

基于几种经典 CNN 模型的眼底近视疾病图像分类问题的分析对比

随着深度学习不断的发展，目前先后在 CV 领域常见的 CNN 网络模型种类主要有 LeNet、AlexNet、GoogleNet、ResNet 等。在眼底疾病分类检测领域，我们先后采用如上几种经典的 CNN 网络模型，分析其在眼底近视疾病图像分类问题上表现效果，进行比较。

一、数据集介绍

[iChallenge-PM](#) 是百度大脑和中山大学中山眼科中心联合举办的 iChallenge 比赛中提供的关于病理性近视（Pathologic Myopia, PM）的医疗类数据集，包含 1200 个受试者的眼底视网膜图片，训练、验证和测试数据集各 400 张。我们利用以上开源数据集进行学习训练。

数据集文件介绍如下：

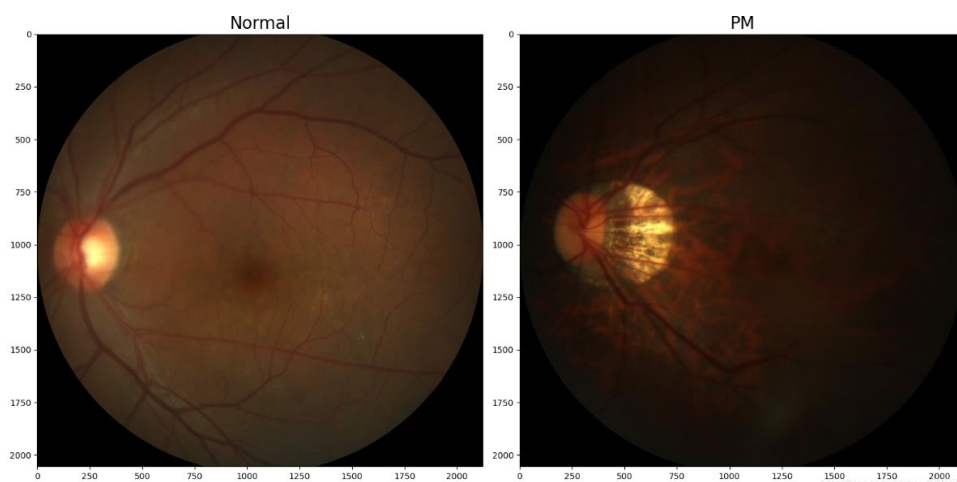
- （1）training.zip：包含训练中的图片和标签
- （2）validation.zip：包含验证集的图片
- （3）valid_gt.zip：包含验证集的标签

iChallenge-PM 中既有病理性近视患者的眼底图片，也有非病理性近视患者的图片，命名规则如下：

- （1）病理性近视（PM）：文件名以 P 开头
- （2）非病理性近视（non-PM）：
 - 高度近视（high myopia）：文件名以 H 开头
 - 正常眼睛（normal）：文件名以 N 开头

二、数据预处理

我们将病理性患者的图片作为正样本，标签为 1；非病理性患者的图片作为负样本，标签为 0。从数据集中选取两张图片，构建分类器，对正负样本进行分类，并将图片显示出来。



接着我们定义数据读取器，使用 OpenCV 读取图片，将图片缩放至 224*224 大小，像素值调整至[-1, 1]之间。



三、LeNet 网络训练

LeNet 网络结构主要是利用多个卷积核与池化层一直对原数据做多次计算，从而达到提取数据特征、减小原数据量的目的。

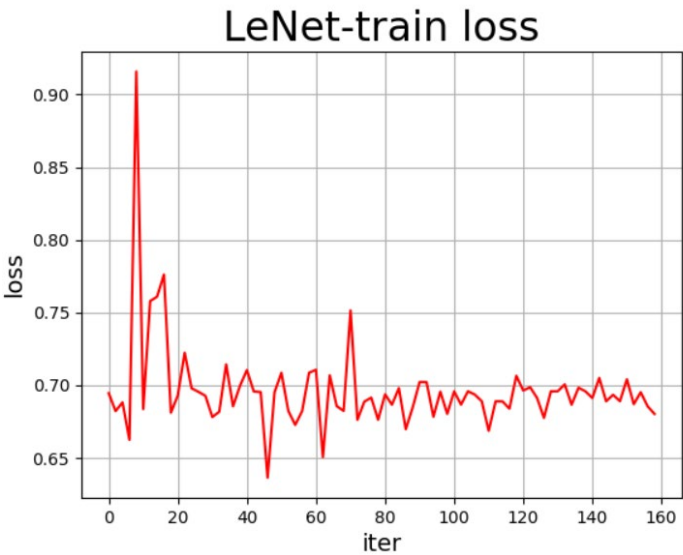
本次测试训练模型的网络结构如下表所示：

输入数据	[3,224,224]
卷积层：6 个，step=1	[3,5,5]
池化层：step=2	[2,2]
卷积层：16 个，step=1	[6,5,5]
池化层：step=2	[2,2]
卷积层：120 个，step=1	[16,4,4]
拉直层	[300000,1]
全连接层	[64,1]

