**北京师范大学珠海分校**

**本科生毕业论文**

论文题目 **深度学习模型研究及应用**

学 院　　 应用数学学院

专 业 数学与应用数学

学 号 1717010022

学 生 姓 名 黄继安

指导教师姓名 李艳

指导教师单位 北师大珠海分校应用数学学院

2020年11月28日

深度学习模型研究及应用

摘要

文献综述

绪论

* 1. 深度学习模型算法发展历史及研究意义

深度学习(Deep Learning)是一种对输入数据进行表示学习或者进行非线性变换的技术，它是不少于2个隐含层的神经网络。因其特殊的层级连接方式可以实现渐进抽象的非线性处理，所以在处理从原始输入数据到期望输出结果的复杂非线性变换上特别出色，并由此实现对原始数据进行表示学习或者非线性建模学习。深度学习特别强调“端到端(end-to-end)”的学习方式，也就是直接从原始输入数据到输出结果的学习过程，这一点相比于传统的机器学习算法需要花费大量的人力资源在特征处理上节省了大量的时间，众所周知，对于传统的机器学习算法，比如决策树(Decision Tree)、K-近邻(k-Nearest Neighbor)等，相关研究人员百分之八十的时间都花在特征处理上面，而使用深度学习模型可以减少花在特征处理上面的时间，让研究人员可以把更多时间专注于模型上，因此深度学习(Deep Learning)通常也称之为表示学习。

深度学习(Deep Learning)本质上是包含了多个隐含层的人工神经网络，而人工神经网络的研究历史并非一帆风顺，具体可追溯到上世纪40年代。Pitts和McCulloch在1943年提出了首个神经元数学模型[1]，Hebb在1949年提出神经元的学习准则[2]，Rosenblatt于1957年提出了感知机(perceptron)模型[3]，自此开启了人工神经网络的第一波热潮。但是因为人工智能领域的知名学者Minsky等指出感知机模型是线性模型，它无法解决异或问题，导致人工神经网络的研究第一次进入了低谷时期。带动人工神经网络第二次研究热潮的是1986年Williams、Rumelhart和Hinton在自然杂志《Nature》发表的著名误差逆传播算法(back propagation,BP)算法[4]，用于训练含有多隐含层的人工神经网络，BP算法使得求解具有非线性学习能力的多层感知机模型变为可能的事情。事实上即使到了现在，BP算法仍作为人工神经网络训练的标准算法，一直是沿用到今天。而进一步激励了非线性感知机的研究发展的是Hornik、Stinchcombe和White在1989年从理论上证明了，神经网络可以逼近任意复杂的连续函数[5]。那么模型有了，训练问题看似也被BP算法解决了，可是不要忽视了其他方面的局限，由于多层神经网络大量的参数需要优化，这决定了计算平台算力一定要高并且训练数据的规模越大越好，但遗憾的是当时的数据规模都还很小，算力也不够，难以支撑多层神经网络的学习，这就导致了多层神经网络在许多领域的表现还没有浅层模型表现的好，研究的热潮曾是一度消沉。

2006年机器学习领域的泰斗加拿大多伦大大学Hinton等在全世界最权威的学术期刊之一的《科学》上发表文章[6,7]，文章指出了多层神经网络比浅层模型在特征学习能力上有着更加出色的能力，并用无监督和分层的预训练可以有效的解决多层神经网络的训练难题。同时期，蒙特利尔大学的Bengio也发表论文强调了分层(lay-wise)训练多层神经网络的重要性[8],多亏了这些学者不懈的努力，重启了深度学习在工业界和学术界的热潮，发展到今日，神经网络已经有上百种模型，在计算机图像识别[9,10]，语音识别[11-13]等许多技术领域大放异彩。

如今，深度学习在许多计算机智能领域上取得了非常广泛而又深远的影响，在很多问题上面取得了意想不到的成绩，这其中不仅包括了自然语言处理(Natural Language Processing)、语音处理(Speech Signal Processing)、视觉计算、搜索、对话、推荐，甚至是在如此复杂的围棋任务上，阿尔法围棋(AlphaGo)是首个击败人类围棋世界冠军的人工机器人，其核心就是深度学习。那么深度学习在如此众多的领域取得巨大的成功，其中最耀眼的要数计算机视觉方面，像在图像分类、图像分割、动作识别、显著性检测等许多视觉处理任务上都是取得了非常显著的进步[14]。

人脸识别，是通过提取人脸的特征信息从而达到身份识别的一项生物识别技术。但由于其收到光线，面部表情，像素等因素的影响使得人脸变化的分布是非线性的且极为复杂。Labels Faces in the Wild(LFW)是由UMass Vision发布的无约束自然场景可供人脸识别模型训练的数据集，也是截止目前位置公开最具挑战的人脸识别数据集，那么在将深度学习的技术运用到此数据集之前，使用传统的基于手工设计的特征提取的方法在LFW数据集上的最好性能是95.17%(平均的分类精度)[15]。深度学习浪潮卷起之后，2014年，Facebook的团队[16]和香港中文大学的团队[17]用深度学习的技术分别报告了97.35%和97.45%的精度。

深度学习是人工智能的一个范畴，而人工智能其实包含了许多技术，比如自然语言处理，计算机图像识别和应用于许多服务行业的智能会话等技术，而这其中许多技术并不是孤立的，都是交叉的，比说会话场景中“机器人”想要准确理解对方所说的话，就要用到自然语言处理的技术，包括意图识别，知识图谱等技术；计算机图像识别技术可以让计算机像人类一样去“认识”世界，像百度飞浆研发的首个口罩识别系统，还有最出名的2012年的图像分类比赛上Hinton的研究团队采用了深度学习的技术获得了ImageNet图像分类比赛[18,19]的冠军,其识别准确率比第二名高达10%以上，冠军的模型AlexNet就是分布式训练模型并行的典型案例。

