

# 实验数据整合

孙凯威

中国地质大学（武汉）

日期：2021 年 4 月 13 日

## 1 实验一-Softmax

### 1.1 固定参数

本实验所固定的参数有各分支的各层激活函数均为 *sigmoid* 函数, 优化器选择使用 *Adam* 优化器, 所设置的学习率  $\epsilon$  为 0.0001, 训练的迭代次数 *epoch* 为 800, 训练的批大小 *mini-batch size* 为 2048, *DBN* 的层数为 5 层, 预训练的批大小 *batch-size* 为 64.

对于空间-光谱分支, 我们使用 5 层 *DBN*, 隐藏层节点数为 1500. 对于地形分支我们固定 *DBN* 层数也为 5 层, 隐藏层节点数需要进行选择。

### 1.2 方案一：无全连接层 + 单输出

在两个分支进行合并后直接送入 *softmax* 分类器进行分类, 分类的输出方式为单输出, 即只输出 20 个二级地物标签的类别。所需调节的超参数为地形分支的各层的节点数, 所选取范围为 [3-1500].

每个节点数运行 5 次, 取平均值和去极值平均两种方法。部分节点的五次结果和平均结果如表1所示, 每个节点的五次平均值可见图1 去极值平均的结果可以见图2. 由图可以看出, 当地形分支节点数为 550 时, 实验的验证集结果最好, 故选定地形分支节点数为 550。

表 1: 实验数据表格片段

	300	350	400	450	500	550	600
1	0.9048	0.9107	0.9139	0.9083	0.8939	0.9133	0.9107
2	0.9083	0.9108	0.9126	0.9098	0.9123	0.9107	0.8974
3	0.9008	0.9104	0.9097	0.9167	0.9077	0.908	0.8717
4	0.8995	0.9136	0.8996	0.9112	0.9148	0.9136	0.8959
5	0.907	0.9034	0.9119	0.911	0.9128	0.9124	0.9113
平均	0.90408	0.90978	0.90954	0.9114	0.9083	0.9116	0.8974
去极值平均	0.9042	0.910633	0.9114	0.910667	0.910933	0.912133	0.901333

