

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

日期: 年 月 日

深度学习模型研究及应用

摘要

深度学习技术自上世纪四十年代就提出，但是一直到二十一世纪才逐渐露出锋芒，在许多领域给出令人惊讶的结果。本文从深度学习的历史作为切入点，介绍了深度学习的研究意义以及目前相关的研究成果。对深度学习领域经典的基础网络 CNN, RNN, GAN 做了详细的阐释，后续提出的各种神经网络很多都是基于这些基础网络做的改进。文章基于国内百度自主研发的深度学习框架 Paddle，将 LeNet 和 AlexNet 应用于 iChallenge-PM 数据集上，经实验对比，原始 LeNet 在手写数据集 MNIST 上能取得 92% 的分类准确率，但在图片尺寸相对较大的 iChallenge-PM 数据集上仅有 53% 的准确率。为解决 LeNet 应用于大尺寸图片上面的不足，通过修改激活函数，增加隐层等方式。最终将原始网络结构的 Sigmoid 激活函数修改为 ReLU 函数，从而有效的解决了由图片尺寸过大引起的梯度弥散问题，结果表明，基于改进的 LeNet 的方法提升了分类模型的分类精度，较原始的 LeNet 模型提升了 40%。

关键词：深度学习；Paddle；CNN；LeNet；AlexNet

Research and application of deep learning model

ABSTRACT

Deep learning technology has been put forward since the 1940s, but it is not until the 21st century that it gradually shows its edge and gives surprising results in many fields. From the history of deep learning as a starting point, this paper introduces the research significance of deep learning and the current related research results. This paper makes a detailed explanation of the classic basic networks CNN, RNN and GAN in the field of deep learning. Many of the subsequent neural networks are based on these basic networks. Based on Baidu's self-developed deep learning framework, Paddle, this paper applies LeNet and AlexNet to the iChallenge-PM dataset. Through experimental comparison, the original LeNet can achieve 92% classification accuracy on the handwritten dataset MNIST, but only 53% accuracy on the iChallenge-PM dataset with relatively large image size. In order to solve the deficiency of LeNet applied to large-size pictures, the activation function is modified and the hidden layer is added. Finally, the Sigmoid activation function of the original network structure is modified to ReLU function, so as to effectively solve the gradient dispersion problem caused by too large image size. The results show that the classification accuracy of the classification model based on the improved LeNet method is improved by 40% compared with the original LeNet model.

Key words: Deep Learning; Paddle; CNN; LeNet; AlexNet

目录

1. 绪论.....	6
1.1. 深度学习模型算法发展历史及研究意义.....	6
1.2. 国内外研究现状.....	8
1.3. 论文组织架构.....	9
2. 主流的深度学习算法模型.....	9
2.1. 卷积神经网络.....	9
2.2. 循环神经网络.....	14
2.3. 生成对抗网络.....	16
3. 深度学习算法模型的应用.....	18
3.1. 百度研发的开源深度学习平台——飞浆.....	18
3.2. iChallenge-PM 数据集.....	19
3.3. LeNet 应用于眼疾数据 iChallenge-PM.....	20
3.4. 基于 LeNet 的改进.....	21
3.5. AlexNet 应用于眼疾数据 iChallenge-PM.....	24
4. 总结.....	25
5. 展望.....	25
6. 致谢.....	26
参考文献.....	27

1. 绪论

1.1. 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多大学 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，比如说会话场景中“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。

1.2. 国内外研究现状

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。

1.3. 论文组织架构

本文的组织结构分为六个章节，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。

第四章：总结。对本文的内容做一个总结。

第五章：展望。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。

第六章：致谢。

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，其网络结构如图 1 卷积神经网络结构所示。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 3×3 的图像矩阵转换为 9×1 的矢量，如图 2 图矩阵转矢量，使用该方法几乎没有准确性可言。