机器无法像人类一样非常容易的去理解图片或者视频中的内容，计算机所能看到的就是一堆没有规则的二维矩阵，而我们需要让计算机能从这些二维矩阵中总结出人类难以发现的规律，而这就是深度学习的用武之地。

* 1. 国内外研究现现状

2021年，国内矿冶集团的鲁恒润等人[20]将基于深度学习的图像识别方法应用于识别剥锌机锌阴极板的残留物，使用基于改进的YOLOv3算法在对阴极板残留物识别准确率上达到95%的准确度，比原始的YOLOv3算法提高3%。他们将网络的输入由原本的256x256x3修改成为512x512x3，在分辨率上面提上了一倍以便留存更多有用的信息，使用深度可分离的Xception卷积网络替换原本的darknet53网络，并对正负采样不平衡做了处理，考虑到YOLOv3采用的是锚点机制，如果出现正样本少的情况下会产生大量的负样本，在负样本远大于正样本的情况下负样本对网络的影响会很大，从而降低了模型的准确度。所以鲁恒润调整生成候选框机制，当达到一定训练次数以后适当的降低负样本的权重，使模型更加关注正样本，从而有效解决了正负样本失衡的问题。

2021年，哈尔滨工业大学机电工程学院的于凌涛等人[21]将卷积神经网络模型和迁移学习算法应用于乳腺癌病例图像分类上，同时提出了针对高分辨率的图像使用基于四叉树图像分割的方法分割图像以提升数据量防止模型过拟合。相比较于用手工提取图像特征数据结合传统机器学习算法的结果提高了9.9%到12.2%的识别准确率；比单纯用卷积神经网络提取特征、用翻转等方法增强数据的方法提高2.3%到5.1%的识别准确率；比用窗口滑动结合卷积神经网络的方法提高5%到11.5%的识别准确率。于等人在文中明确指出，通过卷积神经网络进行特征提取的方法，在分类任务上要优于手工特征提取，还节省了大量人力物力。

2021清华大学电子工程系的高宸等人[22]针对现有的视频推荐系统所存在的问题提出了基于图神经网络的推荐算法(VGCN)，将用户观看视频的行为进行图结构建模，使用图神经网络进行表征学习，从而解决了使用卷积神经网络进行隐式建模的方式缺乏对用户切换视频行为关系的显示建模的问题，也考虑到用户长时间和短时兴趣的行为。VGCN相较于卷积序列嵌入推荐模型(CASER)和深度兴趣网络(DIN)在推荐性能上面平均提升1.7%，该提升相对于传统的推荐模型来说是显著的。

* 1. 论文主要研究内容

本文主要研究的是深度学习算法相关的神经网络，主要目的是研究几个经典的神经网络架构，对比得出它们各自的优缺点，并且将它们应用在实际的生产中，旨研究算法从理论，到收集数据，模型的训练，优化，最后应用的一系列过程。论文主要的研究内容有：

1. 实验数据集的收集

首先在网络上收集相关用于训练图像识别的数据，对它们进行相关的数据预处理，处理成为可以输入网络的张量。

1. CNN网络结构设计、训练

通过tensorflow框架搭建CNN网络，对网络模型结构及相关参数调优进行对比选择，将模型部署到GPU上进行训练，通过训练得出能够完成图像识别任务的CNN模型。

* 1. 论文组织架构

本文的组织结构分为五个章节，每一章节的主要内容如下安排：

第一章：绪论。绪论介绍了深度学习模型的发展历史，以及相关研究意义；然后讨论了国内外深度学习相关的最新研究成果(国外的有待补充),虽然深度学习在许多领域崭露头角，但是其相关的训练，模型的可解释性等方面还是有待我们的研究。

第二章：主流的深度学习模型算法。第二章会挑选三个经典的深度学习模型来介绍，分别是卷积神经网络，循环神经网络，生成对抗网络。并且将它们各自的优缺点做一个比较，得出每种网络更适合于哪种场景。

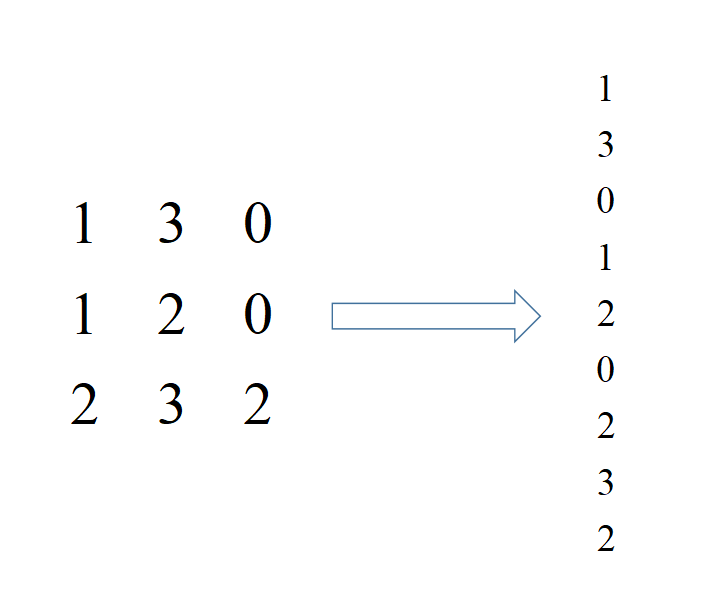
第三章：深度学习的应用。阐述本文所做的有关深度学习在图像识别领域的应用，并对相关的结果进行可视化分析。

