

项目报告

1.项目背景

我们经常遇到这样的场景：一盏灯变成绿色，你面前的车不走。另外，在没有任何意外发生的情况下，前面的车辆突然减速，或者转弯变道。等等这些现象，给道路安全带来了很大的影响。那么造成这样现象的原因是什么，主要有因为司机疲劳驾驶，或者走神去做其他事情，想象身边的例子，开车时候犯困，开始时候打电话，发短信，喝水，拿后面东西，整理化妆的都有。这对道路安全和行车效率形成了极大的影响。该项目的主要目的是通过摄像头采集数据实现驾驶员的行为检测，通过该项目可以实时对图像数据进行分析得到驾驶员行为状态。主要所使用的技术为目前最流行的深度学习中卷积神经网络进行实现。随着LeCun在1989年通过使用卷积神经网络实现手写数字体识别，到2012年Alexnet在Imagenet竞赛中夺得图像分类冠军，深度学习的浪潮就此掀开，鉴于此。本问题使用卷积神经网络完全可以实现。



img_6.jpg



img_115.jpg



img_149.jpg



img_357.jpg



img_448.jpg



img_462.jpg



img_557.jpg



img_623.jpg

2.问题描述

驾驶员可能存在的走神的行为，大概有如下几种：

左右手用手机打字，左右手用手持方式打电话，调收音机（玩车机），喝饮料，拿后面的东西，整理头发和化妆，和其他乘客说话。为此，分类的准确率 **accuracy** 就是衡量解决这个问题好坏的重要指标

该问题简单来说就是一个分类问题，使用深度学习方法进行分类识别。输入为一张彩色图片，输出为十种指定状态的概率。由此可见，如果使用深度学习方法，多分类问题的输出层为**softmax**，通过该层可以预测每一个类别的概率，最终求得最大概率。

3.数据或输入

数据有以下特点：

- 1.训练集的图片数量是 22424，测试集的图片是 79726，测试集的数量远大于训练集。
- 2.训练集司机完全不是测试集司机。

由于数据有这些特点，那么过拟合可能是要遇到的最大的问题。

该数据分为10类，每类约有1000张图片，训练集平衡，整体约20000张图片，可以使用迁移学习把后面几层改为可训练层进行训练。该项目难度中上等，主要难度在于大量的数据训练和模型调试搭建。

4.解决方法描述

这是典型的分类问题，评估指标采用精度 **accuracy** 来评估结果好坏。

Logloss 的评估方式，这也是 kaggle 比赛的评估方式

主要解决方法是使用基于GoogLe net或者Res net 进行分类识别。

5.评估标准

这是典型的分类问题，一般的评估指标采用精度 **Accuracy**来评估结果好坏。

logloss的评估方式，这也是 kaggle 比赛的评估方式：

$$\text{logloss} = -1/N \sum_1^N \sum_1^M y_{ij} \log(p_{ij})$$

相对比这两种方案。**Accuracy**对于判断正确和错误的比重是一样的，也就是对了就多一个，错了就少一个，最终看正确的百分比，也就是判断正确的占总体测试数据的比例。由于**Accuracy**无法显示出函数是否具有显著性，所以选择**logloss**方法进行评判。

6.基准模型

基准模型使用在Lenet,该模型为最早的卷积神经网络模型是一个5层网络，之前用来实现手写数字识别，可以作为baseline进行测试比较。

7.项目设计

项目设计分为以下几个阶段：

1. 进行基准模型搭建，搭建Lenet算法模型，得到基准的准确率。
2. 进行深度学习模型构建，学习搭建GoogLe net 和 Res net进行分类识别。计算准确率。
3. 通过阅读文献对模型结构修改，使用数据增强方法进行进一步微调，提升其准确率。

8.项目实施

8.1算法选择

我整体上采用卷积神经网络 CNN 来解决问题。卷积神经网络是神经网络的一种，把图片的部分相邻像素变成长宽更小，高度越深的单元，从而提炼出局部的特征，一层一层将特征往后映射，之后提取出最终的特征。现在世界上大部分图片处理和计算机视觉问题都是用卷积神经网络 CNN 来解决的。