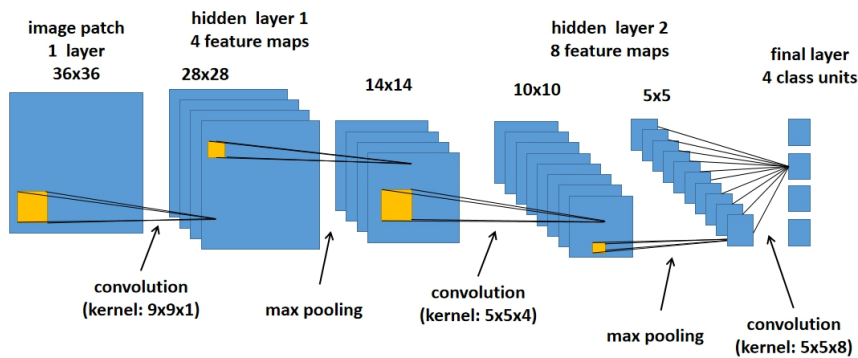


图 1 卷积神经网络结构

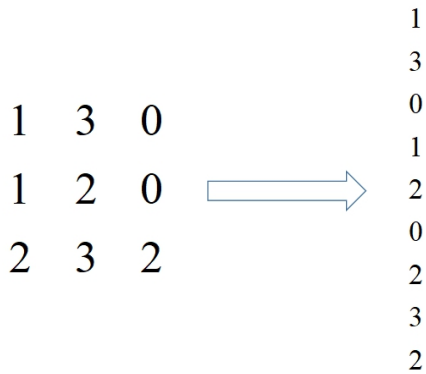


图 2 图矩阵转矢量

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。在图 3 RGB 图像中，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。

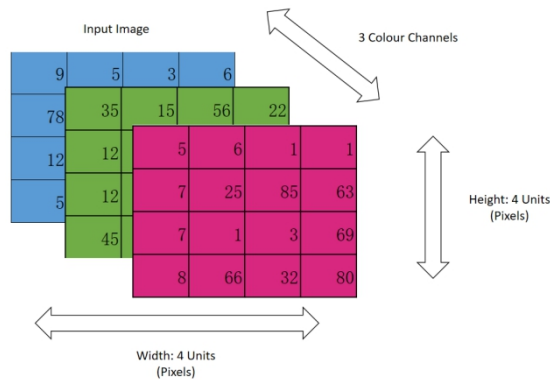


图 3 RGB 图像

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 3×3。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：

$$5(\text{Height}) \times 5(\text{Breadth}) \times 1(\text{Number_of_channels, eg. RGB})$$

图 4 卷积操作一的绿色部分类似于我们输入的 5×5×1 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 K ，在图 4 中以黄色

表示。我们选择 K 为 $3 \times 3 \times 1$ 的矩阵：

$$K = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

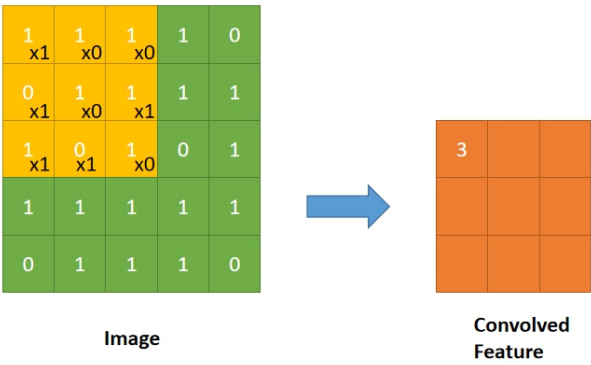


图 4 卷积操作一

卷积核乘上图像的像素值并求和，每次向右位移一个长度，如图 5 卷积操作二。

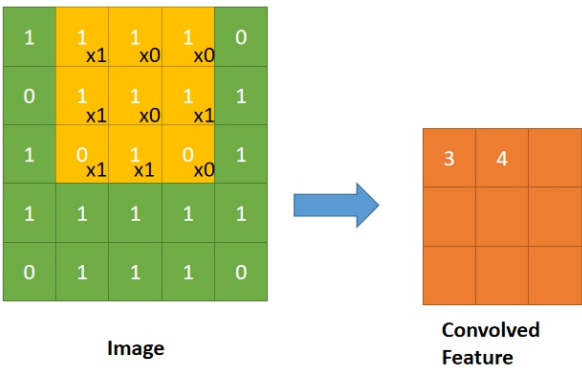


图 5 卷积操作二

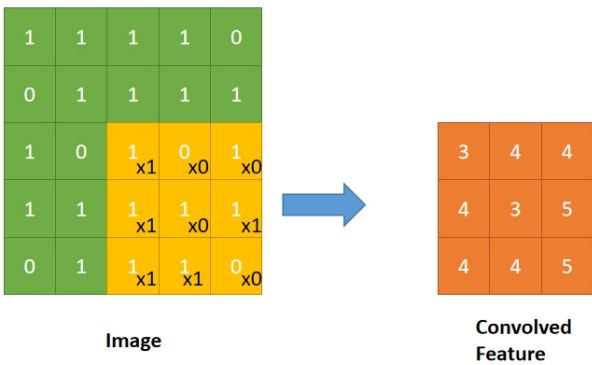


图 6 卷积操作三

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，如图 6 卷积操作三。

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神经网络。

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

图 7 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。

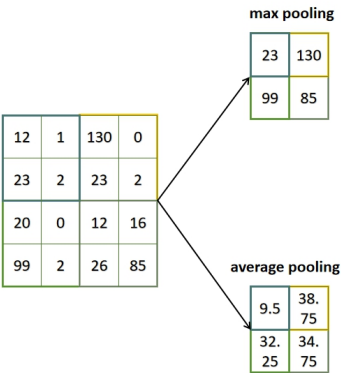


图 7 池化

卷积层和池化层一起形成了卷积神经网络的隐层。根据图像的复杂度，可以增加这种层的数量来捕获更低级的细节，但是相应的会以更多的算力作为代价。当完成上述的过程，模型已经可以捕获图像中的特征信息。下面，我们将整理最终的输出并将其输入常规的神经网络进行分类。