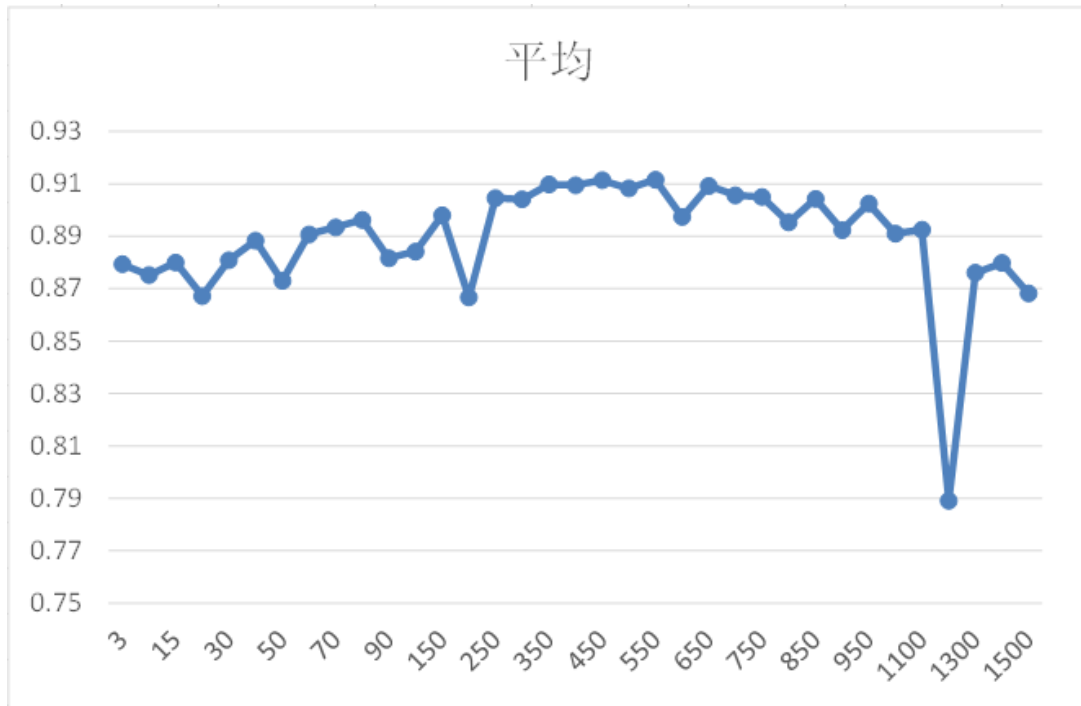


图 1: 实验数据的平均值

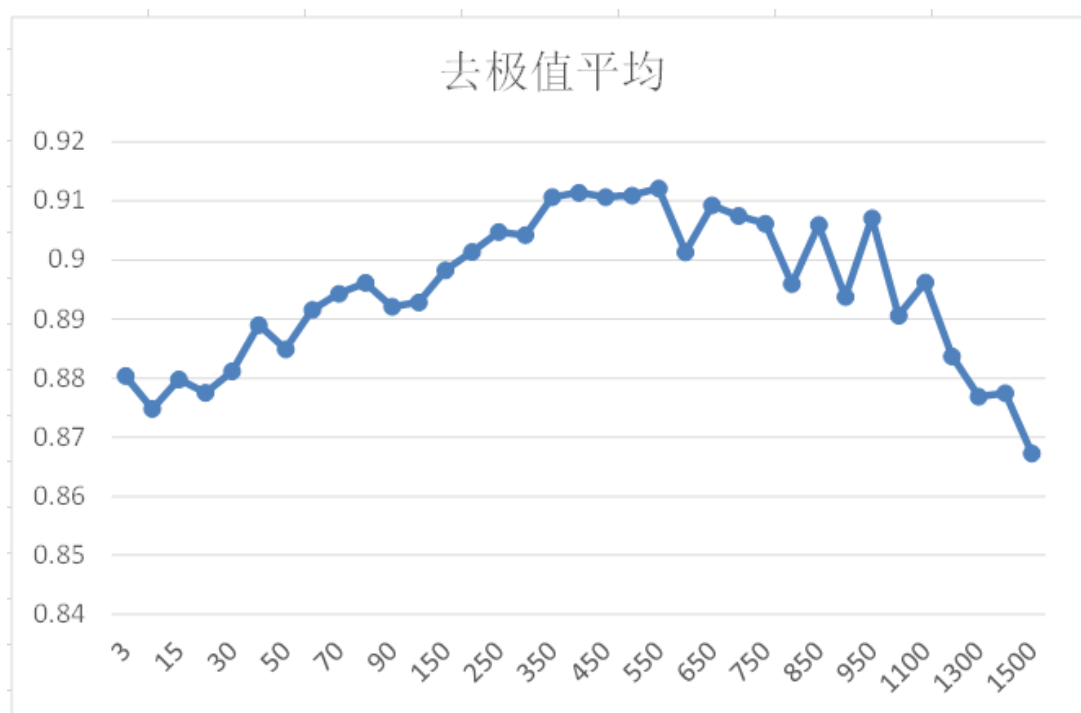


图 2: 实验数据的去极值平均值

我使用地形分支节点数为 550 训练出来的模型在测试集上进行预测，并获取了训练集和验证集的正确率和损失值随着迭代次数变化的图，如图3和4

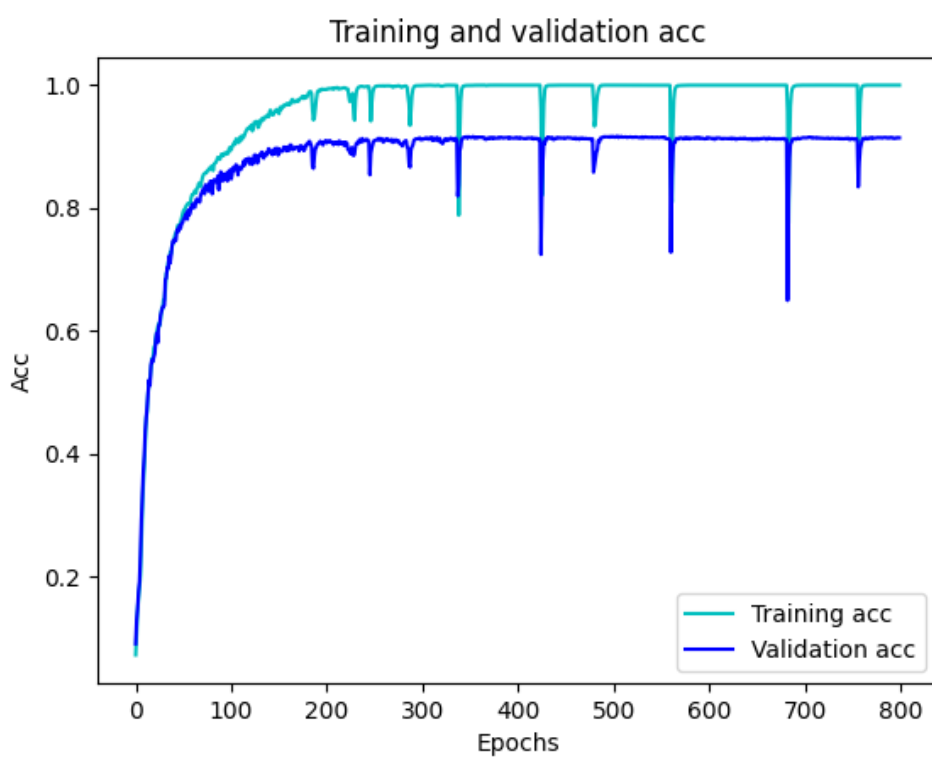


图 3: 地形分支节点数为 550 的训练集和验证集的实验精度

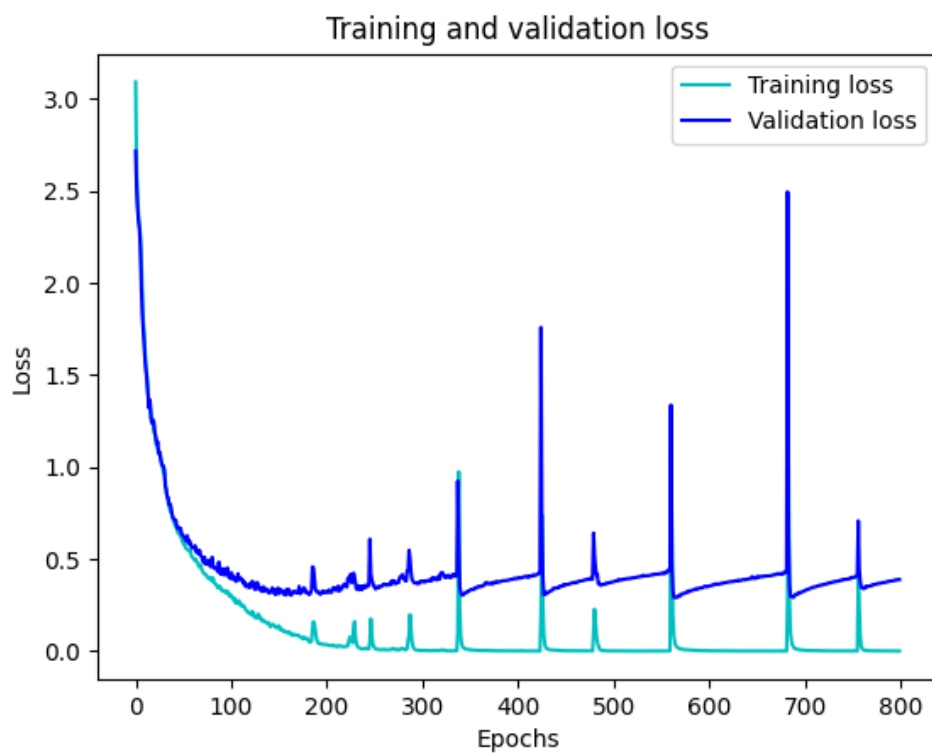


图 4: 地形分支节点数为 550 的训练集和验证集的损失值

### 1.3 方案二：无全连接层 + 多输出

在两个分支进行合并后直接送入 *softmax* 分类器进行分类，分类的输出方式为多输出。即输出 7 个一级地物标签和 20 个二级地物标签。本实验二的地形分支的节点数选取方法一中所得到的 550。本实验需要调节的超参数为一级地物标签的损失值和二级地物标签的损失值的加权系数。二级地物的系数固定为 1，一级地物的系数的所选取的范围为 0-1(一步为 0.1) 和 1-10(一步为 1)。

每个加权系数运行 5 次，最后取平均值和去极值平均值。部分加权系数的 5 次运行所得到的验证集结果如表2所示。各个加权系数 5 次运行的正确率平均值可见图5，去极值平均的结果可见图6。由图5和6可知，一级地物标签的加权系数为 0.3 时，实验的验证集正确率最高，故选定一级地物标签的损失值加权系数为 0.3。

表 2: 部分加权系数的 5 次运行结果

	0.1	0.2	0.3	0.4	0.5
1	0.911	0.9023	0.9113	0.91	0.8985
2	0.9113	0.9138	0.9135	0.9058	0.9111
3	0.9113	0.9087	0.9118	0.916	0.9077
4	0.9131	0.9136	0.9125	0.8999	0.8815
5	0.9131	0.9112	0.914	0.914	0.9124
平均	0.91196	0.90992	0.91262	0.90914	0.90224
去极值平均	0.9119	0.911167	0.9126	0.909933	0.905767