第四章：总结和展望。对本文的研究做一个总结，并指出当前深度学习领域的空白点，给出未来深度学习算法研究的重点方向。

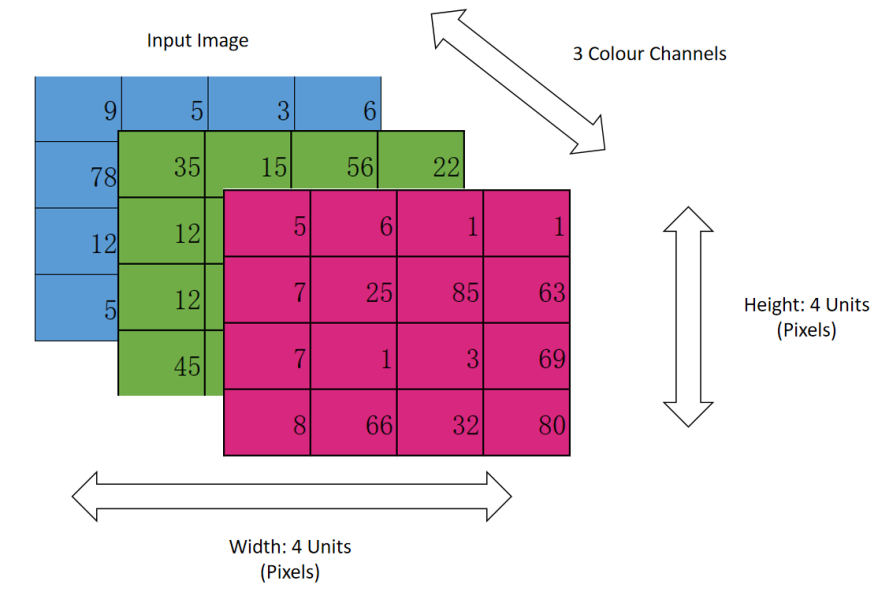
**2 主流的深度学习算法模型**

2.1 卷积神经网络

卷积神经网络(Convolutional Neural Network,CNN)在处理具有网格形状数据的时候表现非常好，如时间序列和图像数据。其在许多领域都取得了优异的成绩。卷积神经网络是最经典的深度学习模型，从20世纪80年代末提出至今人们对其的研究一直没有停止过，其网络结构图1所示。卷积神经网络的架构与人脑中的神经元的连接模型相似，其灵感来自于视觉层的组织。单个神经元仅在被称为感受野的视野受限区域内对刺激做出反应。这些反应的集合将重叠以覆盖整个可视区域。那么我们在对图像数据做处理的时候为什么要选择卷积神经网络而不是前馈神经网络。虽然图像是像素值的矩阵，但是对于涉及像素依赖性的复杂图像，如果仅仅将图像展平，比如将的图像矩阵转换为的矢量，使用该方法几乎没有准确性可言。



通过使用相关过滤器，卷积神经网络能够成功捕获图像中的时空依存关系。 由于所涉及参数的数量减少以及权重的可重用性，该体系结构对图像数据集的拟合效果更好。也就是说，可以训练网络以更好地理解图像的复杂性。在下图中，我们有一个RGB图像，该图像已被其三个颜色平面(红色，绿色和蓝色)分开。图像中存在许多这样的色彩空间-灰度，RGB，HSV，CMYK等。设想，一旦图像达到8K(7680×4320)的尺寸，那么图像矩阵的数据量将是非常巨大的。卷积神经网络的作用是将图像缩小为易于处理的形式，但是又不会丢失重要的数据信息。当我们要设计一种架构时，这一点很重要，该架构不仅要善于学习数据中的规律，而且还可以扩展到海量数据集。



特征抽取器和分类器是卷积神经网络的基本架构，特征提取器一般由多个卷积层和池化层一起叠加组成，卷积、池化的过程会对特征图进行连续地缩小，缩小了特征图就会导致特征图的数目有所增加。最后一个特征提取器的后面，将输入模型的所有特征图展开并排列成一个特征向量，这个特征向量的维度是一，最后这个一维的特征向量作为输入连接分类器。由多层感知机组成的分类器一般连接在特征提取器的后面。

