

深度学习高手该怎样炼成？这位拿下阿里天池大赛冠军的中科院博士为你规划了一份专业成长路径

时间 2017-11-13 14:38:06 51CTO

http://ai.51cto.com/art/201711/557545.htm?utm_source=tuicool&utm_medium=referral

深度学习本质上是深层的人工神经网络，它不是一项孤立的技术，而是数学、统计机器学习、计算机科学和人工神经网络等多个领域的综合。深度学习的理解，离不开本科数学中最为基础的数学分析(高等数学)、线性代数、概率论和凸优化;深度学习技术的掌握，更离不开以编程为核心的动手实践。没有扎实的数学和计算机基础做支撑，深度学习的技术突破只能是空中楼阁。

深度学习本质上是深层的人工神经网络，它不是一项孤立的技术，而是数学、统计机器学习、计算机科学和人工神经网络等多个领域的综合。深度学习的理解，离不开本科数学中最为基础的数学分析(高等数学)、线性代数、概率论和凸优化;深度学习技术的掌握，更离不开以编程为核心的动手实践。没有扎实的数学和计算机基础做支撑，深度学习的技术突破只能是空中楼阁。

所以，想在深度学习技术上有所成就的初学者，就有必要了解这些基础知识之于深度学习的意义。除此之外，我们的专业路径还会从结构与优化的理论维度来介绍深度学习的手上，并基于深度学习框架的实践浅析一下进阶路径。

最后，本文还将分享深度学习的实践经验和获取深度学习前沿信息的经验。

数学基础

如果你能够顺畅地读懂深度学习论文中的数学公式，可以独立地推导新方法，则表明你已经具备了必要的数学基础。

掌握数学分析、线性代数、概率论和凸优化四门数学课程包含的数学知识，熟知机器学习的基本理论和方法，是入门深度学习技术的前提。因为无论是理解深度网络中各个层的运算和梯度推导，还是进行问题的形式化或是推导损失函数，都离不开扎实的数学与机器学习基础。

- 数学分析：在工科专业所开设的高等数学课程中，主要学习的内容为：微积分。对于一般的深度学习研究和应用来说，需要重点温习函数与极限、导数(特别是复

合函数求导)、微分、积分、幂级数展开、微分方程等基础知识。在深度学习的优化过程中,求解函数的一阶导数是最为基础的工作。当提到微分中值定理、**Taylor** 公式和拉格朗日乘子的时候,你不应该只是感到与它们似曾相识。这里推荐同济大学第五版的《高等数学》教材。

- **线性代数**:深度学习中的运算常常被表示成向量和矩阵运算。线性代数正是一门以向量和矩阵作为研究对象的数学分支。需要重点温习的包括向量、线性空间、线性方程组、矩阵、矩阵运算及其性质、向量微积分。当提到 **Jacobian** 矩阵和 **Hessian** 矩阵的时候,你需要知道确切的数学形式;当给出一个矩阵形式的损失函数时,你可以很轻松的求解梯度。这里推荐同济大学第六版的《线性代数》教材。
- **概率论**:概率论是研究随机现象数量规律的数学分支,随机变量在深度学习中有许多应用,无论是随机梯度下降、参数初始化方法(如 **Xavier**),还是 **Dropout** 正则化算法,都离不开概率论的理论支撑。除了掌握随机现象的基本概念(如随机试验、样本空间、概率、条件概率等)、随机变量及其分布之外,还需要对大数定律及中心极限定理、参数估计、假设检验等内容有所了解,进一步还可以深入学习一点随机过程、马尔可夫随机链的内容。这里推荐浙江大学版的《概率论与数理统计》。
- **凸优化**:结合以上三门基础的数学课程,凸优化可以说是一门应用课程。但对于深度学习而言,由于常用的深度学习优化方法往往只利用了一阶的梯度信息进行随机梯度下降,因而从业者事实上并不需要多少“高深”的凸优化知识。理解凸集、凸函数、凸优化的基本概念,掌握对偶问题的一般概念,掌握常见的无约束优化方法如梯度下降方法、随机梯度下降方法、**Newton** 方法,了解一点等式约束优化和不等式约束优化方法,即可满足理解深度学习中优化方法的理论要求。这里推荐一本教材, **Stephen Boyd** 的《**Convex Optimization**》。
- **机器学习**:归根结底,深度学习只是机器学习方法的一种,而统计机器学习则是机器学习领域事实上的方法论。以监督学习为例,需要你掌握线性模型的回归与分类、支持向量机与核方法、随机森林方法等具有代表性的机器学习技术,并了解模型选择与模型推理、模型正则化技术、模型集成、**Bootstrap** 方法、概率图模型等。深入一步的话,还需要了解半监督学习、无监督学习和强化学习等专门技术。这里推荐一本经典教材《**The elements of Statistical Learning**》。