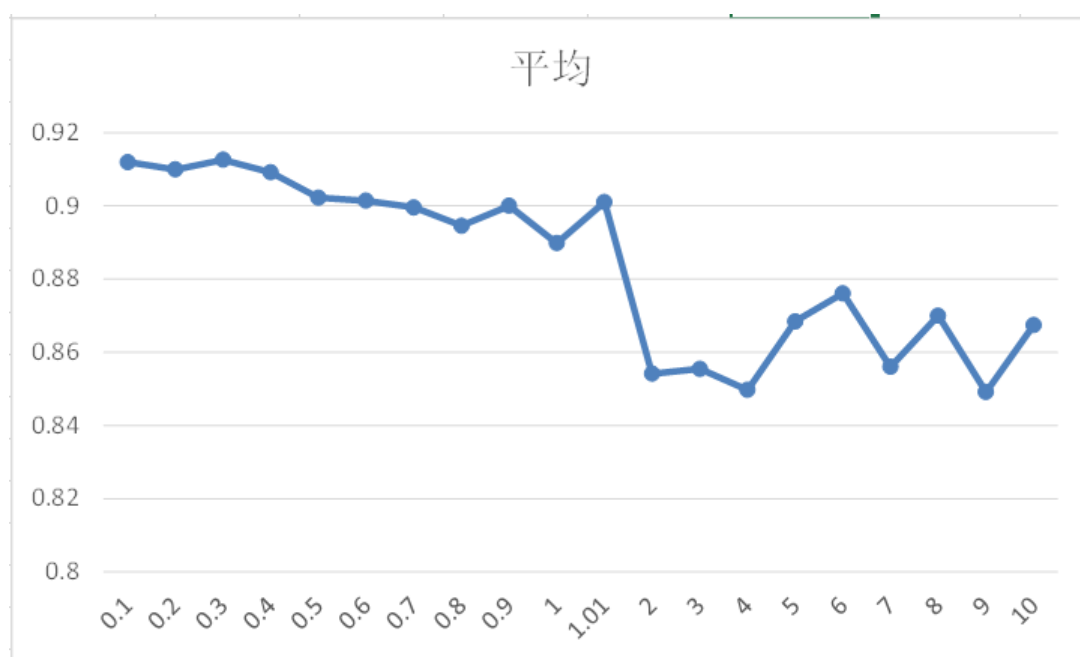


图 5: 各个加权系数 5 次运行的正确率平均值

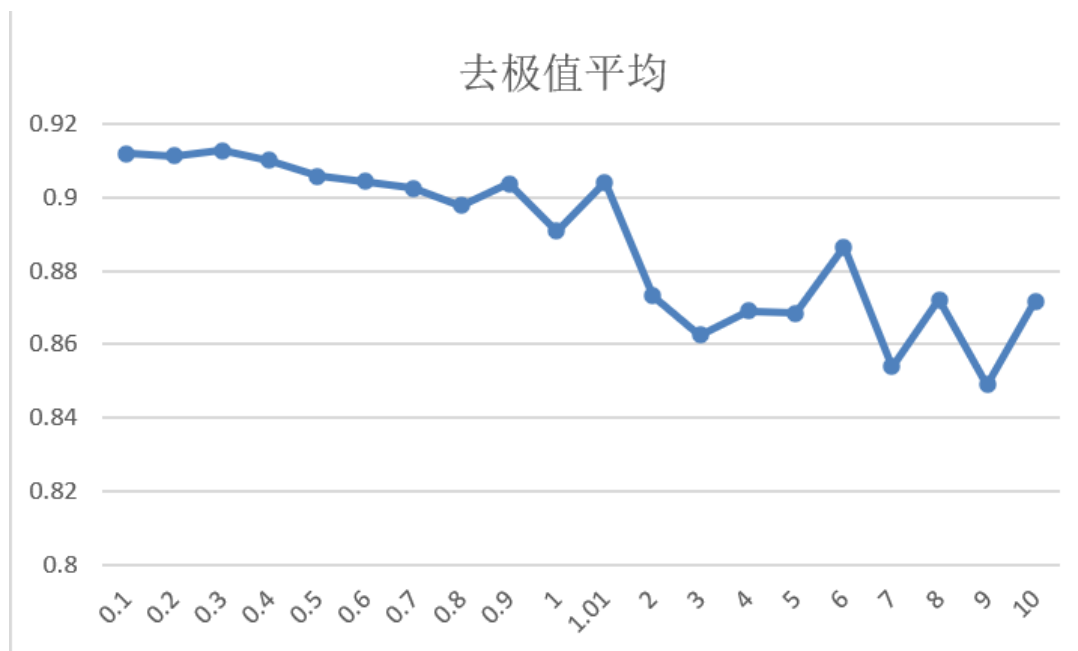


图 6: 各个加权系数 5 次运行的正确率去极值平均值

我使用一级地物加权系数为 0.3 训练得到的模型对测试集进行预测，并获取了其在训练集和验证集上的正确率和损失值随着迭代次数变化的图，其中验证集的变化如图7和8所示

