

原 Deep Learning (深度学习) 学习笔记整理系列之 (二)

分类: 机器学习 Deep Learning 机器视觉

2013-04-09 00:28 1622人阅读

Deep Learning (深度学习) 学习笔记整理系列

zouxy09@qq.com

<http://blog.csdn.net/zouxy09>

作者: **Zouxy**

version 1.0 2013-04-08

声明:

- 1) 该Deep Learning的学习系列是整理自网上很大牛和机器学习专家所无私奉献的资料的。具体引用的资料请看参考文献。具体的版本声明也参考原文献。
- 2) 本文仅供学术交流, 非商用。所以每一部分具体的参考资料并没有详细对应。如果某部分不小心侵犯了大家的利益, 还望海涵, 并联系博主删除。
- 3) 本人才疏学浅, 整理总结的时候难免出错, 还望各位前辈不吝指正, 谢谢。
- 4) 阅读本文需要机器学习、计算机视觉、神经网络等等基础(如果没有也没关系了, 没有就看看, 能不能看懂, 呵呵)。
- 5) 此属于第一版本, 若有错误, 还需继续修正与增删。还望大家多多指点。大家都共享一点点, 一起为祖国科研的推进添砖加瓦(呵呵, 好高尚的目标啊)。请联系:
zouxy09@qq.com

目录:

一、概述

二、背景

三、人脑视觉机理

四、关于特征

4.1、特征表示的粒度

4.2、初级(浅层)特征表示

4.3、结构性特征表示

4.4、需要有多少个特征？

五、Deep Learning的基本思想

六、浅层学习（Shallow Learning）和深度学习（Deep Learning）

七、Deep learning与Neural Network

八、Deep learning训练过程

8.1、传统神经网络的训练方法

8.2、deep learning训练过程

九、Deep Learning的常用模型或者方法

9.1、AutoEncoder自动编码器

9.2、Sparse Coding稀疏编码

9.3、Restricted Boltzmann Machine(RBM)限制波尔兹曼机

9.4、Deep Belief Networks深信度网络

9.5、Convolutional Neural Networks卷积神经网络

十、总结与展望

十一、参考文献和Deep Learning学习资源

接上

因为我们要学习的是特征的表达，那么关于特征，或者说关于这个层级特征，我们需要了解地更深入点。所以在说Deep Learning之前，我们有必要再啰嗦下特征（呵呵，实际上是看到那么好的对特征的解释，不放在这里有点可惜，所以就塞到这儿了）。

