**前深度学习时代的计算机视觉**

在深度学习算法出来之前，对于视觉算法来说，大致可以分为以下5个步骤：**特征感知，图像预处理，特征提取，特征筛选，推理预测与识别**。早期的机器学习中，占优势的统计机器学习群体中，对特征是不大关心的。

传统的计算机识别方法把特征提取和分类器设计分开来做，然后在应用时再合在一起，比如如果输入是一个摩托车图像的话，首先要有一个特征表达或者特征提取的过程，然后把表达出来的特征放到学习算法中进行分类的学习。

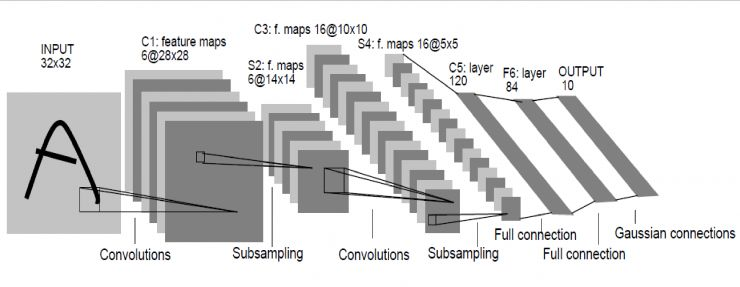
过去20年中出现了不少优秀的特征算子，比如最著名的SIFT算子，即所谓的对尺度旋转保持不变的算子。它被广泛地应用在图像比对，特别是所谓的structure from motion这些应用中，有一些成功的应用例子。另一个是HoG算子，它可以提取物体，比较鲁棒的物体边缘，在物体检测中扮演着重要的角色。

1981年诺贝尔医学生理学奖颁发给了David Hubel，一位神经生物学家。他的主要研究成果是发现了视觉系统信息处理机制，证明大脑的可视皮层是分级的。他的贡献主要有两个，一是他认为**人的视觉功能一个是抽象，一个是迭代**。抽象就是把非常具体的形象的元素，即原始的光线像素等信息，抽象出来形成有意义的概念。这些有意义的概念又会往上迭代，变成更加抽象，人可以感知到的抽象概念。

### LeNet网络

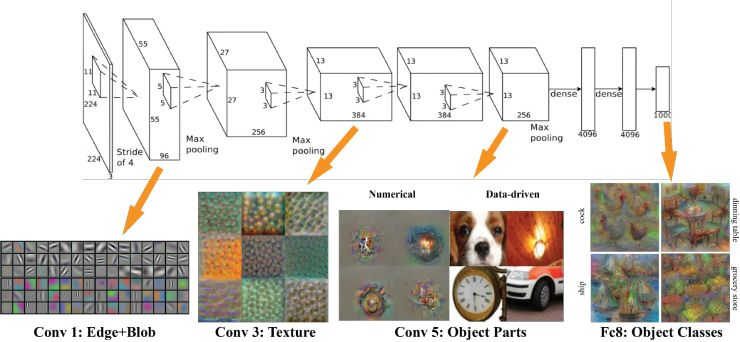
Le顾名思义就是指人工智能领域的大牛Lecun。这个网络是深度学习网络的最初原型，因为之前的网络都比较浅，它较深的。LeNet在98年就发明出来了，当时Lecun在AT&T的实验室，他用这一网络进行字母识别，达到了非常好的效果。

怎么构成呢？输入图像是32×32的灰度图，第一层经过了一组卷积和，生成了6个28X28的feature map，然后经过一个池化层，得到得到6个14X14的feature map，然后再经过一个卷积层，生成了16个10X10的卷积层，再经过池化层生成16个5×5的feature map。



### 压在骆驼身上的最后一根稻草AlexNet

AlexNet由多伦多大学几个科学家开发，在ImageNet比赛上做到了非常好的效果。当时AlexNet识别效果超过了所有浅层的方法。此后，大家认识到深度学习的时代终于来了，并有人用它做其它的应用，同时也有些人开始开发新的网络结构。