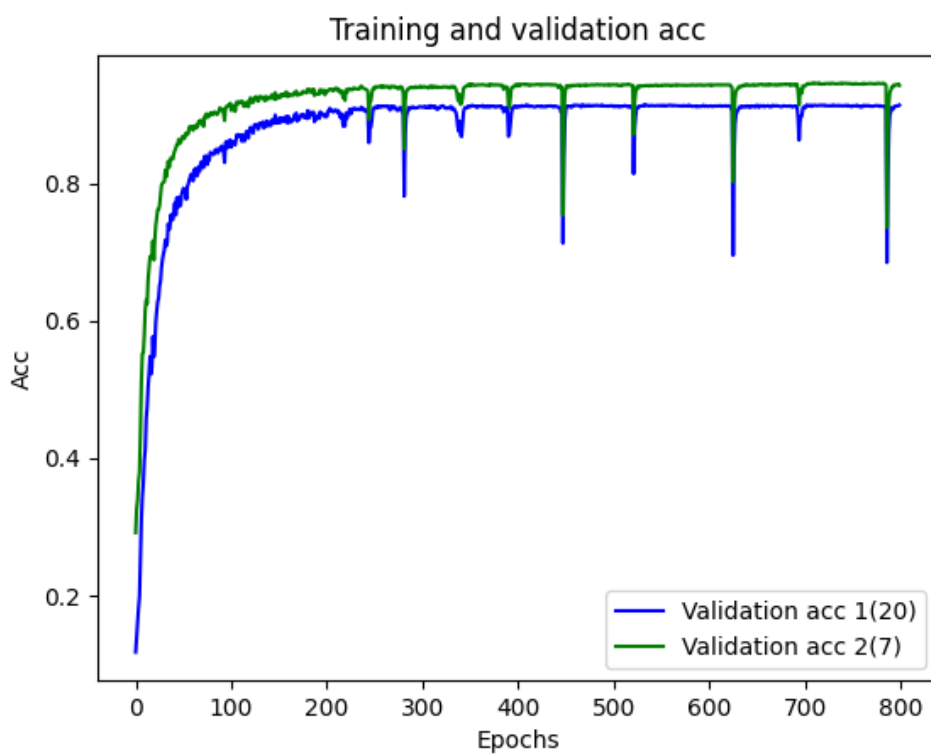


图 7: 一级地物的加权系数为 0.3 的验证集正确率

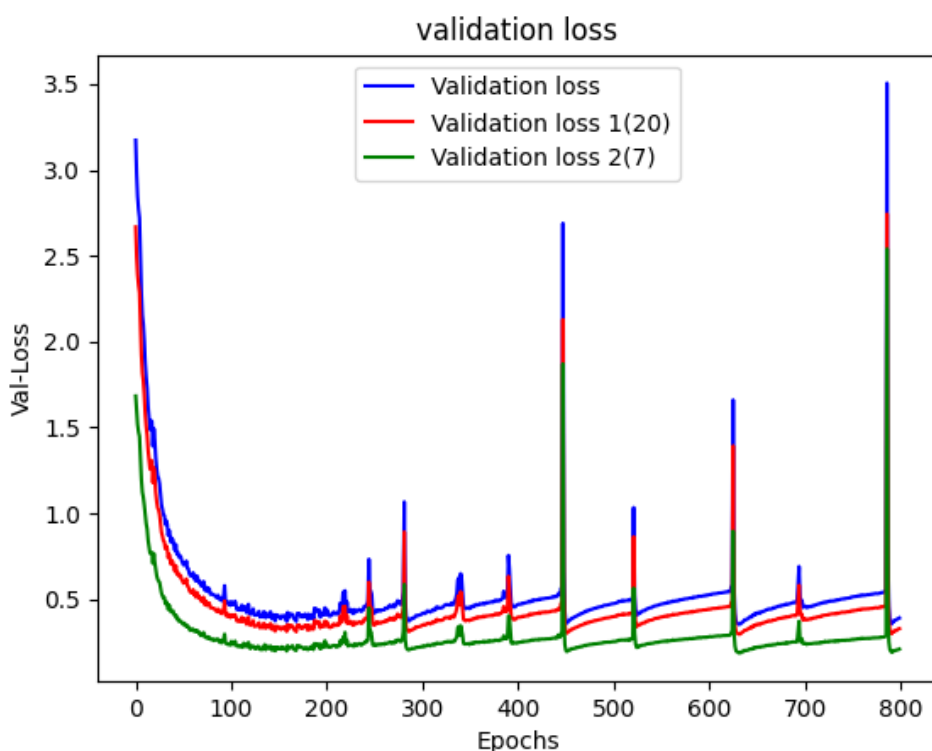


图 8: 一级地物的加权系数为 0.3 的验证集损失值

#### 1.4 方案三：一个全连接层 + 单输出

本实验在两个分支进行合并后再输入到一层全连接层，然后再送入 *softmax* 分类器进行分类，分类的输出方式为单输出，即只输出 20 个二级地物标签的类别。本实验没有需要调节的超参数，地形分支节点数为方法一中所选取的 550。全连接层的节点数为空间-光谱分支的节点数与地形分支的节点数的加和，即  $1500 + 550 = 2050$ 。

使用该模型对测试集结果进行预测，并获取到其在训练集与验证集上的精度与损失值随着迭代次数的变化情况的图，如图9和10所示。

#### 1.5 方法四：一个全连接层 + 多输出

本实验在两个分支进行合并后再输入到一层全连接层，然后再送入 *softmax* 分类器进行分类，分类的输出方式为多输出，即输出 7 个一级地物标签和 20 个二级地物标签。本实验四的地形分支的节点数为方法一中所选取的 550，一级地物的损失加权系数为方法二中所选取的 0.3。全连接层的节点数为空间-光谱分支的节点数与地形分支的节点数的加和，即  $1500 + 550 = 2050$ 。

