

数字图形处理大作业

(查阅文献)

课程名称	数字图形处理
项目名称	高光谱遥感图像分割
专业班级	计算机技术
姓 名	孙熙春
学 号	182306057026

前言：

当今时代是一个信息化的时代，它以其信息量大、知识爆炸和传播速度快等特点不断对人们的思维方式、学习方式提出挑战。面对日新月异的技术和瞬息万变的知识，人们需要更好的组织思维、更好的意义建构，以实现對知识内容更深层次的理解。

深度学习源于对知识的深度理解，而不只是机械记忆、简单运用。深度学习（deep learning）作为机器学习算法中的一个新兴技术，其动机在于建立，模拟人脑进行分析学习的神经网络。深度学习的本质是对观察数据进行分层特征表示，实现将低级特征进一步抽象成高级特征表示，而这一切都是通过神经网络来执行的。

一、深度学习的概念

深度学习的概念源于人工神经网络的研究。含多隐层的多层感知器就是一种深度学习结构，是机器学习中一种基于对数据进行表征学习的方法。观测值（例如一幅图像）可以使用多种方式来表示，如每个像素强度值的向量，或者更抽象地表示成一系列边、特定形状的区域等。而使用某些特定的表示方法更容易从实例中学习任务（例如，人脸识别或面部表情识别）。深度学习的好处是用非监督式或半监督式的特征学习和分层特征提取高效算法来替代手工获取特征。

深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征表示。概念由 Hinton 等人于 2006 年提出。基于深度置信网络(DBN)提出非监督贪心逐层训练算法，为解决深层结构相关的优化难题带来希望，随后提出多层自动编码器深层结构。此外 Lecun 等人提出的卷积神经网络是

第一个真正多层结构学习算法，它利用空间相对关系减少参数数目以提高训练性能。

深度学习是机器学习研究中的一个新的领域，其动机在于建立、模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如图像，声音和文本。

同机器学习方法一样，深度机器学习方法也有监督学习与无监督学习之分。不同的学习框架下建立的学习模型很是不同。例如，卷积神经网络（Convolutional neural networks，简称 CNNs）就是一种深度的监督学习下的机器学习模型，而深度置信网（Deep Belief Nets，简称 DBNs）就是一种无监督学习下的机器学习模型。

二、深度学习研究现状

1. 语音识别领域

长期以来，语音识别系统大多是采用高斯混合模型（GMM）来描述每个建模单元的概率模型。由于这种模型估计简单，方便使用大规模数据对其训练，该模型有较好的区分度训练算法，保证了该模型能够很好的训练。在很长时间内占据了语音识别应用领域主导性地位。但是 GMM 实质上一种浅层学习网络模型，特征的状态空间分布不能够被充分描述。而且，使用 GMM 建模数据的特征维数通常只有几十维，这使得特征之间的相关性不能被充分描述。最后 GMM 建模实质上是一种似然概率建模方式，即使一些模式分类之间的区分性能够通过区分度训练模拟得到，但是效果有限。

从 2009 年开始，微软亚洲研究院的语音识别专家们和深度学习领军人物 Hinton 合作。2011 年微软公司推出基于深度神经网络的语音识别系统，这一成果将语音识别领域已有的技术框架完全改变。采用深度神经网络后，样本数据特

征间相关性信息得以充分表示，将连续的特征信息结合构成高维特征，通过高维特征样本对深度神经网络模型进行训练。由于深度神经网络采用了模拟人脑神经架构，通过逐层的进行数据特征提取，最终得到适合进行模式分类处理的理想特征。

2. 图像识别领域

对于图像的处理是深度学习算法最早尝试应用的领域。2012 年的 ImageNet 竞赛中，Krizhevsky 等人训练了一个大型的深度卷积神经网络 AlexNet，它包含了 8 个学习层：5 个卷积层和 3 个全连接层。Krizhevsky 等人成功借助了 ImageNet 这个足够大的数据集，借助了 GPU 的强大计算能力，借助了深层的 CNN 网络，借助了随机梯度下降 (SGD) 和 Dropout 等优化技巧，借助了训练数据扩充策略，最终赢得了 2012 年的 ILSVRC，实现了 top-5 的错误率为 15.3% 的好成绩。无论如何，他们靠着深度学习震惊了机器学习领域，从此大量的研究人员开始进入这个领域，一发不可收拾。2013 年的 ZFNet[3] 错误率降到了 11.2%。2014 年 Christian Szegedy 等人设计的 GoogLeNet 是一个 22 层的卷积神经网络，在 ILSVRC2014 上 top-5 的错误率降到了 6.7%。2015 年，微软亚洲研究院的 Kaiming He 等人设计了一个多达 152 层的 ResNet 架构，它比之前的深度神经网络都深，但却降低了复杂度，从而更易于训练，ResNet 的错误率也降到了 3.6%。而人类的水平也仅仅是 5%-10%，由此可见，深度学习网络已经跟人类的分辨能力相当了。

