

**《机器学习》课程论文**

**[](http://bysj.zjgsu.edu.cn/)**

题目：cifar-100数据集上的cnn算法研究

**学 院：统计与数学学院**

**专 业：应用统计**

**学 号：22020040175**

**学生姓名：朱晨**

**二○二三 年 四 月**

cifar-100数据集上的cnn算法研究

摘要

Cnn卷积神经网络是常被用于处理图像相关任务的一种深度学习模型，其拥有参数共享和稀疏连接两个特性，参数量远远小于全连接神经网络，且与图像的平移不变性相适应，在图片上能够取得很好的成绩，本文将基于cnn，在cifar100数据集上展开算法研究。

由于课程实验主要针对卷积神经网络的结构（卷积核大小，卷积层数量，激活函数和全连接层等）开展研究和对比，因此本文就在此基础上重点对于优化算法，学习策略展开研究和对比，最后，我还选择了几个比较先进的网络模型在cifar100上测试其效果，并取得了比较不错的结果。

**关键词：**图像分类，卷积神经网络，优化算法

**目录**

[一、 数据与模型简介简述 １](#_Toc11860954)

[二、 优化器选择 １](#_Toc11860955)

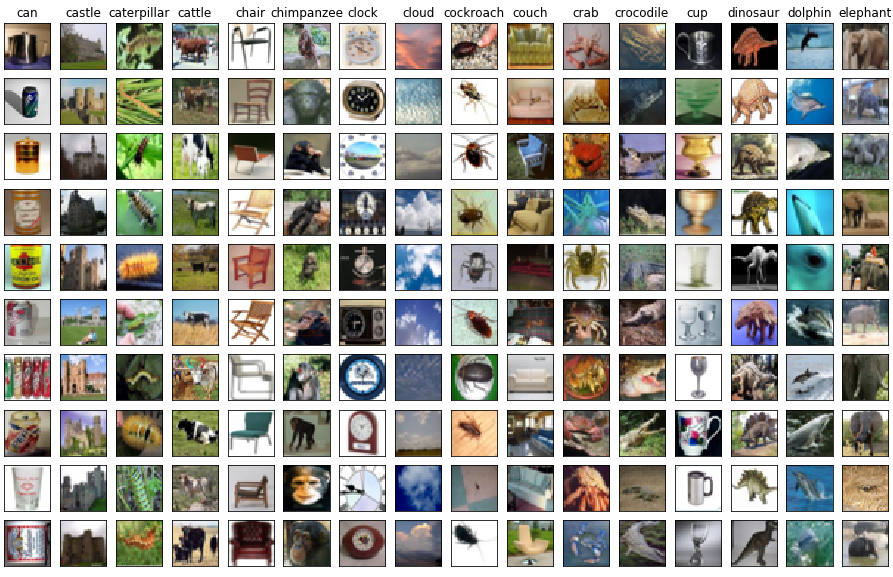
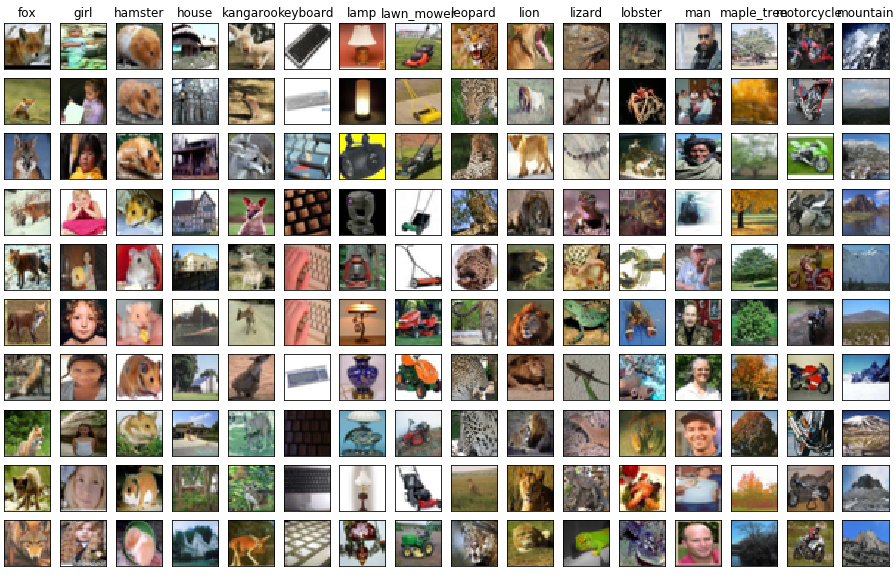
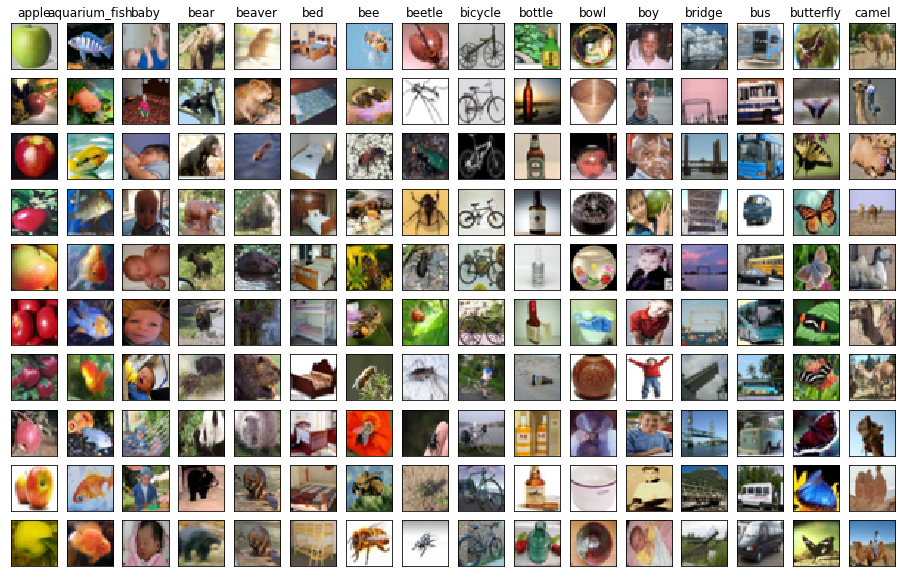
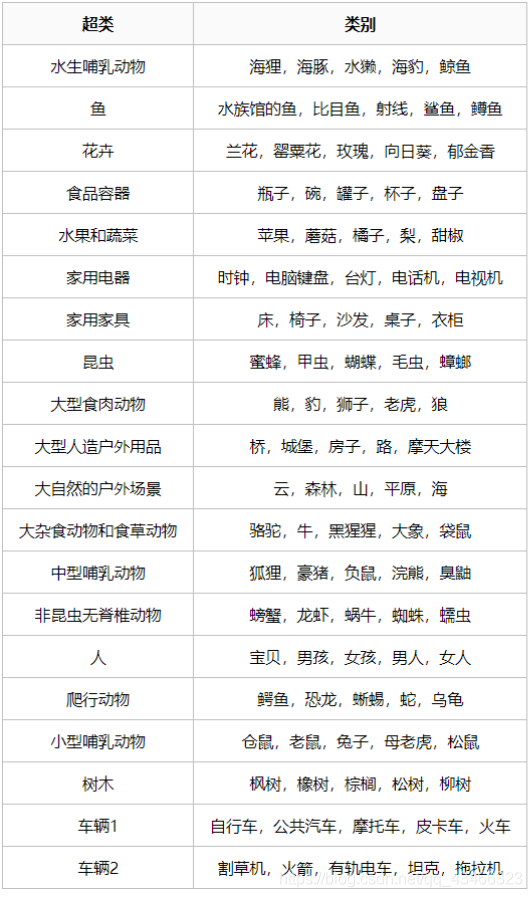
[三、 动态学习率调整 ２](#_Toc11860956)

