




遼寧科技大學
University of Science and Technology Liaoning

深度学习技术及其在模式识别中应用



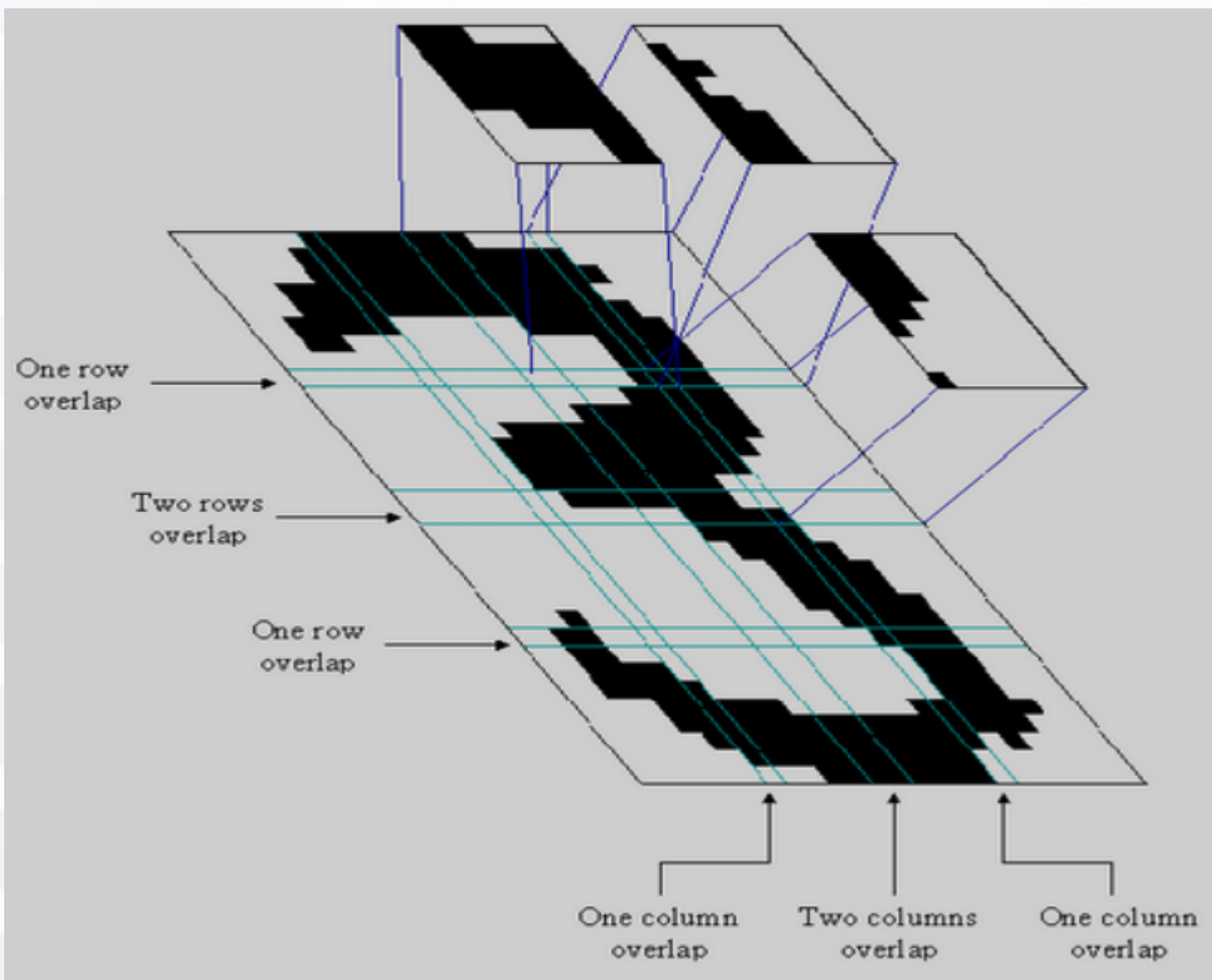
报告人基本信息

- 姓名：田莹
 - 专业技术职务：教授
 - 最高学历（学位）：博士
 - 毕业院校：沈阳工业大学
 - 研究领域：模式识别，生物特征识别，数字图像处理，机器学习等。
- 

模式识别——智能程序的诞生

- 人工智能是人类一个非常美好的梦想，跟星际漫游和长生不老一样。我们想制造出一种机器，使得它跟人一样具有一定的对外界事物感知能力，比如**看见世界**。
- 当我们人眼看到一幅画时，我们能够很清晰的知道其中哪里是动物，哪里是山，水，人等等。人眼又是如何识别和分辨的呢，其实人类也是在**先验知识**和对以往多个此类事物的具体实例进行观察的基础上得到的对此类事物整体性质和特点的认识的。
- 模式识别：是一种从大量信息和数据出发，在**专家经验**和**已有认识**的基础上，利用计算机和数学推理的方法对形状、模式、曲线、数字、字符格式和图形自动完成识别的过程，是信息科学和**人工智能**的重要组成部分。

- 广义的模式识别属计算机科学中智能模拟的研究范畴，内容非常广泛，包括声音和语言识别、文字识别、指纹识别、声纳信号和地震信号分析、照片图片分析、化学模式识别等等。
- 模式识别强调的是如何让一个计算机程序去做一些看起来很“智能”的事情，例如识别“3”这个数字。而且在融入了很多的智慧和直觉后，人们也的确构建了这样的一个程序。



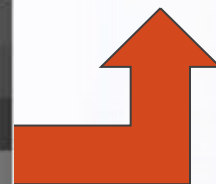
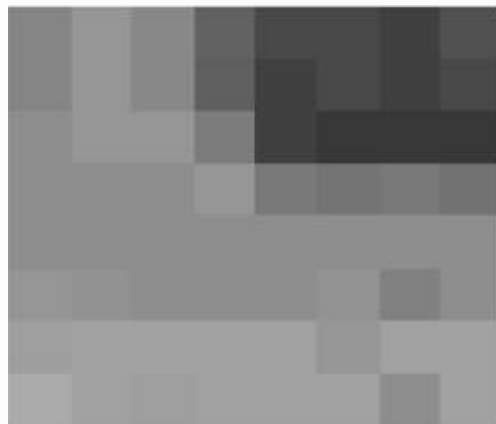
机器是否具有真正的智能？

□ 人眼中图像



□ 计算机看到的图像

130	146	133	95	71	71	62	78
130	146	133	92	62	71	62	71
139	146	146	120	62	55	55	55
139	139	139	146	117	112	117	110
139	139	139	139	139	139	139	139
146	142	139	139	139	143	125	139
156	159	159	159	159	146	159	159
168	159	156	159	159	159	139	159