本次测试采用 paddle 开源环境进行训练，模型运行 20~40epochs，未完全稳定，运行完后 LeNet 的 loss 为 0.691962，准确率约为 52.46%，效果较差。

从运行结果观察，在眼疾筛查数据集 iChallenge-PM 上，LeNet 的 loss 很难下降，模型没有收敛。这是因为 MNIST 数据集的图片尺寸比较小（28*28），而眼疾筛查的数据集图

片尺寸较大（原始尺寸：2000*2000，经缩放后为 224*224），LeNet 模型很难进行有效分类，说明在图片尺寸比较大时，LeNet 在图像分类上存在局限性。模型 loss 可视化曲线如下：



四、AlexNet 网络训练

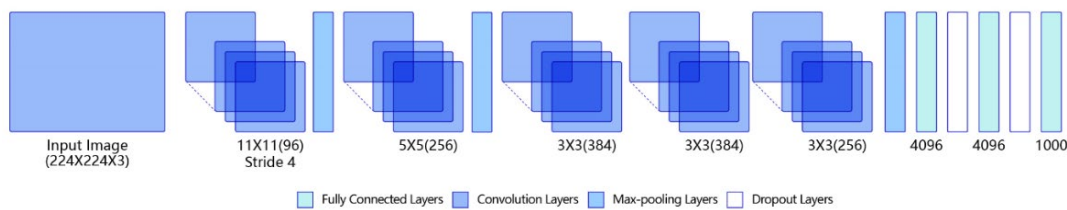
AlexNet 与 LeNet 相比，具有更深的网络结构，包含 5 层卷积和 3 层全连接，同时使用了如下三种方法改进模型的训练过程：

（1）数据增广：深度学习中常用的一种处理方式，通过对训练随机加一些变化，比如平移、缩放、裁剪、旋转、翻转或者增减亮度等，产生一系列跟原始图片相似但又不完全相同的样本，从而扩大训练数据集。通过这种方式，可以随机改变训练样本，避免模型过度依赖于某些属性，能从一定程度上抑制过拟合。

（2）使用 Dropout 抑制过拟合。

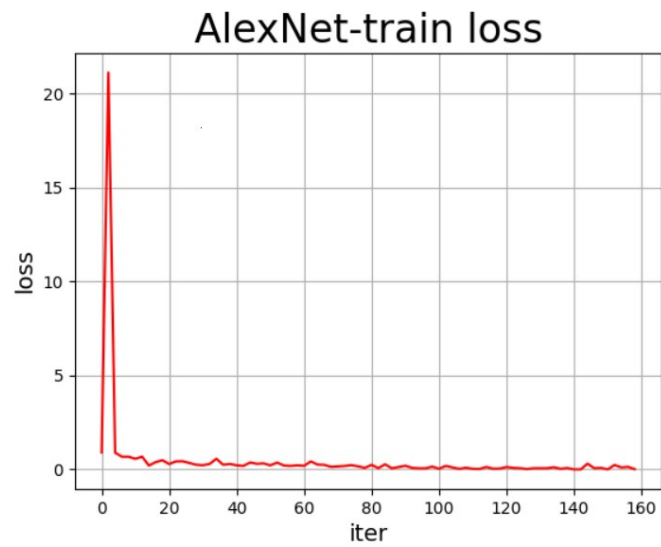
（3）使用 ReLU 激活函数减少梯度消失现象。

本次测试训练模型的网络结构如下图所示：



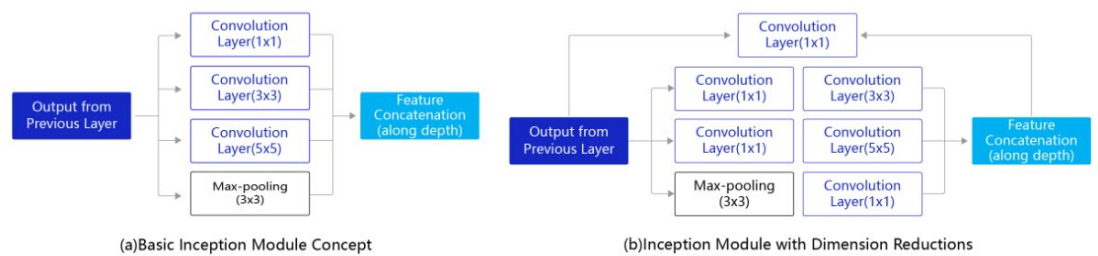
在眼疾筛查数据集 iChallenge-PM 上使用 AlexNet，loss 能有效下降，进行 20epochs 的

