
LeNet

林子清

2212279

智能科学与技术

nkuailzq@gmail.com

1 Training

相较于MLP, LeNet 是一种更深的网络, 其参数量更大, 训练难度更大, 尝试使用SGD($lr=1e-1$)的优化器进行训练, 观察到 loss 在前几个 step 剧烈震荡, 转而使用SGD($lr=1e-2$)的优化器进行训练, 观察到 loss 在前几个 epoch 几乎不下降, 因此LeNet 对于学习率极其敏感。这里我们采用AdamW($lr=0.001, betas=(0.9, 0.999)$)这个对学习率极其不敏感的优化器进行训练, 得到了较好的结果。训练后期提升较慢, 因此我们均训练 10epoch, 即使模型依旧没有完全稳定。所有实验均设置batch_size 为 256。

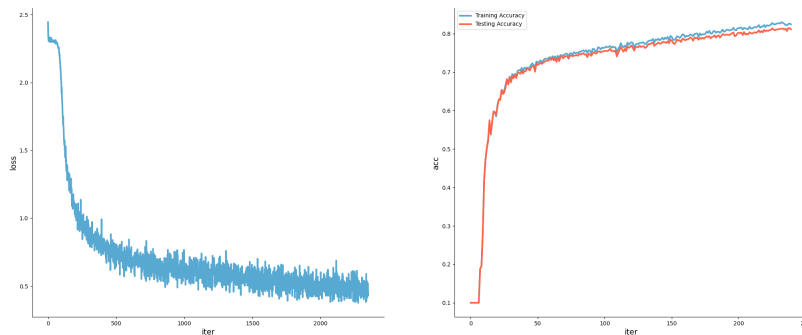


图 1: LeNet

1.1 ReLU

这里我们使用ReLU 替代sigmoid 激活函数, 模型在前期的收敛速度大大提升, 这是因为sigmoid 函数在远离零点时拥有很小的导数, 导致梯度下降速度很慢, 而ReLU 有着恒定导数, 这也使得 loss 下降更平滑。这与AlexNet 论文中所提及的ReLU 加速效果相当。¹

¹代码已经发布在<https://github.com/nkulzq/DLLab.git>

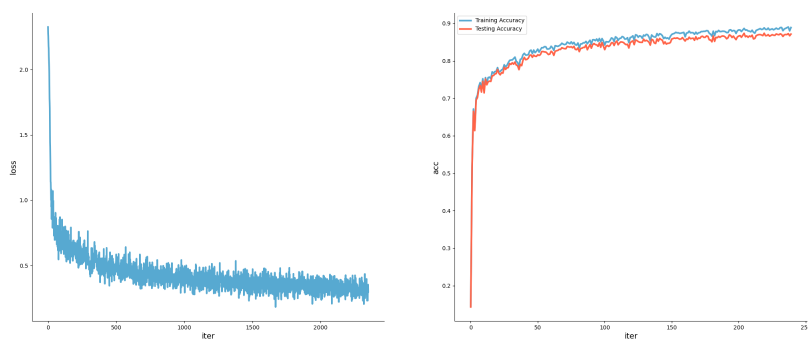


图 2: LeNet_ReLU

12 1.2 Batch Normalization

13 这里我们尝试使用 $f(x) = BN(\text{sigmoid}(x))$ 的方法，但没有观察到明显提升，这可能
 14 是FashionMINST 数据集样本分布较为集中，同时batch_size 设置较大，使得模型本身性能
 15 能已经很好。

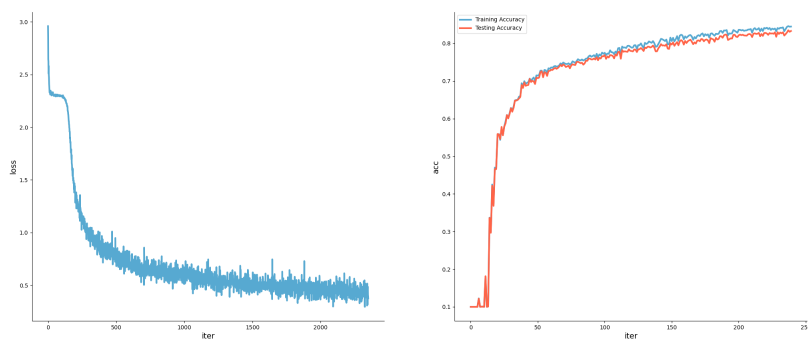


图 3: LeNet_BN

16 1.3 Size of Kernels

17 这里我们将两个大小为 5 的卷积核改为两个大小为 7 的卷积核。增大卷积核时除了使模型更
 18 关注大范围特征，也同时增加了模型的参数量，这无疑会极大的提升模型的性能，使准确率
 19 从 82% 提升至 90%。
 20 特征提取部分参数量： $1 \times 6 \times 5 \times 5 + 6 \times 16 \times 5 \times 5 = 2550 \Rightarrow 1 \times 6 \times 7 \times 7 + 6 \times 16 \times 7 \times 7 = 4998$

