

深度学习 Lab2

PB20111686 黄瑞轩

1 实验目标

使用 pytorch 或者 tensorflow 实现卷积神经网络 CNN，在 CIFAR-10 数据集上进行图片分类。研究 dropout、normalization、learning rate decay、卷积核大小、网络深度等超参数对分类性能的影响。

2 实验过程和关键代码展示

注：为了展示核心功能，某些代码中一些无关紧要的部分在报告中被删去了。

2.1 数据集说明

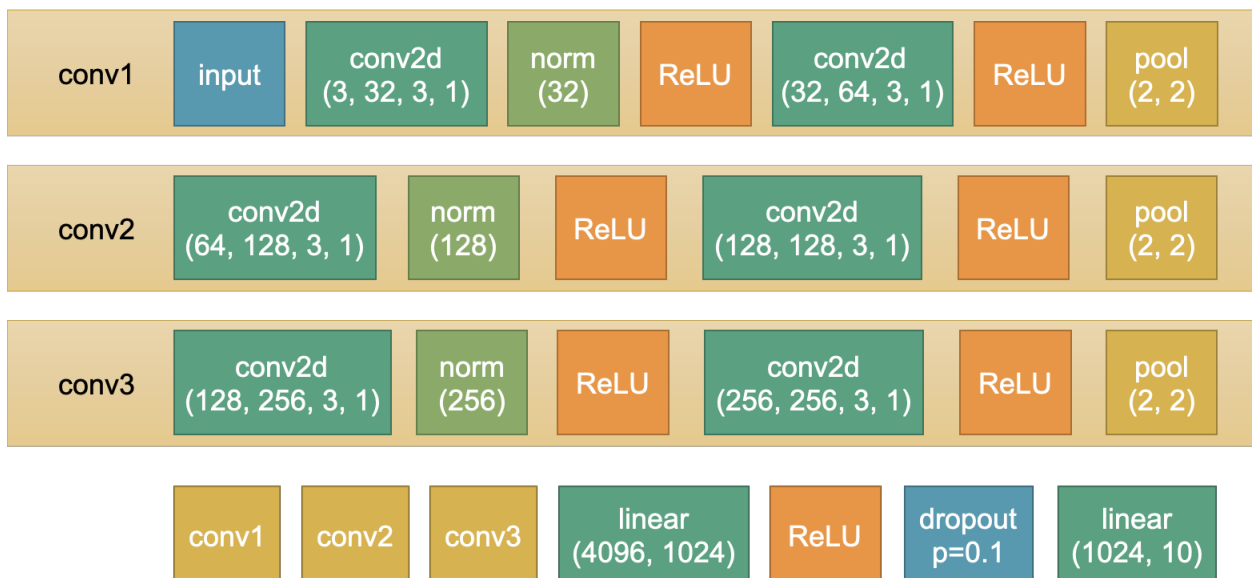
本次实验的数据集为 CIFAR-10 数据集，数据集已被分成训练集和测试集两部分，按照要求，再将训练集按 4:1 的比例分割为训练集和测试集。

pytorch 提供了专门载入 CIFAR-10 数据集的接口，可以通过 `train` 参数指定载入训练集或数据集。

```
1 data_raw = torchvision.datasets.CIFAR10(  
2     root="./",  
3     train=True,  
4     transform=torchvision.transforms.ToTensor(),  
5     download=False,  
6 )  
7  
8 # 数据集大小  
9 len_raw = len(data_raw)  
10 divide_rate = 0.8  
11 # 训练集大小  
12 len_train = int(len_raw * divide_rate)  
13  
14 # 使用 random_split 划分为训练集 train_set 和验证集 val_set  
15 train_set, val_set = random_split(data_raw, [len_train, len_raw - len_train])
```