图像识别（classification）：  
输入：图片  
输出：物体的类别  
评估方法：准确率

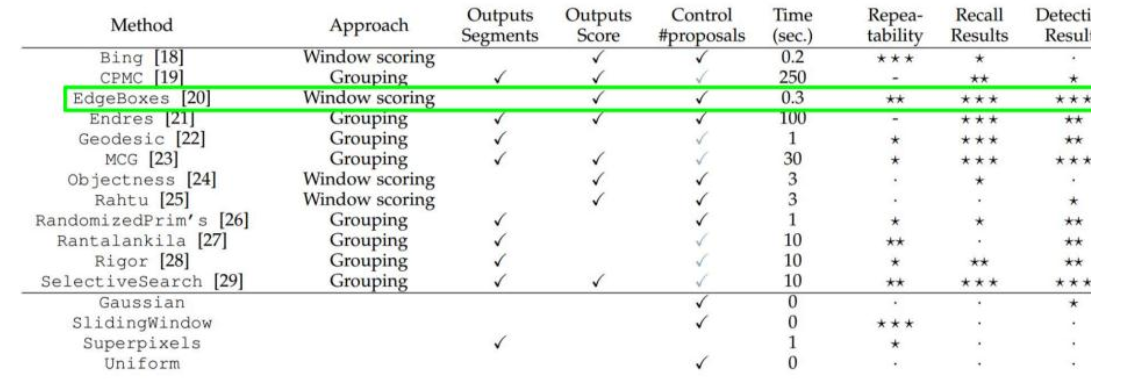
定位（localization）：  
输入：图片  
输出：方框在图片中的位置（x,y,w,h）  
评估方法：检测评价函数 intersection-over-union ( IOU )

卷积神经网络CNN已经帮我们完成了图像识别（判定是猫还是狗）的任务了，我们只需要添加一些额外的功能来完成定位任务即可。

定位的问题的解决思路有哪些？  
**思路一：看做回归问题**  
看做回归问题，我们需要预测出（x,y,w,h）四个参数的值，从而得出方框的位置。

**思路二：取图像窗口**

大牛们发明好多选定候选框的方法，比如EdgeBoxes和Selective Search。  
以下是各种选定候选框的方法的性能对比。



Yolo

网络的输出结果为一个张量，维度为：S∗S∗(B∗5+C)