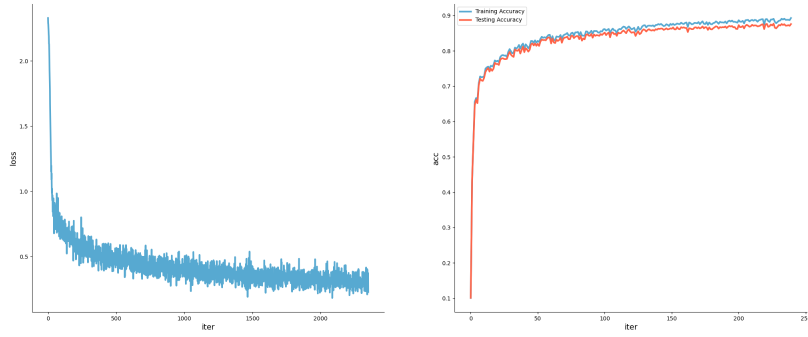


图 4: LeNet_kernel7

21 1.4 Number of Channels

22 增大通道数可以使输入分类部分的特征信息量更大, 特征提取更充分, 从而达到更好的效果。
 23 同时与前一部分增大卷积核相比参数量更大。特征提取部分参数量: $1 \times 6 \times 5 \times 5 + 6 \times 16 \times$
 24 $5 \times 5 = 2550 \Rightarrow 1 \times 16 \times 5 \times 5 + 16 \times 32 \times 5 \times 5 = 13200$

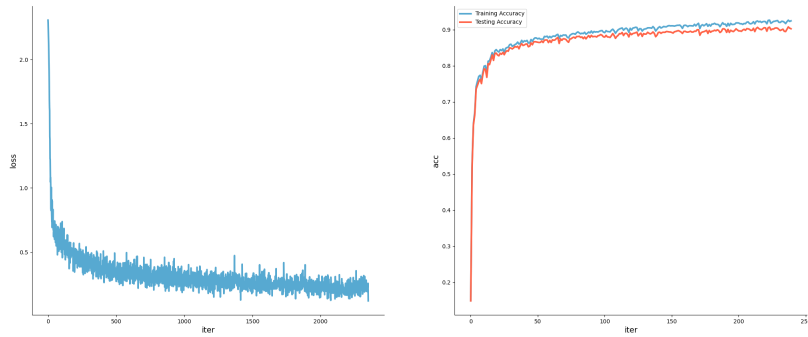


图 5: LeNet_channel32

25 1.5 Max Pooling

26 此处将AvgPool2d 改为MaxPool2d 模型效果有明显提升, 这主要是因为AvgPool2d 更关注模
 27 型整体特征, 而MaxPool2d 更关注模型的纹理特征, 而FashionMINST 数据集图片整体特征
 28 类似, 其主要依靠纹理来分辨。

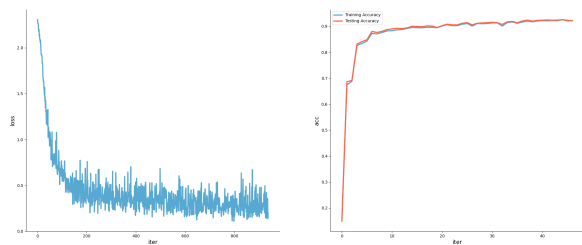


图 6: L2

2 Other Convolutional Neural Networks

在这部分我们分别使用VGG, NiN, GoogLeNet, ResNet 这四种模型运行同样的任务, 考虑到这四种模型都较大, 我们均只训练 3epoch, 同时为了使用这四种模型的结构, 我们将图片resize 成大小为 224 和 96 的图片。

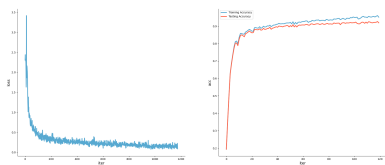


图 7: VGG

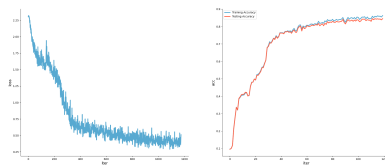


图 8: NiN

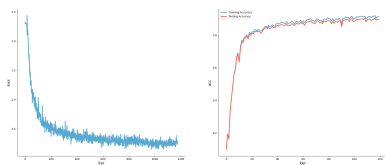


图 9: GoogLeNet

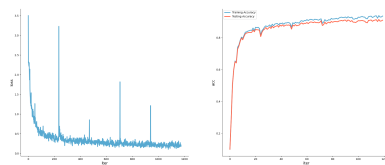


图 10: ResNet

3 Feature Visualization

对于LeNet 的特征可视化, 我们将特征提取部分的所有卷积核用灰度图画出。



图 11: kernel1



图 12: kernel2

35 这里我们并看不出如AlexNet 中一般的特征，这是一方面是因为，这里的卷积核大小较小，
 36 另一方面是因为这里的数据集FashionMINST 特征过于相似，通过对其分类任务的训练并不
 37 能在充分提取图像的底层特征，而智能提取某些较为局限的特征。这里我们给出某一输入的
 38 特征图。²



