卷积神经网络研究综述

姓名：叶珺明

学号：19335253

课程：智能算法及应用

2022 年 4 月 16 日

卷积神经网络研究综述

叶珺明

摘要

自 McCulloch 和 Pitts 提出了首个带有神经元的数学模型以来，许多研究者在这一领域进行了深度的探索。随着人工神经网络的迭代更新，一些特殊的人工神经网络如卷积神经网络已经较为成熟。卷积神经网络已经广泛应用于我们的生活中，如人脸识别，自动驾驶等。基于前研究者对卷积神经网络的研究成果和自己所学对卷积神经网络进行探讨，了解网络在发展过程中的模型结构更改、算法更迭，对卷积神经网络有了更为全面的理解和新的想法。

关键词：卷积神经网络；网络模型；深度学习

1 引言

卷积神经网络（Convolutional Neural Network, CNN）属于特殊的人工神经网络（Artificial Neural Network，ANN）。在探讨卷积神经网络之前，我们需要对人工神经网络有一定了解。人工神经网络这一概念最早于 1943 年由心理学家 McCulloch 和数理逻辑学家 Pitts 在《A logical calculus of the ideas immanent in nervous activity》提出[1]，开创了人工神经网络研究的时代。他们提出了 MP 模型，首次融合了人工神经元的数学模型。后来，美国计算机科学家 Rosenblatt 提出感知器（Perceptron）的概念[2]，但随后单层感知机模型被 Minsky 证明不能解决简单的异或等线性不可分问题[3]。

在 1975 年，Werbos 提出误差反向传播算法（Back Propagation Network，BP 算法）[4]，1986 年 Rumelhart 等人[5]利用该算法训练网络，成功解决了单层感知机无法解决的问题。基于 Hubel 和 Wiesel 在猫脑视觉皮层研究总结的感受野理论，Fukushima 提出了新认知机（Neocognitron）和权值共享的卷积神经层[6]，被视为卷积神经网络的雏形。

在 1989 年，LeCun 结合 BP 算法和卷积神经层发明了卷积神经网络[7]，即是 LeNet 的最初版本，该网络被运用到邮局的手写字符识别系统中。后来在 1998 年，LeCun 又提出卷积神经网络的经典网络模型 LeNet-5[8]，增加了卷积层和全连接层，更进一步提高了手写字符识别的正确率。

随着对神经网络和机器学习融合研究的深入，在有限计算力的环境下，支持向量机模型的简单结构的训练速度更快更容易实现。而 BP 算法的神经网络训练对象要求是有标注的数据集，训练过程中在隐藏层之间的来回传播导致训练速度过慢，网络结构容易陷入局部最优。此时的卷积神经网络还不适合在大规模数据集上训练，受硬件等的限制。

进入新世纪后，卷积神经网络迎来了第一个突破点，是在工程上，借助 GPU 的高计算力，CNN 的实现速度较在 CPU 上提高了 4 倍。到 2012 年，卷积神经网络出现了历史突破，AlexNet[9]在 ILSVRC 比赛中以优异的表现获得广泛关注，它由 Alex Krizhevsky 等人设计实现。AlexNet 网络模型标志着神经网络的复苏和深度学习的崛起，计算技术的提高更进一步推动了深度网络的发展。

在这之后，有更多的卷积神经网络模型被提出，主要分为四个方向：

1. 网络加深：VGG-16、VGG-19 MSRANet

2. 增强卷积模块功能：NIN、GoogLeNet、Inception V3

3. 从分类任务搭配检测任务：SPP-Net、R-CNN、Fast R-CNN

4. 增加新的功能单元：Inception V2、FCN、STNet、CNN+RNN

卷积神经网络的另一个历史突破是在 2015 年，由 He 等人提出的残差网络 ResNet[10]，并在当年的 ILSVRC 比赛中取得了冠军，首次超过人类水准。为追求网络模型的准确率，我们会增加神经网络深度，而随着网络深度的增加，模型准确率先提高后出现大幅度降低，HE 等人将这一现象称为“退化”。网络退化的原因是随着网络深度增加，梯度在传播过程中逐渐消失，无法对前面网络的权重进行有效调整。针对这一问题，他们提出了“快捷连接”（Shortcut Connection），在神经网络中增加线性转换分支，平衡了非线性转换和线性转换，使得神经网络深度突破 100 层，准确率也得到提升。2017 年 Huang 等人提出的 DenseNet[11]，参数和计算成本都比 ResNet 更少但性能更优。原因在于：

- 建立的前面所有层与后面层的密集连接
- 通过特征在通道上的连接来实现特征重用。因此，DenseNet 的网络结构更窄，使用的参数更少，特征和梯度传递的效果更好，更容易训练。

总的来说，CNN 的发展如图 1所示，卷积神经网络的发展经历了网络结构的改进和完善，算法的选择，训练的技巧，也得益于计算力的发展，容许更大的神经网络训练。如今卷积神经网络较为成熟，被广泛地应用于人脸识别，自动驾驶，智慧医疗等领域。

