

小作业 基于 ResNet 的 MedMINST 数据集图像分类实验

王怡闻 519030910367 曹嘉航 519030910347

June 16, 2021

1 完成情况

我们使用 Pytorch 复现了 ResNet-18 与 ResNet-50 在 MedMINST 数据集上的分类，并对实验结果做出分析：网络越深对训练集拟合越准确，但过拟合越明显；残差块的设置能够避免梯度消失；数据集越小过拟合越严重。

针对小数据集过拟合严重的问题，我们使用组卷积减少参数量，验证其可以一定程度上缓解过拟合现象，并降低随机性的影响；虽然参数大量减少，但这种方法对准确率与类别区分度的影响甚微，堪称本次实验的亮点。

我们基于曲线下面积（AUC）和准确率（ACC）分析了十个数据集上的训练与测试结果，部分值得展示的统计结果如下：

模型	测试集 平均 AUC	训练测试 平均 AUC 差值	测试集 平均 ACC	训练测试 平均 ACC 差值
普通 ResNet-18	0.892	0.089	0.779	0.202
普通 ResNet-50	0.877	0.088	0.771	0.188
使用组卷积的 ResNet-50	0.876	0.082	0.771	0.105

2 实验数据与网络结构

2.1 MedMINST 数据集简介

MedMINST 为十个医疗领域图像数据集的整合，数据为各类医疗设备产生的影像，各数据集的标签含义各不相同，包括但不限于是否为恶性病变、病变程度及组织所在器官。所有图像被预标注，且从 100 至 100,000 不等的规模统一压缩至 28×28 像素，标签被数字化。数据集规模从 10^2 量级到 10^5 量级不等，其上的任务包括二分类、多分类、多标签二分类、有序回归四种类型，由于其无需专业知识的特性，广泛应用于测试医学图像识别系统。[3]

2.2 残差神经网络 ResNet 的原理

ResNet 由何凯明于 2015 年提出，通过残差块结构避免了梯度消失，增多网络参数量。

图1为一个残差块 (Residual Block)[1]，这是残差神经网络的基本结构。残差块将权重-激活单元的输入和输出之和传递给下一层神经网络，记输入 x 依次通过参数为 W_1 的权重层，函数为 σ 的激活层，参数为 W_2 的权重层，得到的输出为

$$\mathcal{F}(x) = W_2 \sigma(W_1 x) \quad (1)$$

那么附加激活函数的残差块将输出

$$z = \sigma \mathcal{H}(x) = \sigma(\mathcal{F}(x) + x) \quad (2)$$

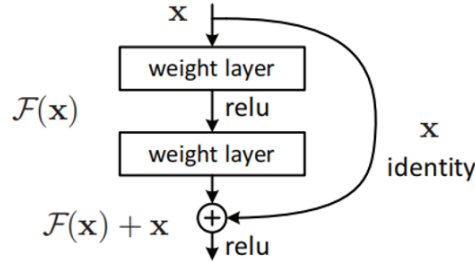


Figure 2. Residual learning: a building block.

Figure 1: 残差块的结构

下面考虑反向传播阶段，若将上一层输出 z^i 传入残差块，残差块层输出 z^{i+1} ，误差函数 \mathcal{J} 经过残差块时的梯度应为

$$\frac{\partial \mathcal{J}}{\partial z^{i+1}} = \sigma'(W_2^T(\sigma' W_1^T) + 1) \frac{\partial \mathcal{J}}{\partial z^i} \quad (3)$$

上式中的常数项 1 来源于残差块中输入的直接链接，保证了梯度在经过多层后不会消失，网络深度受到的限制更小。实验证明，每个残差块至少包含两个权重层的结构可以增强拟合效果。[1]

ResNet 根据权重层层数的不同分为 ResNet-18, ResNet-50, ResNet-101 等网络，本实验选用 ResNet-18 和 ResNet-50 在 MedMINST 数据集上进行训练与预测，比较性能差异。

2.3 ResNet-18 与 ResNet-50 的结构

ResNet 的传统结构由五层结构链接而成，第一层为 7×7 卷积层，使用 3×3 最大值池化进行降维，第二至第五层依次通过若干数目的残差块扩充通道数，五层结构后经平均值池化层、全连接层、Softmax 层得到预测向量。残差块权重层均使用 1×1 卷积，所有卷积层不含偏置且激活前进行批量归一化。

在本实验中，第三至第五层每层的第一个卷积块设置为残差块。除第一层外，所有卷积层的激活函数使用 ReLU。

Table 1: ResNet-18 和 ResNet-50 对 MedMINST 数据的处理

层	输出 维度	ResNet-18	ResNet-50
输入	28×28		
第一层	14×14	$7 \times 7, 64, s=2$	
	7×7	$3 \times 3, \text{MaxPool}, s=2$	
第二层	7×7	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
第三层	4×4	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
第四层	2×2	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
第五层	1×1	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
预测	2	$1 \times 1, \text{AdaptiveAvgPool}$ FC, 2	
可训练 参数量		$1.12e7$	$2.35e7$

本实验采用两种网络参数及对 28×28 输入的处理见表1.