2.2 实验过程说明

首先，根据 CNN 的经典结构搭建一个三层的小型 CNN 模型，这个模型的结构如图所示：



这个三层网络仅在输出层使用了一层 $p = 0.1$ 的 dropout，在各卷积层都引入了 normalization，学习率固定为 $lr = 0.001$ 、没有 decay，卷积核大小为 3。

对于一组特定的超参数，指定 batch 大小为 8、训练轮数为 20，在每轮训练之后都在验证集上预测获得 accuracy，然后画出训练过程中的训练集 loss 和验证集 accuracy 变化曲线，以 accuracy 收敛时的取值作为当前超参数的 metric。

在这组超参数基础上，实验过程将对 dropout、normalization、learning rate decay、卷积核大小、网络深度等超参数进行调参，以获取最佳效果。

这组超参数的结果如下：



训练过程中 loss 一直在下降，稳定时的精度为 78.9%。

3 超参数的调节

本实验采用单一变量的对照法，即在一组训练结果不错的超参数上每次选择一个超参数进行调参，然后将每个超参数最好的取值组合在一起。

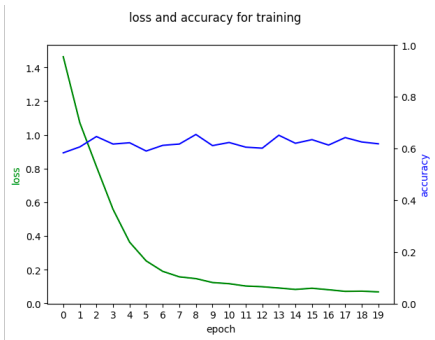
各种参数下的结果以最终 accuracy 为指标，在 accuracy 相近时训练的耗时（训练 accuracy 大约收敛的 epoch 序号）也将作为参考。

3.1 网络深度的影响

基础超参数具有三个卷积层，最大卷积层输出维度为 256，卷积层的数量会影响训练的速度，可以尝试 1、2、4、5 个卷积层时的效果。

卷积层数	最大卷积层输出维度
1	64
2	128
3（默认）	256
4	512
5	512

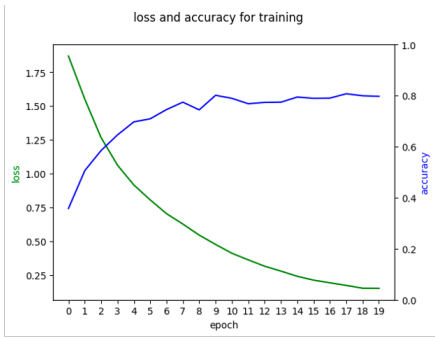
3.1.1 不同卷积层数的效果



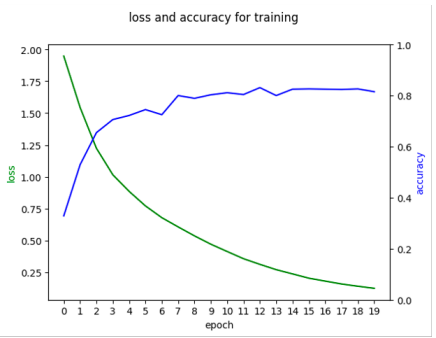
(a) 1 层卷积层结果



(b) 2 层卷积层结果



(c) 4 层卷积层结果



(d) 5 层卷积层结果

3.1.2 不同卷积层数的影响总结

卷积层数	最大卷积层输出维度	稳定时精度
1	64	61.8%
2	128	70.5%
3（默认）	256	78.9%
4	512	79.8%
5	512	81.5%