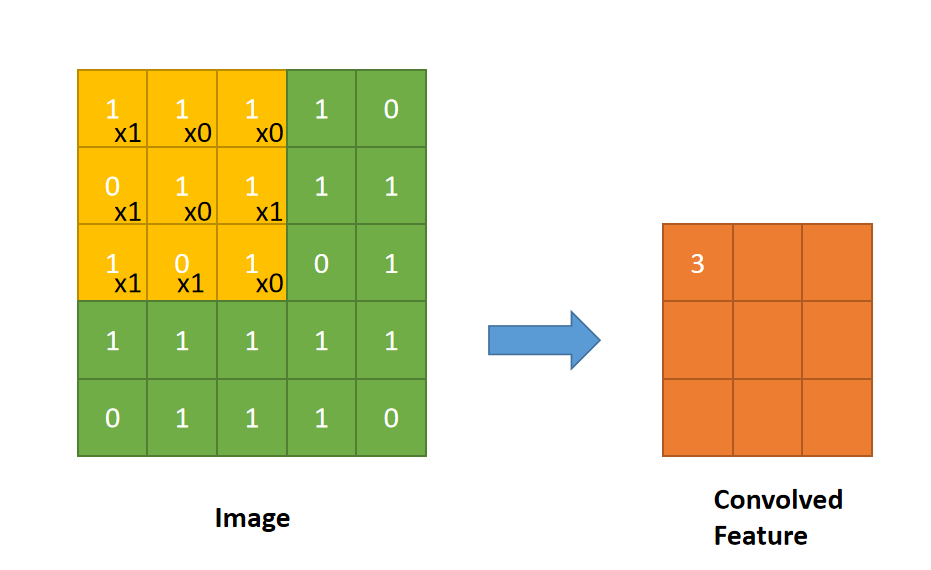
下面说一说有关参数的选择，这里以卷积核的选取为例子，在相同感受野的前提下，小卷积核和多卷积层的计算量要比大卷积核和少卷积层少，但是要注意，卷积核如果小于1则无法起到提升视野的作用了。在进行滑动卷积的时候，为保证锚点恰好在中间，通常选择奇数的卷积模板，容易产生位置偏移。卷积核的大小一般为。卷积的层数则需要根据数据集、模板的大小、网络尺寸以及模型训练的效果来动态的进行优化。

卷积层，我们假定输入的图像维度是：

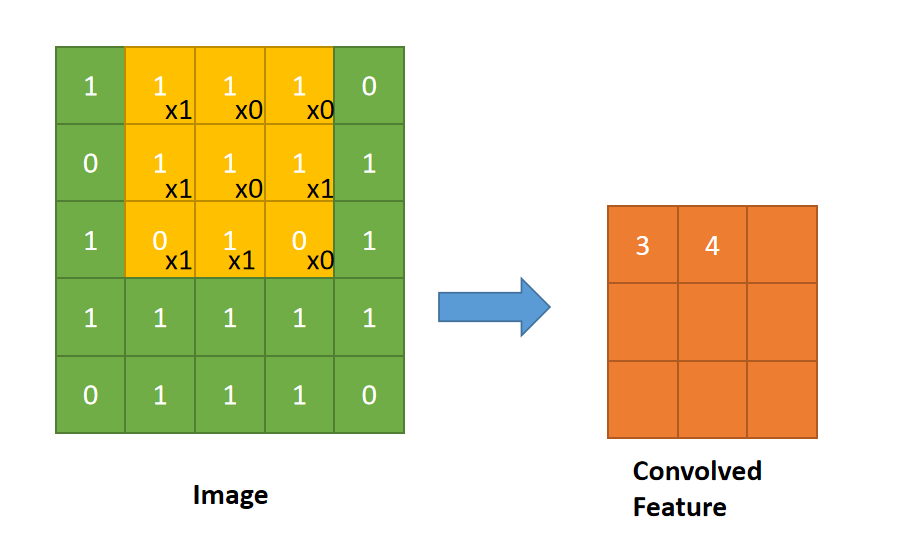


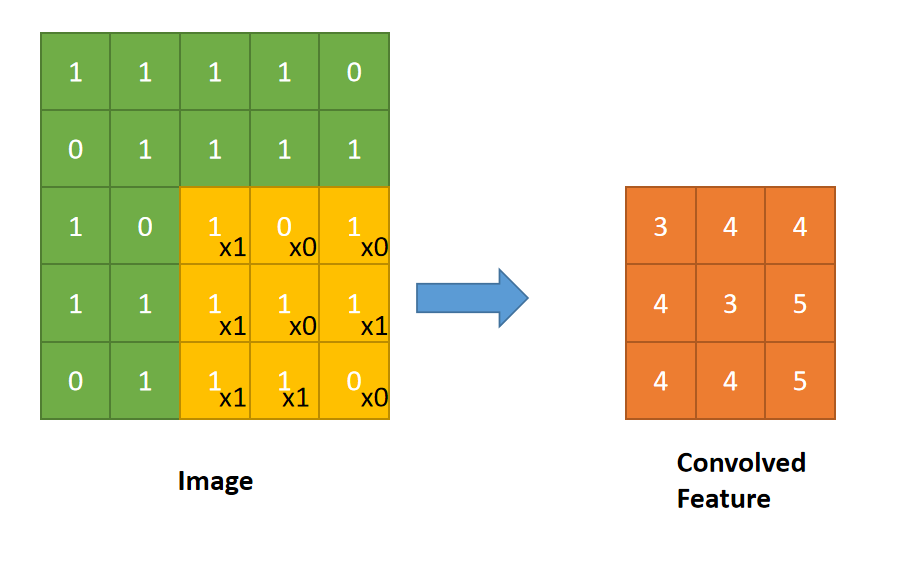
下图的绿色部分类似于我们输入的的图像。在卷积层的第一部分执行卷积运算所涉及的元素我们称之为卷积核或者过滤器，在下图中以黄色表示。我们选择为的矩阵：