计算机基础

深度学习要在实战中论英雄，因此具备 GPU 服务器的硬件选型知识，熟练操作 Linux 系统和进行 Shell 编程，熟悉 C++ 和 Python 语言，是成长为深度学习实战高手的必备条件。当前有一种提法叫“全栈深度学习工程师”，这也反映出了深度学习对于从业者实战能力的要求程度：既需要具备较强的数学与机器学习理论基础，又需要精通计算机编程与必要的体系结构知识。

- **编程语言：**在深度学习中，使用最多的两门编程语言分别是 C++ 和 Python。迄今为止，C++ 语言依旧是实现高性能系统的首选，目前使用最广泛的几个深度学习框架，包括 Tensorflow、Caffe、MXNet，其底层均无一例外地使用 C++ 编写。而上层的脚本语言一般为 Python，用于数据预处理、定义网络模型、执行训练过程、数据可视化等。当前，也有 Lua、R、Scala、Julia 等语言的扩展包出现于 MXNet 社区，呈现百花齐放的趋势。这里推荐两本教材，一本是《C++ Primer 第五版》，另外一本是《Python 核心编程第二版》。
- **Linux 操作系统：**深度学习系统通常运行在开源的 Linux 系统上，目前深度学习社区较为常用的 Linux 发行版主要是 Ubuntu。对于 Linux 操作系统，主要需要掌握的是 Linux 文件系统、基本命令行操作和 Shell 编程，同时还需熟练掌握一种文本编辑器，比如 VIM。基本操作务必要做到熟练，当需要批量替换一个文件中的某个字符串，或者在两台机器之间用 SCP 命令拷贝文件时，你不需要急急忙忙去打开搜索引擎。这里推荐一本工具书《鸟哥的 Linux 私房菜》。
- **CUDA 编程：**深度学习离不开 GPU 并行计算，而 CUDA 是一个很重要的工具。CUDA 开发套件是 NVidia 提供的一套 GPU 编程套件，实践当中应用的比较多的是 CUDA-BLAS 库。这里推荐 NVidia 的官方在线文档 <http://docs.nvidia.com/cuda/>。
- **其他计算机基础知识：**掌握深度学习技术不能只满足于使用 Python 调用几个主流深度学习框架，从源码着手去理解深度学习算法的底层实现是进阶的必由之路。这个时候，掌握数据结构与算法(尤其是图算法)知识、分布式计算(理解常用的分布式计算模型)，和必要的 GPU 和服务器的硬件知识(比如当我说起 CPU 的 PCI-E 通道数和 GPU 之间的数据交换瓶颈时，你能心领神会)，你一定能如虎添翼。

