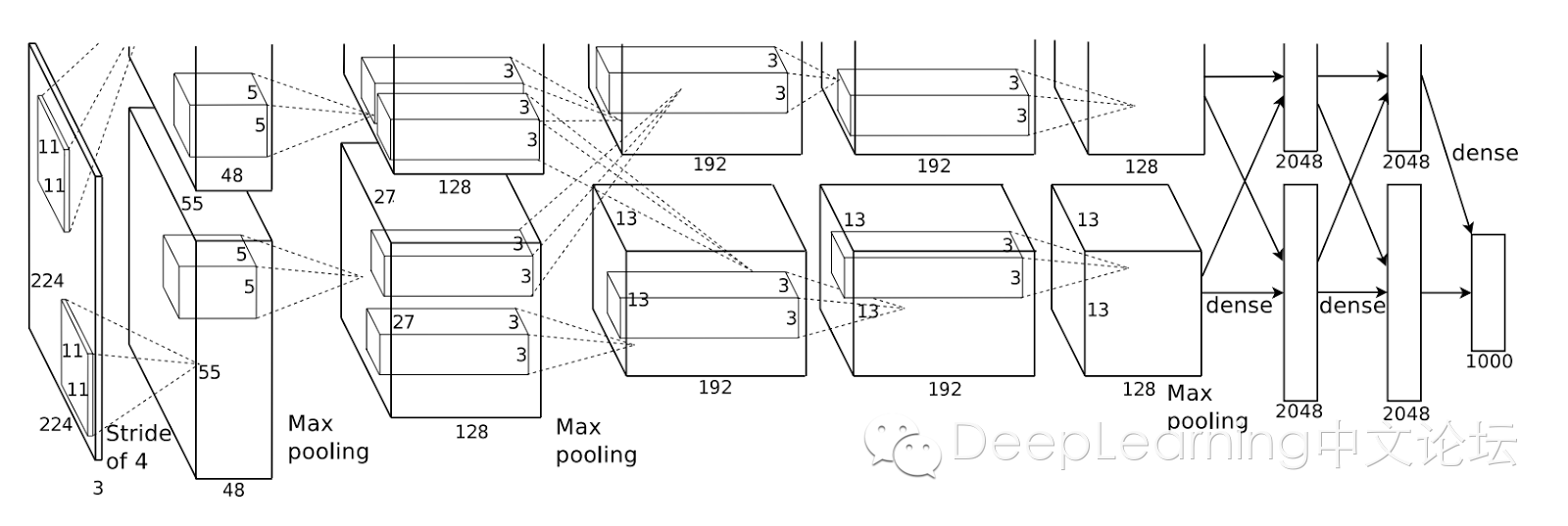
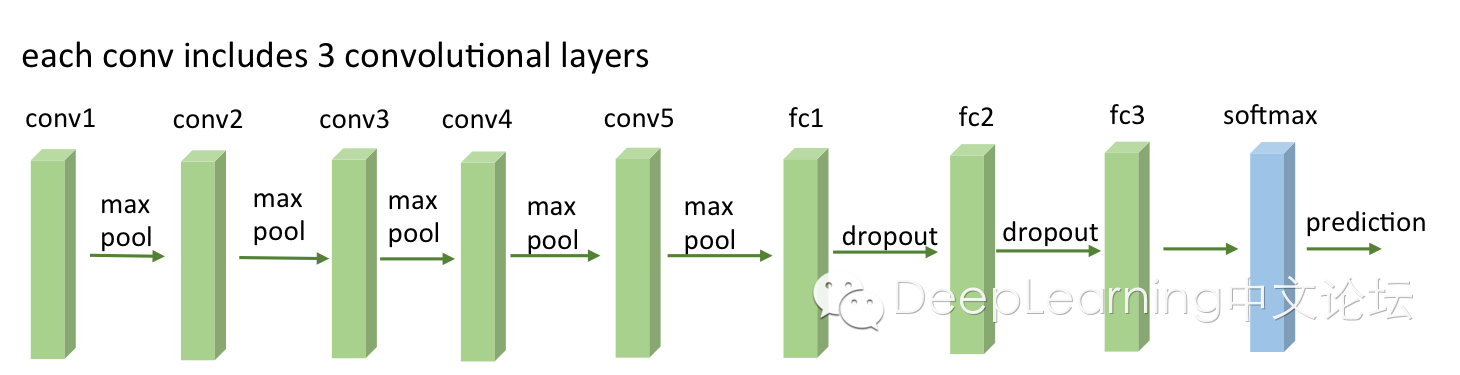
CVPR 2015 之深度学习篇 Part 1 － AlexNet 和 VGG-Net