全连接层。添加全连接层是学习卷积层输出高级特征的非线性组合的简单方法。全连接层会学习该空间中可能存在的非线性函数。如图 8 SoftMax，我们已经

将输入的图像转换为适用于多层感知机的列向量。展平的输出被输入到前馈神经网络，在每次训练迭代中应用反向传播。在每一轮的训练中，改模型能够区分图像中的主要特征和某一些低级的特征，并使用 *Soft max* 分类技术对其进行分类。

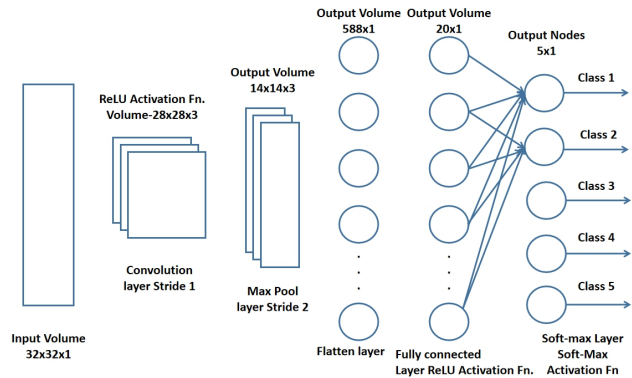


图 8 SoftMax

2.2. 循环神经网络

前馈神经网络 (Feed-Forward Nerual Network) 的结构如图 9 前馈神经网络结构所示，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。

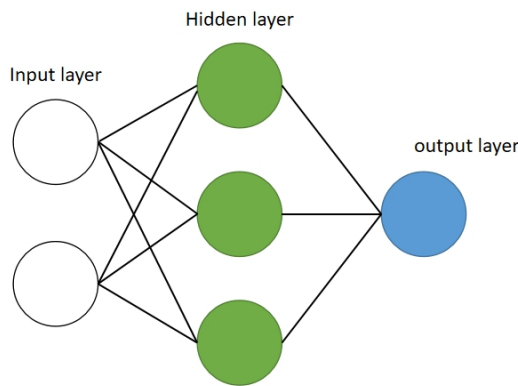


图 9 前馈神经网络结构

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。其网络结构原理如图 10 循环神经网络结构所示，图片中，有许多不同的组件，其中最重要的是：

- X ：输入，它可以使句子中的单词或者其他类型的序列数据；
- O ：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；
- h ：RNN 中主要的模块，它包含了网络的权重和激活函数；
- V ：表示从一个时间节点到另一个时间节点的通信。