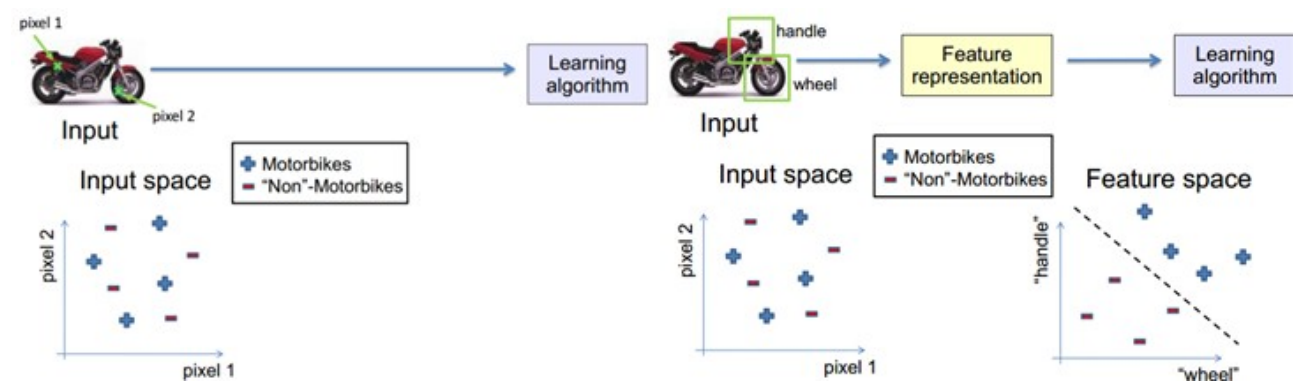
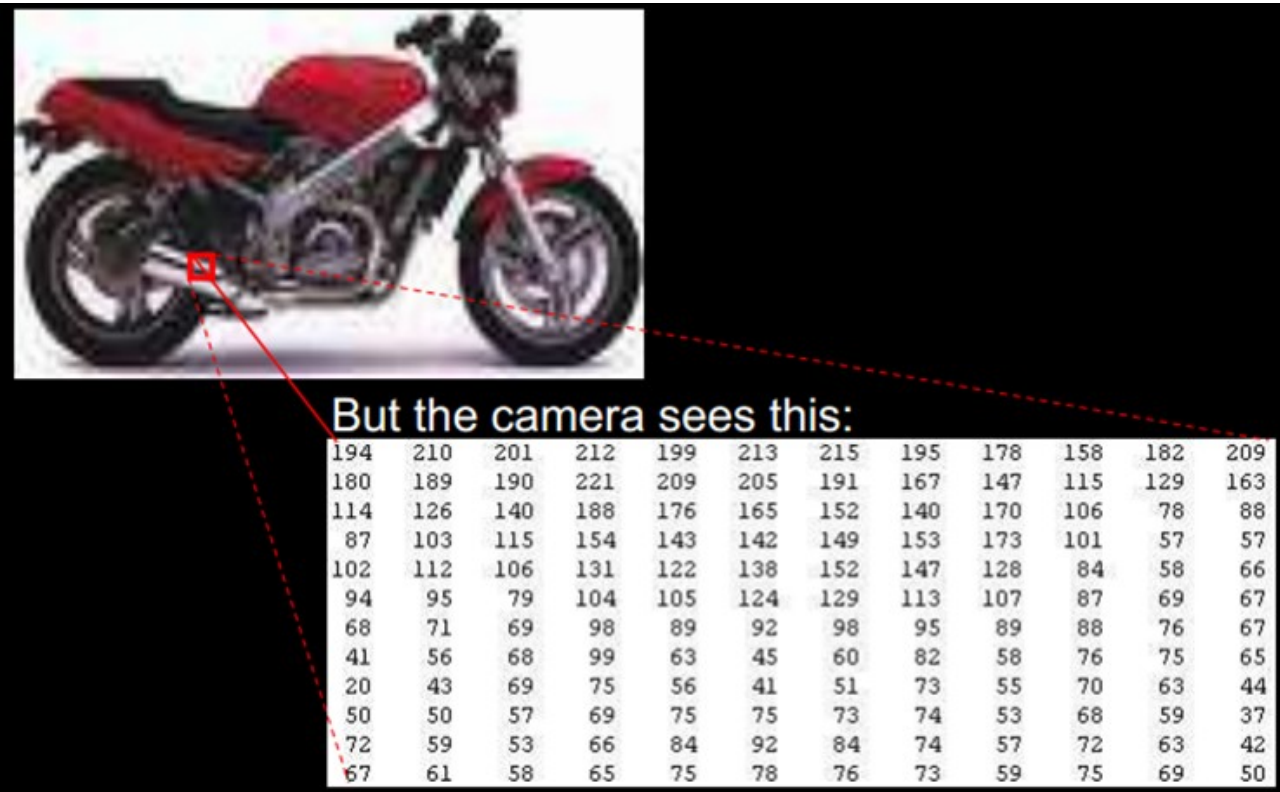
四、关于特征