8.2Loss函数选择

由于是多分类问题，所以采用 categorical_crossentropy 的计算原理来使用 loss，下面简单介绍一下 categorical_crossentropy 的数学公式：

$$\text{logloss} = -1/N \sum_1^N \sum_1^M y_{ij} \log(p_{ij})$$

该函数具有一个显著特点，下面举例简述下：

1.该函数是对每个分类的概率分别求和。

2. **logloss** 的评估方式对判断有明显性，如果正确了， $P_{ij}=1 \Rightarrow \log(P_{ij})=0$ ，而 $P_{ij}=0.999 \Rightarrow \log(P_{ij})=-0.001$ 。最后增加的 \log 差不多。但如果判断错误，如 $P_{ij}=0 \Rightarrow \log(P_{ij})=-\infty$ 。 $P_{ih}=0.001 \Rightarrow \log(P_{ij})=-6.9$ 也就是判断错误一个，对得分的影响会非常大，所以用 **logloss** 评估对比 **accuracy**，更能反映模型和算法的能力。

8.3 优化函数选择

我最终选择 Adam 来做主优化器，然后再用学习率极低的 RMSprop 来补充优化。

下面简单介绍一些优化器的由来：

RMSprop 优化器，可以算作 Adadelta 的一个特例：

$$\rho=0.5 \quad E|g^2|_t = \rho * E|g^2|_{t-1} + (1 - \rho) * g_t^2$$

就变为了求梯度平方和的平均数。

如果再求根的话，就变成了 RMS(均方根)：

$$RMS|g|_t = \sqrt{E|g^2|_t + \epsilon}$$

此时，这个 RMS 就可以作为学习率 η 的一个约束：

$$\Delta x_t = -\frac{\eta}{RMS|g|_t} * g_t$$

RMSprop 的特点是：

1.RMSprop 依然依赖于全局学习率。

2.RMSprop 算是 Adagrad 的一种发展，和 Adadelta 的变体，效果趋于二者之间适合处理非平稳目标。

Adam：

Adam 优化器(Adaptive Moment Estimation) 本质上是带有动量项的 RMSprop，它利用梯度的一阶矩估计和二阶矩估计动态调整每个参数的学习率。Adam 的优点主要在于经过偏置校正后，每一次迭代学习率都有个确定范围，使得参数比较平稳。公式如下：

$$m_t = u * m_{t-1} + (1 - u) * g_t$$

$$n_t = v * n_{t-1} + (1 - v) * g_t^2$$

\$\$

Adam 的特点是：

1. 结合了 Adagrad 善于处理稀疏梯度和 RMSprop 善于处理非平稳目标的优点
2. 对内存需求较小
3. 为不同的参数计算不同的自适应学习率
4. 也适用于大多非凸优化 - 适用于大数据集和高维空间

由于 Adam 的这些特点，我在做 Fine-tune 的时候选择 Adam 作为主优化器，在优化到差不多的时候，再采用 RMSprop 进行超低学习率的调优。由于单模型的特征已经学习过了，所以在混合模型的训练的时候，我直接 Adam 用超低学习率来进行学习。