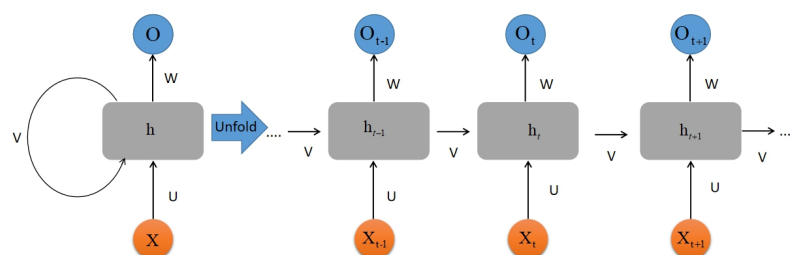


图 10 循环神经网络结构

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) 的网络结构如图 11 生成对抗网络结构所示。

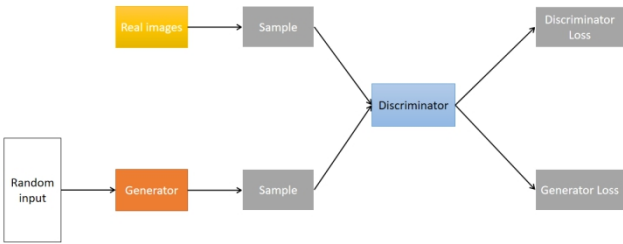


图 11 生成对抗网络结构

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，见图 12 生成器学习过程一：

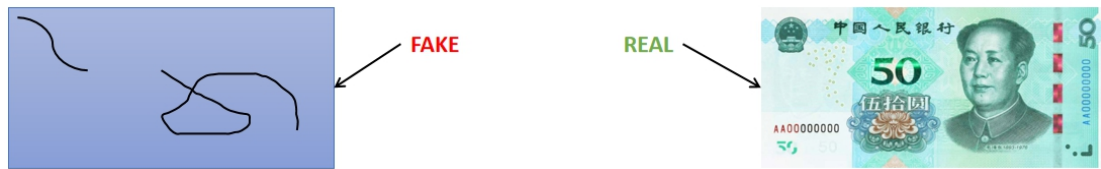


图 12 生成器学习过程一

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，见图 13 生成器学习过程二：

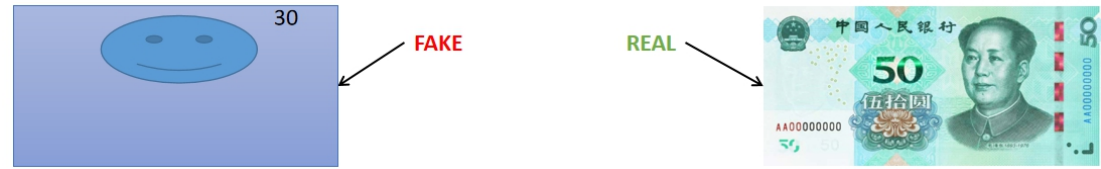


图 13 生成器学习过程二

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，如图 14 生成器学习过程三，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。

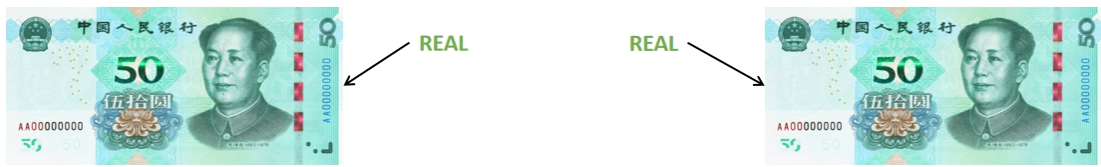


图 14 生成器学习过程三

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。

下图 15 判定模型反向传播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。

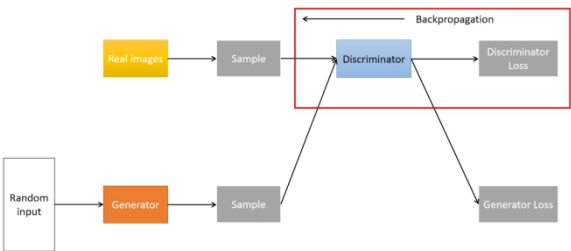


图 15 判定模型反向传播

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。

图 16 生成模型反向传播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。

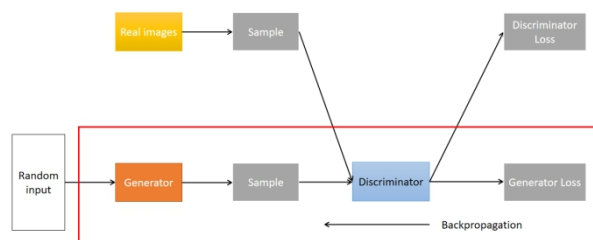


图 16 生成模型反向传播

3. 深度学习算法模型的应用

3.1. 百度研发的开源深度学习平台——飞桨

通常工程师们使用深度学习技术进行产品研发的时候，并不会从底层写起，通常都是基于某一些深度学习框架展开工作。深度学习框架处于硬件层和应用层之间。具有代表性的开源深度学习框架有 Google 研发的 Tensorflow，Facebook 团队研发的 PyTorch 等等这里就不赘述了。

tensorflow 可能是最受欢迎的工具，TensorFlow 基于 JavaScript，并配备了广泛的工具和社区资源，可简化训练和部署 ML/DL 模型。TensorFlow 需要较大编码的量，并且需要使用静态计算图进行操作。因此，首先需要定义图形，然后运行计算。如果模型架构发生任何更改，则必须重新训练模型。TensorFlow 的优势：TensorFlow 最适合开发 DL 模型和尝试深度学习架构。它用于数据集成功能，包括输入图，SQL 表和图像。

PyTorch 基于 Torch 库，其设计的主要目的是加快从研究原型到生产部署的整个过程。PyTorch 不同于 Tensorflow, PyTorch 使用动态更新的图进行操作，这意味着可以在训练过程中对模型体系结构进行必要的更改。PyTorch 的优势：它非常适合训练，构建，部署小型项目和原型。它广泛用于深度学习应用程序，例如自然语言处理和计算机视觉。