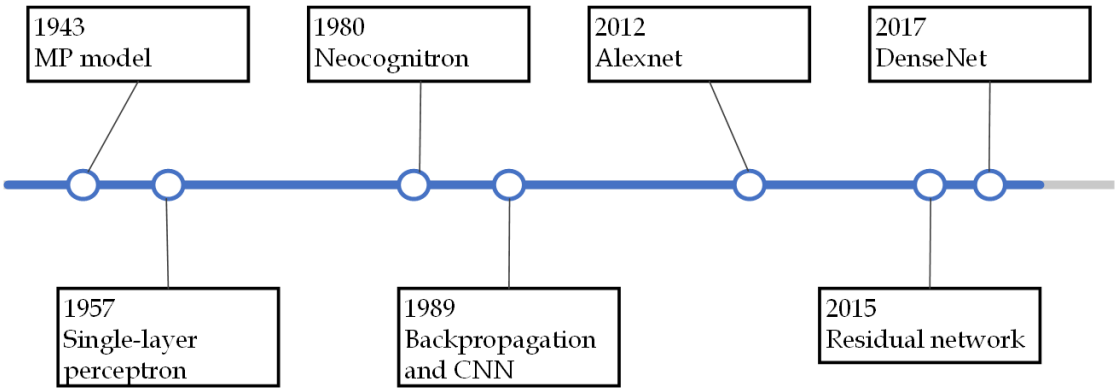

图 1: 卷积神经网络发展历程

2 算法介绍

我们将介绍 CNN 的发展过程中的几个经典模型，LecNet[7]，LeNet-5[8]最初将卷积神经网络模型运用到实际生活中，然后是 AlexNet[9]，一个重要的历史突破，后来不同发展方向中的 VGG[12], Inception V2[13]，以及解决梯度消失问题的 ResNet[10]和性能更优的 DenseNet[11]。

2.1 LeNet

LetNet 有 5 层如图 2，两个卷积层，两个池化层和一个全连接层。输入 28×28 的图像，先经过一个卷积层，得到 4 通道 24×24 的 feature map，进行均值池化后，再经过卷积层得到 12 通道 8×8 的 feature map，再次均值池化后通过一个全连接层提取特征，输出十个类别，数字 0-9 的概率。LeCun 结合反向传播算法与权值共享的卷积神经层发明了 LeNet，通过局部连接降低参数，权值共享的卷积核将卷积核参数共享到整个图片上，卷积操作能够利用图片空间上的局部相关性自动提取特征，解决传统机器学习手工提取特征的问题。

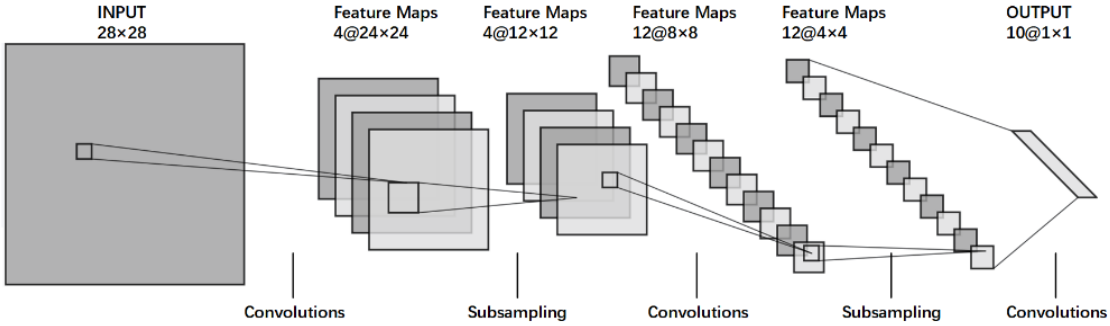

图 2: LeNet 网络结构

2.2 LeNet-5

LeNet-5 有 7 层如图 3，两个卷积层，两个池化层和三个全连接层。输入 32*32 的图像，同样也是经过卷积操作后再进行均值池化得到 feature map 后，再进行一次卷积操作和均值池化得到 16 通道 5×5feature map，最后经过三个全连接层提取特征。需要注意的是，从 S2 到 C3 的映射并非是一一对应的，通过调整 C3 中 feature map 的输入打破不同 feature map 间的对称性，增强特征的鲁棒性。在当时的计算环境下，网络并不深，作者选择了 tanh 作为激活函数，为了更快收敛。LeNet-5 比 LeNet 的网络模型更深，更多的 feature map 和增加的全连接层使得字符识别错误率进一步降低。此后，许多与 CNN 相关的研究发展都是几乎都是基于 LeNet-5 的网络模型开展的。