在上世纪50年代，数学家图灵提出判断机器是否具有人工智能的标准：图灵测试



从图灵测试提出来到本世纪初，50多年时间有无数科学家提出很多机器学习的算法，试图让计算机具有与人一样的智力水平

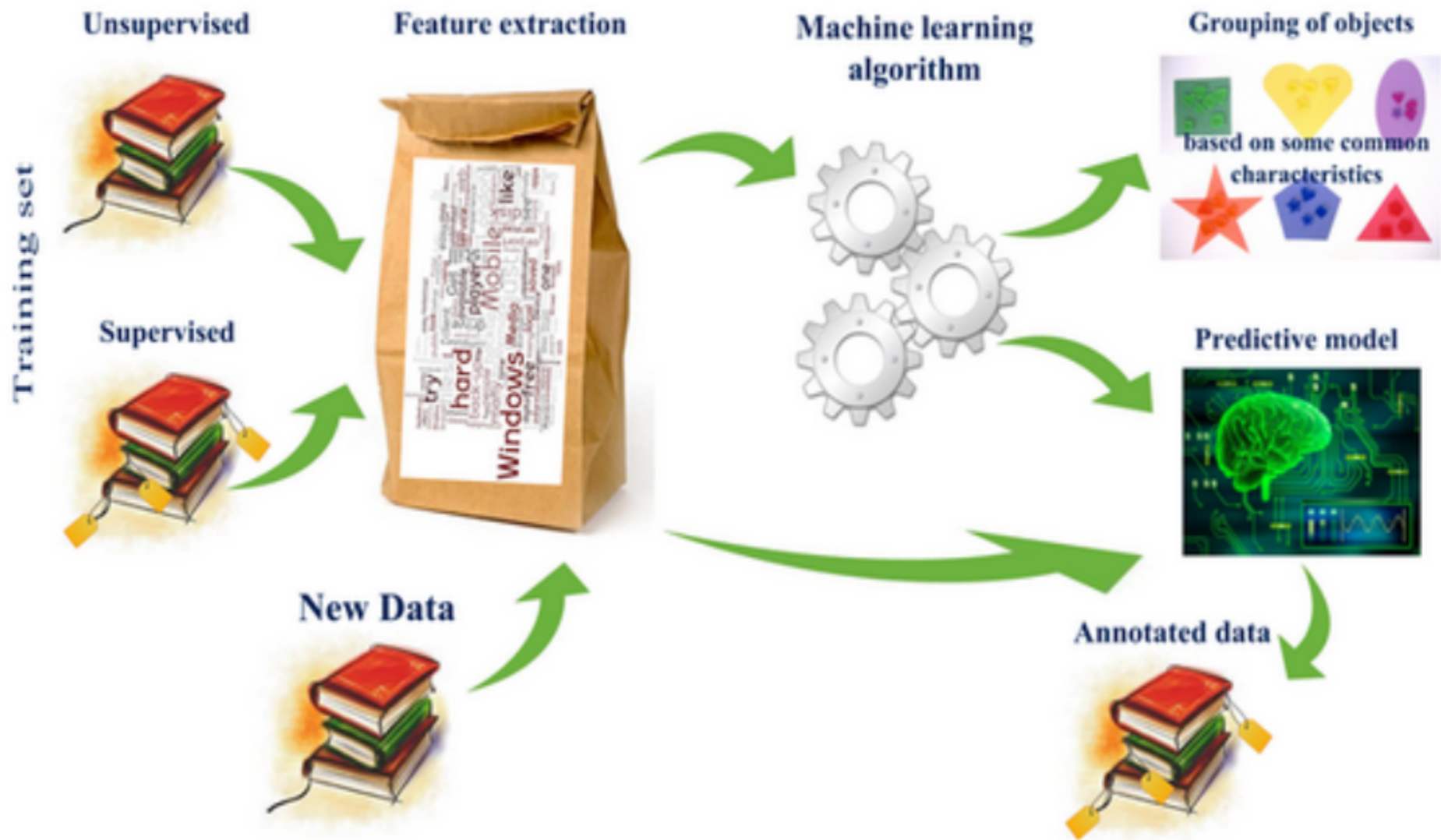



机器学习——从样本中学习的智能程序

- 在90年代初，人们开始意识到一种可以更有效地构建模式识别算法的方法，那就是用数据（可以通过廉价劳动力采集获得）去替换专家（具有很多图像方面知识的人）。
- 因此，我们搜集大量的人脸和非人脸图像，再选择一个算法，然后冲着咖啡、晒着太阳，等着计算机完成对这些图像的学习。这就是机器学习的思想。
- “机器学习”强调的是，在给计算机程序（或者机器）输入一些数据后，它必须做一些事情，那就是**学习**这些数据，而这个学习的步骤是明确的。就算计算机完成学习要耗上一天的时间，也会比手工为这个任务设计一些分类规则要好。

- ❑ 机器学习（Machine Learning）是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。
- ❑ 机器学习专门研究计算机怎么模拟或实现人类的学习行为，其目的在于使得机器可以根据数据进行自动学习，通过算法使得机器能从大量历史数据中学习规律从而对新的样本做出决策。
- ❑ 包括相互关联的两个阶段，即学习阶段和实现阶段，
学习阶段是对样本进行特征选择，寻找分类的规律；
实现阶段是根据分类规律对未知样本集进行分类和识别。
- ❑ 机器学习是人工智能的一个分支，它也是人工智能的核心，是使计算机具有智能的根本途径，其应用遍及人工智能的各个领域。

Machine learning workflow



- 在21世纪中期，机器学习成为了计算机科学领域一个重要的研究课题，计算机科学家们开始将这些想法应用到更大范围的问题上，不再限于识别字符、识别猫和狗或者识别图像中的某个目标等等这些问题。
 - 研究人员开始将机器学习应用到机器人、基因数据的分析和金融市场的预测中。每一个机器人专家都“无奈地”成为了机器学习专家，同时，机器学习也迅速成为了众人渴望的必备技能之一
 - 然而，“机器学习”这个概念对底层算法只字未提。我们已经看到凸优化、核方法、支持向量机和Boosting算法等都有各自辉煌的时期。再加上一些人工设计的特征，那在机器学习领域，我们就有了很多的方法，很多不同的思想流派。
- 

机器学习五大流派

- 符号主义：（逻辑学） 使用符号、规则和逻辑来表征知识和进行逻辑推理，最喜欢的算法是：**规则和决策树**
- 联结主义：（反向传播） 使用概率矩阵和加权神经元来动态地识别和归纳模式，最喜欢的算法是：**神经网络**
- 进化主义：（进化生物学） 生成变化，然后为特定目标获取其中最优的，最喜欢的算法是：**遗传算法**
- 贝叶斯主义：（统计学） 获取发生的可能性来进行概率推理，最喜欢的算法是：**朴素贝叶斯或马尔可夫**
- 行为类推主义：（心理学） 根据约束条件来优化函数（尽可能走到更高，但同时不要离开道路），最喜欢的算法是：**支持向量机**

然而，对于一个新人来说，对特征和算法的选择依然一头雾水，没有清晰的指导原则。

但，值得庆幸的是，这一切即将改变……



深度学习----一统江湖的架构

● 如果不手动设计特征，不挑选分类器，有没有别的方案呢？能不能同时学习特征和分类器？即输入某一个模型的时候，输入只是图片，输出就是它自己的标签。

● 深度学习（Deep Learning）理论提出了一种让计算机自动学习出模式特征的方法，并将特征学习融入到了建立模型的过程中，不需要人参与特征的选取过程，不需要挑选分类器，同时学习特征和分类器。

● 深度学习是近年来发展起来的一种神经网络模型，神经网络曾经在八九十年代非常流行，它试图通过模拟大脑认知的激励，解决各种机器学习的问题。但是它具有学习不充分、深度不足等特点，后来由于种种原因大多数学者在相当长一段时间放弃了神经网络。深度学习则克服了这些缺点可用于模式分类、目标跟踪等领域。

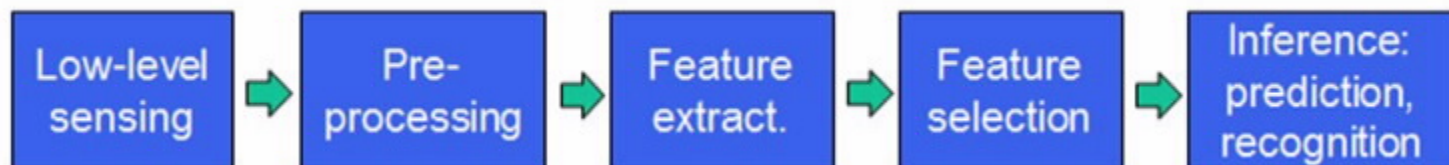
- 深度学习在很多学术领域，比非深度学习算法往往有20-30%成绩的提高。很多大公司也逐渐开始出手投资这种算法，并成立自己的深度学习团队，其中投入最大的就是谷歌，2008年6月披露了谷歌脑项目。2014年1月谷歌收购DeepMind，然后2016年3月其开发的Alphago算法在围棋挑战赛中，战胜了韩国九段棋手李世石，证明深度学习设计出的算法可以战胜这个世界上最强的选手。
- 在硬件方面，Nvidia最开始做显示芯片，但从2006及2007年开始主推用GPU芯片进行通用计算，它特别适合深度学习中大量简单重复的计算量。目前很多人选择Nvidia的CUDA工具包进行深度学习软件的开发。