但是本文要介绍另一个深度学习平台——飞浆(PaddlePaddle)，它是国内首

个也是唯一一个功能完备的开源深度学习平台，基于百度在深度学习领域的长期积累自主研发，并在 2016 年开源^[23]。Paddle 和 PyTorch 一样都是动态图的编程方式。飞浆提供了针对服务器的在线服务(Paddle Serving)和移动端部署库(Paddle Lite)。Paddle Serving 是在线预测的部分，它可以提供深度学习的预测云服务。在移动端，Paddle Lite 支持将模型部署到各种硬件、软件操作系统之上。飞浆开源了智能视觉(PaddleCV)，智能文本处理(PaddleNLP)，智能推荐(PaddleRec)，智能语音(PaddleSpeech)四大领域 70 多个官方模型，再配合飞浆提供的迁移学习组件(PaddleHub)，开发者可以更好的将预训练模型应用到自己的细分场景。

本文将基于 iChallenge-PM 数据集，Paddle 深度学习框架，将几个图像分类领域非常经典的卷积神经网络应用到 iChallenge-PM 数据集上以解决图像分类的问题，网络分别是 LeNet，AlexNet，VGG，GoogleNet，ResNet。

3.2. iChallenge-PM 数据集

目前由于缺乏高质量的数据，用于解释眼科成像的算法的开发面临着巨大的挑战，其中数据标签不正确且没有精细的注释会成为研究人员训练高效算法的绊脚石。iChallenge-PM 数据集是百度大脑和中大眼科中心联合举办的 iChallenge 比赛所提供的数据，其中有一项数据是关于病理性近视(Pathologic Myopia，简称:PM)疾病的，也就是 iChallenge-PM。设置 iChallenge 的目的是共享高质量的眼科标注和注释成像数据，以加强不同研究人员之间的交流，促进人工智能辅助医疗诊断和图像自动分割算法的开发。人工智能有望帮助眼科医生阅读眼科图像。iChallenge-PM 数据集中有病理性近视患者的照片也有非病理性的近视患者照片，图片的标签已经蕴含其文件名中：

- 病理性近视(PM):P 字开头的文件；
- 非病理性近视(non-PM):
 - 高度近视(hight myopia):H 字开头的文件；
 - 正常眼睛(normal):N 字开头的文件。

我们将病理性近视患者的照片作为正例，标签为 1，非病理性患者的照片作为负

例，标签为 0，从数据集中选取三张照片显示出来如图 17 iChallenge-PM 数据集，并且读取照片的形状，约为 2000×2000 。

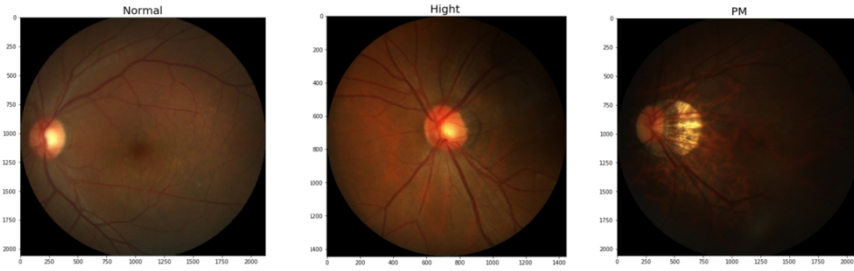


图 17 iChallenge-PM 数据集

3.3. LeNet 应用于眼疾数据 iChallenge-PM

LeNet 是最早的卷积神经网络之一，因为在手写数据集上出色的表现使它一战成名。本文使用 LeNet 不包括输入一共有七层，和 LeNet-5 的结构类似。下面详细介绍各层参数。考虑到眼疾数据集的照片尺寸比较大，本文对数据进行压缩处理，将每张图片压缩到 224×224 ，并将像素值调整至 -1 至 1 之间，下面所有网络的输入都是基于此形状。