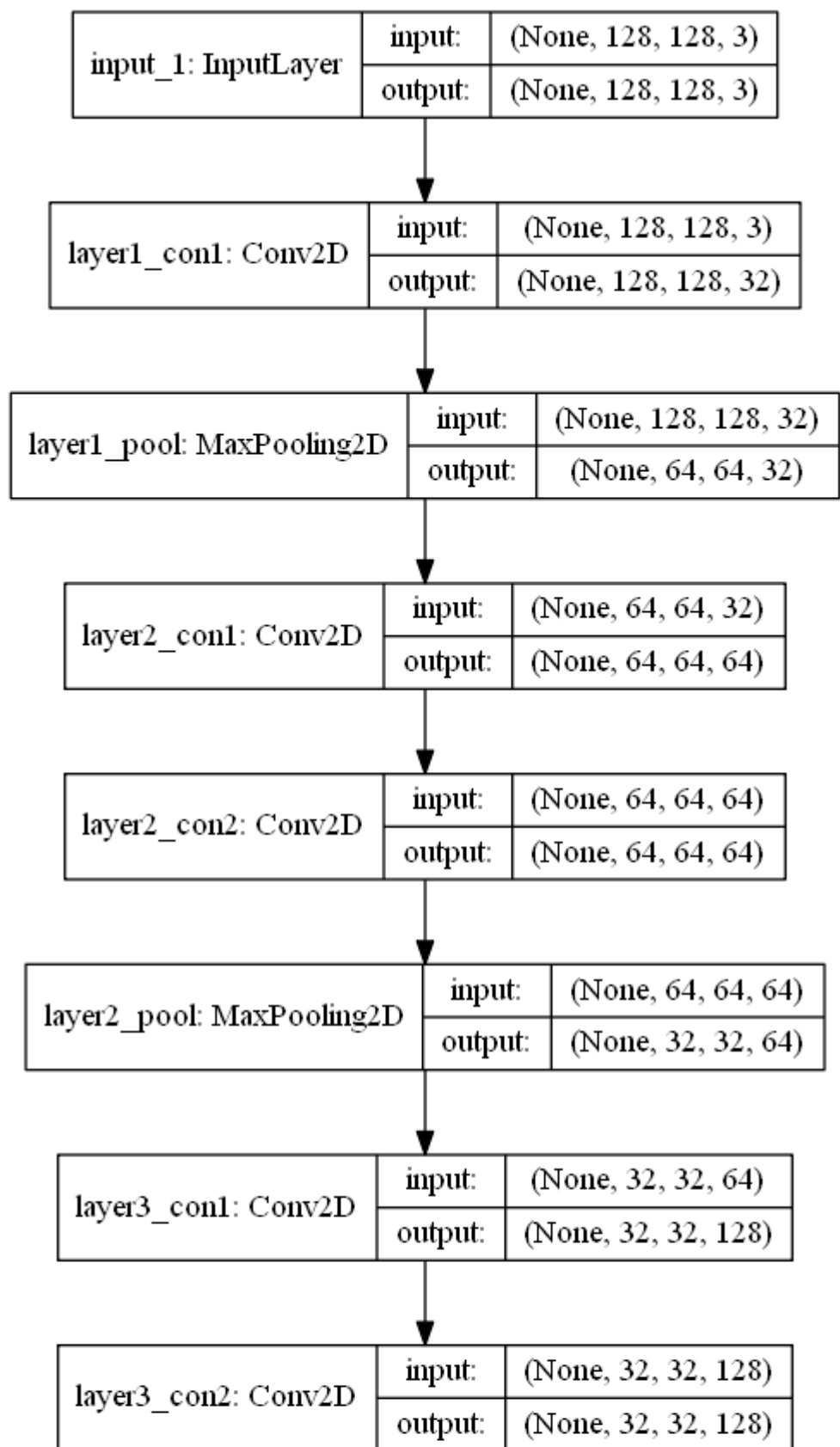
试验后发现，在单模型 Fine-tune 的时候，使用 RMSprop 比起 Adam 优化完成的，要好 0.03~0.05 个 accuracy，LogLoss 要好 0.05~0.2 个点。但是如果全部使用 RMSprop，收敛得比较慢，且也达到开始用 Adam 优化器训练的精度。