传统机器学习算法为什么没有达到深度学习的精度？

传统机器学习去解决图像识别问题思路是这样的：

通过传感器来获得数据，然后经过**预处理**、**特征提取**、**特征选择**，再到**识别分类**、**推理或者预测**。而中间的三部分，概括起来就是**特征表达**，对最终算法的准确性起了非常关键的作用。在采用这些机器学习方法的时候，不得不自己设计前面3个部分。但对任何人来说这都是一个比较难的任务。

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对比与识别中，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。

这些算子还包括Textons, Spin image, RIFT和GLOH，都是在深度学习诞生之前或者深度学习真正的流行起来之前，占领视觉算法的主流。

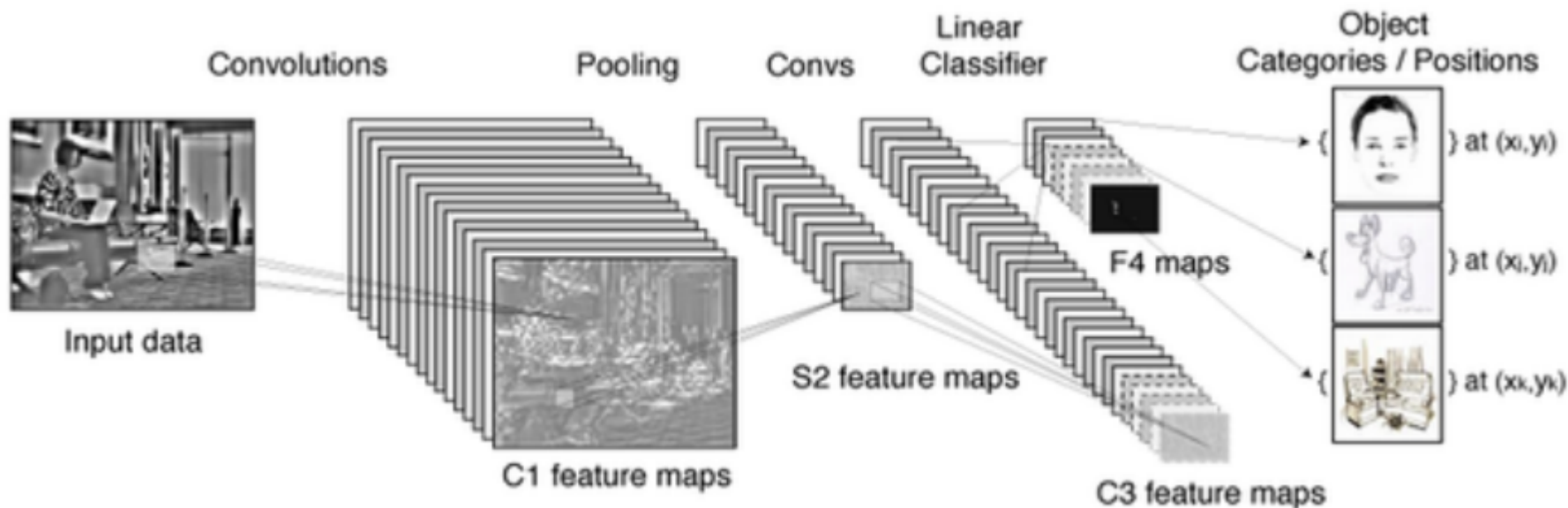
但这种成功例子太少了，因为手工设计特征需要大量的经验，需要你对这个领域和数据特别了解，设计出来特征还需要大量的调试工作。

另一个难点在于，你不只需要手工设计特征，还要在此基础上有一个比较合适的分类器算法。同时设计特征然后选择一个分类器，这两者合并达到最优的效果，几乎是不可能完成的任务。

- ◆对于图像的处理是深度学习算法最早尝试应用的领域。
- ◆深度学习在机器视觉领域最具影响力的突破是在2012年10月，Hinton的研究小组采用深度学习赢得了ImageNet图像分类的比赛。ImageNet是当今计算机视觉领域最具影响力的比赛。它的训练和测试样本都来自互联网图片，训练样本超过百万，任务是将测试样本分成1000类。在2012年的测试中，排名2到4位的小组采用传统模式识别方法，他们的准确率的差别不超过1%，而作为首次参加的Hinton研究小组采用的是深度学习方法，且准确率超出第二名10%以上，这个结果在计算机视觉领域产生极大的震动，也使得对于图像识别的领域研究更进一步。

在深度学习的模型中，受宠爱最多的就是被用在大规模图像识别任务中的卷积神经网络（Convolutional Neural Nets, CNN）

早在1989年，加拿大多伦多大学教授Yan LeCun就提出了卷积神经网络——一种包含卷积层的深度神经网络模型。2006年Hinton等提出了深度置信网络和相应的高效学习算法，成为了其后至今深度学习算法的主要框架。



一个深度学习框架结构

卷积神经网络Convolutional Neural Nets ----CNN

卷积神经网络，(CNN)，是一种对人脑比较精准的模拟。

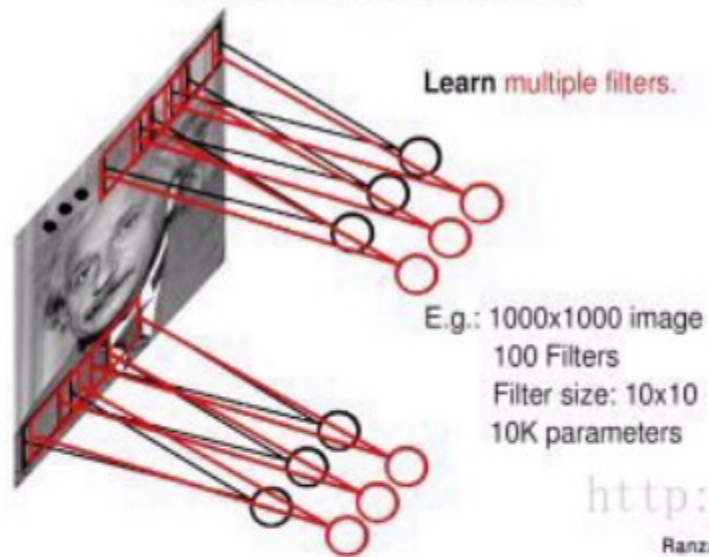
卷积神经网络的主要操作是：**卷积和池化**

卷积就是两个函数之间的相互关系，然后得出一个新的值，他是在连续空间做积分计算，然后在离散空间内求和的过程。实际上在计算机视觉里面，可以把卷积当做一个抽象的过程，就是把小区域内的信息统计抽象出来。

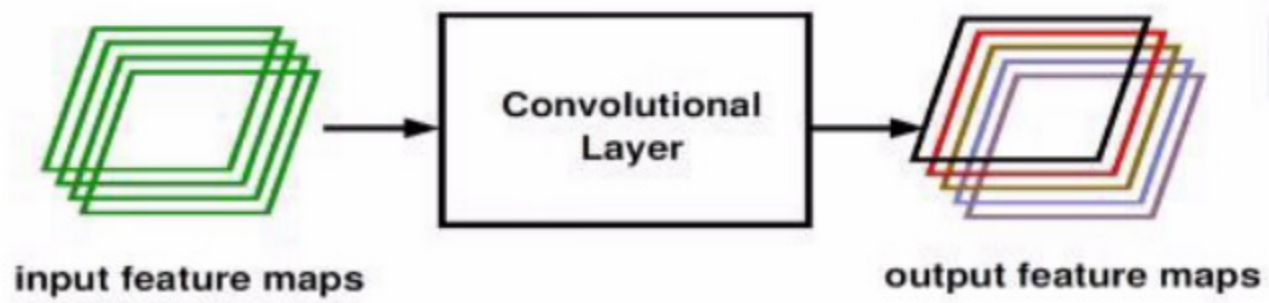
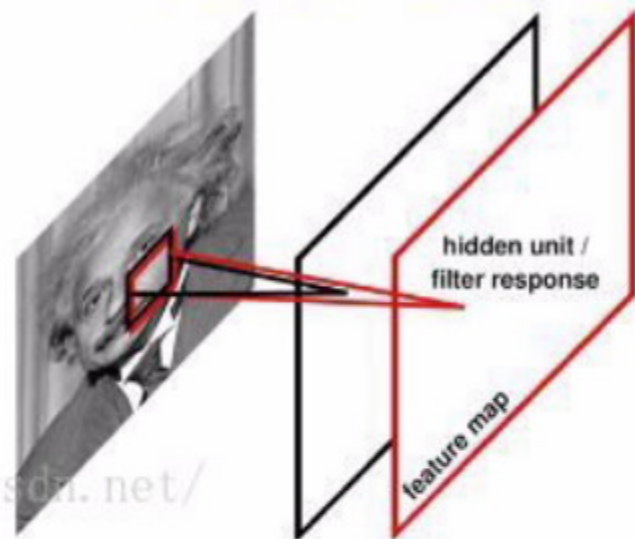
比如，对于一张爱因斯坦的照片，我可以学习n个不同的卷积和函数，然后对这个区域进行统计。可以用不同的方法统计，比如着重统计中央，也可以着重统计周围，这就导致统计的和函数的种类多种多样，为了达到可以同时学习多个统计的累积和。