[四、 更先进的模型 ９](#_Toc11860957)

[参考文献 １０](#_Toc11860958)

1. 数据与模型简介

（一）cifar100数据集介绍

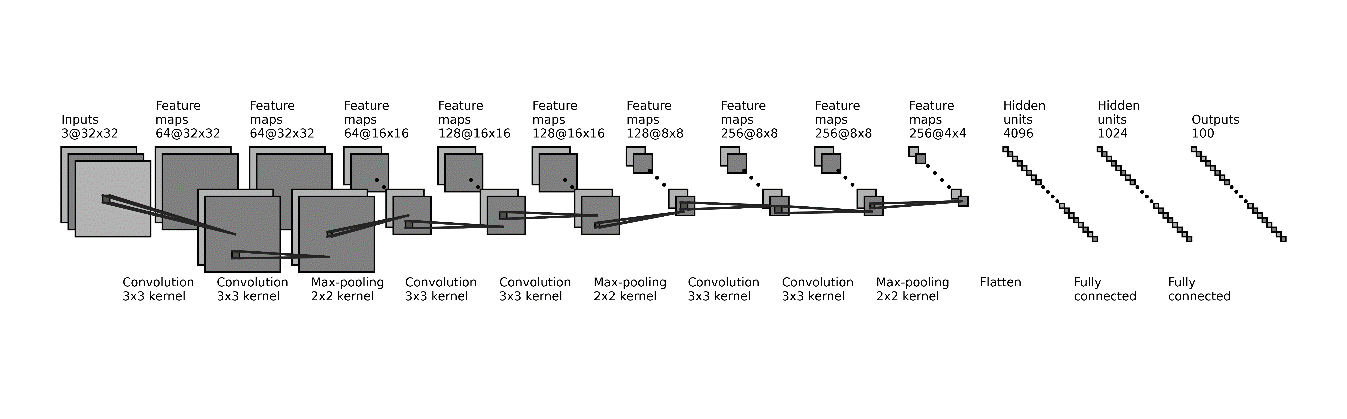
cifar-100类似于cifar-10，除了它有100个类，每个类包含600个图像，每类各有500个训练图像和100个测试图像。cifar-100 中的100个类被分成20个超类。每个图像都带有一个“精细”标签（它所属的类）和一个“粗糙”标签（它所属的超类）。以下是 cifar-100 中的图片浏览和类别列表：

可以看出，cifar-100图片种类多，分辨率低。许多图片即使是人也很难完全正确分辨，因此cifar-100数据集的top-1准确率不会像mnist或是cifar-10那么高。