特征是机器学习系统的原材料，对最终模型的影响是毋庸置疑的。如果数据被很好的表达成了特征，通常线性模型就能达到满意的精度。那对于特征，我们需要考虑

什么呢？

4.1、特征表示的粒度

学习算法在一个什么粒度上的特征表示，才有能发挥作用？就一个图片来说，像素级的特征根本没有价值。例如下面的摩托车，从像素级别，根本得不到任何信息，其无法进行摩托车和非摩托车的区分。而如果特征是一个具有结构性（或者说有含义）的时候，比如是否具有车把手（**handle**），是否具有车轮（**wheel**），就很容易把摩托车和非摩托车区分，学习算法才能发挥作用。



4.2、初级（浅层）特征表示

既然像素级的特征表示方法没有作用，那怎样的表示才有用呢？

1995 年前后，Bruno Olshausen和 David Field 两位学者任职 Cornell University，他们试图同时用生理学和计算机的手段，双管齐下，研究视觉问题。

他们收集了很多黑白风景照片，从这些照片中，提取出400个小碎片，每个照片碎片的尺寸均为 16x16 像素，不妨把这400个碎片标记为 $S[i]$, $i = 0, \dots, 399$ 。接下来，再从这些黑白风景照片中，随机提取另一个碎片，尺寸也是 16x16 像素，不妨把这个碎片标记为 T 。

他们提出的问题是，如何从这400个碎片中，选取一组碎片， $S[k]$ ，通过叠加的办法，合成出一个新的碎片，而这个新的碎片，应当与随机选择的目标碎片 T ，尽可能相似，同时， $S[k]$ 的数量尽可能少。用数学的语言来描述，就是：