卷积核乘上图像的像素值并求和，每次向右位移一个长度。





由于步长= 1(不跨步)，内核每次移位9次，每次在K与内核所徘徊的图像P部分之间执行矩阵乘法运算。

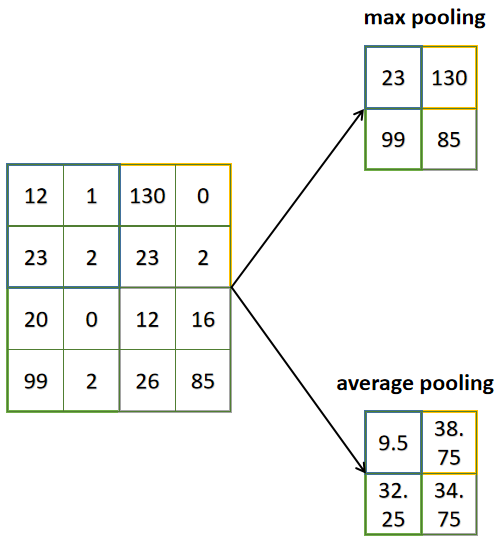
过滤器将以某个“步幅值”向右移动，直到解析出完整的宽度为止。继续前进，它将跳至具有相同“步幅值”的图像的左侧，并重复该过程，直到遍历整个图像为止。

卷积运算的目的是从输入的图像里面提取出诸如边缘之类的隐含在图像中的高级特征。卷积网络不必仅仅局限于一个卷积层。按照惯例，第一个ConvLayer负责捕获图像中的低级特征，例如颜色，边缘，渐变方向等。通过添加图层，这样的体系结构也可以适应高级特征，为我们提供一个对数据集中的图像有全面了解的网络。

池化层与卷积层相似，池化层负责减小卷积特征空间的大小。这是通过降维来减少处理数据所需要的的算力，这对于提取相关旋转和不变位置的主要特征十分有用，从而保持模型训练有效的过程。池化有两种类型：最大池化和平均池化。

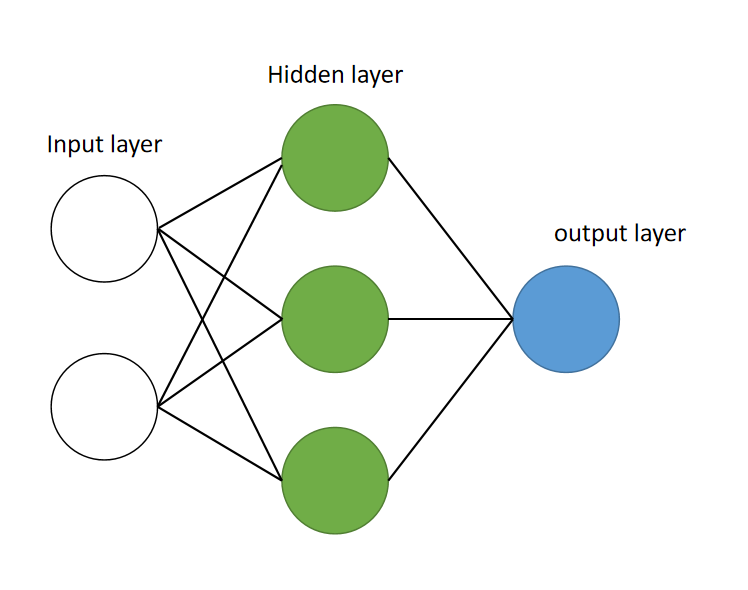
 最大池化从内核覆盖的图像部分返回最大值，平均池化从内核覆盖的图像部分返回所有值的平均值。

最大池化还可以充当噪声抑制器，它完全丢弃了复杂的激活，并且还执行了降维和降噪。而平均池化仅是执行降维作为噪声抑制的机制，所以我们认为最大池化的性能要比平均池化好得多。



2.2 循环神经网络

前馈神经网络(Feed-Forward Nerual Network)的结构如下图所示，它是最简单的神经网模型，只有几个简单的神经元组成，网络包含了输入层，隐层单元，输出层。上一层的神经元与下一层的神经元全连接，同层的神经元之间没有连接，网络之间没有回路，整个网络之间没有反馈，原始数据从输入层逐层通过网络到达输出层。网络中连接权重存储了网络在这些数据中所学到的知识。那么这样的网络结构就会导致FNN无法学习到原始数据中上下文包含的信息，而在现实生活中，许多的数据都是包含丰富的上下文信息，比如语音信息，视频信息，文本信息等。这就导致了FNN在很多包含如上数据的任务上存在很大的局限性。



前馈神经网络通常以其最简单的形式被视为单层感知器。在此模型中，一系列输入进入该层并乘以权重。然后将每个值加在一起，以获得加权输入值的总和。 如果值的总和高于特定阈值(通常设置为零)，则生成的值通常为1，而如果总和低于阈值，则输出值为-1。单层感知器是前馈神经网络的重要模型，通常用于分类任务。此外，单层感知器可以结合机器学习的各个方面。使用称为增量规则的属性，神经网络可以将其节点的输出与预期值进行比较，从而允许网络通过训练来调整其权重以产生更准确的输出值。训练和学习的过程会产生一种形式的梯度下降。在多层感知器中，权重更新的过程几乎类似，但是将过程更具体地定义为反向传播。在这种情况下，将根据输出层产生的输出值来调整网络中的每个隐藏层。