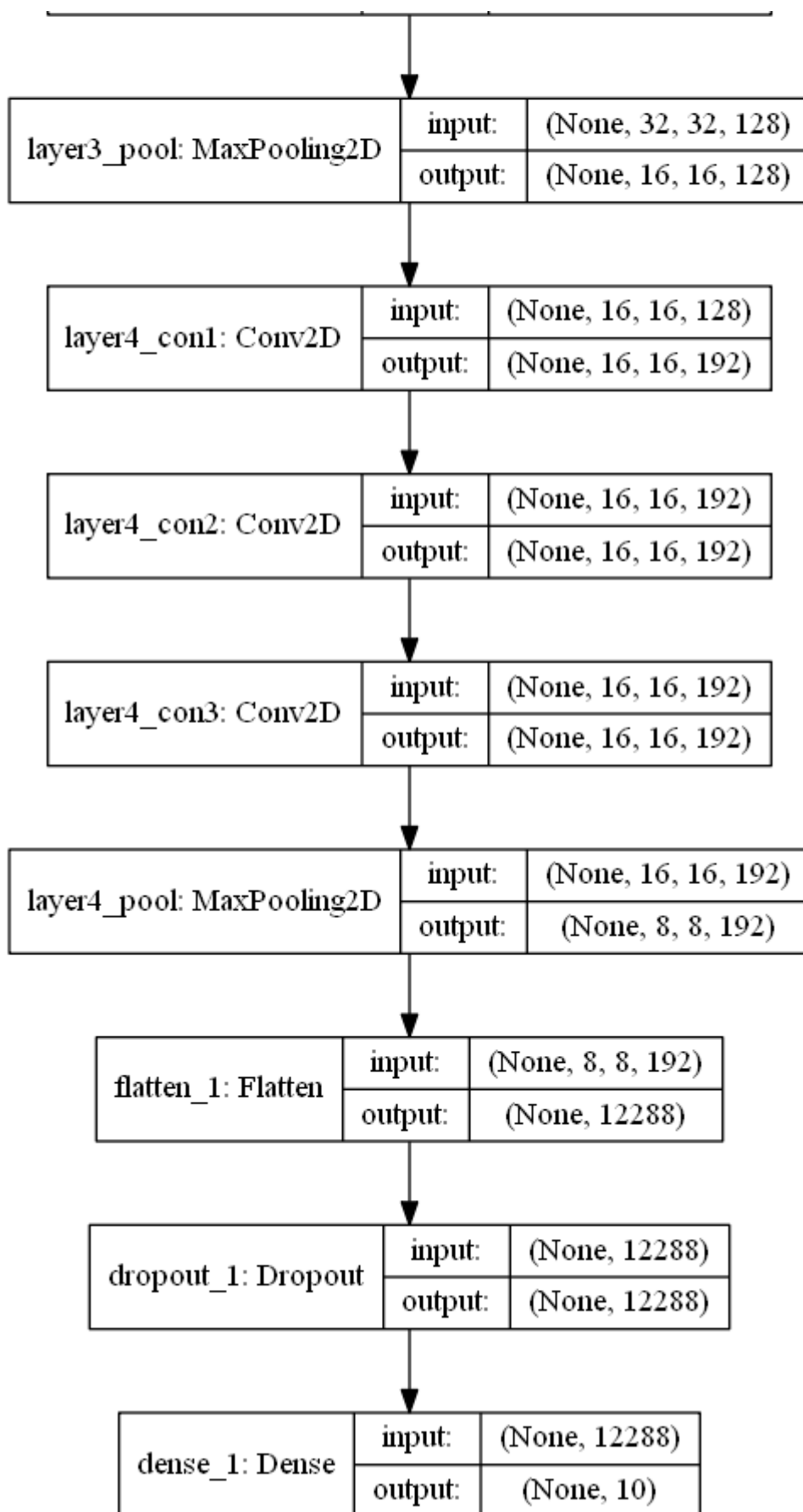
8.4 模型搭建

基于借鉴vggnet和lenet网络模型，我自行搭建了两个神经网络模型。

1.第一个模型使用输入为128*128大小，深度为三通道的图片作为输入，搭建8层卷积层，没有添加全连接层，其目的是减少权重大小，但是经过多次训练结果显示效果并不好，偶尔一次会出现正确率达到90左右的情况，可想而知初始化权重是一件非常重要的事情。