（二）基础模型介绍

由于课程实验主要针对卷积神经网络的结构（卷积核大小，卷积层数量，激活函数和全连接层等）开展研究和对比，因此本文就在此基础上重点对于优化算法，学习策略展开研究和对比，最后，我还选择了几个比较先进的网络模型在cifar100上测试其效果，并取得了比较不错的结果。

本文实验所用的基础模型有三个卷积模块（一个卷积模块包含两次3\*3卷积和一次池化），输出通道数分别为64，128，256，最后是全连接层，隐藏层维度为1024，使用dropout，激活函数使用ReLU函数，使用BatchNorm层，下图是本文网络的示意图。



1. 优化器选择

优化器是在深度学习反向传播过程中，指引损失函数（目标函数）的各个参数往正确的方向更新合适的大小，使得更新后的各个参数让损失函数（目标函数）值不断逼近全局最小。

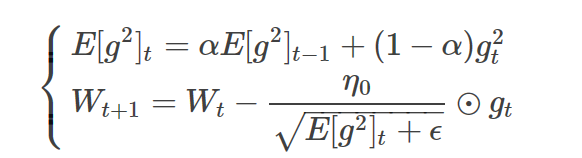
本文所介绍的优化器都是基于梯度下降算法，梯度下降法是最基本的一类优化器，其利用反向传播算法求出的梯度，可以直接找出损失函数下降最快的方向。

1. 常用优化器介绍

SGD（随机梯度下降法）：随机选取一批数据计算梯度并进行迭代优化，是实际应用中最简单的优化算法。

Momentum（动量法）：通过引入一个历史梯度信息，让优化器的更新值不完全又当前点的梯度决定，而是上一次更新值和当前点梯度的加权和，可以避免出现梯度不停来回震荡的情况，使模型的优化更加稳定。