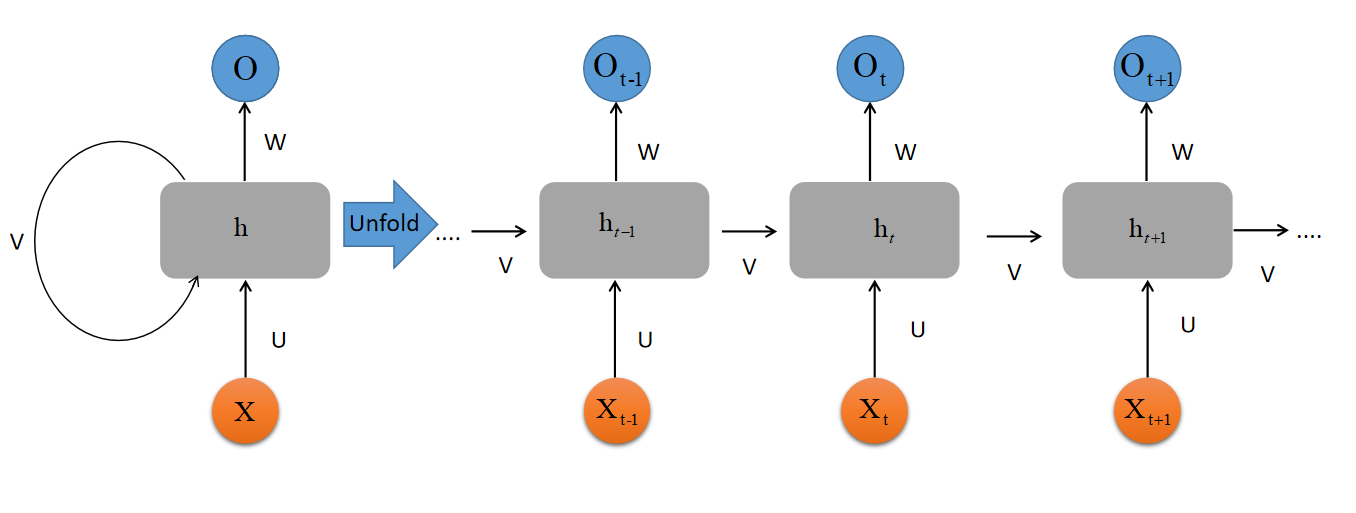
循环神经网络(Recurrent Neural Network,RNN)主要用于序列数据的处理，由于其在自然语言处理(NLP)中的出色成果，它变得越来越受欢迎。在NLP中，它们用于多样化的任务，例如翻译，文本分类，自动文本生成。其网络结构原理如下图所示，图片中，有许多不同的组件，其中最重要的是：

·：输入，它可以使句子中的单词或者其他类型的序列数据；

·：输出，例如，网络认为句子中的下一个单词应该被赋值为之前的单词；

·：RNN中主要的模块，它包含了网络的权重和激活函数；

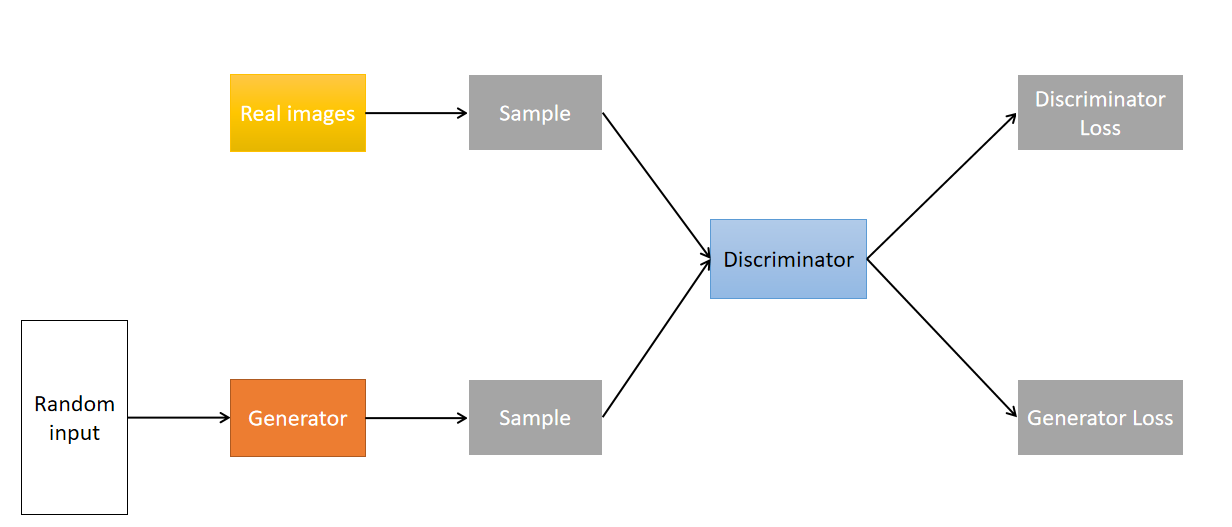
·：表示从一个时间节点到另一个时间节点的通信。



通过下图我们还可以看到，在隐层上面是有回路连接的，其相比于前馈神经网络，RNN神经元在某一时刻的输出结果可以作为输入再次输入到神经元中，这种网络结构对于时间序列数据非常的合适，它能发掘出数据中上下文的依赖关系。展开后的RNN，网络中的参数是共享的，得益于这种设计，大大的减少了需要训练的网络参数。虽然RNN的设计初衷是为了解决学习长期的依赖性问题，但是结果并不如意，标准的RNN无法实现信息的长效存留，RNN的缺点还包括梯度消失和梯度爆炸。

2.3 生成对抗网络

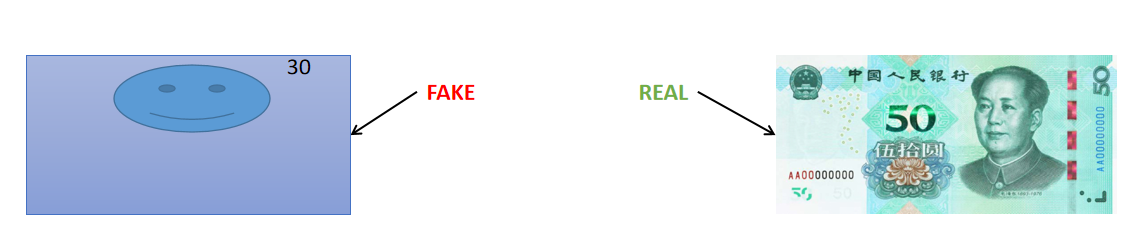
生成对抗网络(Generative Adversarial Network,GAN)的网络结构如下图所示。