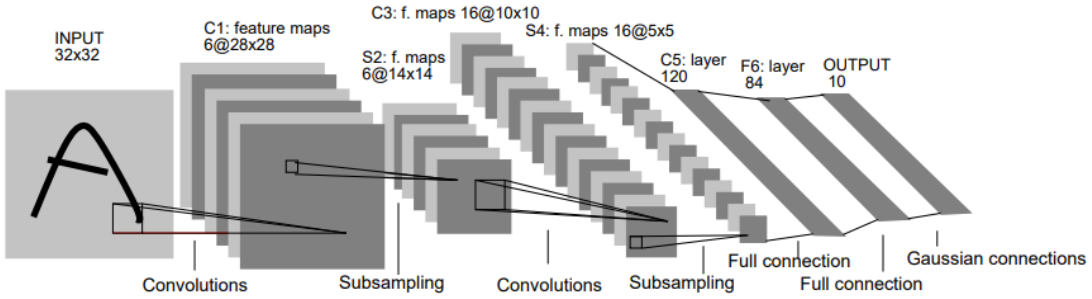

图 3: LeNet-5 网络结构

2.3 AlexNet

AlexNet 有 8 层如图 4，五个卷积层和三个全连接层。输入 3 通道 224×224 的图像，使用 96 个卷积核，使用两片 GPU 并行加速，分别计算 48 个卷积核，经过 5 个卷积层和最大值池化，得到两个 128 通道 13×13 的 feature map，最后经过三个全连接层，得到 1000 个分类。相比于 LeNet-5，AlexNet 在网络结构上没有太大更改。

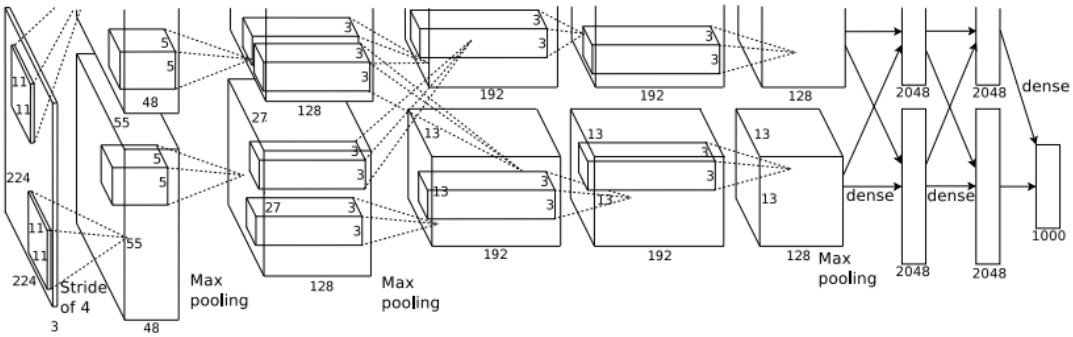

图 4: AlexNet 网络结构

在该网络中，有几个特点：

1. 所有卷积层都选用 ReLU 作为激活函数而不是 tanh，模型的收敛速度更快，解决了网络较深时梯度消失的问题。相同网络下要达到 25% 训练错误率，ReLU 要比 tanh 快 6 倍，如图 5。

2. 采用最大值池化而不是均值池化，最大值池化更能提高所提取特征的鲁棒性。

3. 经过 ReLU 后的归一化处理为局部响应归一化（Local Response Normalization，LRN），经过 LRN 处理后能够有效降低错误率，提高模型的泛化能力。

4. 采用 overlapping Pooling，池化窗口大小小于步长使得相邻池化窗口有重叠部分，有利于避免训练数据过拟合。

5. 在避免过拟合方面，采取了数据增量（Data augmentation）和 dropout 两种方法：

   • 数据增量中：通过镜像反射和随机剪裁增大数据量; 利用 PCA，随机改变训练样本 RGB 通道强度值。

   • Dropout 方法：具有 0.5 的概率将隐藏神经元设置输出为 0，既不参与前向传播也不参与反向传播。每次输入一个样本时，神经网络就尝试了一个新的结构，这些结构共享权重，减少了神经元适应的复杂性。