可以看到随着卷积层数的增加，稳定时的精度也在增加，但是卷积层数为 4 的情况下在 11 轮训练时就已经接近收敛，比卷积层数为 3 的效果好。

经过这层调参后，将卷积层数设置为 4。

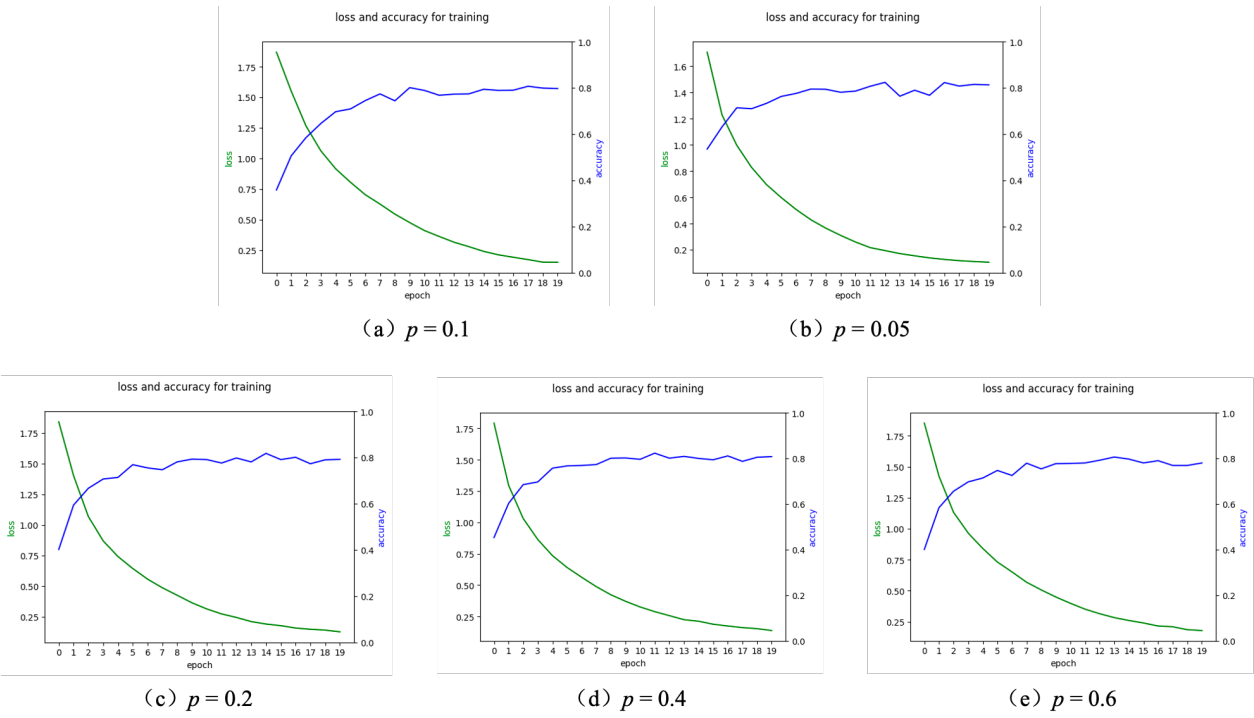
3.2 dropout 的影响

基础超参数仅在输出层使用了一层 dropout，这里先对 dropout 层 p 的取值调参，然后再尝试在卷积层使用 dropout 层。

3.2.1 p 的取值

取 $p = 0.05, 0.1(\text{默认}), 0.2, 0.4, 0.6$ 进行研究。

3.2.1.1 不同 p 的取值的效果



3.2.1.2 不同 p 的取值的影响总结

p 的取值	稳定时精度
0.05	81.2%
0.1 (默认)	79.8%
0.2	79.9%
0.4	80.4%
0.6	79.1%

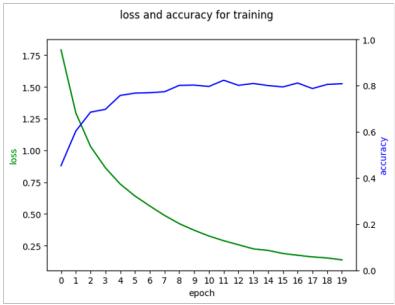
可以看到随着 p 值的增加，稳定时的精度都在 80% 附近，但是 $p = 0.4$ 时在训练轮数为 9 时就已经接近收敛，并且早期精度随着训练轮数的增加的变化较快。