$\text{Sum_k} (a[k] * S[k]) \rightarrow T$ ， 其中 $a[k]$ 是在叠加碎片 $S[k]$ 时的权重系数。

为解决这个问题，Bruno Olshausen和 David Field 发明了一个算法，稀疏编码（Sparse Coding）。

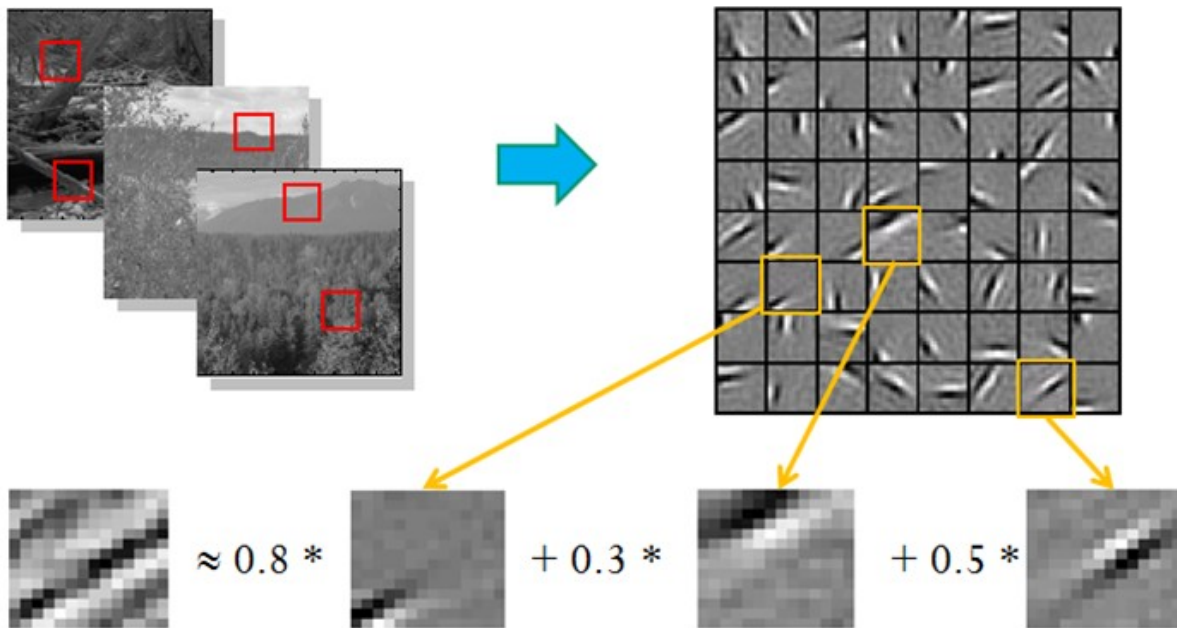
稀疏编码是一个重复迭代的过程，每次迭代分两步：

- 1) 选择一组 $S[k]$ ，然后调整 $a[k]$ ，使得 $\text{Sum_k} (a[k] * S[k])$ 最接近 T 。
- 2) 固定住 $a[k]$ ，在 400 个碎片中，选择其它更合适的碎片 $S'[k]$ ，替代原先的 $S[k]$ ，使得 $\text{Sum_k} (a[k] * S'[k])$ 最接近 T 。

经过几次迭代后，最佳的 $S[k]$ 组合，被遴选出来了。令人惊奇的是，被选中的 $S[k]$ ，基本上都是照片上不同物体的边缘线，这些线段形状相似，区别在于方向。

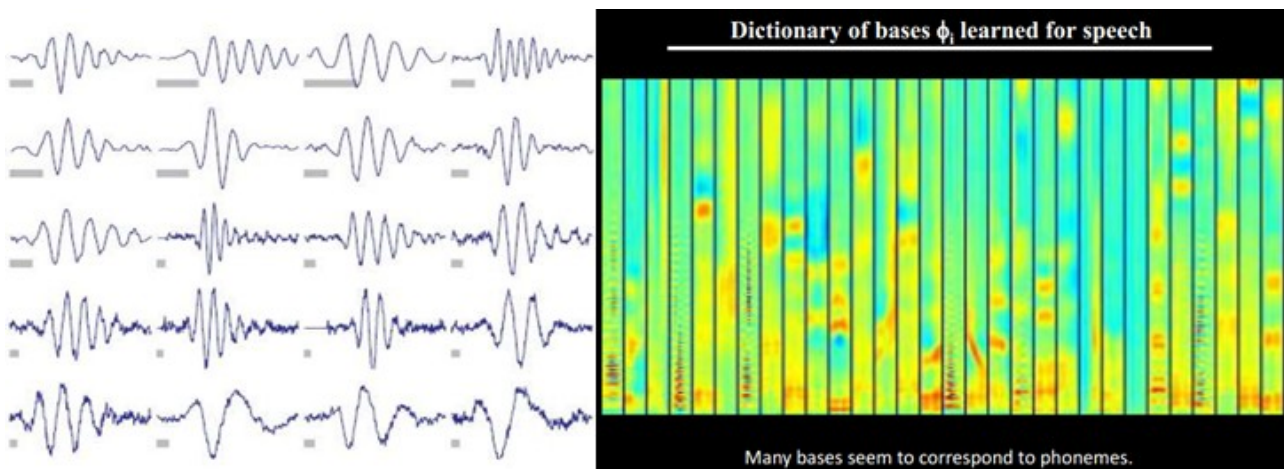
Bruno Olshausen和 David Field 的算法结果，与 David Hubel 和Torsten Wiesel 的生理发现，不谋而合！

也就是说，复杂图形，往往由一些基本结构组成。比如下图：一个图可以通过用 64种正交的edges（可以理解成正交的基本结构）来线性表示。比如样例的x可以用1-64个edges中的三个按照0.8,0.3,0.5的权重调和而成。而其他基本edge没有贡献，因此均为0。