RMSProp：

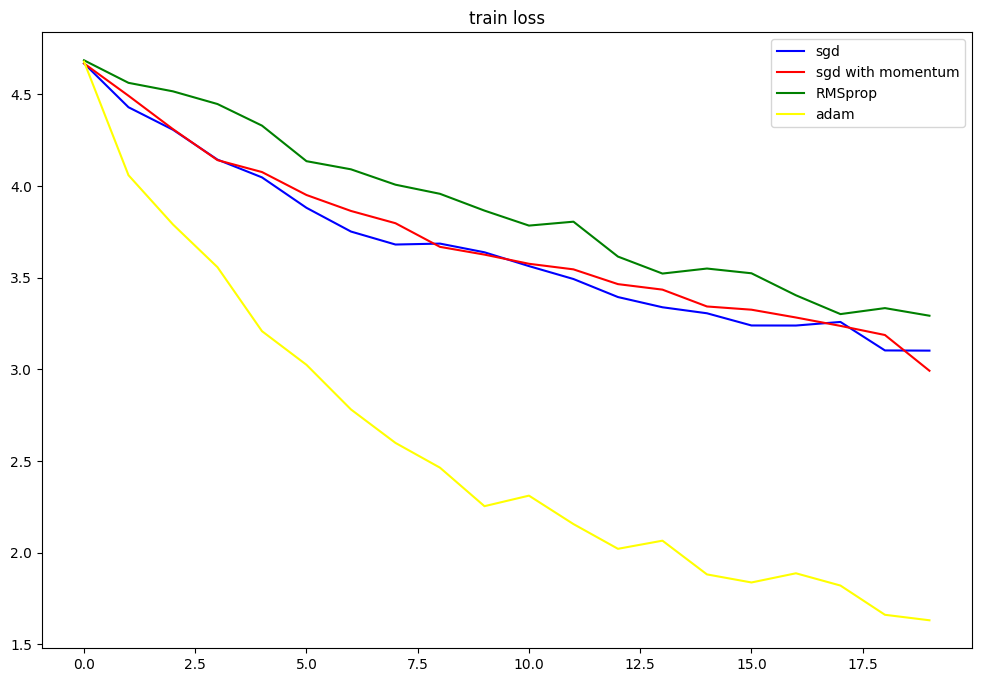


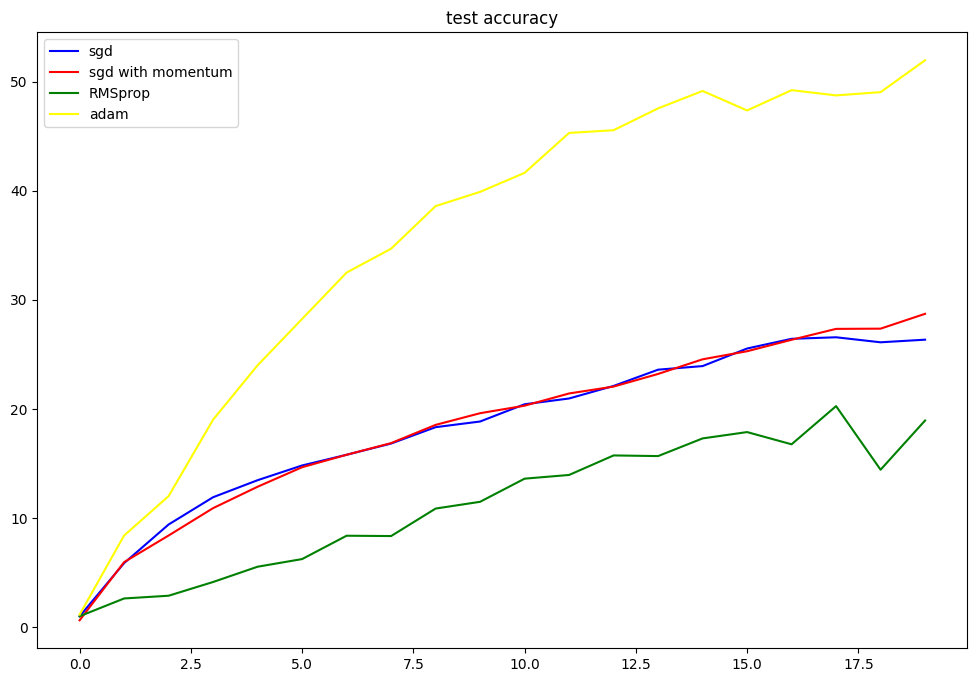
是一种自动调节学习率的方法，他可以自适应的调节学习率的大小，且能避免学习率越来越低的情况，已经被证明是一种有效且实用的深度神经网络优化算法。目前它是深度学习从业者经常采用的优化方法之一。

Adam：Adam算法可以看作是RMSProp和动量算法的结合，是一种非常优秀的优化算法，有模型优化速度快，和对学习率超参数不敏感的优点。

1. 优化器实验结果

本文分别使用以上四种优化器对模型进行优化，结果如下：





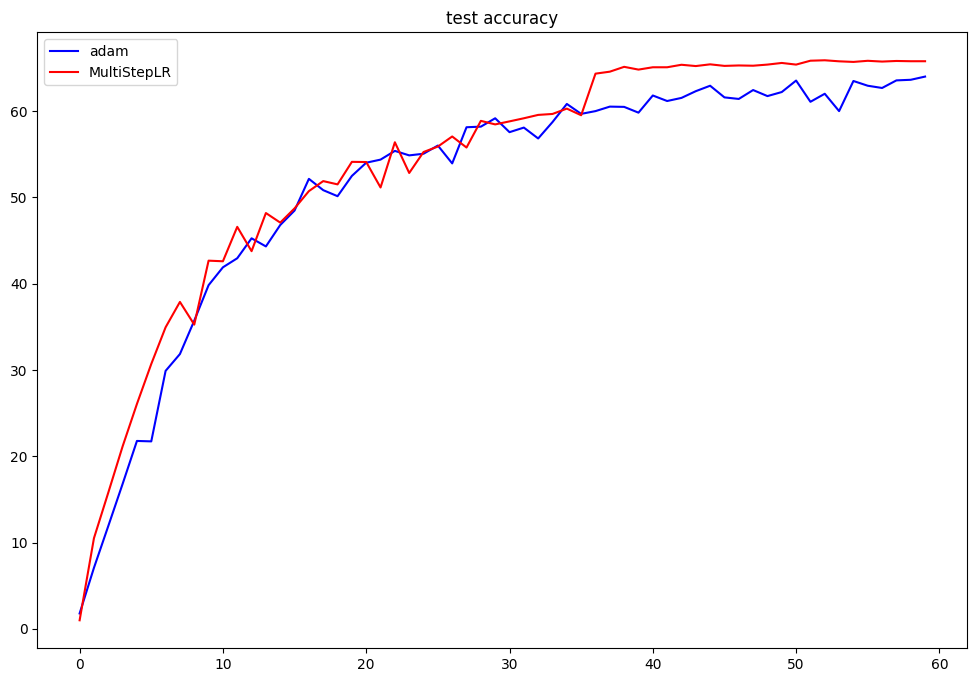
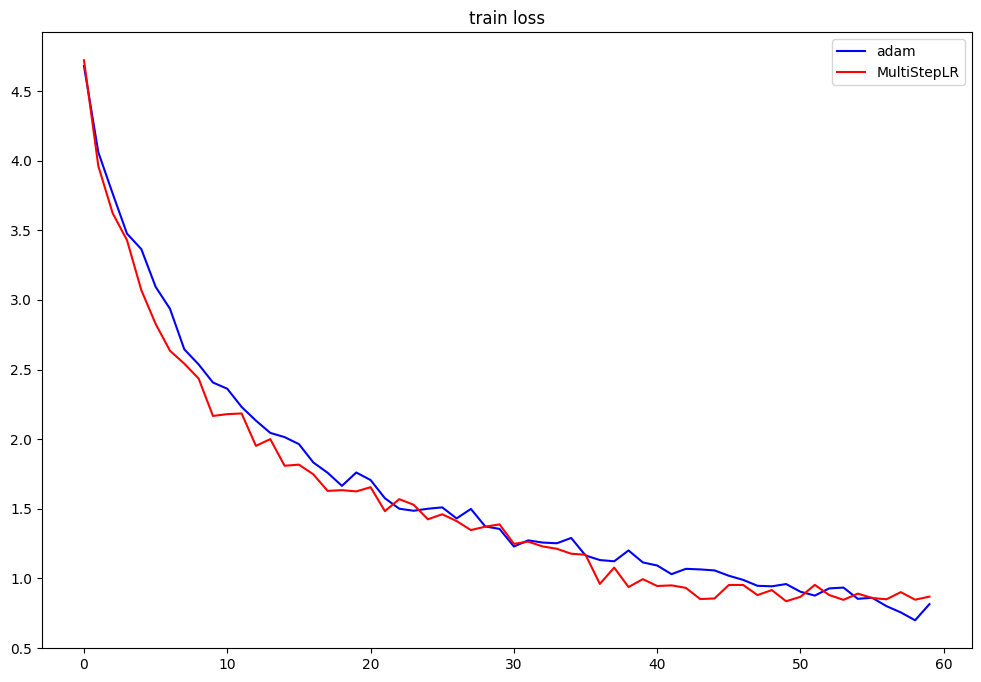
可以看出，Adam优化器相较于其他三种优化器在速度上有着明显的优势，本文后续实验均是在使用Adam优化器的基础上进行的。