训练，loss 基本稳定，准确率达到约 94.19%，效果明显好于 LeNet 模型。



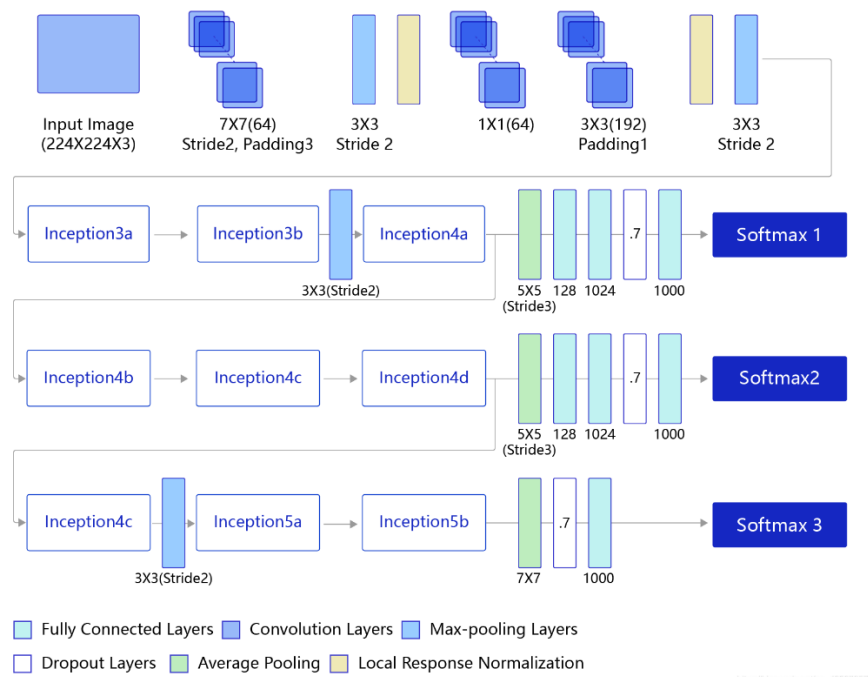
五、GoogLeNet 网络训练

GoogLeNet 是 2014 年 ImageNet 比赛的冠军，它的主要特点是网络不仅有深度，还在横向上具有“宽度”。由于图像信息在空间尺寸上的巨大差异，如何选择合适的卷积核大小来提取特征就显得比较困难了。空间分布范围更广的图像信息适合用较大的卷积核来提取其特征，而空间分布范围较小的图像信息则适合用较小的卷积核来提取其特征。为了解决这个问题，GoogLeNet 提出了一种被称为 Inception 模块的方案。如下图所示。



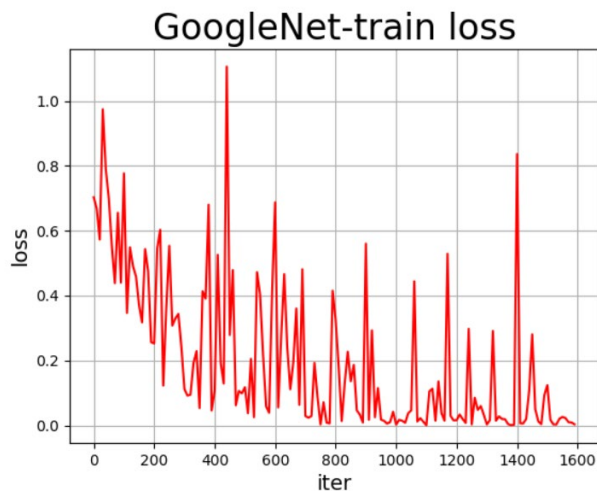
图(a)是 Inception 模块的设计思想，使用 3 个不同大小的卷积核对输入图片进行卷积操作，并附加最大池化，将这 4 个操作的输出沿着通道这一维度进行拼接，构成的输出特征图将会包含经过不同大小的卷积核提取出来的特征。Inception 模块采用多通路(multi-path)的设计形式，每个支路使用不同大小的卷积核，最终输出特征图的通道数是每个支路输出通道数的总和，这将会导致输出通道数变得很大，尤其是使用多个 Inception 模块串联操作的时候，模型参数量会变得非常大。为了减小参数量，Inception 模块使用了图(b)中的设计方式，在每个 3x3 和 5x5 的卷积层之前，增加 1x1 的卷积层来控制输出通道数；在最大池化层后面增加 1x1 卷积层减小输出通道数。基于这一设计思想，形成了上图(b)中所示的结构。