6. 使用了多 GPU 训练，提高了训练规模和训练速度。

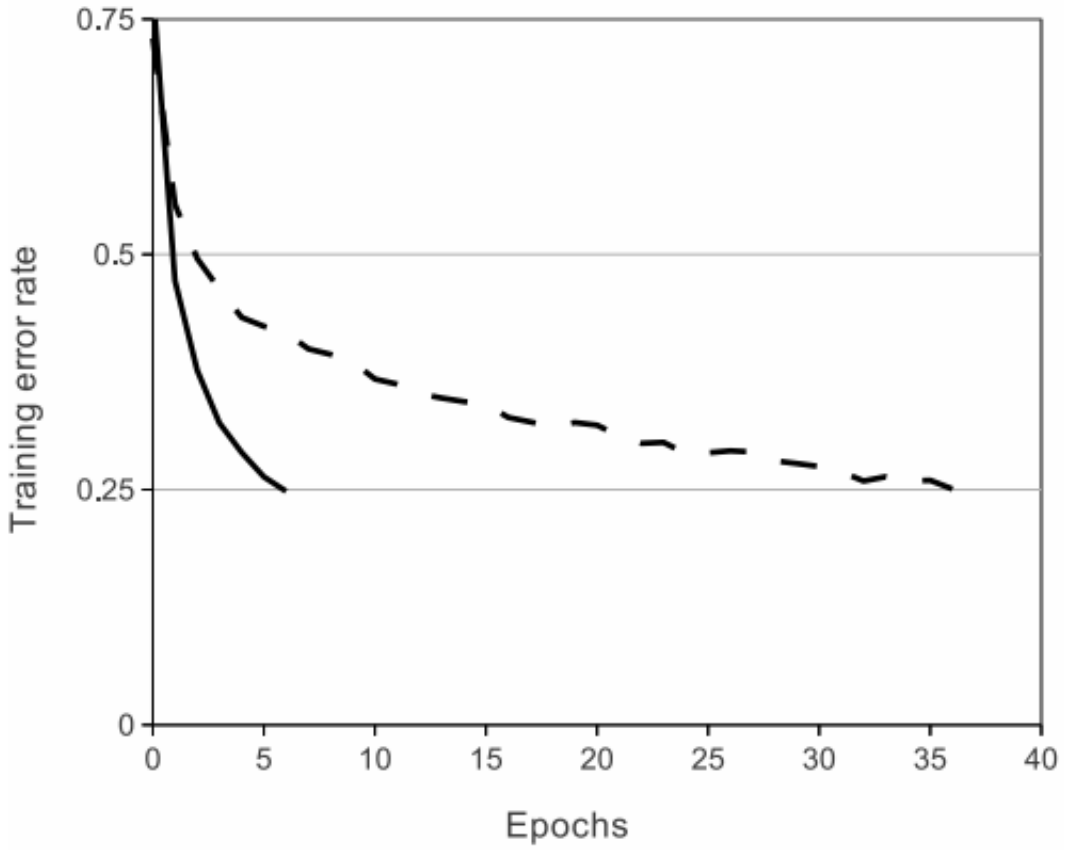

图 5: 收敛速度对比

2.4 VGG16

VGG 由 6 部分组成如图 6，详细结构如图 7. 有 5 个 block 和 3 个全连接层。一个 block 由卷积层和最大池化层组成。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

图 6: VGG 组成部分-其中 D 列代表是 VGG16 的结构，E 是 VGG19 的结构

VGG 模型的特点是结构简单：

1. 卷积层均采用相同的卷积核参数，且卷积核较小。
2. 池化层采用相同的池化参数。
3. 基于前两点，堆叠卷积层和池化层构成 block，有利于加深网络而计算开销缓慢增加。

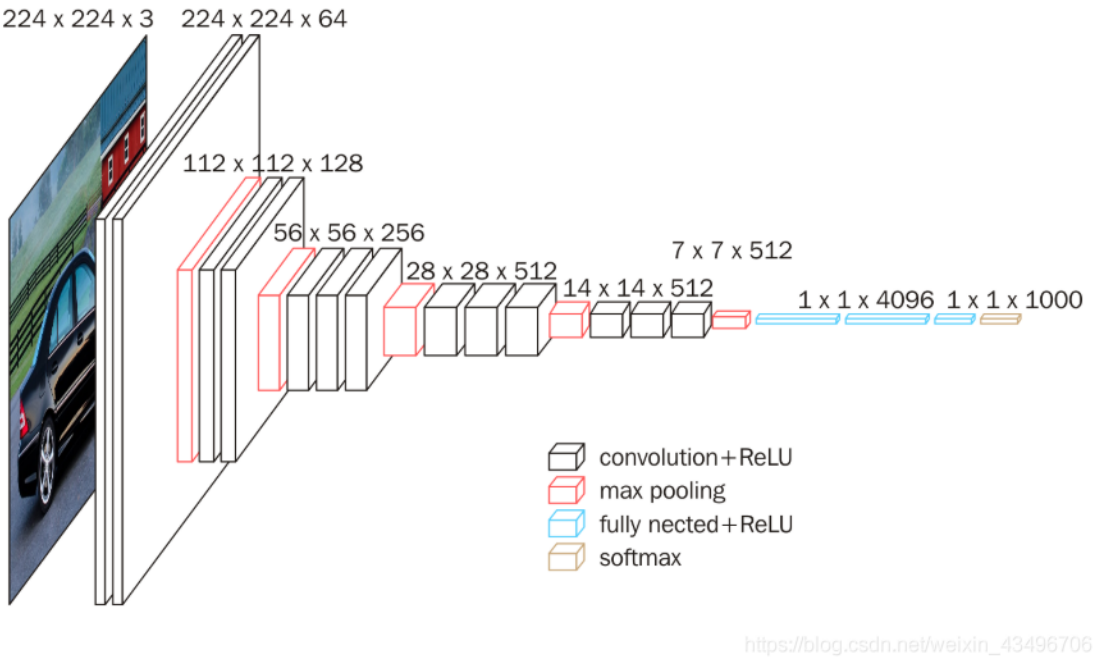

图 7: VGG 网络结构

2.5 Inception V2

Inception 的网络架构中:

1. 使用了 Inception 模块，如图 8，满足网络在训练过程中面对不同输入自动调整参数选择卷积池化操作，先进行 1×1 卷积操作能够减少参数量，增加非线性

2. 引入 Batch。Normalization，主要解决了训练过程中数据分布发生变化的问题。在 BN 算法下，可以设置较大的学习率，网络对参数初始化的依赖降低，训练的收敛速度大大提高了；省去了局部响应归一化、Dropout 和 L2 正则化的步骤，BN 算法也能归一化数据，避免过拟合和提高网络的泛化能力。

