 三、模型设计

在本项目中，我们选择使用 YOLOv11n 作为手写汉字识别的核心模型。YOLOv11 是目标检测领域的新一代深度神经网络模型，具有速度快、精度高、部署轻量化等优点，广泛应用于各种实际场景。本项目所使用的 `yolov11n` 为其中的最小版本（Nano），在保持基本检测能力的前提下，显著降低了参数量与计算开销，非常适用于小目标检测任务与实时推理需求。

YOLOv11 相较于前代版本（如 YOLOv8）在架构上进行了多项关键改进，旨在在保持高精度的同时进一步提升模型的计算效率和部署灵活性。

**C3k2 模块：**该模块是对传统 CSP（Cross Stage Partial）结构的优化，采用两个较小的卷积核替代大型卷积操作，从而减少计算开销并提升特征提取效率。

**C2PSA（Channel-wise and Point-wise Spatial Attention）模块：**引入空间注意力机制，提升模型对关键区域的关注度，从而提高检测精度，尤其在小目标和遮挡场景中具有优势。

**LSKA（Large Kernel Separable Convolution Attention）模块：**通过将二维卷积核分解为水平和垂直的一维卷积，降低计算复杂度的同时增强模型对目标形状的学习能力。

此外，YOLOv11 保留了 YOLO 系列一贯的单阶段检测范式，支持目标检测、实例分割、图像分类等多种计算机视觉任务，具备良好的通用性和扩展性。通过

引入多项架构创新，在保持模型精度的同时显著提升了计算效率和部署灵活性，为本项目的手写汉字识别任务提供了有力的技术支持。