池化操作在统计上的概念更明确，就是一个对一个小区内求平均值或者求最大值的统计操作。

CONVOLUTIONAL NET



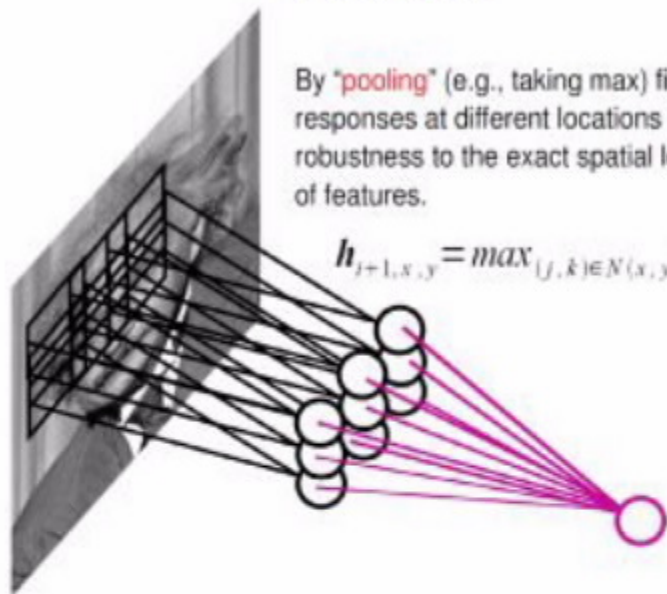
CONVOLUTIONAL NET



POOLING

By "pooling" (e.g., taking max) filter responses at different locations we gain robustness to the exact spatial location of features.

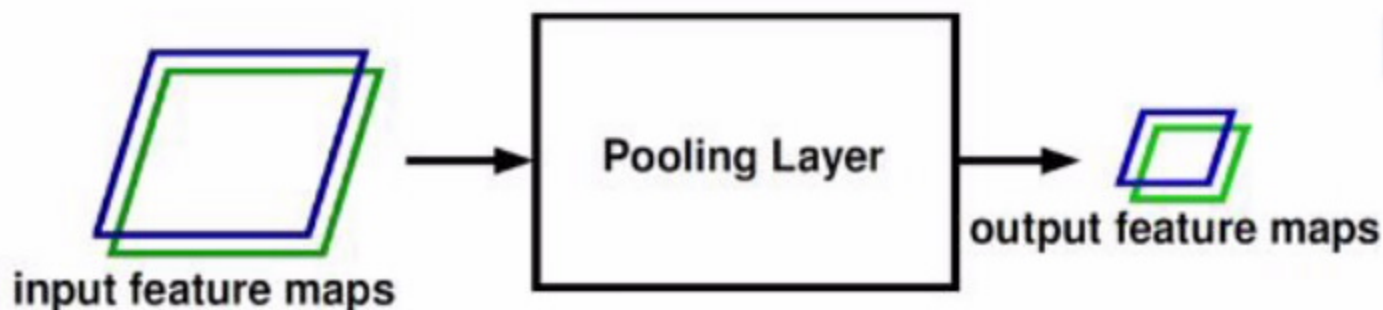
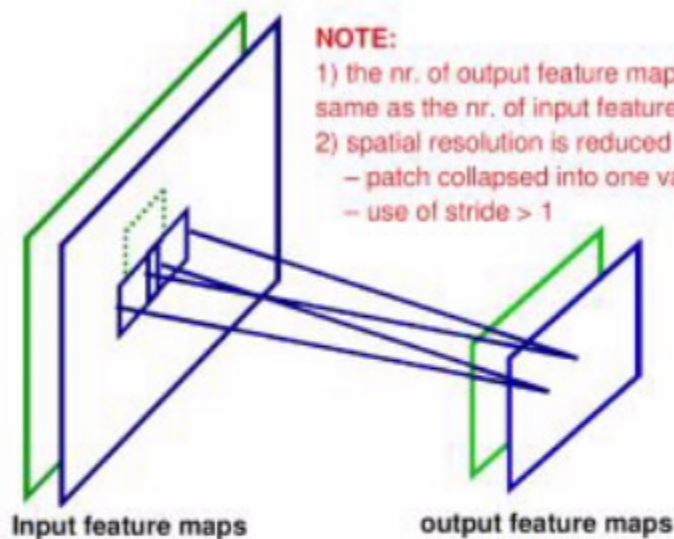
$$h_{l+1,x,y} = \max_{(j,k) \in N(x,y)} h_{l,j,k}$$




POOLING LAYER

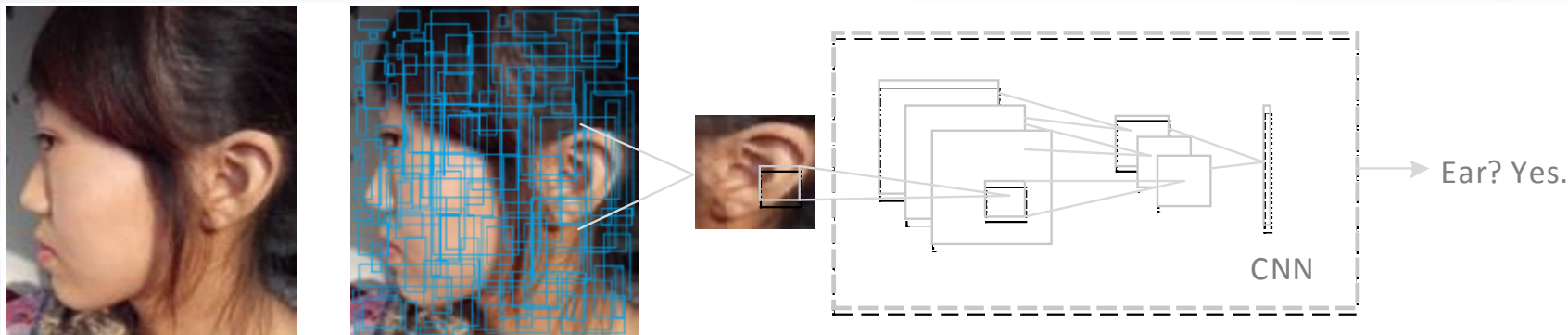
NOTE:

- 1) the nr. of output feature maps is the same as the nr. of input feature maps
- 2) spatial resolution is reduced
 - patch collapsed into one value
 - use of stride > 1



