【1.  数据在网络中的维度顺序是什么?】

进去的时候是Nx1x112x112(N代表batch size，以下忽略)。

经过conv1\_1后，根据(W+2P-K)/S+1=(112+2\*0-5)/2+1=54，所以这步过后是8x54x54（除法那步向下取整）

经过avg\_pool后，根据(W+2P-K)/S+1=（54+0-2）/2+1=27, 所以这步过后是8x27x27

经过conv2\_1后，(27+0-3)/1+1=25，这步过后是16x25x25

经过conv2\_2后，(25+0-3)/1+1=23，这步过后是16x23x23

经过avg\_pool后，(23+0-2)/2+1=12(注意程序这里用了ceil)，这步过后是16x12x12

经过conv3\_1后，(12+0-3)/1+1=10，这步过后是24x10x10

经过conv3\_2后，(10+0-3)/1+1=8，这步过后是24x8x8

经过avg\_pool后，(8+0-2)/2+1=4，这步过后是24x4x4

经过conv4\_1后，(4+2-3)/1+1=4，这步过后是40x4x4

经过conv4\_2后，(4+2-3)/1+1=4，这步过后是80x4x4

经过ip1后，为128

经过ip2后，为128

经过ip3后，为42

【2.  nn.Conv2d()中参数含义与顺序?】

输入的C，出来的C，kernel大小，stride，padding

【3.  nn.Linear()是什么意思?参数含义与顺序?】

输入的C，输出的C

【4.  nn.PReLU()与 nn.ReLU()的区别?示例中定义了很多 nn.PReLU()，能否只定义一个 PReLU?】

输入为负数时不都为0。每个PReLU有个控制斜率的参数，所以不能只定义一个。

【5.  nn.AvgPool2d()中参数含义?还有什么常用的 pooling 方式?】

kernel size和stride。maxpool吧

【6.  view()的作用?

改变了维度，在这里等于flatten

此部分，是一些“程序控制代码”，包括:

1. 如何设置 GPU】

device = torch.device("cuda" if use\_cuda else "cpu")

1. 如何将数据/网络传入 CPU/GPU】

. to(device)

1. 如何读取数据

这个就不拷贝了，程序里挺长的

1. 如何设置 loss】

不清楚什么意思啊，程序里我就直接用了MSE的类了。。。

1. 配合后续周学习，loss 都有哪些。分别有什么作用(常用的即可)】

MSE，cross entropy，L1loss等

1. 如何设置优化器】

SGD的话主要就是lr和momentum了

1. 配合第 8 周内容，常用的优化器有哪些】

SGD,Adam,Adagrad

【2. optimizer.zero()与 optimizer.step()的作用是什么?】

pytorch里头梯度会累积，zero就把所有梯度先赋零；step就是走一步learning step

【3. model.eval()产生的效果?】

梯度就不算了

【4. model.state\_dict()的目的是?】

把model各层用dict表示出来，以方便读写

【5. 何时系统自动进行 bp?】

在计算forward的时候会顺便把梯度算出来

【6. 如果自己的层需要 bp，如何实现?如何调用?】

需要自己把梯度存到weights里，step的时候更新这些weights

1. Finetune 时，有时还要固定某些层不参与训练，请回答如何 freeze 某些层

我是直接将requires\_grad设置成False，不知道有没有更好的办法