3. 参考了 VGG 网络模型，用小卷积核代替大卷积核，降低参数数量，加速运算。

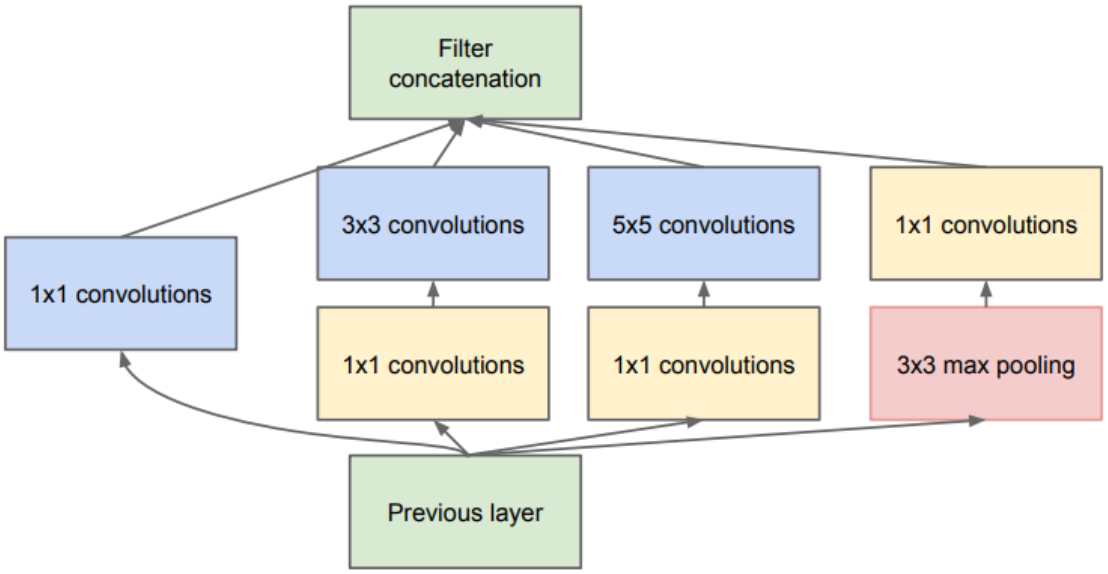

图 8: Inception 模块

2.6 ResNet

ResNet 使用了带有快捷连接（shortcut connection）的 ResNet 的构建块，如图 9.。我们期望的映射为 $H(x)$，当前输入的是 $F(x)$，通过 block，$F(x)$ 需要学习的仅是 $H(x)-x$ 的差，最终用 $F(x)+x$ 来拟合 $H(x)$。

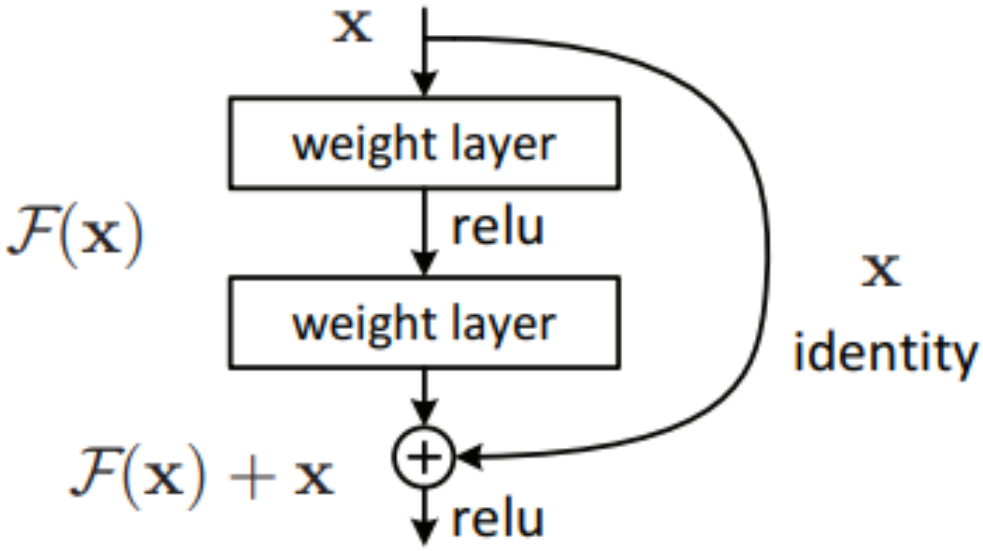

图 9: ResNet 基本模块

在 ResNet 网络架构中：

1. 网络中多了很多线性分支，即 shortcut 路径，如图 10，堆叠两个卷积层组成一个 block。所有的 Residual。Block 都没有池化层，通过卷积操作的 stride 来实现降采样
2. 最终通过平均池化提取特征而不是全连接层。