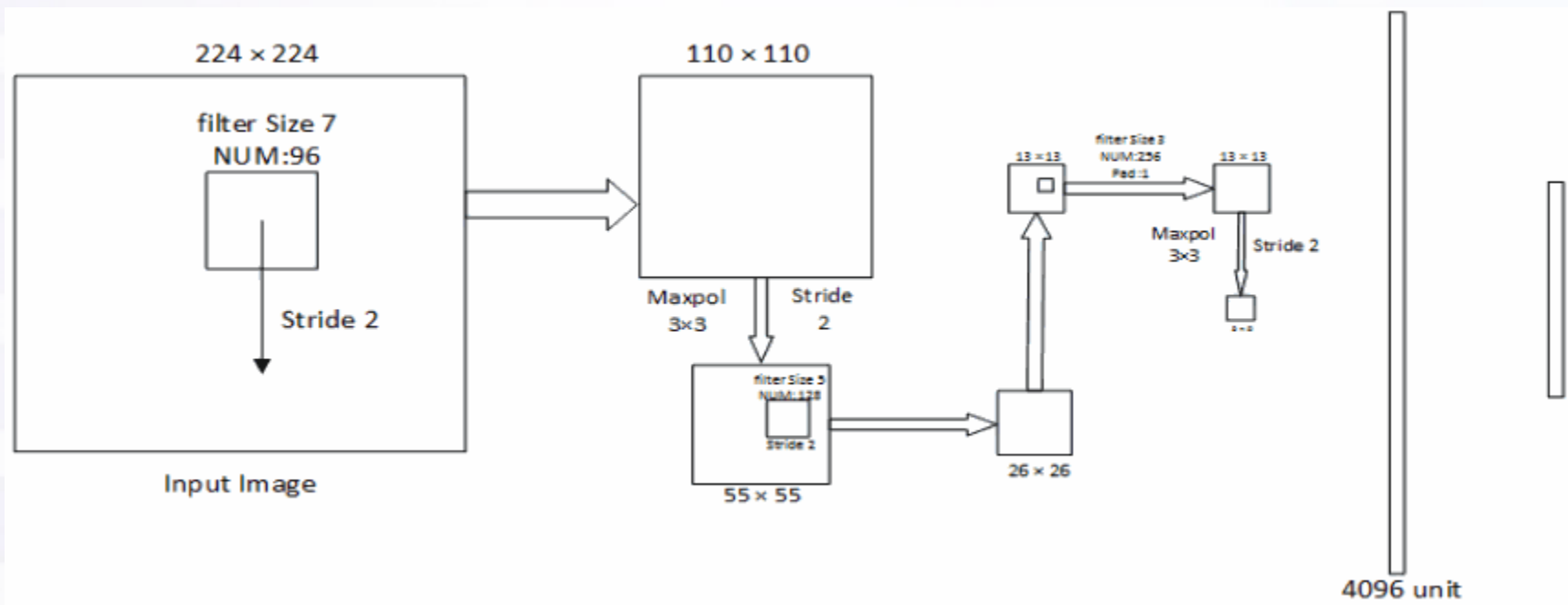
- 目前以深度学习为核心的某些机器学习应用，在满足特定条件的应用场景下，已经达到了超越现有算法的识别或分类性能。
 - 本人从事模式识别与数字图像处理方面的研究近十年，主持辽宁省教育厅项目2项，鞍山市科技项目各1项，作为主要成员参与完成国家自然科学基金项目、辽宁省自然科学基金项目和辽宁省教育厅项目各多项，研究过模式识别领域相关的问题包括人耳识别、人脸识别、字符识别、车牌识别等等。目前已在国内核心期刊外文期刊及国内国际会议上发表论文二十余篇。
 - 近几年开始关注深度学习算法，并进行初步研究，目前已应用在人耳检测识别、人脸识别、车牌检测识别、字符识别等领域有些进展，有3名已毕业研究生的硕士学位论文采用深度学习方法进行模式识别。
- 

基于卷积神经网络的人耳检测



- (1) 为输入图片生成2000个左右的候选区域。
- (2) 将生成的每个候选区域与人工进行标注的区域进行交并比，筛选出正负样本集。
- (3) 设计适用于人耳检测的卷积神经网络，并将两类样本集作为卷积神经网络的输入并提取特征。
- (4) 将特征进入人耳的SVM分类器，根据评分判断是否属于该类别
- (5) 经过分类器选择后得到符合人耳图像的阈值范围内的有限候选框，利用非极大抑制算法去分类结果中的冗余的选框。
- (6) 最后使用回归方法对分类结果进行精细修正，并确认候选框最终位置。

设计的深度学习网络模型



包含4个卷积层，3个池化层 (pooling-layer)，1个全连接层 (fully-connected layer)，作用在每个卷积层和全连接层的激活函数为ReLU 激活函数 (Rectified Linear Units)。

第一层为图像的输入层，根据采集到的样本图像中实际包含人耳部分的分辨率范围，以及我们的硬件环境的计算能力，将输入层的维度设定为 224×224 。

网络的第一层卷积层为C1，卷积核大小为 7×7 ，步长为2，卷积核数量为96个，经过卷积运算得到96个 110×110 的特征图像（feature map）

网络的第二层为池化层，设定为S2：池化窗口大小为 3×3 ，步长为2，得到96个 55×55 的特征图像。

第三层为卷积层C3，128个 5×5 大小卷积核，步长为2，得到128个 26×26 的特征图像。

第四层为池化层S4同样经过 3×3 大小池化窗口得到128个 13×13 的特征图像。

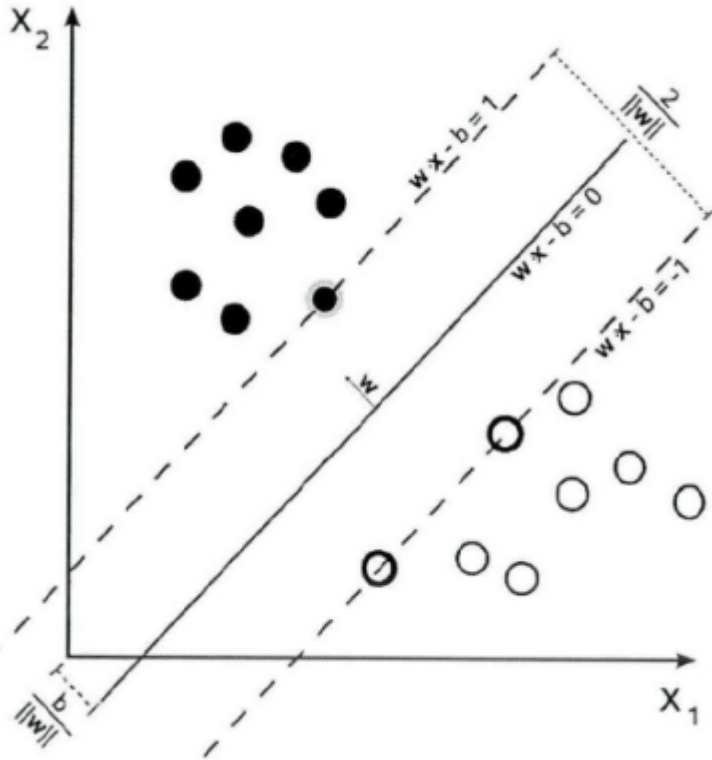
第五层为卷积层C5，256个卷积核大小为 3×3 ， $\text{pad} = 1$ ，得到256个 13×13 的特征图像。

同样卷积层C6存在256个 3×3 大小的卷积核，得到256个 13×13 的特征图像。

S7为最后一个池化层，池化窗口大小为 3×3 ，经过池化操作后得到256个 6×6 大小的特征图。

最后一层为全连接层包含4096个神经元，得到最终的特征向量进入SVM分类器进行人耳与非人耳的检测。

支持向量机



- 一是将 K 分类问题转化为 K 个二分类问题，样本的正负类也与传统二分类相同，每个分类的输出为该目标属于该分类的概率，其最终分类结果为各个二分类输出的最大值。
- 另一是需要构建 $K \times (K-1)$ 个分类器，以两两组合的方式最为二分类依据，每个分类结果作为该分类的选票，其中选票最多的分类为最为本次分类的结果。
- 支持向量机的在二分类领域的识别精度高，并解决了函数映射过程中凸优化问题。

