CNN 问题：

认清楚下面两章图上的猫是同一只对于 CNN 很难。解决图像迁移后的一致识别，是 Capsule 神经网络出现的第一个目的。不止是左右的平移，CNN 对于旋转，加上边框也会难以觉察出其一致性，CNN 会认为下图的三个 R 是两个不同的字母，而这是由网络结构所带来的，这也造成了 CNN 所需的训练集要很大，而数据增强技术虽然有用，但提升有限。

下面的这张图会被 CNN 看成是一张脸，这是因为 CNN 识别脸的时候，是识别脸的几个组成部分，下图中的确有两个眼睛，有鼻子，有嘴，虽然其位置不对，这对于 CNN 来说就够了。

人类在识别图像的时候，是遵照树形的方式，由上而下展开式的，而 CNN 则是通过一层层的过滤，将信息一步步由下而上的进行抽象。这是 Hinton 认为的人与 CNN 神经网络的最大区别。

对比了 CNN 和胶囊神经网络在识别人脸上的区别。CNN 中需要有一层去应对不同翻转角度的脸，而胶囊网络则可以用一个胶囊去完成这个任务

输入是一张手写字的图片。首先对这张图片做了常规的卷积操作，得到 ReLU Conv1；然后再对 ReLU Conv1 做卷积操作，并将其调整成适用于 CapsNet 的向量神经元层 PrimaryCaps

这里使用一个 9x9x256 的卷积核，步长是 2，所以输出向量的维度减半：((20—9) + 1) / 2 = 6。最终可以得到一个含有 256 个输出（6x6）的特征图。然后要处理成胶囊的形式，也就是切分成 32 层，其中每层有 8 个块，也就是说一个胶囊向量中包含了 8 个值。

对于 CNN 中单一的像素标量值来说，我们仅仅能够存储在特定位置是否有一个边角的置信度。数值越大，置信度越高。而对于一个胶囊，我们可以在每个位置存储 8 个值，可以存储更多的信息，比如位置、旋转角度、颜色等。

非线性变换 Squashing：在得到胶囊之后，我们会再对其进行一次非线性变换。该激活函数前一部分是输入输入向量 S 的缩放尺度，后一部分是 S 的单位向量。该激活函数既保留了输入输入向量的方向，又将输入向量的模压缩到[0,1]之间。

一致性路由:接下来的一步是决定将要被传递给下一个层级的信息。在 CNN 中，我们也许会使用类似于「最大池化」的一些方式。然而，在胶囊网络中，使用

称作「一致性路由」的方式，也就是其中每一个胶囊都试图基于它自己猜测下一层神经元的激活情况，其实也就是一种投票的方式。
看到这些预测并且在不知道输入的情况下，也许会把船的信息传下去，因为长方形的胶囊和三角形的胶囊都在船应该是什么样子上达成了一致。但他们并没有在预测房子的样子上达成一致。所以很有可能这个物体不是一个房子。通过一致性路由，我们仅仅将有用的信息传递给下一层并且丢弃掉那些可能使结果中存在噪音的数据。这让我们能够做出更加智能的选择而不仅仅是像最大池化一样选择最大的数字。
真正做的时候，要找到在这所有的预测中和其他预测一致性最高的内容。首先假设这些向量仅仅是 2 维空间中的点。计算所有点的平均值。每个点在最初都被赋予了同样的重要性。接下来我们可以测量每个点和平均值点之间的距离。距离越远的点，其重要程度就越低。然后我们在 normalize 一下，重新计算平均值。随着我们重复进行这个循环，那些和其他点不一致的的点开始消失。而那些相互之间高度一致的点则最终将被传递给激活值最高的的下一层。

达成一致后，我们最终可以得到 10 个 16 维的向量，每个向量都和一个数字相对应。这个矩阵就是我们最终的预测结果。这个向量的长度是数字被找出的置信度——越长越好。这个向量也可以被用于生成输入图片的重构。
重构的意思就是用预测的类别重新构建出该类别代表的实际图像。将最后的那个正确预测类别的向量投入到后面的重构网络中，可以构建一个完整的图像。正确预测类别的向量，即模值最大的向量送入包含三个全连接层的网络解码。这一过程的损失函数通过计算 FC Sigmoid 层的输出像素商店与原始图像像素点的欧氏距离而构建。

下面的图展示了胶囊神经网络在 MINST 数据集下的表现，这里可以看出胶囊神经网络的另一个好处，即可解释性强，你可以看出每个胶囊想做什么。

和全连接神经网络一样，胶囊网络的每一个连接也有权重。在上面图中，W 代表权重，大家需要注意：C 不是权重，它叫耦合系数.
因为每一个胶囊神经元都是向量，即包含多个值，所以每个胶囊神经元的权值 W 也应该是一个向量。
全连接神经网络的输入即线性加权求和，胶囊网络很类似，但是它在线性求和阶段上多加了一个耦合系数 C。

s 是输入，u 是上一层胶囊网络的输出，W 是每个输出要乘的权值，可以看作上一层每一个胶囊神经元以不同强弱的连接输出到后一层的某一个神经元。