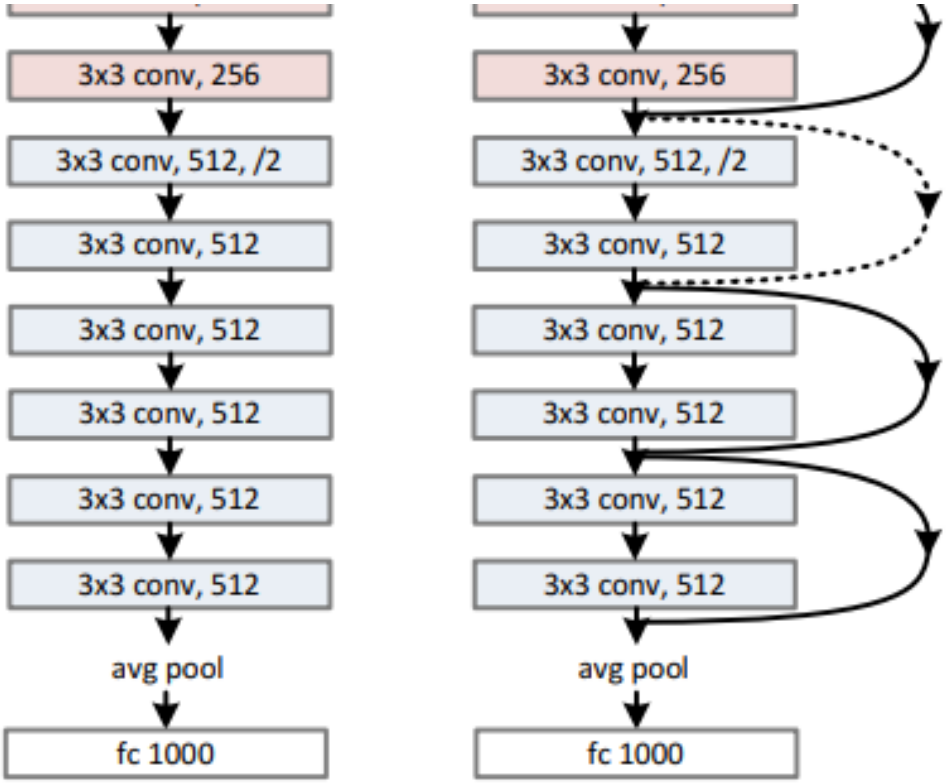

图 10: ResNet shortcut 路径

2.7 DenseNet

DenseNet 的网络结构由多个 Dense Block 构成，Dense Block 中将前面所有层的输出作为下一层的输入，如图 11，使得特征的连接和梯度的传递比 ResNet 要强。

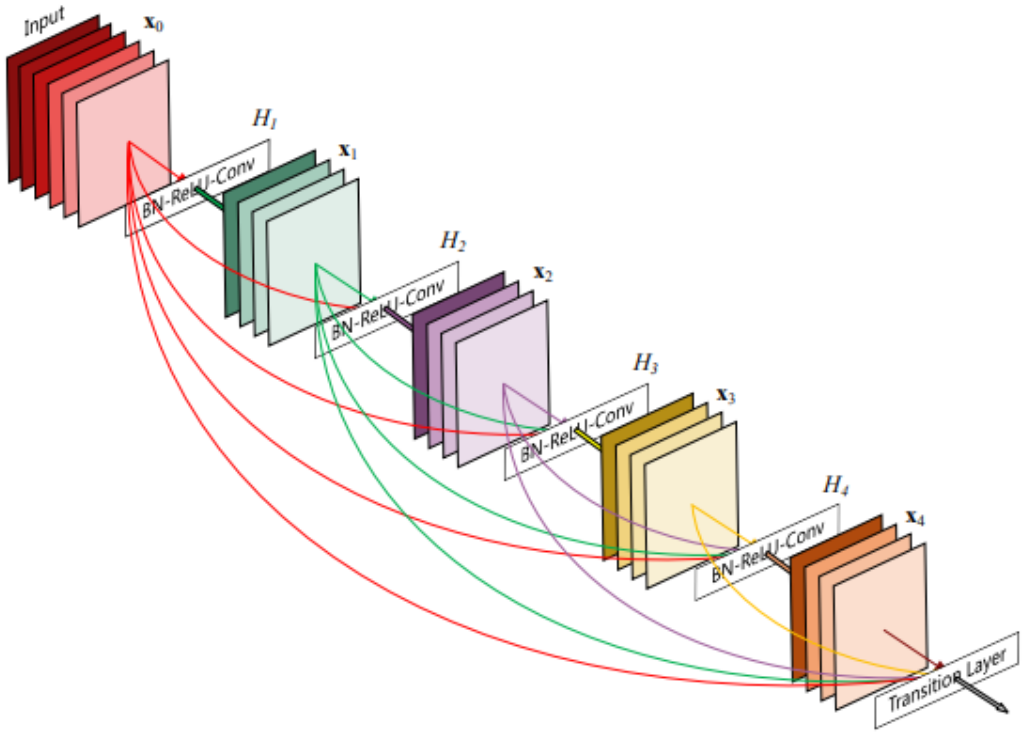

图 11: Dense Block 结构

在 DenseNet 网络结构中，如图 12：

- Dense block 中的有 k 个特征图输出，k 被称为 growth rate。随着层数增加，输入也会变得非常大，在 Dense block 中采用 bottleneck 层减少计算量，主要由 BN 算法、ReLU 激活函数和 1×1 卷积层组成，降低特征数。
- 相邻的 Dense Block 通过 Transition 层连接，Transition 层由 BN 算法、ReLU 激活函数、1×1 卷积层和 2×2 的均值池化层组成，能够压缩模型，降低特征图大小。