经过这层调参后，将 p 取值设置为 0.4。

3.2.2 dropout 层的设置

当前卷积层有 4 层，分别测试在不同层输出时添加 dropout 层的结果。

3.2.2.1 dropout 层设置的效果



(a) 仅输出层



(b) 输出层 + conv4



(c) 输出层 + conv3/4



(d) 输出层 + conv2/3/4



(e) 输出层 + conv1/2/3/4

3.2.2.2 dropout 层设置的影响总结

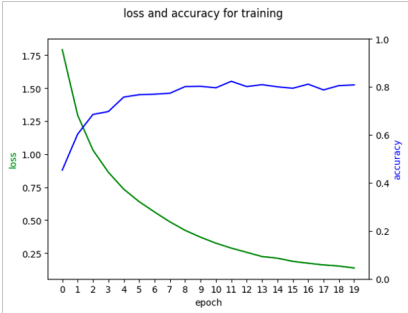
dropout 层设置	稳定时精度
仅输出层（默认）	80.4%
输出层 + conv4	79.4%
输出层 + conv3/4	78.5%
输出层 + conv2/3/4	76.7%
输出层 + conv1/2/3/4	77.1%

可以看到，为卷积层添加多个 dropout 层后，虽然稳定时精度仍然接近 80%，但是训练精度的性质变得差了，不稳定。所以 dropout 层数这一超参数仍使用默认设置，不进行更改。

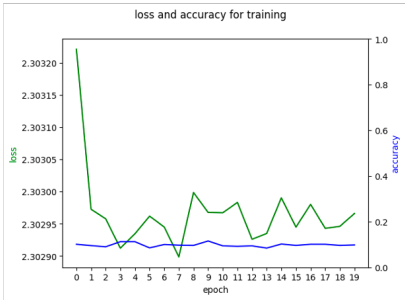
3.3 正则化的影响

基础超参数在每个卷积层都使用了一次正则化，并且使用的是 Batch Normalization，这里测试不加正则化和加 Layer Normalization 的情况。

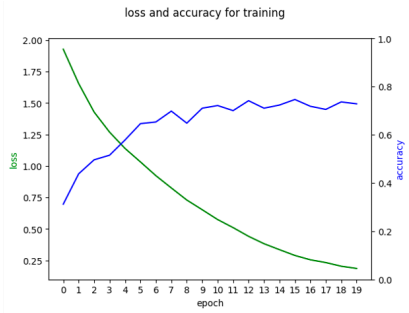
3.3.1 正则化设置的效果



(a) Batch Normalization



(b) No Normalization



(c) Layer Normalization

3.3.2 正则化设置的影响总结

正则化设置	稳定时精度
Batch Normalization（默认）	80.4%
No Normalization	—
Layer Normalization	72.8%

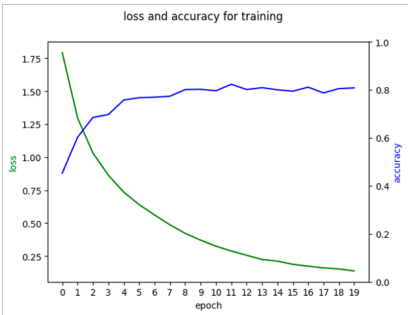
如果不进行正则化，训练将失败，精度和随机猜测相当；如果使用 Layer Normalization，所得到的精度和 loss 下降的性质不如 Batch Normalization。所以仍旧使用（每个卷积层设置一层的）Batch Normalization。