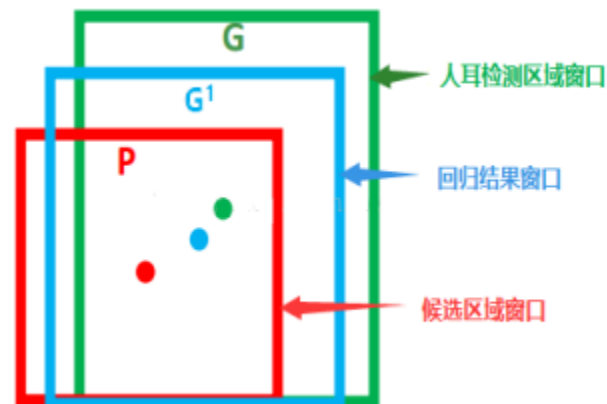
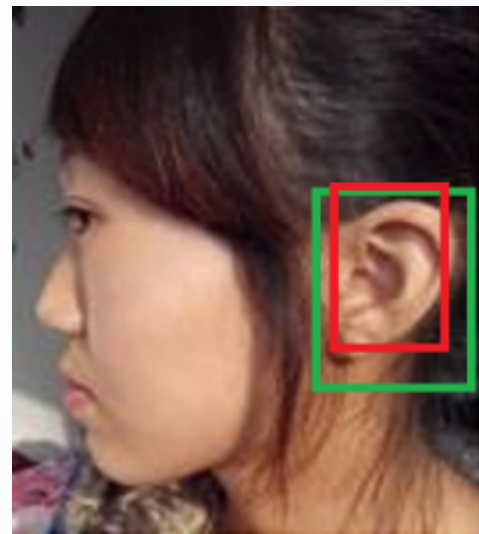
非极大值抑制

- 由于在训练的过程中通过选择搜索生成的候选区域数量较大，且与人工标注的人耳区域 $IoU > 0.7$ 的所有候选区域作为正类样本对卷积神经网络进行训练。所以在测试的过程中会检测出多个人耳相似区域。使用该算法的主要目的就是为了通过迭代-遍历-消除的过程来去掉测试结果中的冗余的选框。
- 假设所有的候选区域经过CNN的特征提取过程后，使用SVM分类器进行分类评分，其中有6个候选区域符合人耳图像的阈值范围，如图，根据SVM分类器的分类概率将这个6个候选区域进行从大到小排列。

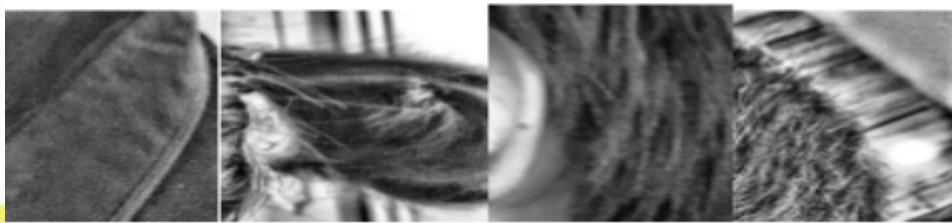
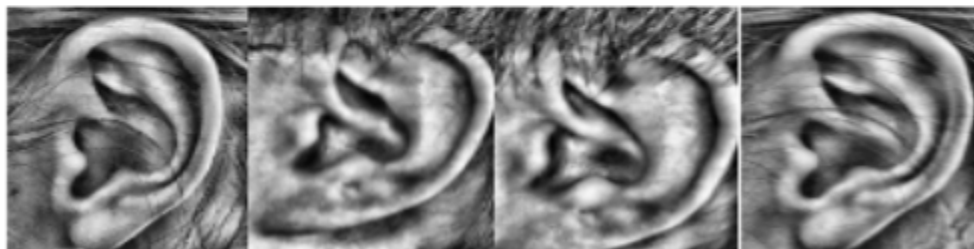
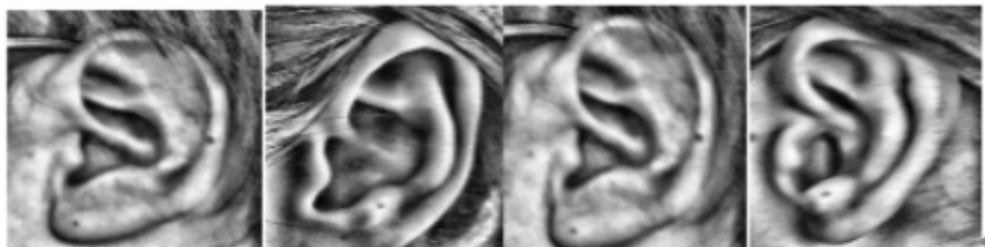


边界框回归

- 通过非极大值抑制算法后，将会剩余一个或者几个符合条件的候选区域，再与分值最高的候选区域的IoU大于设定阈值的其他区域进行重合，得到了一个最为可能的候选框。尽管该候选框的为最终检测区域的概率很高，但也有可能出现定位不够准确的情况。
- 上图，其中绿色框中为真正的人耳检测区域，红色框为经过计算获得的候选区域，所以需要
对红色区域进行微调，使得微调后的区域与人耳检测区域更加接近。针对候选区域的微调建立一个线性回归器，下图。红色的框 P 代表着计算得出的候选区域窗口，绿色的框 G 代表人耳检测区域窗口，线性回归器的目标就是找到一个映射过程，得到一个与人耳检测区域更加接近的回归窗口 G'



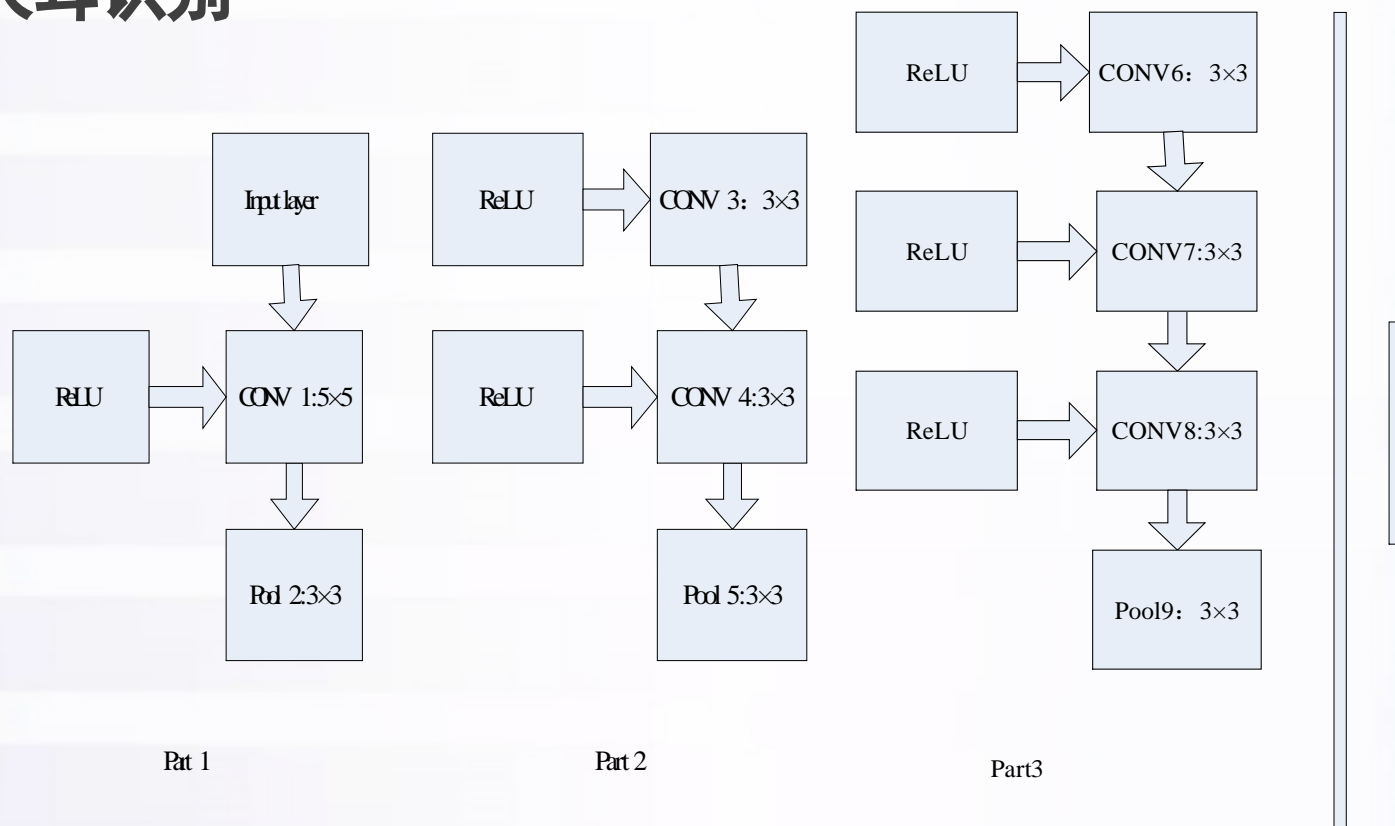
实验用部分正样本和负样本



部分检测结果



人耳识别



人耳识别网络结构图，包含6个卷积层和3个池化层，一个全连接层。选择最优的激活函数ReLU(Rectified Linear Units)。全连接层引入了Dropout技术，其原理是在每次迭代随机更新网络参数，把不工作的神经元节点划分为非网络的一部分，暂停其权重的更新，等下次样本输入，恢复其工作状态。