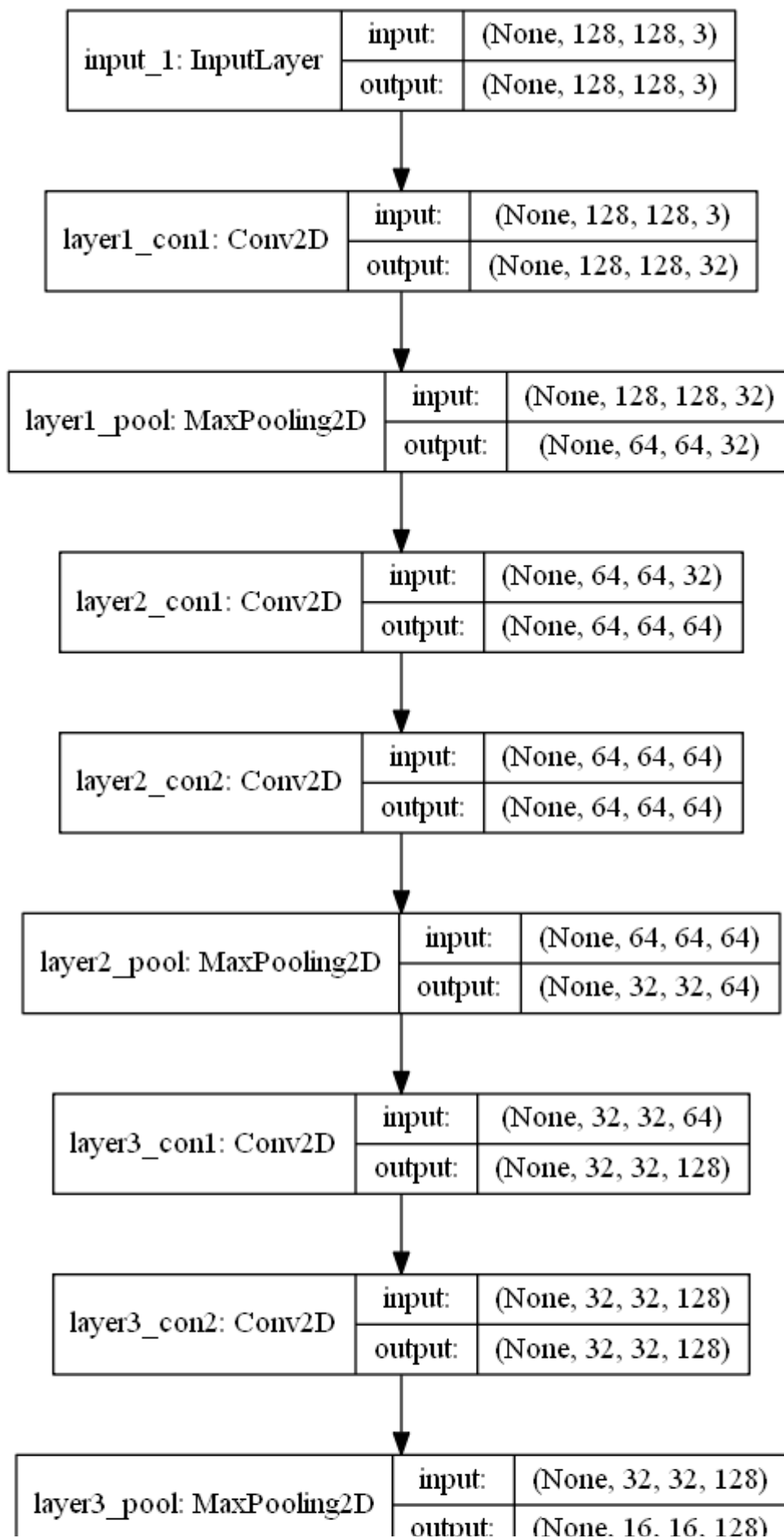
模型1网络结构如图所示：

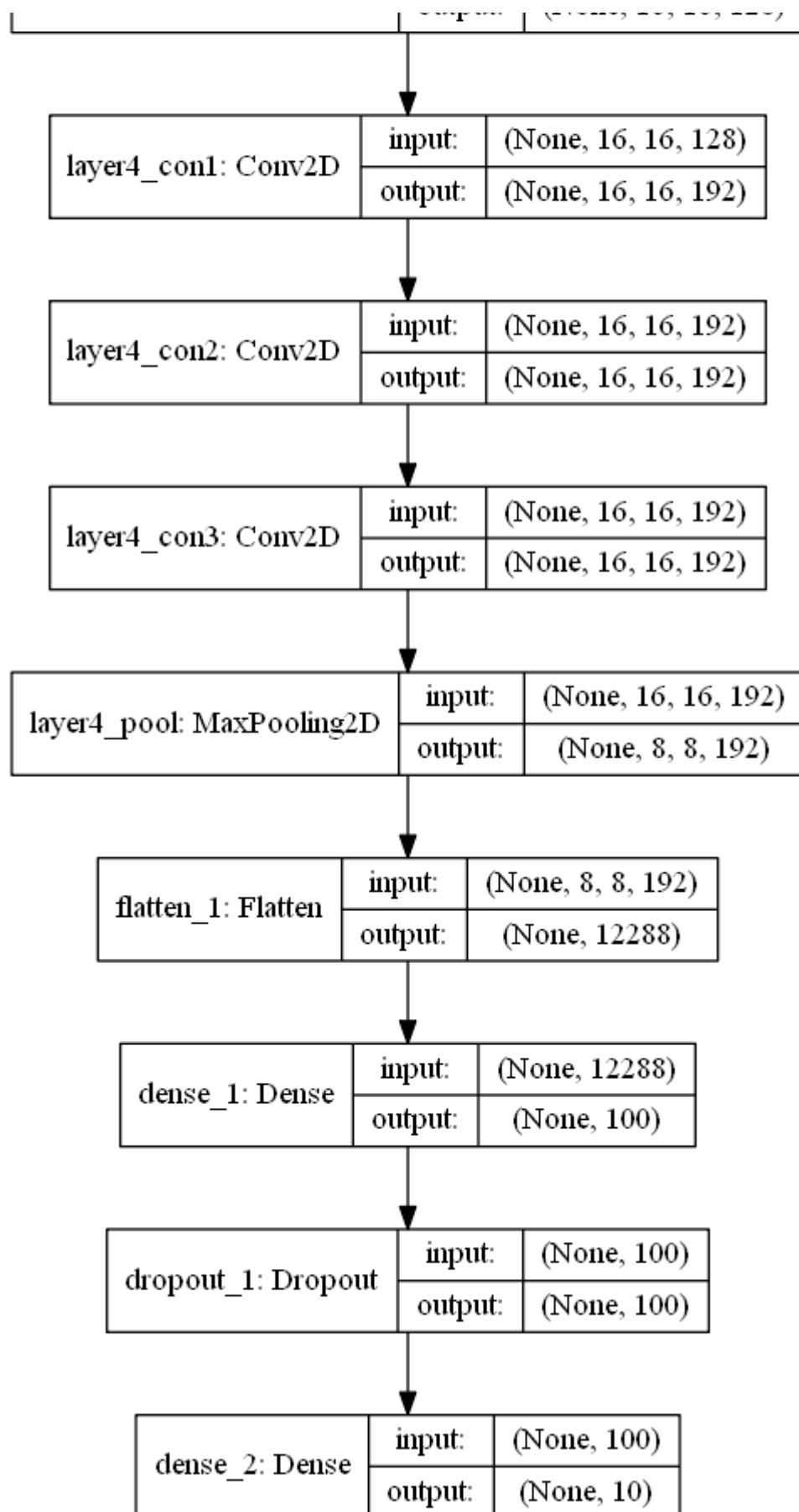




2.模型二主要是在模型一的基础上末端添加了一层全连接，使得具有更好的分类效果，经过测试发现效果果不其然十分显著，在validation数据集上的**Accruacy**达到98%，效果十分好。

网络结构2如图所示：





8.5模型总结

相比于网络上的其他网络模型，如~~Alexnet~~ *VGGnet* *GoogleNet*等网络，该网络具有比较小的运算量，整体权重导出后仅有50MB，最后只有一层全连接层，运算速度快，可以有较高的实时性，且分类效果正确率达到98%，具有较高的现实意义。

Reference

[1]Y. LeCun, B.boser, J.Denker, D.Henderson, R. Howard, W. Hubbard, and L. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Comp.* ,1989. 1

[2]Karen Simonyan, Andrew Zisserman Very Deep Convulutional Networks for Large-Scale Image Recognition 2015