[摘要：CVPR 2015 之深度进修篇 Part 1 － AlexNet 战 VGG-Net 择要－往年的 CVPR 特别很是的水爆，统共有2800多人介入，比拟客岁增添了700多人，那取deep learning的同军崛起是没有无干系的。CVPR 2015 根基是 th]

CVPR 2015 之深度学习篇 Part 1 － AlexNet 和 VGG-Net

摘 要－今年的 CVPR 非常的火爆，总共有2800多人参与，相比去年增加了700多人，这与deep learning的异军突起是不无关系的。CVPR 2015 基本是 “the year of deep learning”，有大概70%的文章是关于deep learning的。  
  
今年的 CVPR 非常的火爆，总共有2800多人参与，相比去年增加了700多人，这与deep learning的异军突起是不无关系的。CVPR 2015 基本是 “the year of deep learning”，有大概70%的文章是关于deep learning的。大会请来了2位 keynote speakers: 计算机学家 Yann LuCun (NYU， Facebook) 和 心理学家、神经学家 Jack Gallan (Berkeley)。 Yann是公认的 Convolutional Neural Network (CNN) 之父，而Jack是为数不多的神经学领域看好deep learning的科学家，他用实验显示了ConvNets不同层和 visual cortex (视觉皮层) 不同层 它们之间的相关性。  
  
显然，deep learning在工程和科学领域都引起了不小的轰动，不过既然CVPR是一个工程领域会议，我们就先聊聊deep learning在工程领域过去2年的进展。  
  
AlexNet和VGG-Net也成为约定俗成的术语  
  
在 过去的十年里，SVM，SIFT，HOG 已成为计算机视觉学者人人皆知的名词，每个计算机视觉学者直接把它们当作专有名词写进文章，而不需要加以解释，甚至到后来也不用注明出处。今年的 CVPR, 由于过去2年 deep learning 在 computer vision 的很多问题上取得了state-of-the-art 的结果，使得这2种 deep architectures: AlexNet & VGG-Net, 成为一个约定俗成的专有名词，大家开始在文中广泛的使用它们，就如大家以前广泛使用 HOG，SVM一样。  
  
AlexNet和VGG-Net 的起源与network结构  
  
AlexNet 是一种典型的 convolutional neural network，它由5层 convolutional layer，2层 fully connected layer，和最后一层 label layer (1000个node, 每个node代表ImageNet中的一个类别) 组成。2012年，deep learning的大牛教授 Geoffrey Hinton (University of Toronto, Google) 的学生Alex Krizhevsky 设计了一个8层的CNN，并把它用于ImageNet的image classification，直接把当时最好算法的错误率差不多减半。这引起了computer vision community 的强烈关注。这篇文章的出现也是 deep learning 开始被 computer vision community 接受的关键转折点。如是后来大家把这个8层的CNN命名为AlexNet。  
  
  
  
VGG- Net同样也是一种CNN，它来自 Andrew Zisserman 教授的组 (Oxford)，VGG-Net 在2014年的 ILSVRC localization and classification 两个问题上分别取得了第一名和第二名，VGG-Net不同于AlexNet的地方是：VGG-Net使用更多的层，通常有16－19层，而AlexNet 只有8层。另外一个不同的地方是：VGG-Net的所有 convolutional layer 使用同样大小的 convolutional filter，大小为 3 x 3。  
  
  
  
AlexNet和VGG-Net 中的 feature  
  
Deep learning 实际上是一种 feature engineering 的方法。不论是AlexNet还是VGG-Net，倒数第二层都可以很好的描述image全局特征，所以它们通常用来当作 input image 新的 feature，用来做 image classification（最后一层N个node的softmax层，代表了N个类别）。那么, AlexNet 和 VGG-Net 的 intermediate layer（中间层）呢？ 他们是否也可以作为特征来使用？ 答案是肯定的。 这些 intermediate convolutional layer 描述了影像的 local feature (局部特征)，而最后的几层（倒数第二、第三层） fully connected layer，描述了影像的 global feature (全局特征)。如果一个 deep architecture 只有 convolutional layer, 那么 input image 可以任意大小，可是一旦在convolutional layer 上面叠加上 fully connected layer，input image大小就需要固定了。今年CVPR有好多文章使用 convolutionally filtered intermediate layer 作为 local feature，典型的代表有 Jitendra Malik (Berkeley) 组的 Hypercolumns for Object Segmentation and Fine-grained Localization，Xiaoou Tang (CUHK) 组的 Action Recognition With Trajectory-Pooled Deep-Convolutional Descriptors。