3.4 学习率衰减的影响

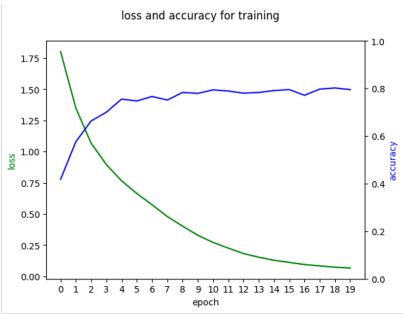
学习率衰减技术可以防止学习率过大而导致目标函数在收敛到全局最优点的时候来回震荡。

基础学习率为 0.001，分别尝试无学习率衰减和指数衰减（衰减率分别为 0.96、0.92、0.88、0.8 和 0.7）时的情况。

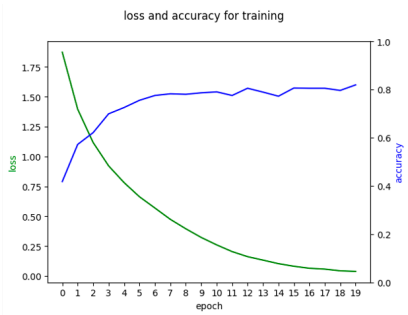
3.4.1 学习率衰减的效果



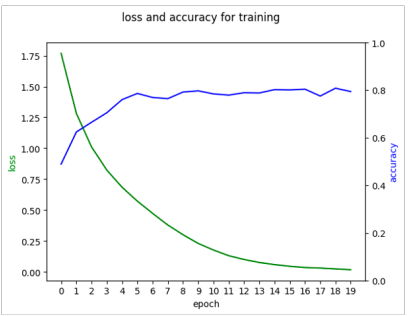
(a) No LearningRate Decay



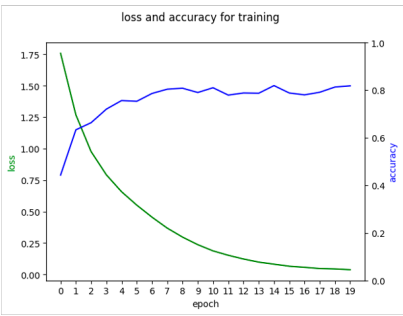
(b) DecayRate = 0.96



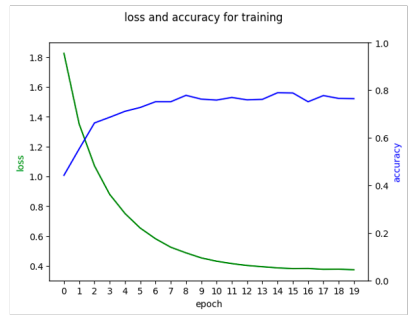
(c) DecayRate = 0.92



(d) DecayRate = 0.88



(e) DecayRate = 0.8



(f) DecayRate = 0.7

3.4.2 学习率衰减的影响总结

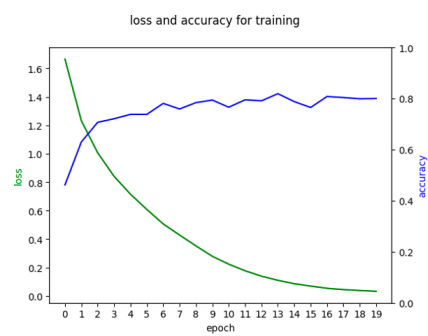
学习率衰减策略 (<i>base lr</i> = 0.001)	稳定时精度
无学习率衰减 (默认)	80.4%
指数衰减, 衰减率 0.96	79.8%
指数衰减, 衰减率 0.92	81.4%
指数衰减, 衰减率 0.88	78.2%
指数衰减, 衰减率 0.8	81.8%
指数衰减, 衰减率 0.7	77.1%

当指数衰减衰减率设置为 0.92 时，稳定时精度达到 81.4%，并且 8 轮左右就收敛，比其他情形的收敛性质好，这一轮调参之后选择指数衰减，衰减率 0.92 作为新的超参数设置。