1. 动态学习率调整

较大的学习率在学习初期有着不错的学习速度，但在后续学习中往往会陷入困境，难以收敛，而较低的学习率则会导致学习速度过慢，加重计算资源负担。因此，在学习的过程中，学习率往往不是相同的，通常在前几个epoch中较大，在后续的epoch中较小。尽管有些优化算法能够自适应的调节学习率，但是很多时候还是需要一个人为指定的学习率调整策略，下面我将介绍几种常用的学习率调整策略和实验结果。

1. MultiStepLR

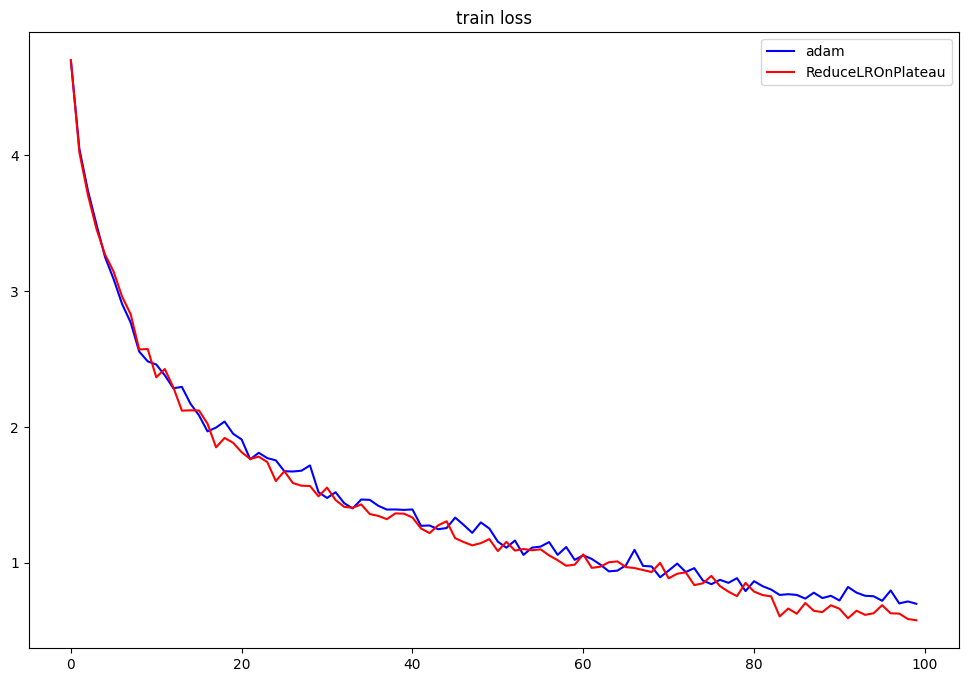
该算法可以在指定的epoch后让学习率衰减。是一种十分简单的学习率衰减策略。下面是是否使用该策略的对比。