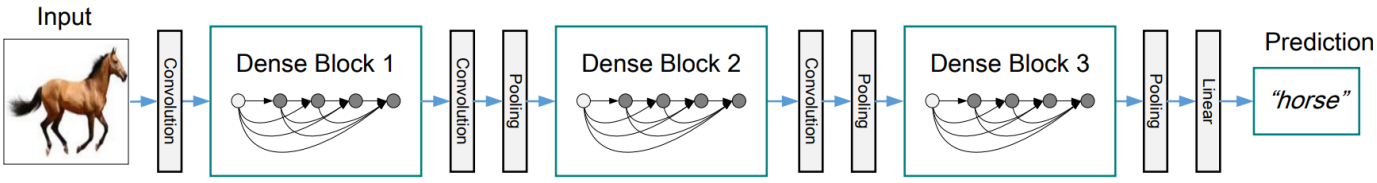

图 12: DenseNet 网络结构

3 效果分析

不同网络结构效果对比如图 1，具体分析如下：

- LeNet 被应用于当时的手写识别系统，用作数字 0-9 的分类。经过 30 次迭代训练后，在训练集上达到 1.1% 的错误率，MSE 为 0.017，在测试集上达到 3.4% 的错误率，MSE 为 0.024。
- 在 LeNet-5 中，训练同样的数据集，经过 19 次迭代训练后，在训练集上的错误率达到 0.35%，在测试集上错误率大约达到 1%。可以看出，LeNet-5 较 LeNet 的准确率有所提高，但是网络模

表 1: ILSVRC 历届网络结构结果

Year	Model	Error Rate(%)	Remark
2012	AlexNet	16.42	5 CNNs
2012	AlexNet	15.32	7CNNs
2013	OverFeat	14.18	7 fast models
2013	OverFeat	13.6	赛后；7 big models
2013	ZFNet	13.51	ZFNet 论文上的结果是 14.8
2013	Clarifai	11.74	
2013	Clarifai	11.20	2011 年的数据
2014	VGG	7.32	7 nets, dense eval
2014	VGG(亚军)	6.8	赛后；2 nets
2014	GoogleNet v1	6.67	7 nets, 144 crops
	GoogleNet v2	4.82	赛后；6 nets, 144 crops
	GoogleNet v3		赛后；4 nets, 144 crops
	GoogleNet v4	3.08	赛后；v4+Inception-Res-v2
2015	ResNet	3.57	6 models
2016	rimps-Soushen	2.99	公安三所
2016	ResNeXt(亚军)	3.03	加州大学圣地亚哥分校
2017	SENet	2.25	Momenta 与牛津大学

型小，应用范围十分受限，并且许多参数需要人工设定调整。LeNet 模型首次结合了权值共享的卷积网络和 BP 算法，为后人研究 CNN 指明了方向。

• 随着计算能力的提高，允许运算更深的网络和处理更大规模的数据集，在 AlexNet 的网络模型中，解决了大规模图像分类问题，该模型在 ILSVRC-2012 比赛中 top-5 达到了 15.3% 错误率。

• 在 VGG 网络模型中，使用小卷积核代替大卷积核，模块化的网络更有利于加深，该模型在 ILSVRC-2014 比赛中 top-5 达到了 7.3% 错误率。

• 在 Inception V2 网络模型中，引入 inception 模块并对数据降维，该模型在 ILSVRC 的数据集中 top-5 达到了 4.8% 错误率。

• 在 ResNet 网络模型中，使用 residual block 解决了由神经网络加深带来的梯度消失问题，模型在 ILSVRC-2015 比赛中 top-5 达到了 3.57% 错误率，首次在该比赛中所得结果低于人眼识别的错误率。

• 在 DenseNet 网络模型中，将前面所有层的输出作为下一层的输入来减轻梯度消失的问题，在相同数据集的情况下，该网络模型表现要比 ResNet 好。具体结果如表 13所示。

Method	Depth	Params	C10	C10+	C100	C100+	SVHN
Network in Network [22]	-	-	10.41	8.81	35.68	-	2.35
All-CNN [32]	-	-	9.08	7.25	-	33.71	-
Deeply Supervised Net [20]	-	-	9.69	7.97	-	34.57	1.92
Highway Network [34]	-	-	-	7.72	-	32.39	-
FractalNet [17]	21	38.6M	10.18	5.22	35.34	23.30	2.01
with Dropout/Drop-path	21	38.6M	7.33	4.60	28.20	23.73	1.87
ResNet [11]	110	1.7M	-	6.61	-	-	-
ResNet (reported by [13])	110	1.7M	13.63	6.41	44.74	27.22	2.01
ResNet with Stochastic Depth [13]	110	1.7M	11.66	5.23	37.80	24.58	1.75
	1202	10.2M	-	4.91	-	-	-
Wide ResNet [42]	16	11.0M	-	4.81	-	22.07	-
	28	36.5M	-	4.17	-	20.50	-
with Dropout	16	2.7M	-	-	-	-	1.64
ResNet (pre-activation) [12]	164	1.7M	11.26*	5.46	35.58*	24.33	-
	1001	10.2M	10.56*	4.62	33.47*	22.71	-
DenseNet ($k = 12$)	40	1.0M	**7.00**	5.24	**27.55**	24.42	1.79
DenseNet ($k = 12$)	100	7.0M	**5.77**	**4.10**	**23.79**	**20.20**	1.67
DenseNet ($k = 24$)	100	27.2M	**5.83**	**3.74**	**23.42**	**19.25**	**1.59**
DenseNet-BC ($k = 12$)	100	0.8M	**5.92**	4.51	**24.15**	22.27	1.76
DenseNet-BC ($k = 24$)	250	15.3M	**5.19**	**3.62**	**19.64**	**17.60**	1.74
DenseNet-BC ($k = 40$)	190	25.6M	-	**3.46**	-	**17.18**	-