3. 自然语言处理领域

自然语言处理问题是深度学习在除了语音和图像处理之外的另一个重要的应用领域。数十年以来，自然语言处理的主流方法是基于统计的模型，人工神经网络也是基于统计方法模型之一，但在自然语言处理领域却一直没有被重视。语

言建模时最早采用神经网络进行自然语言处理的问题。美国 NEC 研究院最早将深度学习引入到自然语言处理研究中，其研究院从 2008 年起采用将词汇映射到一维矢量空间和多层一维卷积结构去解决词性标注、分词、命名实体识别和语义角色标注四个典型的自然语言处理问题。他们构建了一个网络模型用于解决四个不同问题，都取得了相当精确的结果。总体而言，深度学习在自然语言处理上取得的成果和在图像语音识别方面相差甚远，仍有待深入研究。

三、深度学习面临的挑战及解决

深度学习主要利用了神经网络技术，因此深度学习所面临的技术挑战主要来自于神经网络目前所面临的问题。随着层数的不断增加，神经网络中的各种问题开始涌现，总结为以下几点。

1. 可用的带标签数据集的较少

深度学习是一种有监督的学习方法，有监督的方法通过训练数据学习得到一个分类模型，并将该分类模型应用到对未知数据（测试数据）的分类中去。因此，分类模型的好坏强烈依赖于训练数据集的数量和质量，一个好的训练数据集能够得到一个好的分类模型，从而能够对未知数据得到一个较好的分类结果。

目前，可用的带标签的训练数据集仍然比较少，这可能会限制深度学习的进一步发展。2012 年的 ImageNet 竞赛中，提供了超过 22000 类别的多于 1500 万的高分辨率照片，这使得 AlexNet[5]的成功成为可能。但是随着深度神经网络的进一步加深，例如 2015 年的 ILSVRC 竞赛中的 ResNet[2][3]已经达到惊人的 152 层，训练这样的一个深层的神经网络需要更大的数据集。因此，需要获取更大的带标签数据集，使得深度学习的进一步加深成为可能。

而 2015 年 NIPS 的 Ladder Network[1][4]采用半监督的方法，充分利用了无

标签数据加少量有标签数据，相比于以往的方法精度上有了大幅的提升。

2. 网络的复杂度上升和可解释性降低

深度学习的一个很大的局限性是神经网络的可解释性不强，有人形象的把神经网络比喻成一个黑盒，就是因为它的内部知识的表示很不直观，用加权链连接单元的网络表示的知识让人很难解释。因此，提取隐藏在训练后的神经网络中的知识以及如何表示这些知识也是一个很重要的研究方向，包括从网络中提取规则等。

随着深度学习的进一步加深，深度神经网络的复杂度也逐渐上升。以 2012 年的 ILSVRC 竞赛为例，AlexNet[5]的层数达到了 8 层，而其参数和神经元的个数是巨大的，AlexNet[5]具有 6000 万个参数和 65 万个神经元，其复杂度可想而知。并且，越来越多的参数不仅需要更大的数据集来训练，同时需要更强的计算能力，否则神经网络的训练时间将是不可忍受的。

3. 过拟合

神经网络训练的过程其实就是不断调整网络中神经元之间的连接权重，使得网络预测的输出值与实际值之间的误差最小化，这种过程不断进行下去，直到权重收敛。在这个过程中，网络的训练误差不断减小，在未知数据上的测试误差也不断减小，但当训练误差减到某个值之后，网络的预测误差反而增大，说明泛化能力降低，这就是神经网络建模过程中的过拟合现象。

多层神经网络有很强的非线性拟合能力，可映射任意复杂的非线性关系，然而神经网络强大的学习能力同时也是它的一个缺点，过度的学习可能导致神经网络陷入过拟合。最近几届的 ILSVRC 会议上提出的深度神经网络都不同程度的对过拟合提出了解决方案，由此可见过拟合带给神经网络的重大影响。