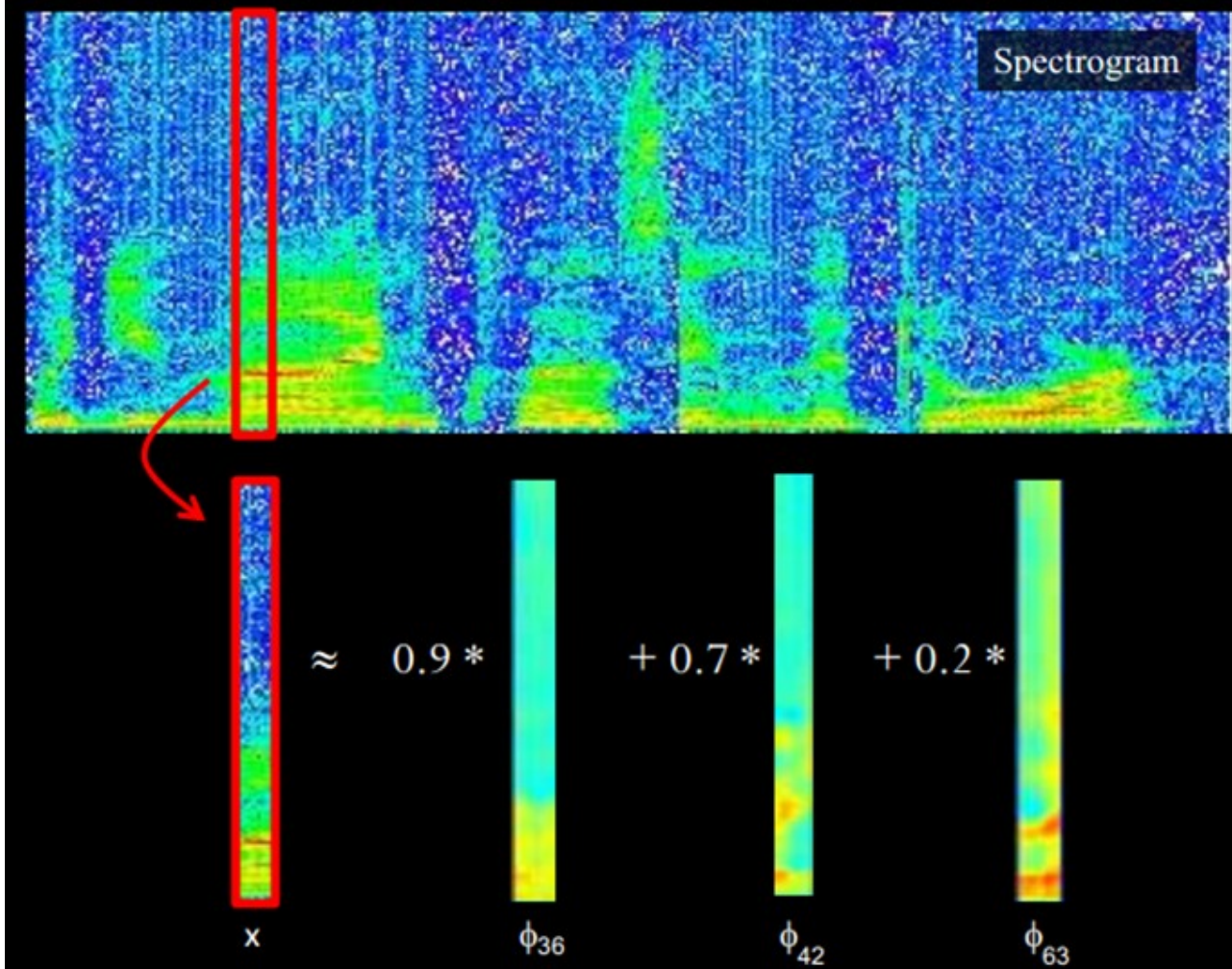


$[a_1, \dots, a_{64}] = [0, 0, \dots, 0, \mathbf{0.8}, 0, \dots, 0, \mathbf{0.3}, 0, \dots, 0, \mathbf{0.5}, 0]$
 (feature representation)