图 13: features1



图 14: feature2

39 但我们依然能看出其中features1 中第二幅图提取了图像中的纹理特征。我们这里使
 40 用VGG16 再次观察其特征图，这解决了训练不够充分的问题。

²这里我们将使用 hook 在运行时抓取特征图

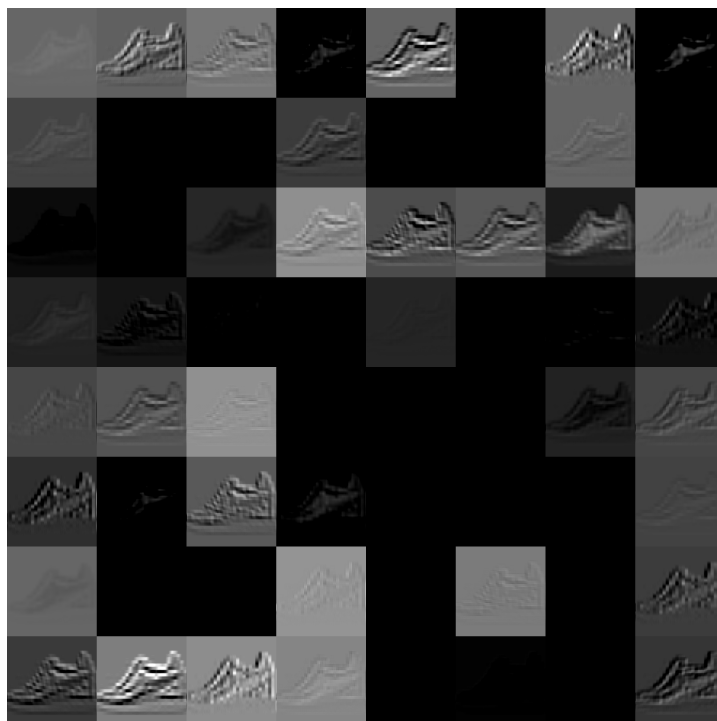


图 15: features_VGG16

41 VGG16 卷积核很小，我们可以观察一下上述实验中，将LeNet 卷积核增大为 11 的实验中训
 42 练得到的卷积核，这里我们设计卷积核为三维，这样正如AlexNet 中一般，可以利用图像的
 43 RGB 显示其三个通道。若通道数大于 3，则我们需要在这里舍弃其多余维度，就如损失函数
 44 可视化一般，我们只能可视化其有限的一部分信息。

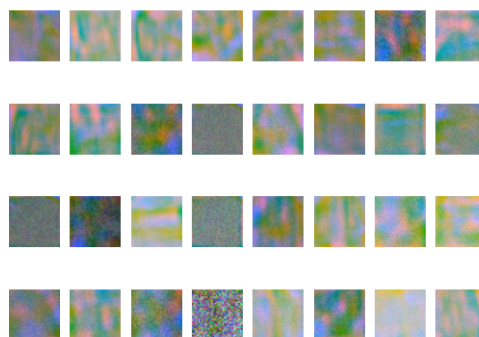


图 16: kernel

45 到此为止我们已经探索了可视化的基本内容，可以对卷积神经网络的运行原理进行一定的归
 46 纳。从课上的学习中，我们可以得知底层的卷积主要学习了图像的纹理等，而上层学习了类
 47 间差异。我们再此更进一步，如在本实验代码中所写，所有网络都具有visual encoder 部分
 48 和classifying head 部分，卷积神经网络是处理视觉编码的一种手段，在上次实验中我们

49 已经实验过MLP 也可以进行视觉编码，而CNN 采用了共享参数的方式减小了网络的参数，同
50 时将视觉信息中很重要的位置信息内嵌进了网络中，这种位置信息相较其他编码的更隐晦的
51 表示更容易被网络所学习。若同传统计算机视觉相比较，底层CNN 其实自主学习到了传统的
52 边缘检测等方法，而分类头则将这些特征组合起来，正如上世纪人们所精心设计的各种特征
53 要求。而深度学习本身就是一种对函数的自动拟合，所有结构只是使得特征更容易被学习。