在神经网络中，数据扩增和 Dropout 是两种常用的正则化方法。数据扩增其实就是获取更多的数据，然而，数据通常是难以获取的和昂贵的。很多时候，收集更多的数据意味着需要耗费更多的人力物力，效率特别低。所以，可以在原始数据上做些改动，获得更多的数据。以图片数据集为例，可以做各种变换，比如讲原始图片旋转一个角度、添加随机噪声、截取原始图片的一部分等。Dropout 每次将一半的隐层单元随机删除，一直迭代，直到训练结束。可以简单地这样解释，运用了 dropout 的训练过程，相当于训练了很多个只有半数隐层单元的神经网络，每一个这样的半数网络，都可以给出一个分类结果，这些结果有的是正确的，有的是错误的。随着训练的进行，大部分半数网络都可以给出正确的分类结果，那么少数的错误分类结果就不会对最终结果造成大的影响。

4. 局部最优解

神经网络的训练存在多个极小点，梯度下降不能保证求出全局最优解。目前，解决这个问题已经提出了很多方案。使用改进的 BP 算法，增加动量项，或者自适应学习率；和别的优化算法组合，例如遗传算法优化初始权值，提前锁定全局最优；重新训练，每次训练的结果都是不同的，下一次的训练很有可能不会陷入局部极小；更改学习函数、训练函数再次训练。

5. 梯度消失现象

具体来说，我们常常使用 sigmoid 作为神经元的输入输出函数。对于幅度为 1 的信号，在 BP 反向传播梯度时，每传递一层，梯度衰减为原来的 0.25。层数一多，梯度指数衰减后低层基本上接受不到有效的训练信号。这就是梯度消失的现象。

引入 Relu 代替 sigmoid 激活函数。采用 sigmoid 等函数，算激活函数时（指

数运算), 计算量大, 反向传播求误差梯度时, 求导涉及除法, 计算量相对大, 而采用 Relu 激活函数, 整个过程的计算量节省很多。对于深层网络, sigmoid 函数反向传播时, 很容易就会出现梯度消失的情况 (在 sigmoid 接近饱和区时, 变换太缓慢, 导数趋于 0, 这种情况会造成信息丢失), 从而无法完成深层网络的训练。而 Relu 会使一部分神经元的输出为 0, 这样就造成了网络的稀疏性, 并且减少了参数的相互依存关系, 缓解了过拟合问题的发生。

四、深度学习未来展望

无监督学习对于重新点燃深度学习的热潮起到了促进的作用, 但是纯粹的有监督学习的成功盖过了无监督学习。在本篇综述中虽然这不是我们的重点, 我们还是期望无监督学习在长期内越来越重要。无监督学习在人类和动物的学习中占据主导地位: 我们通过观察能够发现世界的内在结构, 而不是被告知每一个客观事物的名称。

人类视觉是一个智能的、基于特定方式的利用小或大分辨率的视网膜中央窝与周围环境区域对光线采集成像的活跃的过程。我们期望未来在机器视觉方面会有更多的进步, 这些进步来自那些端对端的训练系统, 并结合 ConvNets 和 RNNs, 采用增强学习来决定走向。结合了深度学习和增强学习的系统正处在初期, 但已经在分类任务中超过了被动视频系统, 并在学习操作视频游戏中产生了令人印象深刻的效果。

在未来几年, 自然语言理解将是深度学习做出巨大影响的另一个领域。我们预测那些利用了 RNNs 的系统将会更好地理解句子或者整个文档, 当它们选择性地学习了某时刻部分加入的策略。

最终, 在人工智能方面取得的重大进步将来自那些结合了复杂推理表示学习

(representation learning) 的系统。尽管深度学习和简单推理已经应用于语音和手写字识别很长一段时间了, 我们仍需要通过操作大量向量的新范式来代替基于规则的字符表达式操作。

参考文献

- [1] Pezeshki, Mohammad, et al. "Deconstructing the ladder network architecture." *International Conference on International Conference on Machine Learning JMLR.org*, 2016:2368-2376.
- [2] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:770-778.
- [3] He K, Zhang X, Ren S, et al. Identity Mappings in Deep Residual Networks[J]. 2016:630-645.
- [4] Rasmus A, Valpola H, Honkala M, et al. Semi-Supervised Learning with Ladder Networks[J]. *Computer Science*, 2015, 9 Suppl 1(1):1-9.
- [5] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.