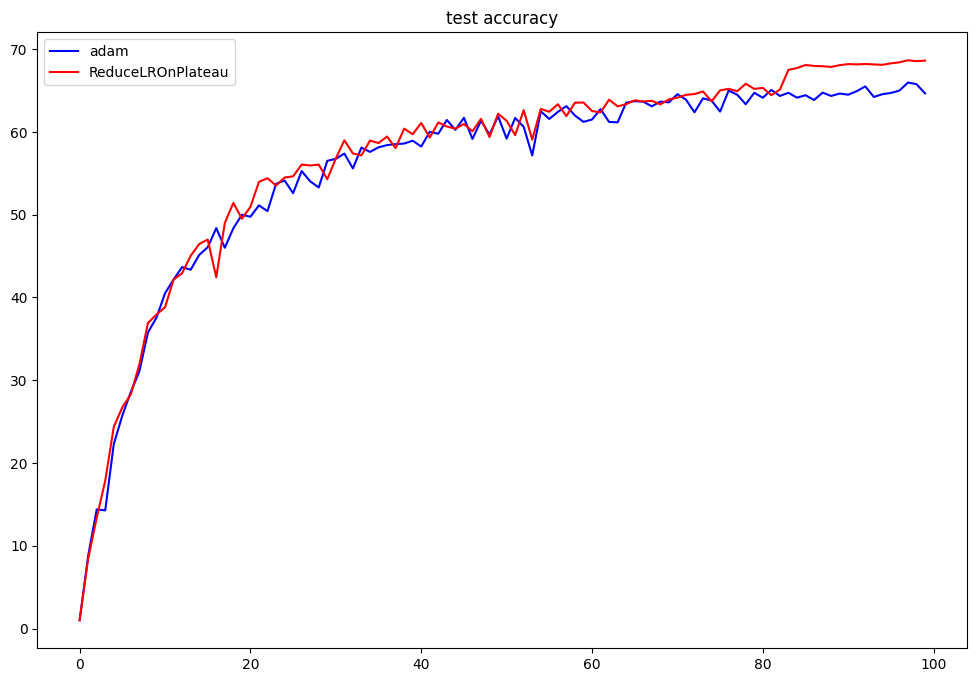


实验设置在35个epoch后进行学习率衰减，可以看到，在第36个epoch的优化速度有了十分显著的提升。如果不采用adam这种自适应学习率的优化器，效果会更加明显。

1. ReduceLROnPlateau

该算法可以指定一个值n，在n次epoch中，loss都没有取得新的最小值，即学习停滞时，自动衰减学习率。下面是是否使用该策略的对比。

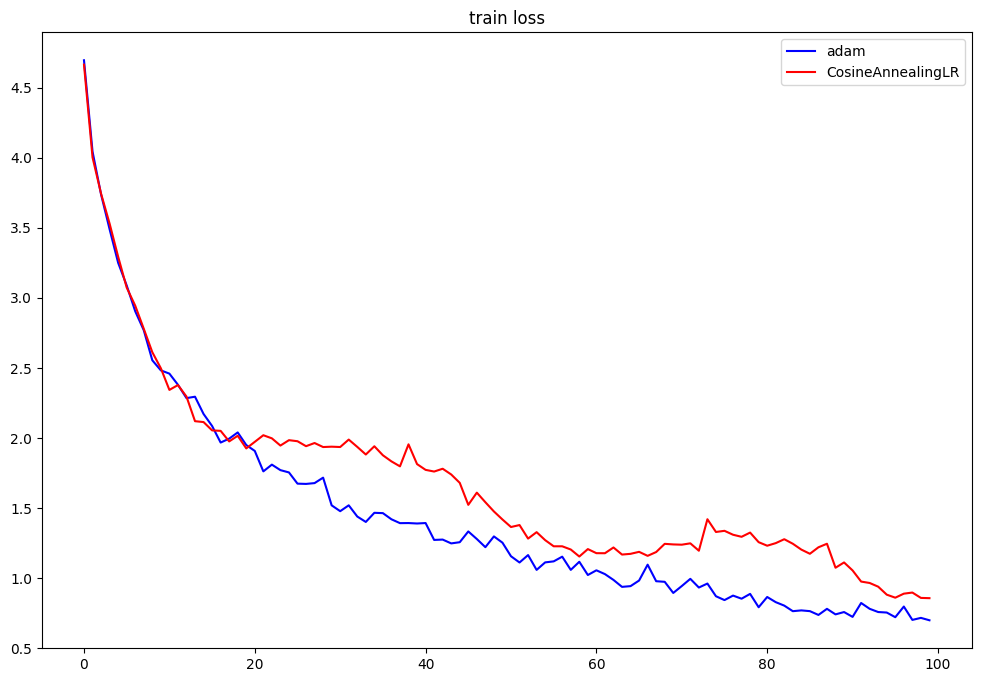


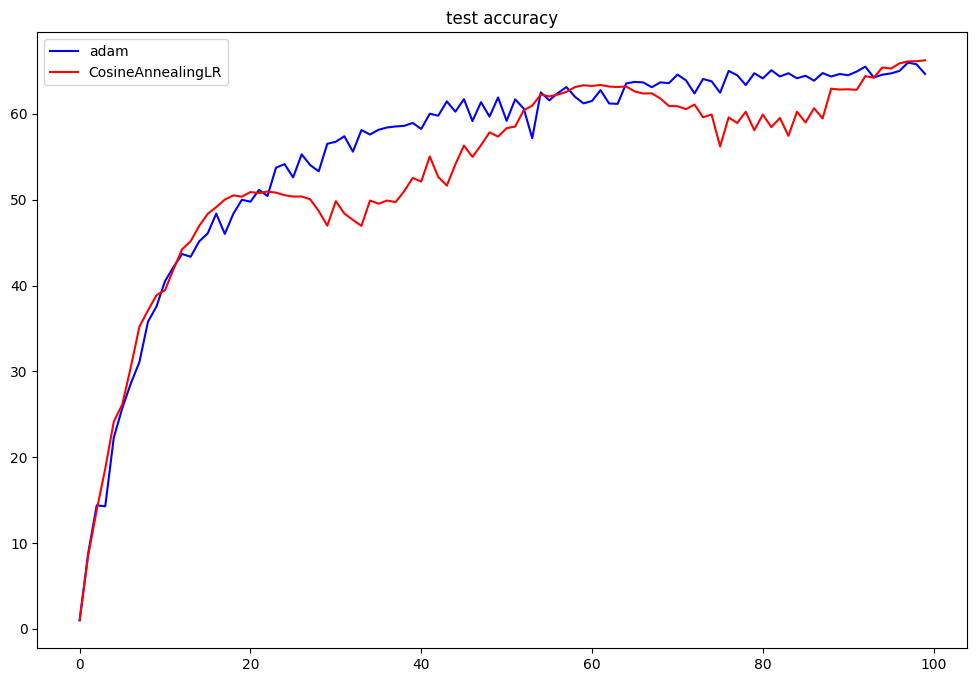


在实验中，学习率在epoch82和epoch96时发生了衰减，可以看到损失和测试集准确率在epoch82后发生了明显改进，说明自动学习率衰减能够适当调整学习率大小。