Softmax分类器

- 定义训练集中的第*i*个样本数据为 x_i ，得分函数会在参数 W 下计算出一个对于所有类的得分结果记作 $f(x_i, W)$ ，其中第*j*个类别的得分结果我们记作 $f(x_i, W)_j$ ， y_i 为实际类别，则对于第*i*张图片，损失函数使用互熵损失/cross-entropy loss，形式：

$$L_i = -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right)$$

- 其中 f_j 代表分向量 f 的第*j*个元素值。总的loss是通过 L_i 遍历训练集之后得到的均值，再加上正则化项 $R(W)$ ，而函数 $f(z_k) = \frac{e^{z_k}}{\sum_j e^{z_j}}$ 为Softmax函数：他的输入值是一个实数向量 Z ，在指数域做了一个归一化（保证和为1）映射为概率。
- 在计算中间项 $e^{f_{y_i}}$ 和 $\sum_j e^{f_j}$ 因为指数运算可能变得非常大，所以我们在分子分母前乘以常数 C ，整理得到：

$$\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} = \frac{C e^{f_{y_i}}}{C \sum_j e^{f_j}} = \frac{e^{f_{y_i}} \log C}{\sum_j e^{f_j} \log C}$$

假设 y 是输入样本 z 对应的类别, $y=0,1,\dots,N$ 。对于 z , 其损失函数定义为:

$$L(y, z) = -\log f(z_y) = -\log \sum_j e^{z_j - z_y}$$

对上式求导, 有: $\frac{\partial L(y, z)}{\partial z} = \frac{e^{z_y}}{\sum_j e^{z_j}} - 1 = f(z_y) - 1$

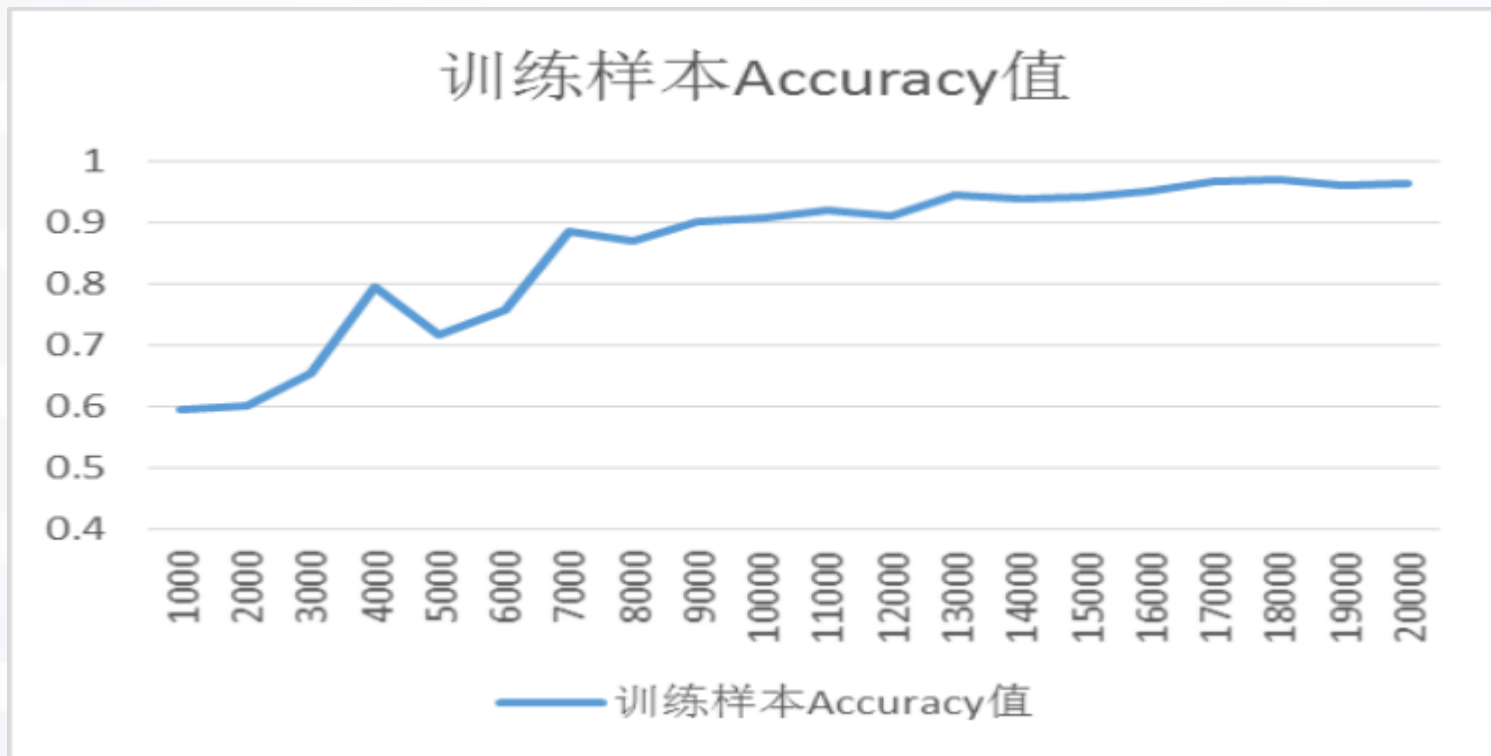
梯度下降方向即为: $-\frac{\partial L(y, z)}{\partial z} = 1 - f(z_y)$

对于Softmax的反向传导设Softmax的输出为 a , 输入为 z , L 为损失函数, 则 $\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \frac{\partial a}{\partial z}$, 其中 $\frac{\partial L}{\partial a}$ 在caffe框架中

要计算 $\frac{\partial a}{\partial z}$, 直接求导可得:
$$\frac{\partial a_i}{\partial z_k} = \begin{cases} a_i - a_k, & i = k \\ -a_i a_k, & i \neq k \end{cases}$$

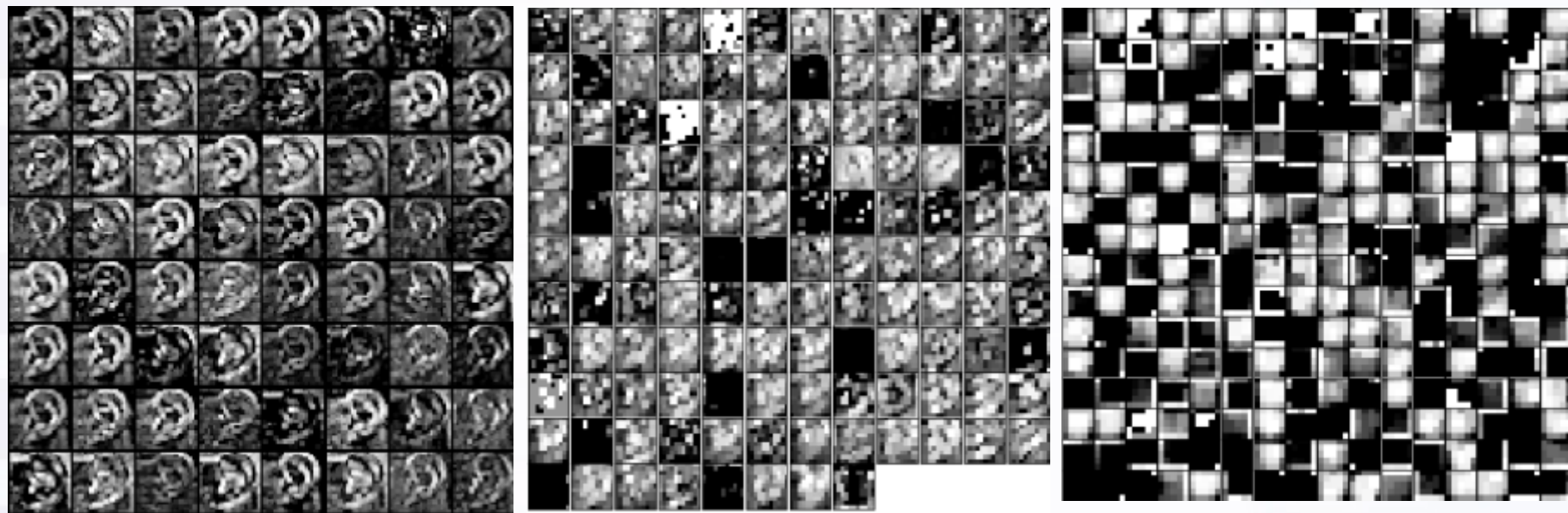
最后得到: $\frac{\partial L}{\partial z_k} = \frac{\partial L}{\partial a} \frac{\partial a}{\partial z_k} = \left(\frac{\partial L}{\partial a}\right) * a_k - \frac{\partial L}{\partial a} * a_k$