网络结构如图 19 LeNet-5 网络结构所示。第一层卷积，输入向量维度 224×224 ，输入图片的数字通道为 3，卷积核大小 5×5 ，6 个卷积核，步长默认为 1，激活函数统一是 Sigmoid。第二层池化层使用最大池化方法，采样区域 2×2 ，步长设置为 2。第三层卷积层数字通道为 6，16 个卷积核，卷积核大小 5×5 。第四层最大池化，参数和第二层池化相同。第五层卷积层，数字通道对应上一卷积层的 16 卷积核设置为 16，120 个卷积核，卷积大小为 4×4 。那么第五层的输出神经元数量将是 $(53 - 4 + 1) * (53 - 4 + 1) * 120 = 300000$ ，第六层全连接层，输入的向量维度既为 300000，输出维度 64，第七层全连接层，输入维度 64，输出维度为 1。一共设置 5 轮训练，每轮的 batch_size 设置为 10。经过训练，LeNet 在 iChallenge-PM 数据集上的表现并不理想，准确率在 52.75% 左右，loss 始终维持在 69.18% 左右无法下降，即使增大训练轮数也无济于事。经实验，相近的参数配置，LeNet 在手写数据集上的准确率能达到 92.99%，loss 在 24.33% 左右。对比手写数据集的输入，是 28×28 ，而本例的输入维度是 224×224 ，说明 LeNet 面对图片尺寸较大的情况时，在分类任务上存在一定的局限性。

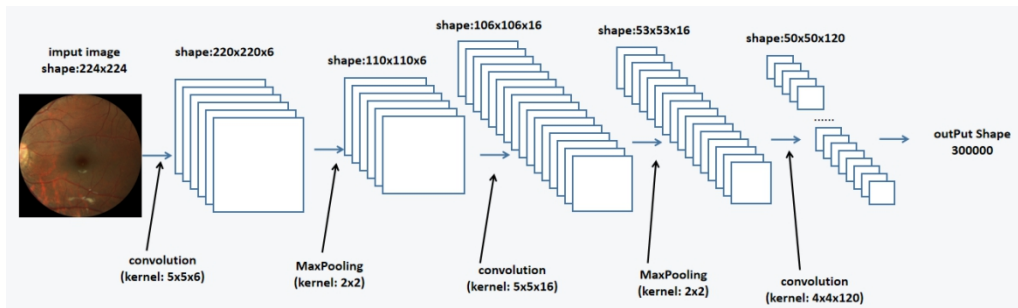


图 18 LeNet-5 网络结构

3.4. 基于 LeNet 的改进

Sigmoid、Tanh 和 ReLU 都是深度学习中常用的激活函数。对于网络中给定的界定，将其乘上节点的权重并相加在一起，此和成为节点的总激活量。然后通过激活函数对总激活量进行转换，并定义节点的特定输出或者激活节点。最简单的激活函数成为线性激活，其不用做任何变换。仅包含线性激活函数的网络非常易于训练，但是其弊端就是无法学习复杂的映射函数。那么在使用神经网络训练时，首选都是非线性激活函数，因为它们可以使网络学习到数据中更复杂的结构。传统上，Sigmoid 和 Tanh 激活函数都是收到广泛使用的。

Sigmoid 函数又称为 logistic 函数。该函数将输入值转换到 0.0 到 1.0 之间的值。远大于 1.0 的输入将被转换为 1.0，类似地远小于 0.0 的值将转换为 0.0，函数的形状为 S 形如图 19 Sigmoid 函数，在很长一段时间它都是神经网络默认的激活方式。

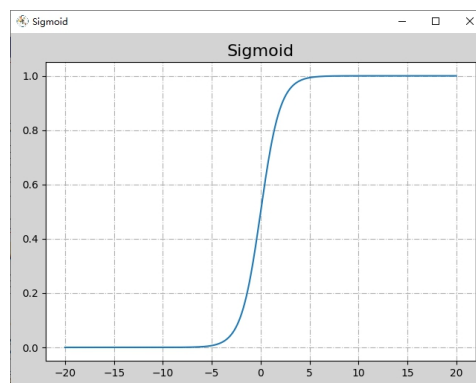


图 19 Sigmoid 函数

Tanh 函数是形状相似的非线性激活函数，如图 20 Tanh 函数。其输出值介于 -1.0 和 1.0 之间。在 1990 年代后期和整个 2000 年代，Tanh 函数比 Sigmoid 函数更受欢迎，因为使用 Tanh 的模型更易于训练，且通常具有更好的预测性能。

但是 Sigmoid 和 Tanh 有一个普遍的问题就是，这两个函数对大值捕捉为 1.0 而小值捕捉为 -1.0 或 0.0，并且这些功能仅对中心点附近的变化非常敏感，Sigmoid 为 0.5，Tanh 为 0.0。

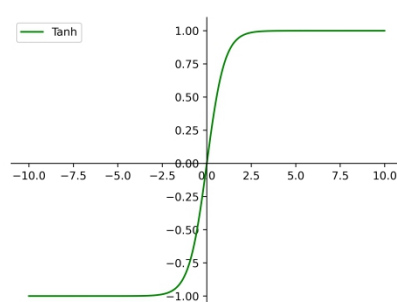


图 20 Tanh 函数

使用这些非线性函数的大型网络中的深层无法接收到有用的梯度信息，错误会通过网络传播回来，并被用于权重更新。给定了所选激活函数的导数，误差的传播量随传播层数的增加而显著降低，这种现象被称为梯度消失问题，是训练多层网络的一个问题。虽然使用非线性激活函数可以使神经网络学习复杂的映射函数，但是它们会使神经网络无法与学习算法一起工作。