 一个GAN包含两个部分：生成器模型(Generative Model)通过训练生成符合分布的数据，生成的数据成为判定模型认为是伪造的实例。判定模型(Discriminative Model)则通过训练，学会将生成器的假数据与真是的数据分开，判定模型会因为产生了非常不合理的数据而对生成器模型进行惩罚。

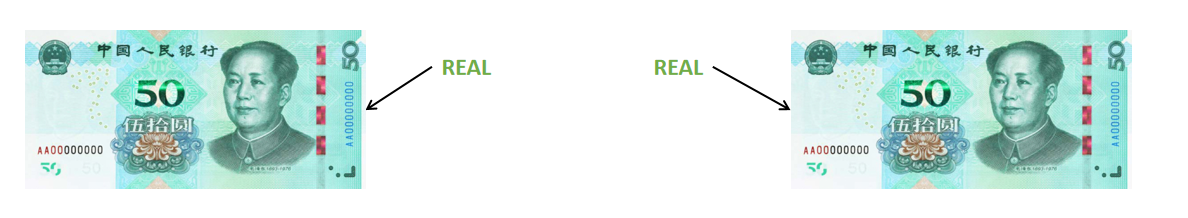
当训练开始的时候，生成器模型会产生非常明显的假数据，判定模型则很快的学会辨别他们是伪造的:



随着训练次数的增加，生成器模型生产出的数据更具有欺骗性：



最终，如果生成器训练顺利的完成，那么判定模型则会难以分辨真假数据，它会开始将伪造的数据分类为真实的数据，其准确性就会降低。

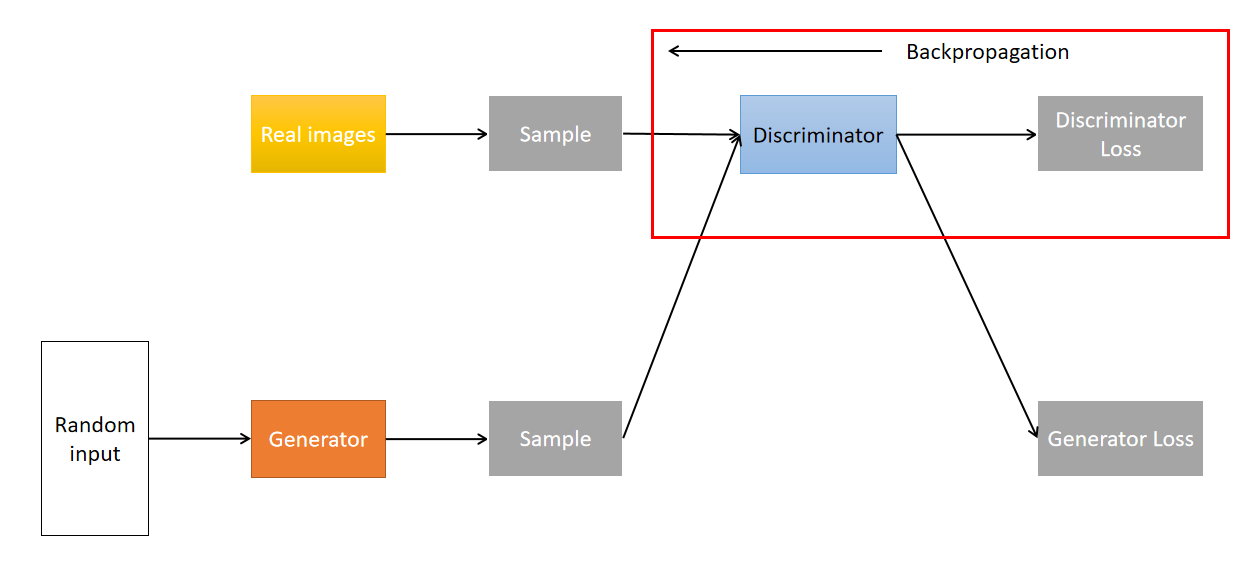


生成器模型和判定模型都是神经网络，生成器模型的输出直接连接到判定模型的输入，通过反向传播，判定模型的分类结果提供了损失，以供生成器模型用来更新其网络结构的参数。

下面我们详细讲述生成器模型和判定模型。判定模型只是一个简单的分类器，它所做的仅仅是尝试将真实数据与生成器创建的数据分开。通过上面GAN的网络结构图可以看到，判定模型的数据来源有两个，一个是真实的数据源，一个是虚假的数据，生产自生成器模型。