训练



网络在训练集迭代次数为12000时趋势逐渐变得平缓，距离训练样本所表现的抽象特征越来越接近。到迭代次数为18000时损失函数的loss值已经降低到0.028左右，损失函数是用来衡量某个类别的概率，loss值越小说明网络的收敛速度越快。

反卷积可视化结果



- ▣ 以上为3个阶段的可视化结果

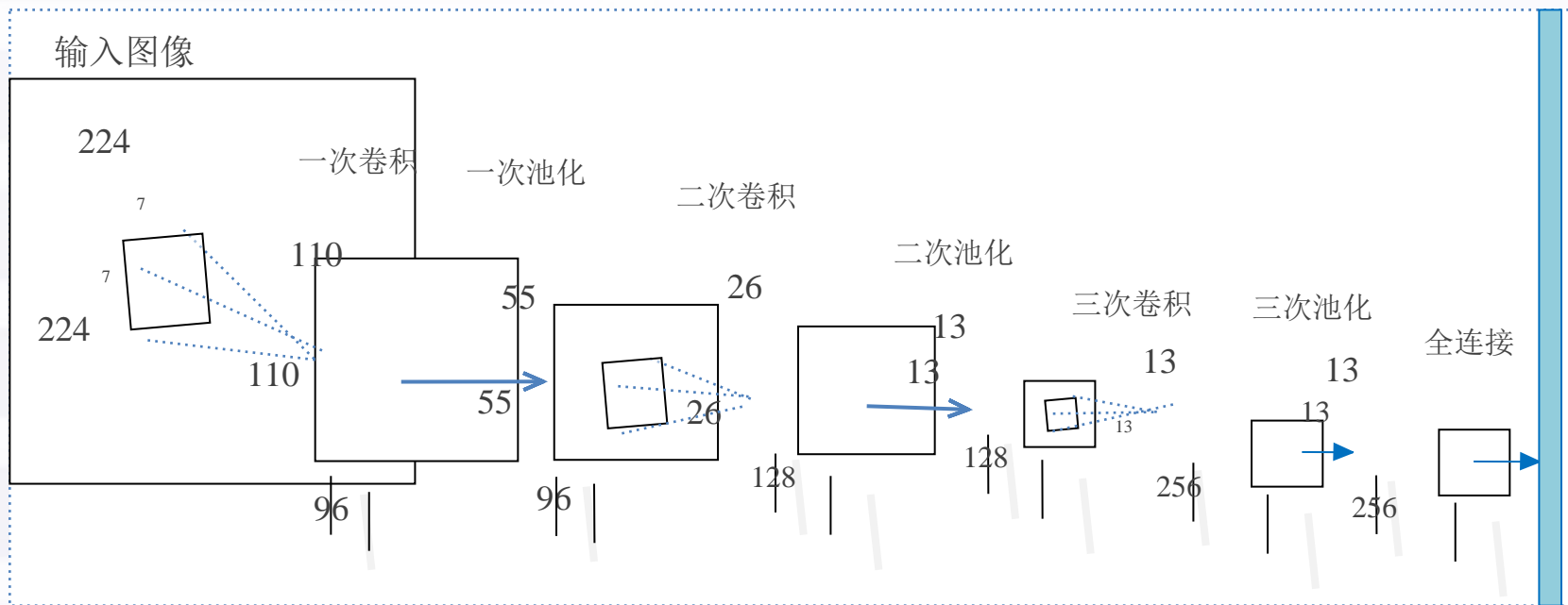
实验结果



方法	0%	10%	20%	30%	40%	50%
HOG	0.9579	0.942	0.880	0.814	0.739	0.524
CNN	0.9820	0.975	0.968	0.927	0.762	0.563

基于深度学习的复杂场景下车牌检测定位

为车牌识别设计的深度学习网络结构



car detections with p(car | box) >= 0.8



car detections with p(car | box) >= 0.8



car detections with p(car | box) >= 0.8



car detections with p(car | box) >= 0.8



car detections with p(car | box) >= 0.8



car detections with p(car | box) >= 0.8



car detections with p(car | box) >= 0.8



car detections with p(car | box) >= 0.8



部分检测结果

基于深度卷积神经网络人脸相似性检测

相似度0.970



导入JPG图片1

E:\琦涛笔记本备份\11\11\测试图片
\Tester_13TrainingPosepose_0.jpg

导入JPG图片2

E:\琦涛笔记本备份\11\11\测试图片
\Tester_13TrainingPosepose_1.jpg

比对

相似度0.972



导入JPG图片1

E:\琦涛笔记本备份\11\11\测试图片
\Tester_19TrainingPosepose_1.jpg

导入JPG图片2

E:\琦涛笔记本备份\11\11\测试图片
\Tester_19TrainingPosepose_2.jpg

比对


深度学习的重点是对模型的运用，模型中需要的参数是通过对大量数据的学习和分析中得到的。

由于深度学习中，图模型的复杂化导致了算法的时间复杂度急剧提升，为了保证算法的实时性，需要更高的并行编程技巧以及更好更多的硬件支持，才能完成一些比较前沿而又实用的应用。

对于深度学习，还存在很多没有解决的问题。既没有完整的关于深度学习有效性的理论，也没有任何一本能超越机器学习实战经验的指南或者书。

目前的深度模型有用于视觉和语音识别的卷积神经网络，以及能够进行自我演绎的深度回归神经网络和会自主玩游戏的深度强化学习模型。深度学习不是万能的，它有足够的理由能日益流行，但始终无法接管整个世界。

要不断增加你的**机器学习**技能，你的饭碗无忧。但也不要对深度框架过于崇拜，不要害怕对这些框架进行裁剪和调整，以得到和你的学习算法能协同工作的软件框架。



近几年发表论文

2016	年份	学术期刊或出版社名称 (影响因子)	卷 (期)	作 (著) 者名次	他引次数
基于L0稀疏约束的近似稀疏解人耳识别	2016	计算机应用研究	33(7)	1	
一种稀疏表示的多姿态人耳识别方法	2014	系统仿真学报	26(9)	1	
基于Gabor小波和ASIFT特征点的人耳识别	2013	计算机应用与软件	30(11)	1	
L0 gradient based image smoothing method for ear identification	2015	International Journal of Signal Processing, Image Processing and Pattern Recognition	8(6)	1	
一种模块化神经网络的人耳识别方法	2016	辽宁科技大学学报	39(3)	1	
Ear recognition based on point feature	2013	International Conference on Vehicle and Mechanical Engineering and Information Technology	380-384	1	
Watermarking Algorithm Based on Quantization Index Modulation and Singular Value Decomposition	2012	Electronic Commerce, Web Application and Communication	2012	1	
Ear Recognition Based on Weighted Wavelet Transform and DCT	2014	26 th Chinese Control and Decision Conference	2014	1	
Image registration based on improved ant colony algorithm	2013	Advanced information and computer technology in engineering and manufacturing. mang	765	1	

谢谢！