为了使用带有误差反向传播的随机梯度下降来训练神经网络，需要一个看起来像线性函数的激活函数，但是实际上是非线性的函数(为了学习数据中复杂的映射关系)。该函数还必须对激活量总和的输入提供更高的灵敏度，并避免过于容易饱和。这个函数便是 ReLU 函数，如图 21 ReLU 函数。采用 ReLU 可以认为是深度学习革命中的几个里程碑之一。ReLU 函数是一种简单的计算，可以直接将输入值作为返回，如果输入小于 0.0 或更小，则返回值为 0.0，对于大于零的值，该函数是线性的，这就意味着在使用反向传播训练神经网络时，它有线性激活函数的许多特性，但它是一个非线性函数，所以输入的负值始终输出为零。

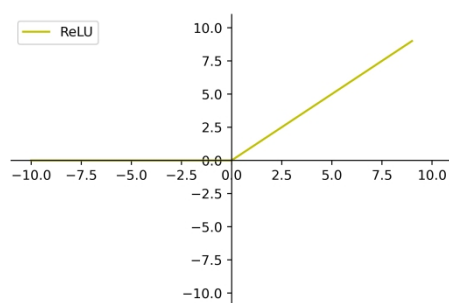


图 21 ReLU 函数

笔者观察到，上面所用的 LeNet 网络结构，每一层的激活函数用的都是 Sigmoid，观察到 loss 始终无法下降，猜测是因为梯度弥散的原因造成的影响，如果使用 relu 的效果也许会比 Sigmoid 更好。经试验：5 轮训练后，准确度能达到 93%，loss 降低到 16%！多么惊人的进步。10 轮训练后，准确率是 92.75%，loss 是 17.77%。以此可以说明 relu 相较于 Sigmoid 更适用于作为图像卷积的激活函数。

而原 LeNet 的网络结构相对于较大型的数据集效果不是很好，那这个能不能通过增加隐层的方法从而使网络学到更多图像中的特征。于是笔者在 3.3 网络结构的基础之上，再增加了两个隐层。在原第五层卷积层后面增加一层最大池化层，设置和之前的一样，第七层卷积层，数字通道为 120，卷积核个数 240，卷积核大小为 6×6 ，那么第八层全连接层的输入变为 $20 \times 20 \times 240 = 96000$ ，输出维度改为 32，第九层全连接层输入 32 维，输出 1 维，经试验，5 轮训练后，准确度能达到 94.00%，loss 降低到 14.06%，10 轮训练以后，准确度降低了一些 91.50%，loss 升到到 22.64%。

网络	激活函数	训练轮数	数据集	准确率	损失
LeNet-7	sigmoid	5	iChallenge-PM	52.75%	69.18%
LeNet-7	sigmoid	10	iChallenge-PM	52.75%	69.16%
LeNet-7	relu	5	iChallenge-PM	93.40%	14.96%
LeNet-7	relu	10	iChallenge-PM	93.25%	17.74%
LeNet-9	sigmoid	5	iChallenge-PM	52.75%	69.17%
LeNet-9	sigmoid	10	iChallenge-PM	52.75%	69.16%
LeNet-9	relu	5	iChallenge-PM	94.50%	14.06%
LeNet-9	relu	10	iChallenge-PM	91.50%	22.64%

表 1 LeNet 应用于 iChallenge-PM

3.5. AlexNet 应用于眼疾数据 iChallenge-PM

AlexNet 也就是 2012 年获得 ImageNet 图像分类比赛冠军的模型。AlexNet 网络结构由八层组成：五个卷积层和三个完全连接层。但这并不是 AlexNet 与众不同的原因。下面是 AlexNet 卷积神经网络所使用的一些创新点：

- RELU Nonlinearity。AlexNet 使用 ReLU 代替当时的标准 tanh 作为激活函数。ReLU 的优势在于训练时间。使用 ReLU 的 CNN 在 CIFAR-10 数据集上达到 25% 错误率的速度比使用 tanh 的 CNN 快六倍。
- Multiple GPUs。多 GPU 分布式训练模型，在 ImageNet 的比赛上，Hinton 的研究团队将模型的神经元一半放在一个 GPU 上，另一半放在另一个 GPU 上，以此来进行多 GPU 训练，这不仅意味着可以训练更大的模型，而且可以减少训练时间。
- Overlapping Pooling。作者引入重叠池化^[19]时，发现误差能减少约 0.5%。