整体 GoogleNet 架构如下图所示。在主体卷积部分中使用 5 个模块（block），每个模块之间使用步幅为 2 的 3×3 最大池化层来减小输出高宽。



（注：原有模型中添加了 softmax1 和 softmax2 两个辅助分类器，此次训练程序做了简化，并未加入辅助分类器）

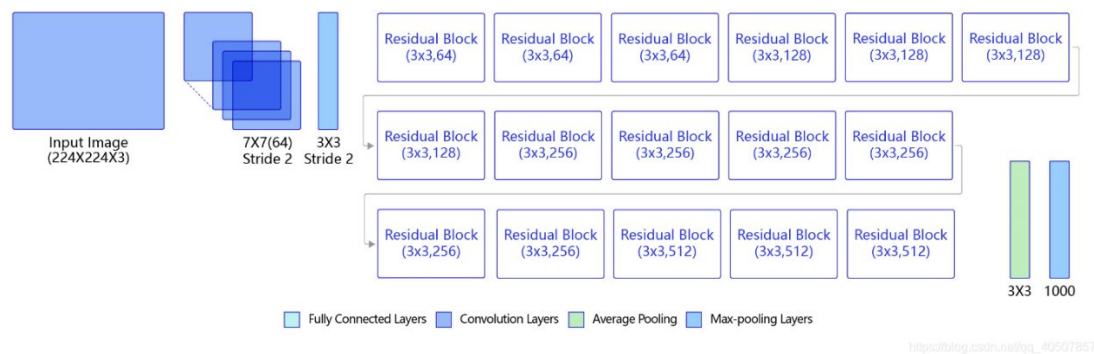
采用进行简化后的 GoogleNet 模型进行训练，结果发现在眼疾筛查数据集 iChallenge-PM 上，loss 能有效的下降，准确率为 95.5%，效果较好。



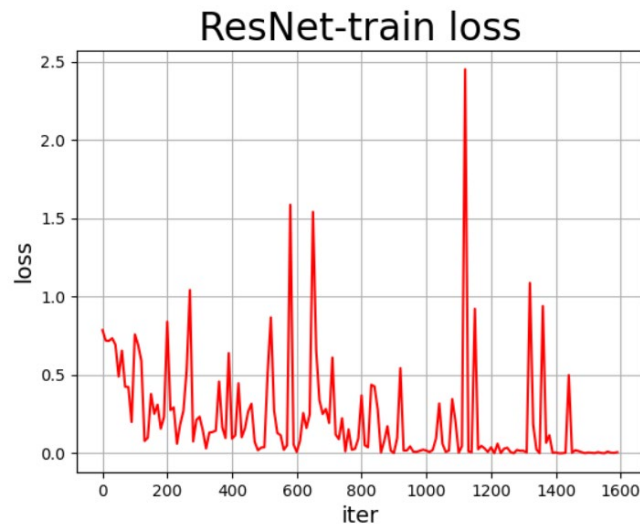
六、ResNet 网络训练

ResNet 网络主要是在结构中加入了残差块。ResNet 网络对数据的卷积计算过程与之前模型的计算过程都类似，只是在在某一层的线性模块之后，非线性模块之前增加前面某层的输出。

下图表示出了 ResNet-50 的结构，一共包含 49 层卷积和 1 层全连接，所以被称为 ResNet-50。



通过运行结果可以发现，使用 ResNet 在眼底筛查数据集 iChallenge-PM 上，loss 能有效的下降，准确率约为 97%，效果显著。



七、模型训练对比

通过对同一眼底近视疾病数据集采用不同的经典 CNN 模型进行分类比较，各个模型均进行了约 20~40epochs 的训练，其 loss 以及 accuracy 数据对比情况如下：

	loss	accuracy
LeNet	0.691962	0.524554
AlexNet	0.166790	0.941964
GoogleNet	0.132089	0.955000
ResNet	0.113273	0.970000

通过上述模型训练对比，我们观察可知传统的 LeNet 模型在处理大尺寸疾病图像上存在缺陷，而 AlexNet，GoogleNet 和 ResNet 模型在进行眼底疾病分类筛查问题方面效果较好，后续我们可以采用如上三种网络模型进行眼底疾病的筛查分类分析。从分类结果上看所选模型能较准确的进行判断，但同时我们也需要更多的数据集进行训练、检验。