1. CosineAnnealingLR

该算法又称余弦退火算法，其可以设置学习率按照余弦曲线的趋势进行变化，在半个周期中下降，并在半个周期中上升。这种算法的特点是不仅可以实现学习率衰减，还可以让学习率增大来跳出局部极小值。下面是是否使用该策略的对比。

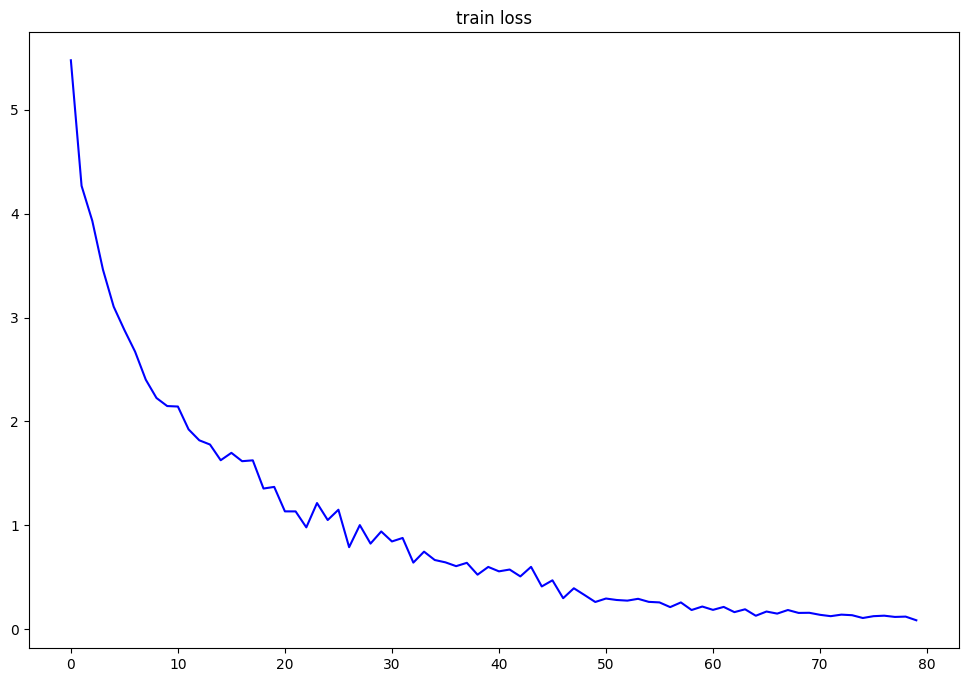


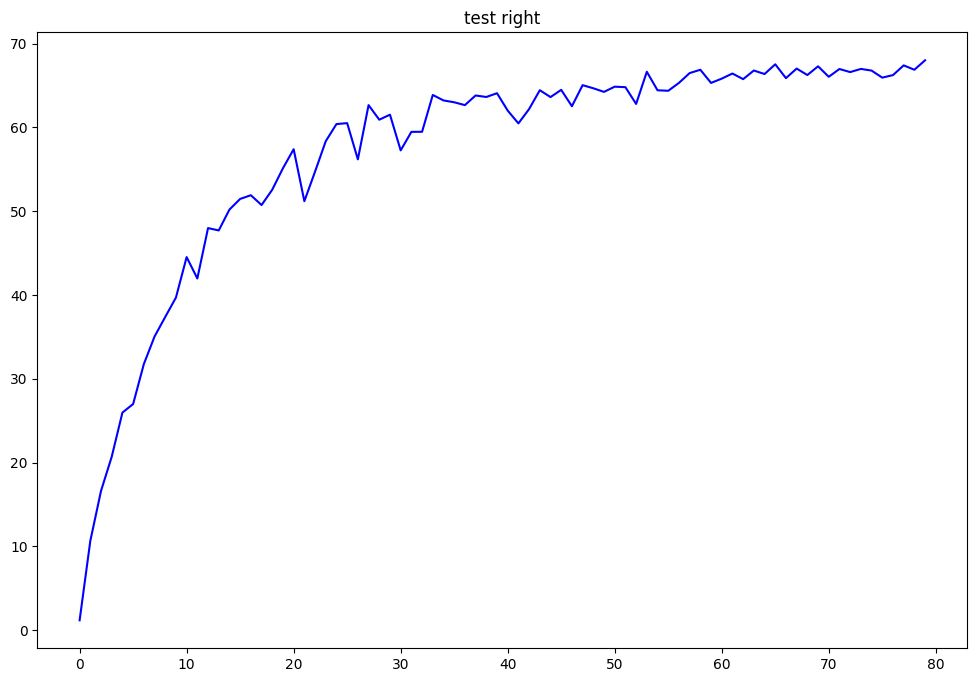


在这次实验中，设置Tmax=20，即余弦周期为40个epoch，可以看到在一个周期和两个周期，也就是大约epoch40和epoch80的位置，学习的速度明显变慢，甚至出现了倒退，但很快随着学习率的减小，能够回到正常水平。

