# 项目需求文档

基于COCO数据集的模型剪枝实现

## 一、项目背景

本项目旨在利用COCO数据集，开发并实现一个基于卷积神经网络（CNN）的模型剪枝算法。通过模型剪枝，优化模型参数，提升模型在资源受限环境下的性能。项目需采用Python编程语言和PyTorch框架进行实现。

## 二、项目目标

1. 构建一个基于COCO数据集的卷积神经网络模型。

2. 实现全局非结构化剪枝算法，以减少模型参数。

3. 在剪枝前后对模型进行训练与评估，以验证剪枝效果。

## 三、数据处理

1. 使用COCO 2017数据集的训练集和验证集：

- 训练集：118,287张图像，约18GB。

- 验证集：5,000张图像，约1GB。

2. 下载地址：

- 训练集图像：[train2017.zip](http://images.cocodataset.org/zips/train2017.zip)

- 验证集图像：[val2017.zip](http://images.cocodataset.org/zips/val2017.zip)

- 标注文件：[annotations\_trainval2017.zip](http://images.cocodataset.org/annotations/annotations\_trainval2017.zip)

3. 数据存储目录为`./coco\_dataset`。

4. 数据预处理包括：

- 图像大小调整到64x64。

- 图像标准化，均值为[0.5, 0.5, 0.5]，标准差为[0.5, 0.5, 0.5]。

5. 数据加载时需进行数据增强处理，以提高模型泛化能力。

## 四、文档目录

project/

│

├── main.py # 主程序入口

├── model.py # 模型定义

├── prune.py # 剪枝算法实现

├── utils.py # 辅助函数

├── data.py # 数据加载和预处理

│

├── coco\_dataset/ # COCO 数据集目录

│ ├── train2017/ # 训练集图像目录

│ ├── val2017/ # 验证集图像目录

│ └── annotations/ # 标注文件目录

│

├── README.md # 项目说明文件

├── requirements.txt # 依赖库列表

└── scripts/ # 脚本目录

├── download\_coco.py # 数据集下载脚本

└── preprocess.py # 数据预处理脚本（可选）

## 五、模型架构

1. 模型采用卷积神经网络（CNN），包含以下层次结构：

- 输入层：大小为64x64的RGB图像。

- 隐藏层：包含两个全连接层，每层包含300个和100个神经元，激活函数为ReLU。

- 输出层：10个神经元，对应10个分类，激活函数为Softmax。

2. 模型在`model.py`文件中定义。

## 六、模型训练

1. 使用交叉熵损失函数。

2. 优化器采用随机梯度下降（SGD），学习率为0.01。

3. 训练参数：

- 批量大小：16。

- 训练轮数：10。

4. 每轮训练结束后，需输出训练损失和验证准确率。

5. 训练和测试过程在`utils.py`文件中实现。

## 七、模型剪枝

1. 实现全局非结构化剪枝算法，使用L1范数进行剪枝。

2. 剪枝比例为20%（即剪除20%的参数）。

3. 剪枝后移除无效参数，并重新训练模型。

4. 剪枝过程和参数移除在`prune.py`文件中实现。

## 八、代码组织

项目代码需按以下文件结构组织：

- `main.py`: 主程序入口，负责调用其他模块进行数据加载、模型训练和剪枝。

- `model.py`: 定义模型架构。

- `prune.py`: 实现剪枝算法。

- `utils.py`: 包含辅助函数，如训练、测试和数据加载。

- `data.py`: 实现数据加载和预处理功能。

## 九、验收

### 9.1验收标准

1. 项目代码结构清晰，模块划分合理。

2. 成功下载并处理COCO数据集，数据预处理合理有效。

3. 模型在训练前后能正常运行，并输出训练损失和验证准确率。

4. 模型剪枝后能正常运行，并重新训练模型，输出剪枝后训练损失和验证准确率。

5. 提供详细的项目文档，包括使用说明和代码注释。

### 9.2交付内容

1. 完整的项目源代码，包括所有必要的依赖文件。

2. 项目文档，包含安装、运行和使用说明。

3. 训练和剪枝的模型及其性能评估结果。

### 9.3时间要求

项目开发周期为4周，自项目启动之日起计算。具体时间安排如下：

1. 第1周：完成数据下载、预处理和初步模型构建。

2. 第2周：完成模型训练和评估。

3. 第3周：实现并测试模型剪枝算法。

4. 第4周：重新训练剪枝后的模型，完成项目文档。

### 9.4其他要求

1. 代码需符合PEP8编码规范，确保代码的可读性和可维护性。

2. 乙方需提供详细的代码注释和使用文档，以便甲方理解和使用代码。

**甲方联系人**

- 姓名：张三

- 职位：项目经理

- 联系方式：zhangsan@example.com

**乙方联系人**

- 姓名：李四

- 职位：开发工程师

- 联系方式：lisi@example.com

请乙方根据以上需求，在规定时间内完成项目开发并提交相关成果。期待与乙方的合作，共同实现项目目标。