以下为判定模型的训练过程：

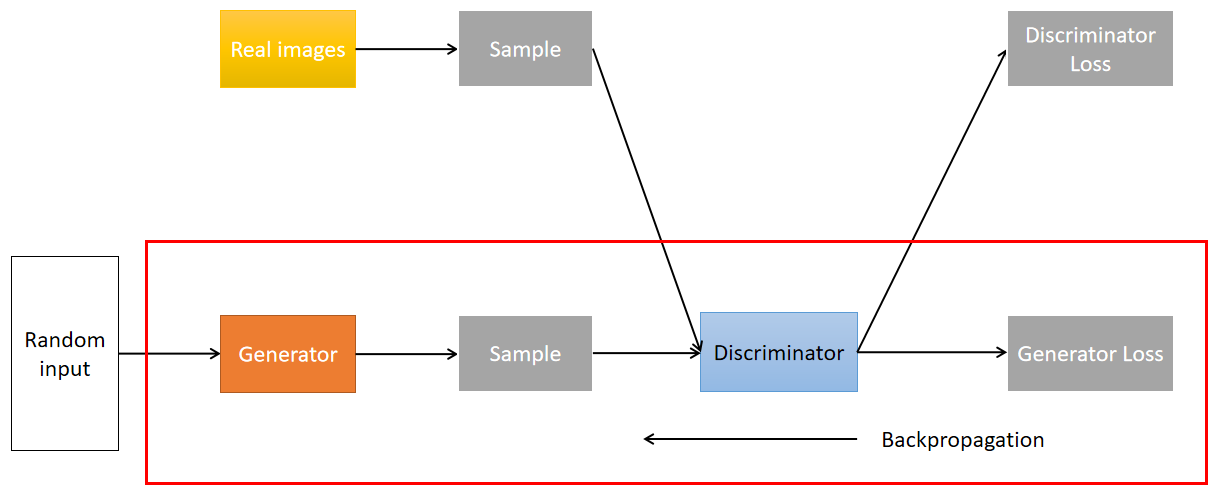
1. 判定模型对来自生成器模型的真实数据和虚假数据进行分类；
2. 如果判定模型将真实数据错误地分类为虚假数据或者将虚假数据错误地分类为真实数据，判定器损失则会对判定模型进行惩罚；
3. 判定模型通过判定器损失的反向传播来更新其权重。



生成模型通过判定模型反馈的结果来学习如何创建虚假数据，它的目的就是让判定模型将其的输出归为真实。

下面为生成模型的训练过程：

1. 采集随机噪声；
2. 根据采样的随机噪声产生输出；
3. 获取判定模型对生成模型的输出的判定结果；
4. 计算来自判定模型分类的损失；
5. 通过判定模型和生成模型的反向传播获得梯度；
6. 使用梯度更新生成模型的权重。



**3 深度学习算法模型的应用**

**参考文献**

[1]Mcculloch W S , Pitts W . A Logical Calculus of the Ideas Immanent in Nervous Activity[J]. biol math biophys, 1943.

[2]Morris R G M . D.O. Hebb: The Organization of Behavior, Wiley: New York; 1949[J]. Brain Research Bulletin, 1999, 50(5-6):437.

[3]Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain.[J]. Psychological Review, 1958, 65:386-408.

[4]Rumelhart D E , Hinton G E , Williams R J . Learning Internal Representations by Error Propagation[M]. 1988.

[5]Hornik K , Stinchcombe M , White H . Multilayer feedforward networks are universal approximators[J]. Neural Networks, 1989, 2( 5):359-366.

[6]Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313: 504-507.

[7]Hinton G E , Osindero S , Teh Y W . A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7):p.1527-1554.

[8]Bengio Y , Lamblin P , Popovici D , et al. Greedy layer-wise training of deep networks[C]// Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 4-7, 2006. DBLP, 2007.

[9]Cireşan, Claudiu D , Meier, et al. Deep, Big, Simple Neural Nets for Handwritten Digit Recognition.[J]. Neural Computation, 2010.

[10]Graves, Alex, Liwicki, et al. A Novel Connectionist System for Unconstrained Handwriting Recognition.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2009.

[11]Mohamed A , Dahl G E , Hinton G . Acoustic Modeling Using Deep Belief Networks[J]. IEEE Transactions on Audio Speech & Language Processing, 2011, 20(1):14-22.

[12]Mohamed A R , Sainath T N , Dahl G E , et al. Deep Belief Networks using discriminative features for phone recognition[C]// IEEE International Conference on Acoustics. IEEE, 2011.

[13]Dahl G E , Yu D , Deng L , et al. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition[J]. IEEE Transactions on Audio Speech & Language Processing, 2011, 20(1):30-42.

[14]山 与变革[J].科技导报,2016,34(14):60-70.

[15]D. Chen, X. Cao, F. Wen and J. Sun, "Blessing of Dimensionality: High-Dimensional Feature and Its Efficient Compression for Face Verification," 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, 2013, pp. 3025-3032, doi: 10.1109/CVPR.2013.389.

[16]Taigman Y , Yang M , Ranzato M , et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2014.

[17]Sun Y , Wang X , Tang X . Deep Learning Face Representation by Joint Identification-Verification[J]. Advances in neural information processing systems, 2014, 27.

[18]Deng J , Dong W , Socher R , et al. ImageNet: A large-scale hierarchical image database[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2009.

[19]Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (June 2017), 84–90.

[20]鲁恒润,李强,杨文旺.基于改进YOLOv3的锌阴极板残留物图像识别方法[J].矿冶,2021,30(01):74-79.

[21]于凌涛,夏永强,闫昱晟,王鹏程,曹伟.利用卷积神经网络分类乳腺癌病理图像[J/OL].哈尔滨工程大学学报:1-6[2021-02-24].http://kns.cnki.net/kcms/detail/23.1390.U.20210209.1511.004.html.

[22]高宸,李勇,金德鹏.基于图神经网络的视频推荐系统[J/OL].中兴通讯技术:1-8[2021-02-25].http://kns.cnki.net/kcms/detail/34.1228.TN.20210120.1010.002.html.