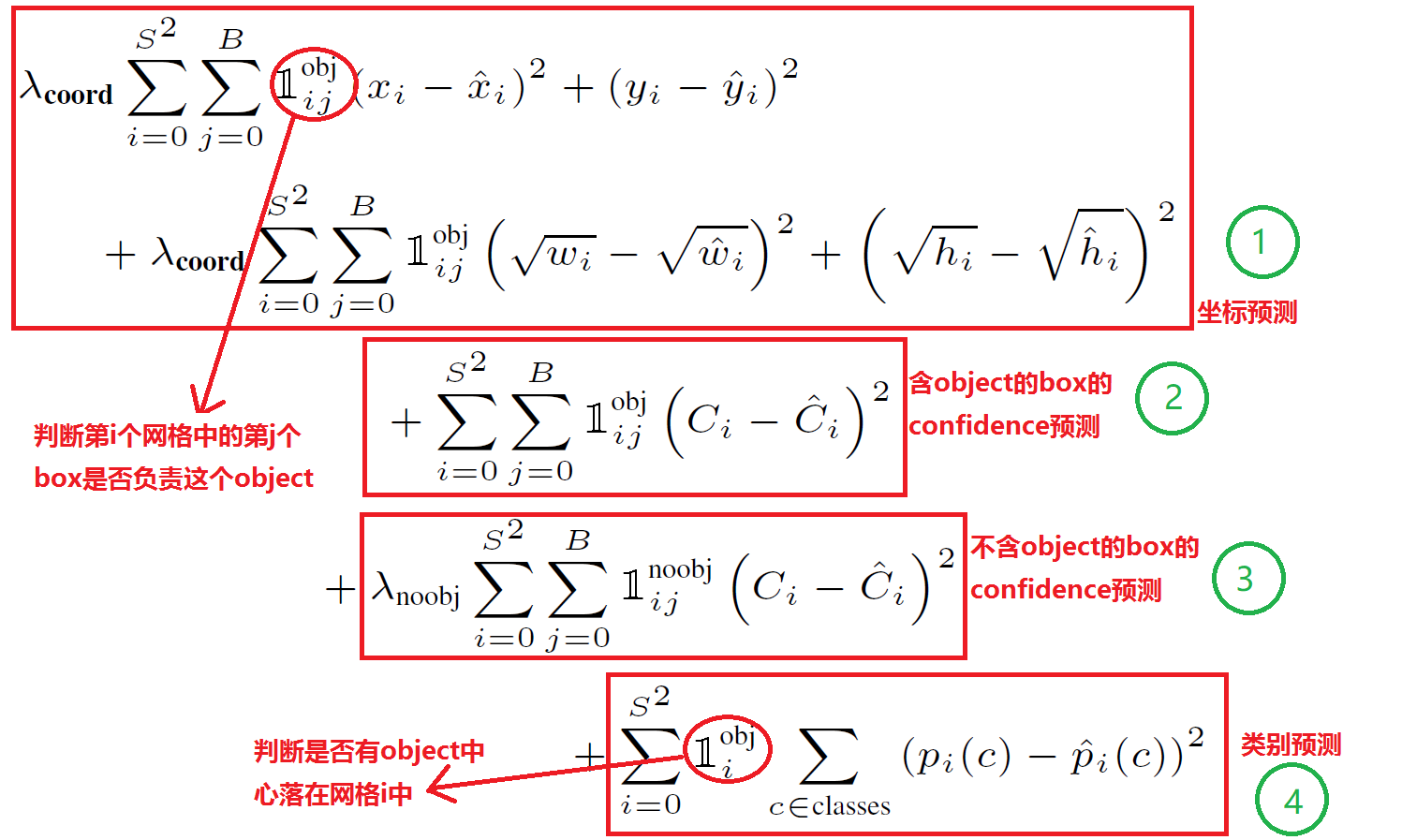
其中，S为划分网格数，B为每个网格负责目标个数，C为类别个数。该表达式含义为：

(1) 每个小格会对应B个边界框，边界框的宽高范围为全图，表示以该小格为中心寻找物体的边界框位置。

(2) 每个边界框对应一个分值，代表该处是否有物体及定位准确度

(3) 每个小格会对应C个概率值，找出最大概率对应的类别P(Class|object)P(Class|object)，并认为小格中包含该物体或者该物体的一部分

因为模型输出结果的改变，损失函数不能使用简单的分类模型的交叉熵，这里使用Yolo损失函数[3]，Loss = λcoord \* 坐标预测误差(1) + 含object的box confidence预测误差 (2)+ λnoobj\* 不含object的box confidence预测误差(3) + 每个格子中类别预测误差(4)



(1) 整个损失函数针对边界框损失(图中1, 2, 3部分)与格子(4部分)主体进行讨论。

(2) 部分1为边界框位置与大小的损失函数，式中对宽高都进行开根是为了使用大小差别比较大的边界框差别减小。例如，一个同样将一个100x100的目标与一个10x10的目标都预测大了10个像素，预测框为110 x 110与20 x 20。显然第一种情况我们还可以失道接受，但第二种情况相当于把边界框预测大了一倍，但如果不使用根号函数，那么损失相同，都为200。

(3) 若有物体落入边界框中，则计算预测边界框含有物体的置信度CiCi和真实物体与边界框IoUCi^Ci^的损失，我们希望两差值越小损失越低。

(4) 若没有任何物体中心落入边界框中，则Ci^Ci^为0，此时我们希望预测含有物体的置信度CiCi越小越好。然而，大部分边界框都没有物体，积少成多，造成loss的第3部分与第4部分的不平衡，因此，作才在loss的三部分增加权重λnobj=0.5λnobj=0.5 。

(5) 对于每个格子而言，作者设计只能包含同种物体。若格子中包含物体，我们希望希望预测正确的类别的概率越接近于1越好，而错误类别的概率越接近于0越好。loss第4部分中，若pi(c)^pi(c)^中c为正确类别，则值为1，若非正确类别，则值为0。