AlexNet 一共有 6000 万个参数，那么面对如此巨大的参数量，AlexNet 使用下面两种方法来抑制过拟合：

- Dropout。Dropout 技术是以预定概率，例如 50% 来(临时)删除神经元，可以有效增强模型的泛化能力，抑制过拟合，但是相应的会增加模型的训练时间。
- Data Augmentation。作者使用数据增广的方法来使训练数据更多样化，数据增广通常是在保留原始数据的标签情况下，对数据进行平移，水平映射，旋转，裁剪，缩放等操作，以此产生一系列和原始图片相似但是又完全不一致的样本。通过这种方式，可以有效避免模型过度依赖某一些属性，从而达到抑制过拟合的效果。

本文将 AlexNet 模型应用于 iChallenge-PM 数据集上，看相对于 LeNet 模型能不能提升分类精度。下面介绍本文所做的 AlexNet 网络参数。输入依然是 224×224 ，第一层卷积数字通道为 3，卷积核大小为 11，一共 96 个卷积核，步长设置为 4，padding 填充设置为 5，激活函数全部是 RELU。第二层最大池化，采样区域 2×2 ，步长为 2。第三层卷积数字通道对应上一卷积层的 96，一共 256 个卷积核，卷积核大小为 5，步长使用 1，padding 设置为 2。第四层最大池化，设置和第二层一样，下面所提到的池化层均与第二层一样设置。第五层卷积层数字通道为 256，一共 384 个卷积核，卷积核大小为 3，步长为 1，padding 为 1。第六层卷积层数字通道为 384，一共 384 个卷积核，大小为 3，步长和 padding

均为 1。第七层卷积层数字通道为 384，一共 256 个卷积核，大小为 3，步长和 padding 均为 1。第八层最大池化。第九层全连接层，输入维度 12544，输出 4096，设置 dropout 系数为 0.5。第十层全连接层，输入维度 4096，输出维度为 1，dropout 系数设置为 0.5。经过 5 轮训练，AlexNet 网络在 iChallenge-PM 数据集上能达到 91.75% 的准确率，loss 为 22.45%。经过 10 轮训练，准确率可以达到 93.25%，loss 降低到 17.58%，由此可见在图片尺寸较大的情况下，AlexNet 相较于原作者提出的 LeNet 结构有更好的分类效果。

4. 总结

本文从深度学习的历史作为开篇，简单介绍了深度学习的原理和研究意义以及国内外研究现状。随后重点介绍了三个非常经典的神经网络，分别是卷积神经网络、循环神经网络以及生成对抗神经网络，后来发表的许多模型如 LeNet，AlexNet，VGG 等等都是基于这些经典模型的思想去改进的。随后使用了 paddle 深度学习框架进行神经网络的搭建，将 AlexNet 和 LeNet 应用到图像分类上，在 iChallenge-PM 和手写数据集上面验证了网络的准确率。并基于 LeNet 原作者的网络将激活函数修改为更适用于图像卷积的 ReLU 函数，且增加两层隐层，从而使得 LeNet 模型的精度有所提升。

5. 展望

深度学习在计算机视觉领域，推荐，语音，自然语言处理都取得了不俗的成绩，文中所做的应用也只是抛砖引玉，读者如果对其感兴趣可以使用 Paddle 框架进行非常便捷的开发应用。实际上深度学习还是存在许多问题影响其应用。所以，未来深度学习的研究放行应该集中在下面几个方向：

(1) 模型参数选择。神经网络模型应用于各种不同的任务时关键问题就是如何选择最优的网络结构，比如隐层的数量，卷积核的大小，池化层数，步长，优化函数等等各种参数组合起来就是非常庞大的数量，而且深度学习模型还是以黑盒的方式运转，无法给予研究人员启发，所以目前主要还是考专业人员根据相关经

验和不断的尝试得出相对较优的参数组合。

(2)深度学习内部可视化。目前深度学习的内部仍是黑箱子，比如 PyramidBox 模型就是用了 VGG16 的预训练参数进行的参数初始化，而 VGG16 的意思就是网络有 13 个卷积层，3 个全连接层，中间涉及的参数数量非常庞大，训练过程无法解释，这决定了网络的输出不够有解释性，那么在相对比较严谨的领域(如医疗诊断)就会不够说服力。那么为了能够将深度学习更好的应用到各行各业，对深度学习内部可视化的研究也是非常重要的。

6. 致谢

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>.
- [23]马艳军, 于佃海, 吴甜, 王海峰. 飞桨: 源于产业实践的开源深度学习平台[J]. 数据与计算发展前沿, 2019, 1(05):105-115.