注：每层输出都采用了批量归一化。

3 ResNet-18 和 ResNet-50 在 MedMINST 数据集上的拟合实验

3.1 代码描述

3.1.1 数据预处理

数据按照预先提供的比例分划训练集、验证集、测试集，并在每个集合中随机打乱。以批量规模 128 对输入数据进行归一化，将值规范到 $[-1, 1]$ 范围。

3.1.2 网络训练

使用随机梯度下降（SGD）算法训练 ResNet，损失函数统一使用交叉熵损失函数（CE），其中多标签二分类问题特化为二分类交叉熵 BCE。

3.1.3 网络测试

测试时固定当前网络的 BN 与 Dropout 参数，取消随机性影响。再将测试数据输入网络得到预测结果，由预测结果与标签计算预测准确率（ACC）和曲线下面积（AUC），衡量分类精准度与区分度。为了检验拟合效果，分别将训练集与测试集数据传入网络计算以上指标。

3.2 实验结果

两种模型在十个数据集测试集上的 AUC 及 ACC 如表2所示，对该表格的数据分析结果列于第 1 节。

Table 2: ResNet-18 和 ResNet-50 拟合效果对比

数据集	数据量		ResNet-18				ResNet-50			
			AUC		ACC		AUC		ACC	
	训练	测试	训练	测试	训练	测试	训练	测试	训练	测试
PathMINST	89,996	7,180	1.000	0.954	1.000	0.811	1.000	0.948	1.000	0.771
ChestMINST	78,468	22,433	0.822	0.704	0.950	0.947	0.766	0.693	0.949	0.947
DermaMNIST	7,007	2,005	0.991	0.883	0.912	0.710	1.000	0.870	0.996	0.721
OCTMNIST	97,477	1,000	1.000	0.909	0.998	0.694	1.000	0.898	0.996	0.694
PneumoniaMNIST	4,708	624	1.000	0.936	1.000	0.864	1.000	0.932	1.000	0.873
RetinaMNIST	1,080	400	0.999	0.712	0.974	0.483	0.883	0.680	0.679	0.478
BreastMNIST	546	156	1.000	0.894	1.000	0.859	1.000	0.830	0.998	0.769
OrganMNISTAxial	34,581	17,778	1.000	0.988	1.000	0.872	1.000	0.983	0.999	0.870
OrganMNISTCoronal	13,000	8,268	1.000	0.984	0.999	0.868	1.000	0.984	0.999	0.872
OrganMNISTSagittal	13,940	8,829	0.999	0.956	0.981	0.687	0.999	0.954	0.973	0.713

ResNet 在大部分数据集上均取得了较好的训练效果与测试效果，ResNet-50 在训练集上的准确率普遍强于 ResNet-18。证实了残差神经网络可以通过加深提高拟合能力。

在数据量较大的数据集上，ResNet 预测准确。但无论哪种模型，在 RetinaMNIST 上效果欠佳，且随着网络的加深准确率反而降低。该数据集是多标签二分类问题，使用不经任何处理的多标签二分类不能很好表征训练效果，因此训练和测试效果都不佳。

在 RetinaMNIST、DermaMNIST、BreastMNIST 三个数据集上，网络出现了过拟合。查阅数据 [3] 得知这些数据集规模非常小，训练数据与测试数据间关联性低方差大，这是过拟合的主要原因。

OCTMNIST 数据集测试数据太少，受随机性影响强烈，虽然分类区分度很高，准确率不及其它数据集。

4 对 ResNet 在 MedMINST 上拟合性能提升的尝试

4.1 组卷积

减少参数可以加快训练并减缓过拟合，组卷积是卷积神经网络中常用的减少参数方法：将多通道数据分为共享卷积核的多组，在加宽网络的同时降低了卷积核通道数，降低参数数量 [2]。本实验将 ResNet-50 每一层中的第二个卷积单元分为 32 组进行卷积，在最大程度上减少冗余参数。

4.2 提升效果

Table 3: 使用组卷积后的网络性能

数据集	ResNet-18			
	AUC		ACC	
	训练	测试	训练	测试
PathMINST	1.000	0.947	0.999	0.777
ChestMINST	0.778	0.689	0.949	0.947
DermaMNIST	1.000	0.863	0.999	0.720
OCTMNIST	1.000	0.904	0.997	0.699
PneumoniaMNIST	1.000	0.943	1.000	0.827
RetinaMNIST	0.809	0.683	0.568	0.528 (+0.050)
BreastMNIST	1.000	0.825	0.996	0.801 (+0.032)
OrganMNISTAxial	1.000	0.984	1.000	0.863
OrganMNISTCoronal	1.000	0.979	0.999	0.848
OrganMNISTSagittal	1.000	0.947	0.989	0.700

使用组卷积后的性能如表3，在原本过拟合的数据集上测试性能均出现了略微的提升。过拟合得到缓解。而原本没有过拟合的数据集训练集上准确率没有明显变化，测试集上准确率略微下降。

观察收敛时训练轮数等数据发现，组卷积除能一定程度缓解过拟合外，还能使梯度下降更平缓，降低随机性的影响。且虽然梯度下降更慢，但由于参数减少，训练总时间仍然得到降低。

References

- [1] HE, K., ZHANG, X., REN, S., AND SUN, J. Deep residual learning for image recognition, 2015.
- [2] XIE, S., GIRSHICK, R., DOLLÁR, P., TU, Z., AND HE, K. Aggregated residual transformations for deep neural networks, 2017.

- [3] YANG, J., SHI, R., AND NI, B. Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis, 2021.