

train2

吴琪

July 2024

本次培训的题目在于探究 ResNet 模型中残差连接的作用。ResNet——即残差神经网络的主要贡献是发现了“退化现象 (Degradation)”，并针对退化现象提出了“残差连接 (Shortcut connection)”，这极大的消除了由于网络深度过大导致的训练困难的问题。在深度学习训练中，一般情况下，随着网络层不断的加深，模型的准确率会不断的提高，但是当达到最大值后，靠近输入层的权重更新会受到靠近输出层的权重更新的影响，所以出现网络深度的继续增加，模型准确率反而会出现大幅度降低的情况，ResNet 将这一现象称为“退化 (Degradation)”。为了解决这种问题，ResNet 提出了“残差连接”，这个连接直接将前几层的输入加到后几层的输出上，允许某一层的输出可以跳过一个或多个层，直接连接到后续层的输入。尽管有些层之间没有做实际的变换，但它们仍然可以传递之前层的信息，并且避免对梯度产生过多的损失。这种深度学习框架可以用公式表示为：

$$H(x) = F(x) + x \quad (1)$$

在论文中残差块的结构如下图所示：

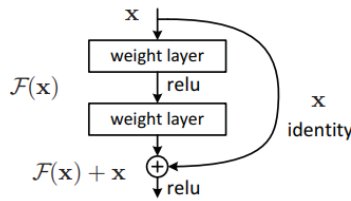


图 1: Residual learning: a building block.

ResNet18 的基本含义是：网络的基本架构是 ResNet，网络的深度是 18 层。假设输入的数据为 $224 \times 224 \times 3$ ，首先经过一个卷积层，卷积核大小为 7×7 ，步长为 2，padding 为 3，输出通道为 64，数据卷积后大小变为 $112 \times 112 \times 64$ 。随后连接一个池化层，数据变为 $56 \times 56 \times 64$ 。接下来是第一个残差结构，卷积核个数统一为 64，因此通过该结构后数据大小不发生变化。在进入第二个残差结构前，对数据先进行下采样，即 128 个 1×1 的卷积层，将数据大小变为 $28 \times 28 \times 128$ 。第二个残差结构后依旧需要下采样，由 256 个 1×1 的卷积层组成，数据大小变为 $14 \times 14 \times 256$ 。依此类推，每经过一次下采样，数据的通道增加一倍，通道大小变为一半。经过四个残差结构和三次下采样后，数据大小变为 $7 \times 7 \times 512$ ，最后将残差层的输出结果经过全局

平均池化后放入全连接层，得到分类结果， $1*1*512$ 。部分残差层结构如下图所示：

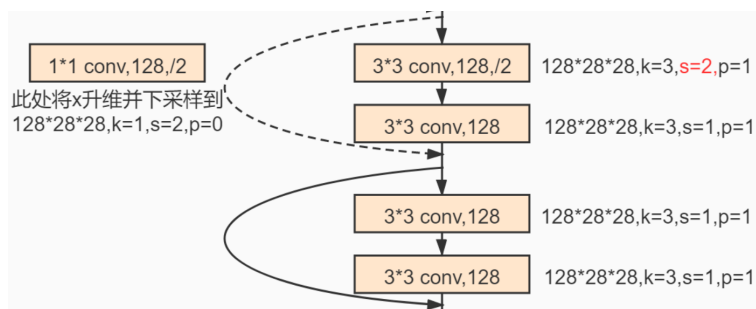


图 2: The structure of the second residual layer.

接下来进行复现 ResNet 网络，分别创建带有残差连接的残差结构 ResidualBlock 以及不含残差连接的普通结构 BasicBlock。残差连接定义为 shortcut，当步长 stride 不等于 1 或者输入通道数不等于输出通道数时，在网络后加入新的卷积层。由前者残差结构构成的网络定义为 net，后者普通结构组成的网络定义为 unnet。训练所使用的损失函数同样采用了先前的交叉熵损失函数，即 `nn.CrossEntropyLoss`。优化器采用了 SGD，随机梯度下降法。加载训练集与测试集，获取设计好的 Net。随后对两个网络进行 15 个 epoch 的训练，每个网络训练过程中的 loss 以及训练集预测准确率分别如下图所示：



图 3: Loss of two models during training.