使用该模型对测试集结果进行预测，并获取到其在训练集与验证集上的精度与损失值随着迭代次数的变化情况的图，验证集精度和损失变化如图11和12所示。

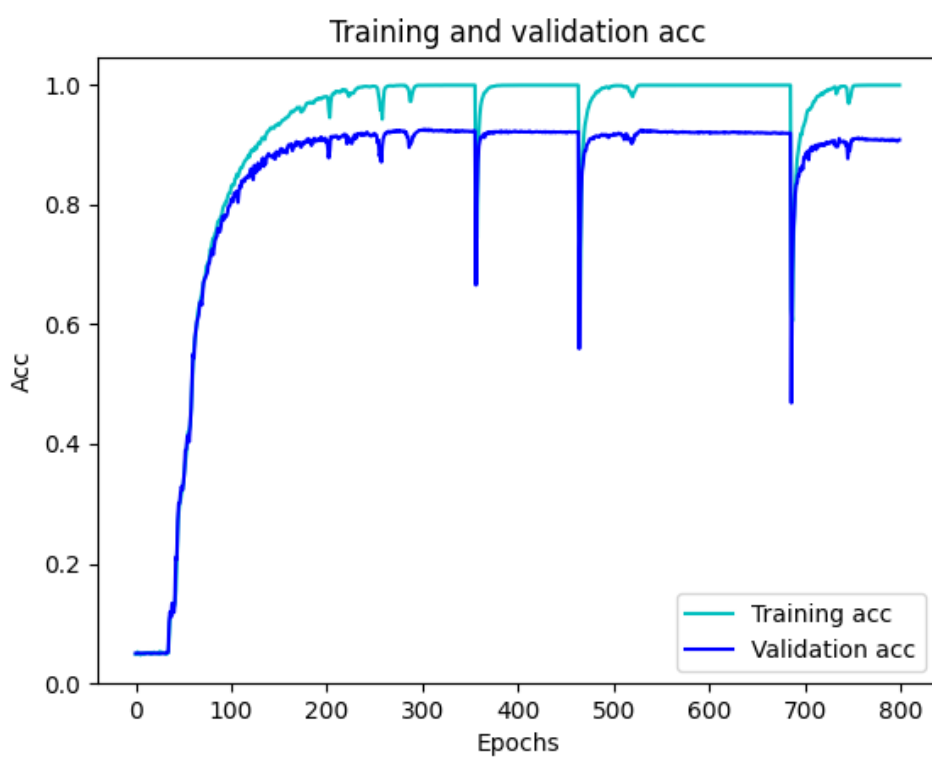


图 9: 方法三在训练集与验证集上的精度变化

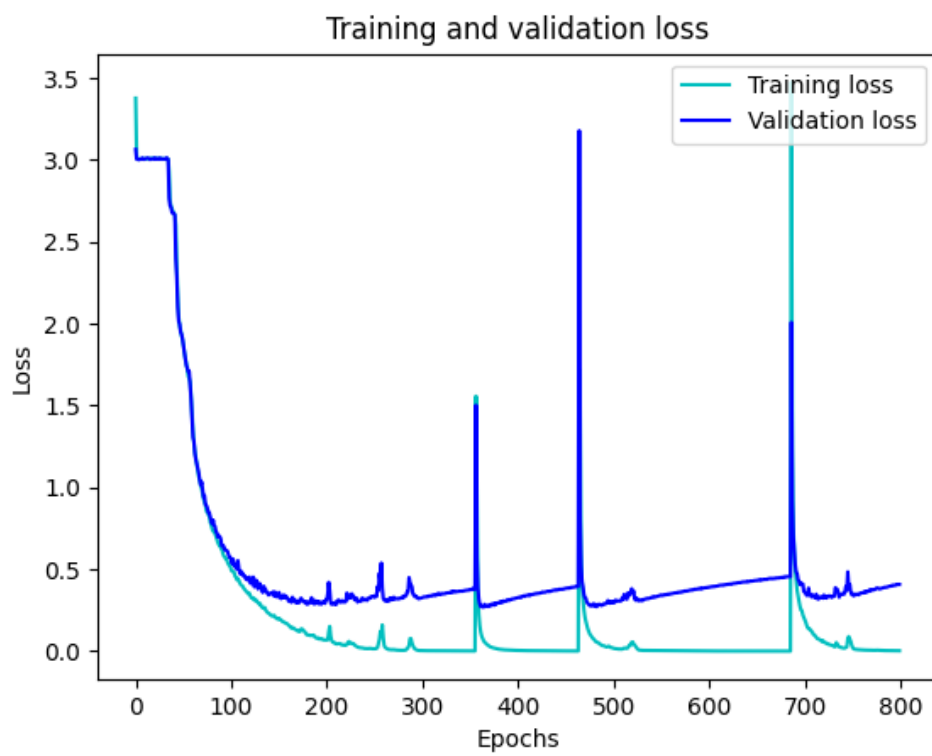


图 10: 方法三在训练集与验证集上的损失值变化

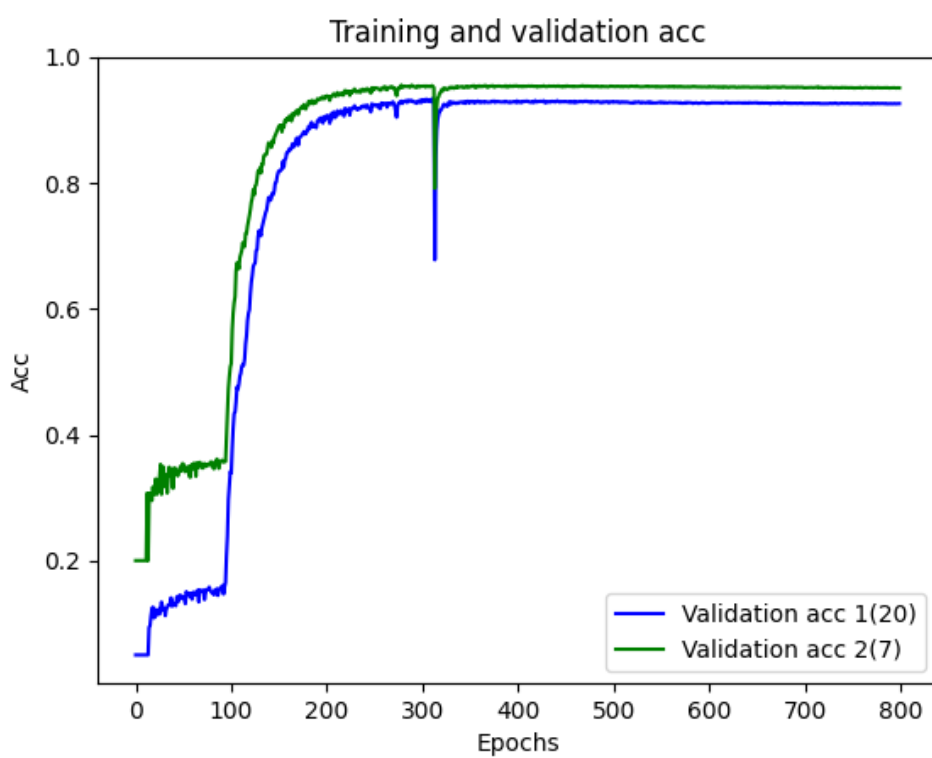