1. 更先进的模型
2. resnet18

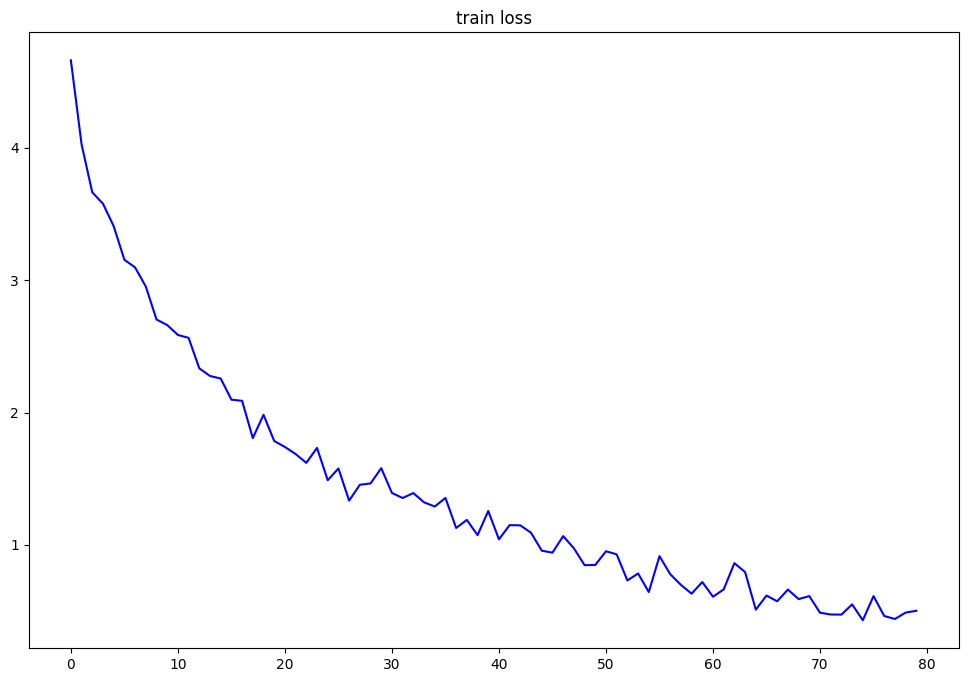
resnet是一个较为经典的网络，其提出了残差连接，使得深层网络的训练成为可能。Resnet18在实验中取得最高的准确率约为67%

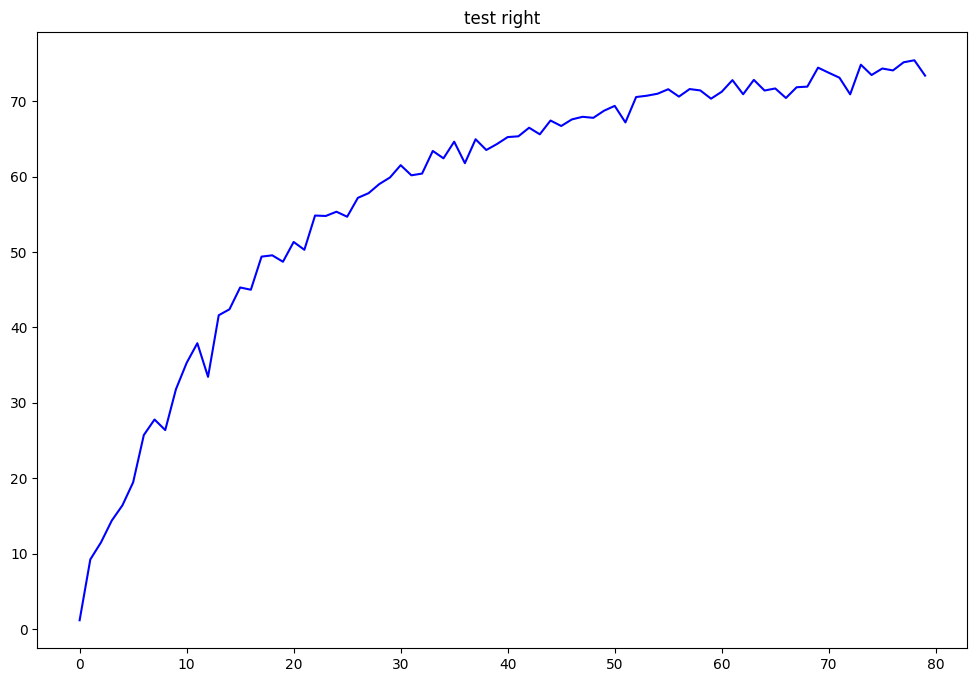




1. EfficientNet v2

EfficientNet v2是一个比较新的网络，其参数量不大的同时，在分类任务上具有很好的效果。使用这个网络，我们的实验最终取得了73%左右准确率





参考文献

[1] https://blog.csdn.net/qq\_41554005/article/details/119879911

[2] https://zhuanlan.zhihu.com/p/365969139

[3] http://www.spytensor.com/index.php/archives/32/