Hinton对池化的不满。

卷积网络对模糊性图片的精度不高（语义分割）

卷积网络对要素齐全但相对关系很乱的图片会认为是同一种。

卷积网络对图片整体的大旋转的反应不灵。(CNN的旋转不变性依赖于图片局部特征的旋转不变。)

组件的朝向和空间上的相对关系CNN是不敏感。转一下基本就识别不出来了。

对单个组件内部而言是关心相对位置的。（毕竟你转一下就认不出来了。即连旋转不变性也没有。）

完全的旋转不变性，你也不能想象把6认成9到底多可怕，也就是说这和任务有关。

此处我们更关心一种视角上的不变性。

CNN的策略：最大池化或连续卷积层。内部表现形式不依赖视角

Capsnet：2017（思想的话零几年就有了。）

本质：

要了解的重要一点是，高级特征将低级特征作为加权总和：前一层的激活与下一层神经元的权重相乘并相加，然后传递给激活非线性。在此设置中，组成更高级别特征的简单特征之间没有姿势（平移和旋转）关系。（一个高阶特征例如有无鼻子）

CNN解决此问题的方法是使用最大池化或连续卷积层，这些层会减小流经网络的数据的空间大小，从而增加高层神经元的“视野”，从而使它们能够检测神经元中的高阶特征。输入图像的较大区域。最大池化是使卷积网络工作异常出色的关键，在许多领域都有超人的表现。但是，不要被它的性能所迷惑：尽管CNN的性能比之前的任何模型都要好，但是最大池化却在丢失有价值的信息。

An important thing to understand is that higher-level features combine lower-level features as a weighted sum: activations of a preceding layer are multiplied by the following layer neuron’s weights and added, before being passed to activation nonlinearity.

Nowhere in this setup there is pose (translational and rotational) relationship between simpler features that make up a higher level feature. CNN approach to solve this issue is to use max pooling or successive convolutional layers that reduce spacial size of the data flowing through the network and therefore increase the “field of view” of higher layer’s neurons, thus allowing them to detect higher order features in a larger region of the input image. Max pooling is a crutch that made convolutional networks work surprisingly well, achieving superhuman performance in many areas. But do not be fooled by its performance: while CNNs work better than any model before them, max pooling nonetheless is losing valuable information.

例如前三个特征表现了他在（50，50）鼻子，他（100，100），（0，100）有眼睛，他有嘴巴(0,50)它们值都很大，对于下一个普通全连接神网而言他其实没法那么关心相对区位关系。

它们之间的关系被缩小了。

第二层又做个卷积

胶囊本身是一组神经元或者小网络，当然胶囊之间还是全连接就没意义了。

旨在检测这块区域的特定目标。

不同在于可解释性。关键在于你要有可解释性的结构。比如像注意力这种东西。

对一个八维输出的胶囊而言，输出的向量的模是代表该物体存在的概率（即模会限制到0-1之间），方向随对象的旋转，移位，缩放而略有变化。

基本胶囊都是以一个小区域为输入，而高层的是路由胶囊。

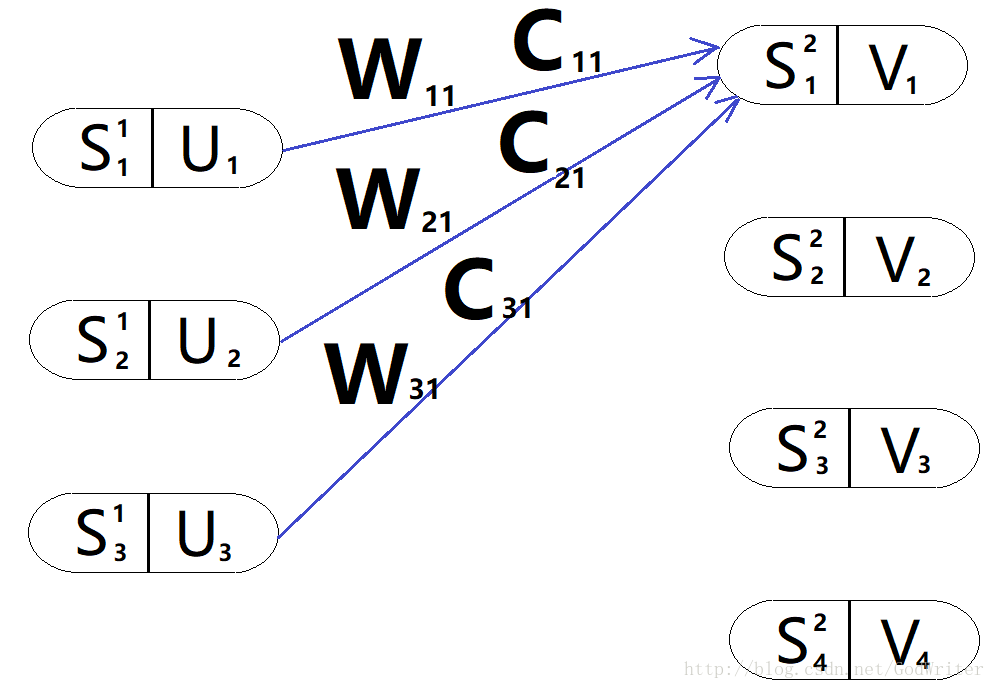
计算机图形学：用一些矩阵3d显示一个对象

胶囊网络的思想：把一个对象得到它的3d表示。他希望一个微型柱体得到的是多维实体。

* 由于没有内在的坐标框架，因此违反了生物形状感知；
* 提供不变性（丢弃位置信息）而不是相等性（解散该信息）；
* 忽略了基于线性流形的图像之间的许多变化；
* 静态路由，而不是将潜在的“发现”传达给可以欣赏它的功能；
* 通过删除它们依赖的信息来损坏附近的特征检测器。

注意力是对相关关系显示建模，而这里这是对旋转关系有一种显示建模。会按照头一层的输出与第二层的输出的内积有一个打分

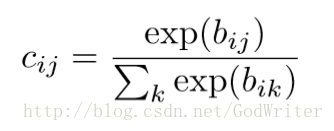
首先每个神经元对





s到u之间还有个网络，u，v是两层分别的输出。

IMG_256



c可以说是某种意义上的注意力。。

由于v与b有互相迭代关系。。即b会影响v，v也会影响b，所以就迭代（动态路由。）

Nlp中如果有类似的就好玩了。

生成词向量求内积出样本。。甚至可以加扰动。