另外，大牛们还发现，不仅图像存在这个规律，声音也存在。他们从未标注的声音中发现了**20种**基本的声音结构，其余的声音可以由这**20种**基本结构合成。

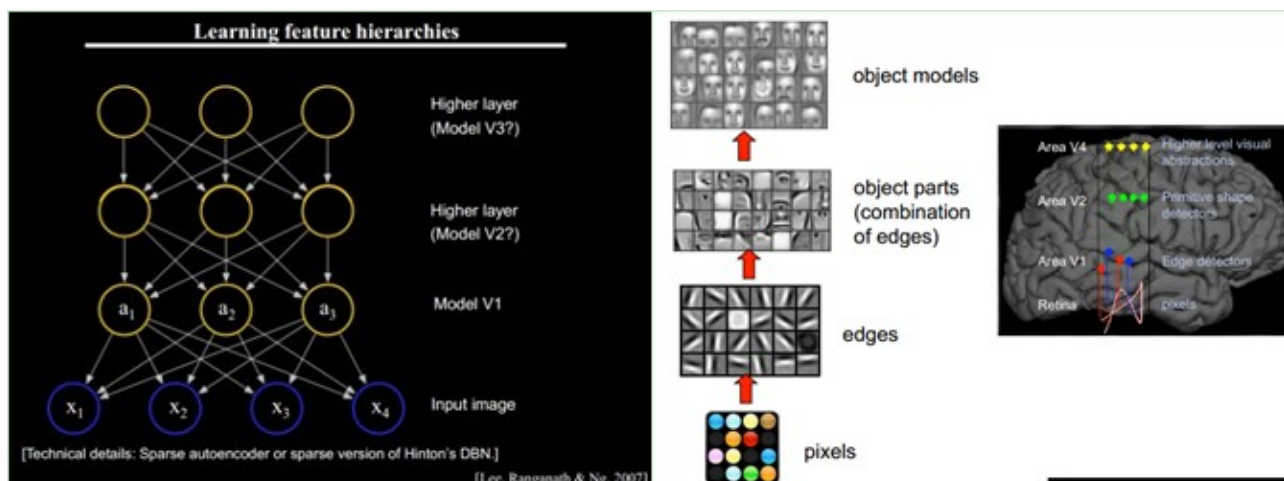


Sparse coding on audio (speech)



