**CNN训练调优——优化器、学习率与正则化策略深度分析**

1. **实验背景与目的**

本实验旨在基于 CIFAR-10 数据集，系统性比较不同优化器（如SGD、Adam等）、学习率调整策略（如StepLR、Cosine Annealing等）以及正则化技术（如weight decay），深入探讨这些训练策略对CNN模型在收敛速度、最终精度、泛化能力与资源消耗等方面的影响。通过多组控制变量实验，评估模型在不同配置下的表现，提升训练技巧的实际掌握能力。

二、数据集与数据增强策略

**数据集**：CIFAR-10，包含10类RGB图像，共60,000张（训练集50,000张，测试集10,000张），每张图像为32×32分辨率。

**统一的数据增强策略**：随机裁剪（RandomCrop）、随机水平翻转（RandomHorizontalFlip）、标准化处理（对均值与方差进行标准化）

三、模型选择与结构说明

本实验固定使用ResNet-18标准结构，随机初始化权重（pretrained=False），以确保不同训练策略对性能的影响更加可控与明显。

四、对比实验设计

为控制变量分析，本实验对以下两个因素（学习率调度器、优化器）进行组合实验设计，如下表所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学习率调度器 | 优化器 | lr | momentum | weight\_decay |  |
| StepLR | SGD | 0.001 | 0.9 | 5e-4 | step\_size:30  gmma:0.1 |
| CosineAnnealingLR | SGD | 0.001 | 0.9 | 5e-4 | T\_max:50 |
| PlateauLR | SGD | 0.001 | 0.9 | 5e-4 | mode:min  factor:0.5  patience:5 |
| StepLR | Adam | 0.001 | 0.9 | 5e-4 | step\_size:30  gmma:0.1 |
| CosineAnnealingLR | Adam | 0.001 | 0.9 | 5e-4 | T\_max:50 |
| PlateauLR | Adam | 0.001 | 0.9 | 5e-4 | mode:min  factor:0.5  patience:5 |
| StepLR | RMSprop | 0.001 | 0.9 | 5e-4 | step\_size:30  gmma:0.1 |
| CosineAnnealingLR | RMSprop | 0.001 | 0.9 | 5e-4 | T\_max:50 |
| PlateauLR | RMSprop | 0.001 | 0.9 | 5e-4 | mode:min  factor:0.5  patience:5 |

五、训练配置

训练轮数：50 epochs

Batch Size：16

损失函数：CrossEntropyLoss

训练平台：NVIDIA RTX 4060 GPU、PyTorch 2.0.1

六、实验指标

为全面评估不同策略的效果，采用以下4个指标：Top-1准确率（Accuracy）、训练时间（Time）、浮点运算量（FLOPs）、参数量（Params）、模型大小（Model Size）。

七、实验结果与分析

根据实验结果分析，不同优化器和学习率调度策略对模型性能和训练效率影响显著。Adam优化器在准确率方面表现优越，尤其是配合Cosine Annealing学习率调度时，达到最高的84.69%准确率，虽训练时间较长但效果最佳。SGD结合Cosine Annealing同样表现出较高准确率（82.89%）且训练时间更短，体现了良好的性能与效率平衡。相比之下，RMSprop无论搭配何种学习率策略，准确率均较低且训练时间较长，表现最差。综上，推荐在保证较高准确率的前提下优先采用Adam或SGD优化器，配合Cosine Annealing调度策略，以实现训练效率和模型性能的最佳权衡。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 训练策略 | Acc(%) | Time(s) | FLOPs(M) | Params(M) | Size(MB) |
| SGD + StepLR | 71.53 | 2217.40 | 37.25 | 11.18 | 42.73 |
| Adam + StepLR | 83.53 | 2481.84 | 37.25 | 11.18 | 42.73 |
| RMSprop + StepLR | 69.29 | 2471.46 | 37.25 | 11.18 | 42.73 |
| SGD + CosineAnnealing | 82.89 | 2225.59 | 37.25 | 11.18 | 42.73 |
| Adam + CosineAnnealingLR | 84.69 | 2464.61 | 37.25 | 11.18 | 42.73 |
| RMSprop + CosineAnnealingLR | 65.91 | 2445.37 | 37.25 | 11.18 | 42.73 |
| SGD + PlateauLR | 65.79 | 2208.69 | 37.25 | 11.18 | 42.73 |
| Adam + PlateauLR | 77.82 | 2450.51 | 37.25 | 11.18 | 42.73 |
| RMSprop + PlateauLR | 56.01 | 2515.82 | 37.25 | 11.18 | 42.73 |

|  |  |  |  |
| --- | --- | --- | --- |
| Step |  |  |  |
| Cosine |  |  |  |
| Plateau |  |  |  |
|  | SGD | Adam | Rmsprop |

图 Epoch-（Loss、Acc）曲线

八、结论

在模型训练中，优化器与学习率调度策略的组合对最终性能与训练效率具有决定性影响。其中，Adam优化器配合Cosine Annealing调度策略在准确率上表现最佳，适用于对性能要求较高的场景；而SGD配合Cosine Annealing在保持较高准确率的同时训练时间更短，在性能与效率之间实现了良好平衡。相比之下，RMSprop在本次实验中整体表现不佳，既未实现较高准确率，训练效率也较低。因此，建议在实际应用中优先选择Adam或SGD优化器，并搭配Cosine Annealing学习率调度策略，以在保证模型性能的同时提高训练效率。