图 4: Accuracy of two models during training.

训练过程中，在每个 epoch 后，都将模型在测试集上进行预测。同时，为了更加清晰的展示数据集中不同类别预测准确率，使用 list 记录了每个分类预测正确个数与总个数。在每

个 epoch 中，网络预测的总结果如下图所示：

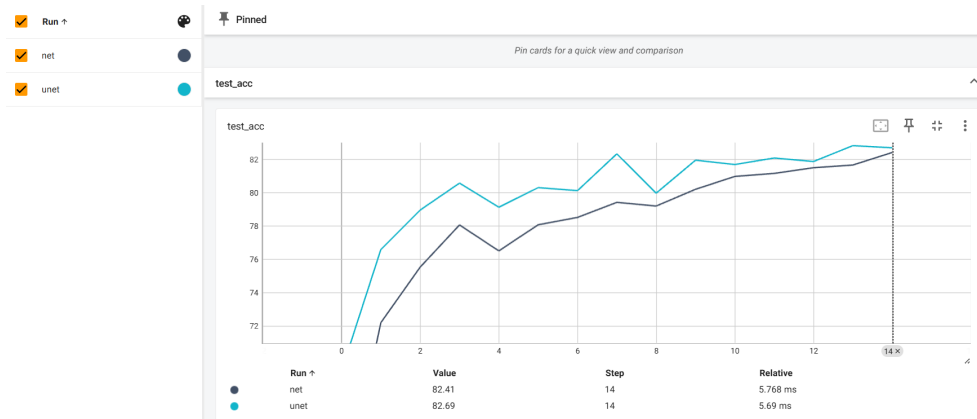


图 5: Prediction accuracy of two models on test set during training.

两个模型在每个类别上的 accuracy 如下表所示：

表 1: Accuracy of the two models in each category

category	net(%)	unnet(%)
plane	81	84
car	86	90
bird	72	70
cat	59	66
deer	73	78
dog	66	69
frog	83	82
horse	83	83
ship	87	87
truck	84	86

分析时间，在 ResNet18 中，残差连接可以缓解梯度消失问题，深层的神经网络更易产生梯度消失，导致训练效果不佳，而残差连接可以通过在网络层之间增加"捷径"，使梯度在反向传播过程中更有效，缓解梯度消失问题，进一步提升模型性能。正如训练过程中的曲线所示，在最终的 epoch 时，包含残差连接的 net 比不包含残差连接的 unnet 效果更好。同时，残差连接可以帮助模型更快地收敛到最优解，同样在训练的结果里面可以看到，net 比 unnet 到达最好效果所需的 epoch 更少。而分析测试集，残差连接减弱了准确率骤降的情况，尽管目前 unnet 比 net 效果更高，但是可能因为训练 epoch 次数不足，猜测仅从曲线来看，包含残差连接的 net 曲线斜率要大于 unnet，当继续训练增加 epoch 时，最终的效果可能还是 net 网络模型占优。

为了更清楚的比较效果，将 epoch 增加为 30，将 batchsize 修改为 64，再对训练集进行

训练, 同时每个 epoch 后对测试集进行预测, 得到两个模型在测试集上的 acc 如下图所示:



图 6: New accuracy rates tested on the test set

可以发现没有残差连接的网络效果还是优于有残差连接的网络, 可能由于网络模型在去除残差连接时, 没有正确的对 ResNet18 进行修改; 同时也可以因为 ResNet 的优势在于能够训练更加深层的网络, 而 ResNet18 的网络还不够深, 残差连接可能无法完全发挥其优势。还有原因可能是 CIFAR10 数据集相对简单, 对于简单数据集, 较浅的网络就能达到较好的性能, 而过于复杂的残差网络可能会过拟合。