**基于CRNN的中文场景文本识别系统**

**一、项目背景**

在当今数字化信息爆炸的时代，文本识别技术在诸多领域展现出了重要的应用价值，如文档数字化、图像搜索、自动驾驶等。本项目是一个端到端的中文场景文本识别系统,旨在实现一个高效准确的文本识别系统，采用卷积循环神经网络（CRNN）架构，结合 CTC（Connectionist Temporal Classification）损失函数，实现对自然场景图片中不定长中文文本的识别。系统使用360万中文数据集训练，提供完整的训练、验证和模型管理功能。

**二、项目目标**

本项目的主要构建一个基于 CRNN 的文本识别模型，能够对特定数据集（360CC数据集）中的图像文本进行准确识别。具体包括：

实现数据加载与预处理，确保输入数据的质量和一致性。

构建 CRNN 模型，结合卷积层和循环层，提取图像特征并进行序列建模。

定义合适的损失函数和优化器，训练模型以提高识别准确率。

对训练好的模型进行评估和测试，验证其性能。

**三、项目结构与代码模块**

1、数据处理模块

数据集：

使用Synthetic Chinese String数据集，Synthetic Chinese String为中文字符识别数据集，包含360多万张训练图片，5824个字符。图片场景简单，白底黑字。

\_360cc.py：

定义了\_360CC数据集类，负责加载和预处理 360CC 数据集。包括读取图像和标签文件，将图像进行尺寸调整、归一化等操作，并将标签转换为字符串形式。

utils.py：

包含了一些数据处理和工具函数，如get\_batch\_label用于获取批量数据的标签，strLabelConverter用于字符串和标签之间的转换，get\_char\_dict用于读取字符字典。

2、模型构建模块

crnn.py：

定义了 CRNN 模型的架构，包括卷积层和双向 LSTM 层。CRNN类通过一系列卷积和池化操作提取图像特征，然后将特征输入到双向 LSTM 层进行序列建模，最后输出文本识别结果。BidirectionalLSTM类实现了双向 LSTM 层的具体结构。

3、训练与评估模块

train.py：

主训练脚本，负责加载配置文件、创建输出文件夹、初始化模型、定义损失函数和优化器，以及进行模型训练和验证。通过function.train和function.validate函数完成训练和验证过程，并保存最佳模型。

function.py：

包含了训练和验证过程中的核心函数，如train函数用于模型训练，validate函数用于模型验证，AverageMeter类用于计算和存储平均值和当前值。

4、测试模块

test.py：

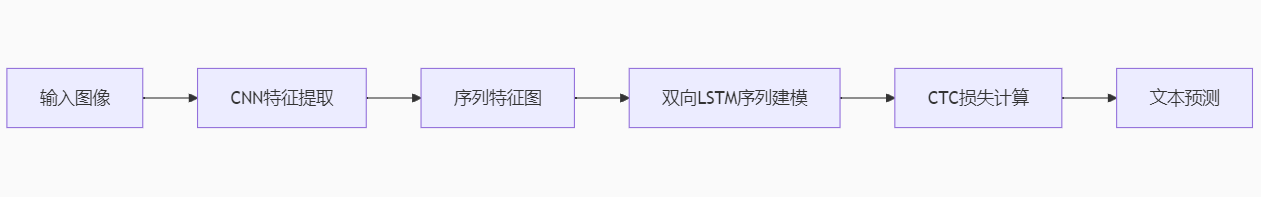
用于对训练好的模型进行测试，包括解析命令行参数、加载图像、进行预处理、模型推理和结果解码。

5、配置文件

360CC\_config.yaml：

包含了项目的所有配置信息，如 GPU ID、数据集路径、训练参数、模型参数等。

1. **模型架构**



1. CRNN 模型

采取CNN特征提取 + BiLSTM序列建模 + CTC损失

CRNN 模型结合了卷积神经网络（CNN）和循环神经网络（RNN）的优点，能够有效地处理图像序列数据。具体结构如下：

卷积层：通过七层卷积和最大池化进行四次下采样操作，提取图像的特征图。参数包括卷积核大小、步长、填充等，通过convRelu函数实现。

特征序列转换：采用双向 LSTM 层对卷积层输出的特征序列进行建模，捕捉序列中的上下文信息。

输出层：通过线性层将 LSTM 层的输出映射到字符类别空间，最后使用F.log\_softmax函数进行分类。

2、损失函数

使用 CTC 损失函数，它能够处理序列长度不一致的问题，无需对输入序列和输出序列进行对齐。

**五、项目流程**

1、数据准备

准备 360CC 数据集，包括图像文件和标签文件。

配置360CC\_config.yaml文件中的数据集路径和相关参数。

2、模型训练

运行train.py脚本，加载配置文件，创建输出文件夹。

初始化 CRNN 模型，定义损失函数和优化器。

进行模型训练，通过function.train函数更新模型参数。

在每个 epoch 结束后，使用function.validate函数对模型进行验证，保存最佳模型。

3、模型测试

运行test.py脚本，加载训练好的模型和测试图像。

对测试图像进行预处理，输入到模型中进行推理。

解码模型输出，得到最终的文本识别结果。

**六、实验结果**

1、最佳模型：mixed\_second\_finetune\_acc\_97P7.pth

2、验证集准确率：97.7%

**七、项目总结**

本项目实现了一个基于 CRNN 的文本识别系统，通过合理的数据处理、模型构建和训练策略，能够对图像中的文本进行准确识别。未来可以考虑引入更多的数据集进行训练，优化模型架构和损失函数，提高模型的泛化能力和识别准确率。