Table 2: Error rates (%) on CIFAR and SVHN datasets. k denotes network's growth rate. Results that surpass all competing methods are **bold** and the overall best results are **blue**. "+" indicates standard data augmentation (translation and/or mirroring). $*$ indicates results run by ourselves. All the results of DenseNets without data augmentation (C10, C100, SVHN) are obtained using Dropout. DenseNets achieve lower error rates while using fewer parameters than ResNet. Without data augmentation, DenseNet performs better by a large margin.

图 13: ResNet 和 DenseNet 在不同数据集上的表现

4 讨论和未来展望

卷积神经网络起源于生物视觉皮层研究，发展至今，从最初的单层感知机到如今扩展为千百层的神经网络，从勉强完成小任务到如今被应用于各领域，这其中经历了多方面的迭代更新。

4.1 研究融合

生物领域视觉皮层研究的启发人们发明了新感知机和权值共享的网络，再有 LeCun 结合 BP 算法和权值共享网络发明卷积神经网络。同样，2012 年提出的 AlexNet 网络模型融合不少前人的研究成果，如使用的 ReLU 激活函数、数据增强。过去的研究并不会因为时代更新而被遗忘，在往前探索寻求突破时依然需要回顾前人的研究，能得到的不仅仅他们的研究成果，也有可能获得新的研究思路。

4.2 硬件支持

硬件是促进卷积神经网络发展的重要一环，容量更大的算力资源允许人们训练更大的神经网络，完成更大的任务。虽然 LeCun-5 表现优秀，但受算力资源限制，无法处理大任务，在当时支持向量机（support vector machines）更受人们青睐，卷积神经网络的发展因此停滞了一段时间，直到人们利用 GPU 加速了 CNN。在更多的算力资源下，神经网络变得更深，人们可以训练更大更复杂的数据，训练准确率也得到提升。

4.3 模块化网络

为了增加卷积神经网络深度，也更容易更改网络模型，人们提出模块化的网络结构。在 VGG，Inception V2、ResNet 等网络模型中都将部分网络组成模块化，使得网络更为简洁，有助于网络更改和提高性能。

如今，卷积神经网络已经被应用于各个领域，如图像识别，物体追踪等，从最初的简单网络模型逐渐变深变广，然后再将网络模型压缩符合移动设备的要求。但是卷积神经网络还是黑盒模型，未来是否能够实现中间过程图的可视化，可解释化机器还是一大难点。另外，卷积神经网络的许多改进是

再数学或计算机领域上进行，未来当人们对生物视觉研究更深，有可能对卷积神经网络中的神经元结构和连接进行完善。

5 总结

借助卷积神经网络的卷积核池化带来的不变性、本地连接、权值共享等优点，卷积神经网络被深度研究并广泛使用。卷积神经网络的发展同时也带动了深度学习和机器学习的进步，我主要从卷积神经网络的结构变化、算法更迭、研究成果等方面从浅到深探讨卷积神经网络算法，详细介绍了卷积神经网络在发展历程中的几种经典模型。卷积神经网络发展至今，其表现已经十分优秀，但是依然存在未解决的问题，如对中间过程特征图的可视化，打开卷积神经网络的黑盒模型等。

参考文献

[1] MCCULLOCH W S, PITTS W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of mathematical biophysics, 1943, 5(4): 115-133.

[2] ROSENBLATT F. The perceptron, a perceiving and recognizing automaton Project Para[M]. Cornell Aeronautical Laboratory, 1957.

[3] MINSKY M, PAPERT S. Perceptron: an introduction to computational geometry[Z]. 1969.

[4] HECHT-NIELSEN R. Theory of the backpropagation neural network[G]//Neural networks for perception. Elsevier, 1992: 65-93.

[5] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.

[6] FUKUSHIMA K, MIYAKE S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition[G]//Competition and cooperation in neural nets. Springer, 1982: 267-285.

[7] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541-551.

[8] LECUN Y, et al. LeNet-5, convolutional neural networks[J]. URL: http://yann. lecun. com/exdb/lenet, 2015, 20(5): 14.

[9] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[J]. ArXiv preprint arXiv:1602.07360, 2016.

[10] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[11] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[12] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. ArXiv preprint arXiv:1409.1556, 2014.

[13] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.