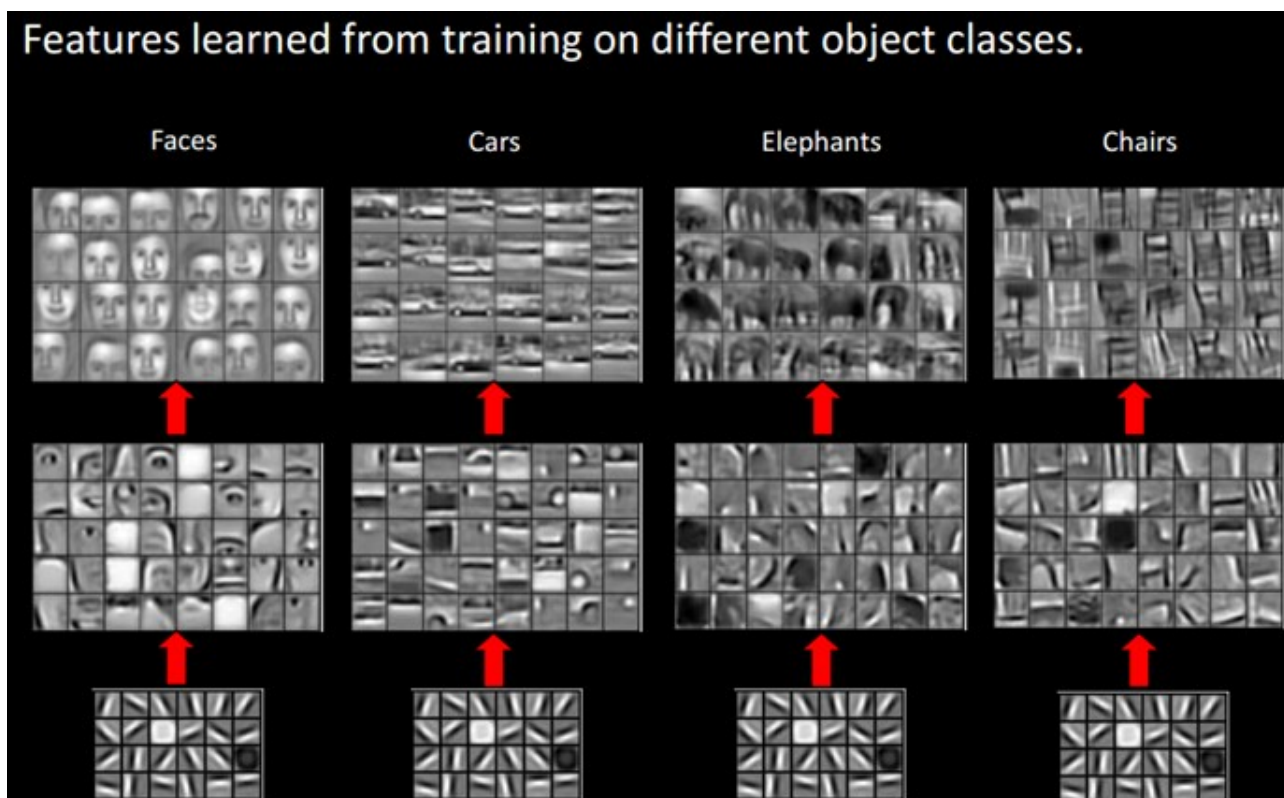
4.3、结构性特征表示

小块的图形可以由基本edge构成，更结构化，更复杂的，具有概念性的图形如何表示呢？这就需要更高层次的特征表示，比如V2，V4。因此V1看像素级是像素级。V2看V1是像素级，这个是层次递进的，高层表达由底层表达的组合而成。专业点说就是基basis。V1取提出的basis是边缘，然后V2层是V1层这些basis的组合，这时候V2区得到的又是高一层的basis。即上一层的basis组合的结果，上上层又是上一层的组合basis.....（所以有大牛说Deep learning就是“搞基”，因为难听，所以美其名曰Deep learning或者Unsupervised Feature Learning）



直观上说，就是找到make sense的小patch再将其进行combine，就得到了上一层的feature，递归地向上learning feature。

在不同object上做training是，所得的edge basis 是非常相似的，但object parts 和models 就会completely different了（那咱们分辨car或者face是不是容易多了）：