图 11: 方法四在验证集上的精度变化

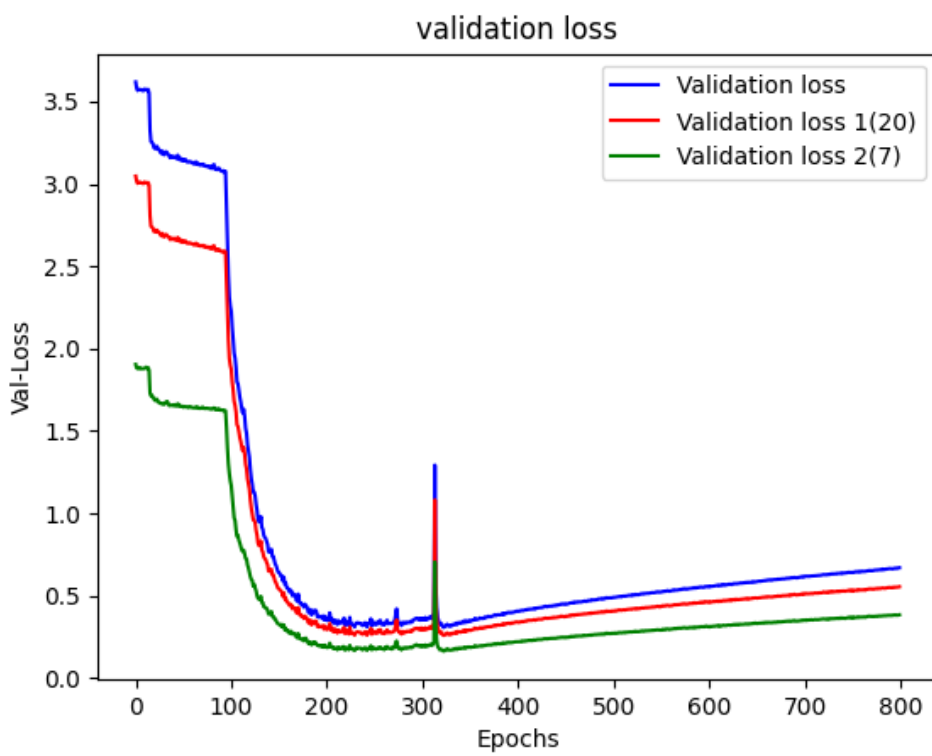


图 12: 方法四在验证集上的损失值变化



## 1.6 方法对比

我们根据上述四种方法确定了地形分支的节点数为 550，在多输出方面，一级地物损失值的及加权系数为 0.3. 接着，我们将训练好的四种方法的模型应用到测试集上进行对比，对比使用五组测试集，并求得平均进行比较，实验结果如表3所示。