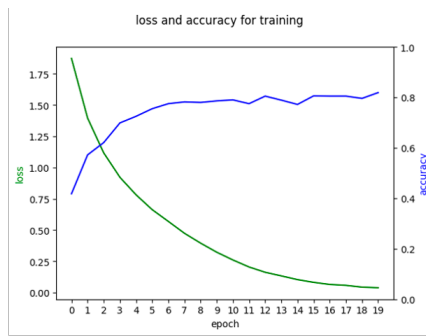
3.5 卷积核大小的影响

当前卷积核大小为 3，因为 CIFAR-10 图片大小均为 32×32 ，因此尝试卷积核大小为 2、3、5、7 的情况，注意：修改 kernel-size 时需要一并修改 padding。

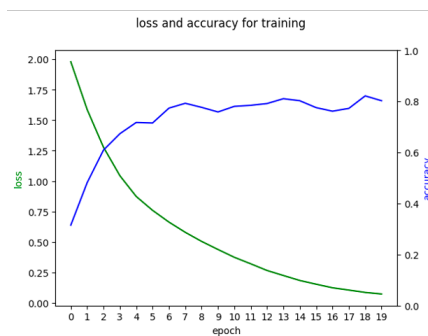
3.5.1 卷积核大小设置的效果



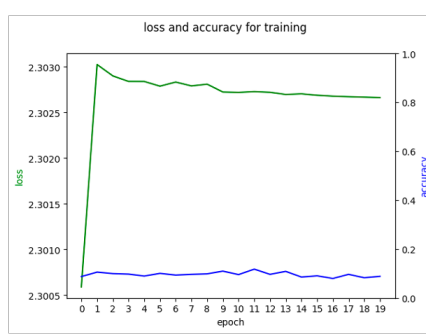
(a) KernelSize = 2



(b) KernelSize = 3



(c) KernelSize = 5



(d) KernelSize = 7

3.5.2 卷积核大小设置的影响总结

卷积核大小	稳定时精度
2	79.9%
3	81.4%
5	79.4%
7	—

当卷积核大小为 2、5 时，虽然最终精度都能达到 80% 左右，但是大小为 2 时产生了明显的过拟合，表现为精度曲线发生了频繁的震荡，训练集上 loss 数量级来到了 10^{-5} 数量级；卷积核大小为 5 时 loss 下降地慢，精度曲线不稳定。

当卷积核大小为 7 时，训练失败了。

所以最终还是选择卷积核大小为 3。

4 最终参数及测试

经过调参，最终选定的参数如下：

Batch 大小：8

网络深度：4 层卷积层

dropout 设置：

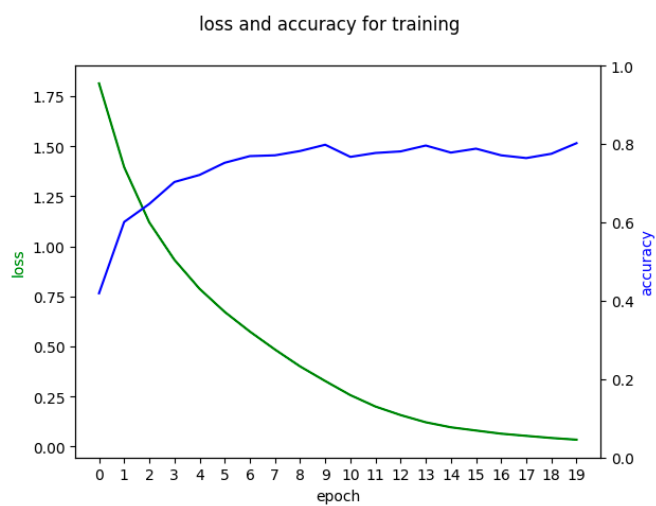
- $p = 0.4$
- 只在输出层使用 dropout

正则化设置：使用 Batch Normalization

学习率衰减设置：基础学习率 0.001，指数衰减，衰减率 = 0.92

卷积核大小：3

训练结果：



测试集上精度：79.2%