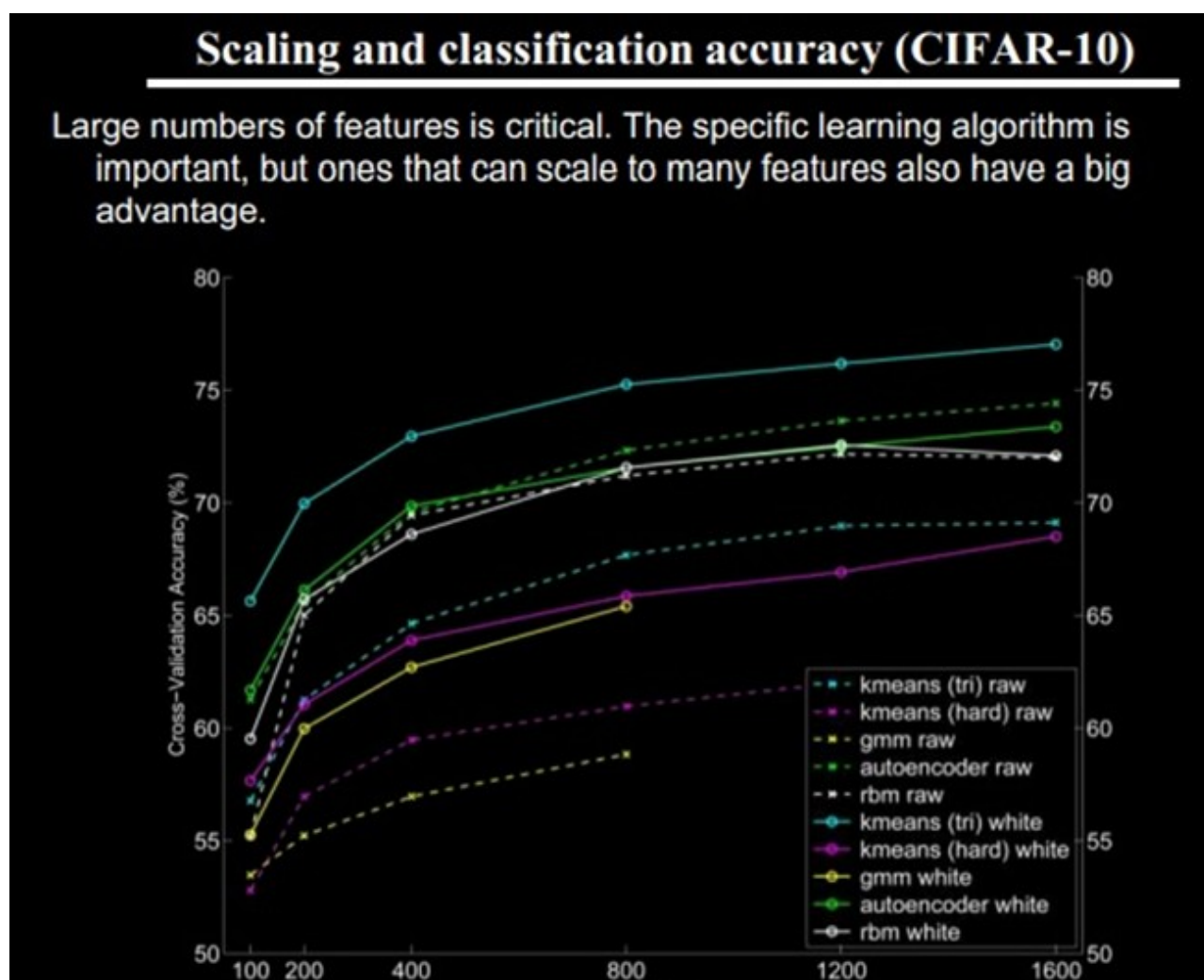
从文本来说，一个doc表示什么意思？我们描述一件事情，用什么来表示比较合适？用一个一个字嘛，我看不是，字就是像素级别了，起码应该是term，换句话说每个doc都由term构成，但这样表示概念的能力就够了嘛，可能也不够，需要再上一步，达到topic级，有了topic，再到doc就合理。但每个层次的数量差距很大，比如doc表示的概念->topic（千-万量级）->term（10万量级）->word（百万量级）。

一个人在看一个doc的时候，眼睛看到的是word，由这些word在大脑里自动切词形成term，在按照概念组织的方式，先验的学习，得到topic，然后再进行高层次的learning。

4.4、需要有多少个特征？

我们知道需要层次的特征构建，由浅入深，但每一层该有多少个特征呢？

任何一种方法，特征越多，给出的参考信息就越多，准确性会得到提升。但特征多意味着计算复杂，探索的空间大，可以用来训练的数据在每个特征上就会稀疏，都会带来各种问题，并不一定特征越多越好。



好了，到了这一步，终于可以聊到Deep learning了。上面我们聊到为什么会有Deep learning（让机器自动学习良好的特征，而免去人工选取过程。还有参考人的分层视觉处理系统），我们得到一个结论就是Deep learning需要多层来获得更抽象的特征表达。那么多少层才合适呢？用什么架构来建模呢？怎么进行非监督训练呢？