表 3: 方法对比

	测试集 1	测试集 2	测试集 3	测试集 4	测试集 5	平均值
方法 1	0.9129	0.9181	0.9172	0.9203	0.919	0.9175
方法 2	0.9139	0.9141	0.919	0.9162	0.9134	0.91532
方法 3	0.9333	0.9296	0.9357	0.9355	0.9346	0.93374
方法 4	0.9259	0.9251	0.9274	0.9287	0.9267	0.92676

由表3可以看出，方法 3 一个全连接层加单输出的方式在五组测试集上结果最好实验平均值达到了 0.93374. 其次是方法 4 一个全连接层加多输出的方式，达到了 0.92676. 由此可见，在原来的模型上加上一个全连接层，可使当前的分类结果更好。对于多输出模型的精度小于单输出模型精度的，可能是由多输出模型在训练集上过拟合导致的。

## 2 实验二-SVM

### 2.1 固定参数

我们在本实验中使用的固定参数与实验一相同，并且在实验一的基础上，我们得到了地形分支的节点数为 550，并且在多输出方面，一级地物的损失值加权系数设为 0.3。

### 2.2 模型选择

我们使用由方法一得到的四个模型来获取深度特征，并将其输入到 SVM 分类器中进行分类。选取在测试集上分类效果最好的模型来对 SVM 分类器进行调整参数，实验结果如表4所示。

表 4: 模型选择

	无	SVM
方法 1	0.9129	0.9491
方法 2	0.9139	0.9497
方法 3	0.9333	0.9378
方法 4	0.9259	0.9324

由表4可以看出，方法 2 的模型所得到的结果最好，故选取方法二的模型作为 SVM 的超参数调整模型。

## 2.3 参数调整

对于 DBN-SVM 模型，对 cost 和 gamma 进行了优化。总共使用了 30 种组合，即 cost 选取 6 个值 (0.5,2,8,32,128,512), gamma 选取 5 个值 ( $2^{-11}$ ,  $2^{-9}$ ,  $2^{-7}$ ,  $2^{-5}$ ,  $2^{-3}$ )。我们获得了该 30 个组合在验证集上的结果，由表5可以看出，gamma 值可以确定选取  $2^{-5}$ ，对于 cost 值，因为当值为 2, 8, 32, 128, 512 时，其精度相差不大，故选取该五个值均在测试集上获取结果。

表 5: 30 个组合在验证集上的精度

cost \ gamma	0.00049	0.00195	0.00781	0.03125	0.125
0.5	0.8952	0.9146	0.9289	0.9468	0.7859
2	0.9133	0.9221	0.9347	0.95	0.882
8	0.9179	0.9237	0.9331	0.9497	0.8821
32	0.9199	0.923	0.9319	0.9498	0.882
128	0.92	0.9223	0.932	0.9497	0.882
512	0.9194	0.9221	0.9318	0.9497	0.882

在6中，我们可以看到在测试集上精度最好的的 cost 值是 8，其所得到的精度为 0.95. 由此，我们得到了 SVM 的最佳超参数组合为 cost=8, gamma= $2^{-5}$ , 使得最后的分类精度达到了 95%.

表 6: 5 个 cost 参数在测试集上的精度

cost \ gamma	0.03125
2	0.9497
8	0.95
32	0.9499
128	0.9499
512	0.9499

## 2.4 实验对比

和实验一的对比相同，我们选取方法一到方法四的模型并与 SVM 分类器结合来预测测试集分类结果。其中 SVM 分类器的超参数选择前面所得到的最优参数组合 cost=8, gamma= $2^{-5}$ . 同样的，使用 5 组测试集，并对最后求得的精度进行平均来对比，实验结果如表7所示。

由7可以看出，方法一的模型加上 SVM 可以使得分类结果最好，其结果为 95.386%，其次方法所得到的结果依次递减。其中可能的原因为，和 softmax 不同，不加全连接层所得到的深度特征，能够使得 SVM 分类器得到更好的结果。

表 7: 模型对比

	测试集 1	测试集 2	测试集 3	测试集 4	测试集 5	平均值
方法 1+svm	0.9491	0.9533	0.9547	0.9573	0.9549	0.95386
方法 2+svm	0.95	0.9476	0.9527	0.9538	0.9515	0.95112
方法 3+svm	0.9378	0.9367	0.9419	0.9409	0.939	0.93926
方法 4+svm	0.9324	0.9302	0.9347	0.9359	0.9319	0.93302