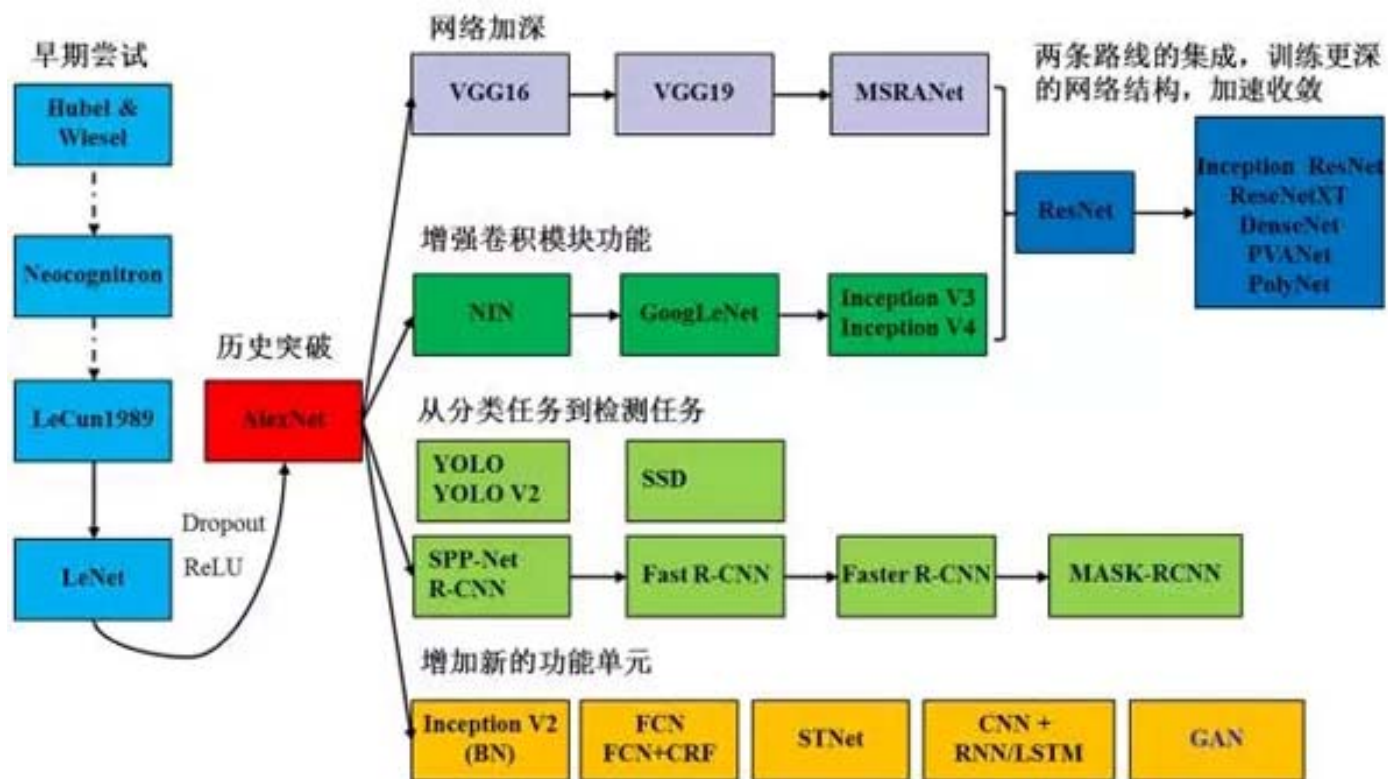
深度学习入门

接下来分别从理论和实践两个角度来介绍一下深度学习的入门。

- 深度学习理论入门：我们可以用一张图(图 1)来回顾深度学习中的关键理论和方法。从 MCP 神经元模型开始，首先需要掌握卷积层、Pooling 层等基础结构单元，Sigmoid 等激活函数，Softmax 等损失函数，以及感知机、MLP 等经典网络结构。接下来，掌握网络训练方法，包括 BP、Mini-batch SGD 和 LR Policy。最后还需要了解深度网络训练中的两个至关重要的理论问题：梯度消失和梯度溢出。



以卷积神经网络为例，我们用图 2 来展示入门需要掌握的知识。起点是 Hubel 和 Wiesel 的对猫的视觉皮层的研究，再到日本学者福岛邦彦神经认知机模型(已经出现了卷积结构)，但是第一个 CNN 模型诞生于 1989 年，1998 年诞生了后来被大家熟知的 LeNet。随着 ReLU 和 Dropout 的提出，以及 GPU 和大数据所带来的历史机遇，CNN 在 2012 年迎来了历史性的突破——诞生了 AlexNet 网络结构。2012 年之后，CNN 的演化路径可以总结为四条：1. 更深的网络；2. 增强卷积模的功能以及上诉两种思路的融合 ResNet 和各种变种；3. 从分类到检测，最新的进展为 ICCV 2017 的 Best Paper Mask R-CNN；4. 增加新的功能模块。



- 深度学习实践入门：掌握一个开源深度学习框架的使用，并进一步的研读代码，是实际掌握深度学习技术的必经之路。当前使用最为广泛的深度学习框架包括 Tensorflow、Caffe、MXNet 和 PyTorch 等。框架的学习没有捷径，按照官网的文档 **step by step** 配置及操作，参与 GitHub 社区的讨论，遇到不能解答的问题及时 **Google** 是快速实践入门的好方法。

初步掌握框架之后，进一步的提升需要依靠于具体的研究问题，一个短平快的策略是先刷所在领域权威的 **Benchmark**。例如人脸识别领域的 LFW 和 MegaFace，图像识别领域与物体检测领域的 ImageNet、Microsoft COCO，图像分割领域的 Pascal VOC 等。通过复现或改进别人的方法，亲手操练数据的准备、模型的训练以及调参，能在所在领域的 **Benchmark** 上达到当前最好的结果，实践入门的环节就算初步完成了。

后续的进阶，就需要在实战中不断地去探索和提升了。例如：熟练的处理大规模的训练数据，精通精度和速度的平衡，掌握调参技巧、快速复现或改进他人的工作，能够实现新的方法等等。

深度学习实战经验

在这里，分享四个方面的深度学习实战经验。

1. 充足的数据。大量且有标注的数据，依旧在本质上主宰着深度学习模型的精度，每一个深度学习从业者都需要认识到数据极端重要。获取数据的方式主要有三种：开放数据(以学术界开放为主，如 **ImageNet** 和 **LFW**)、第三方数据公司的付费数据和结合自身业务产生的数据。
2. 熟练的编程实现能力。深度学习算法的实现离不开熟练的编程能力，熟练使用 **Python** 进行编程是基础。如果进一步的修改底层实现或增加新的算法，则可能需要修改底层代码，此时熟练的 **C++**编程能力就变得不可或缺。一个明显的现象是，曾经只需要掌握 **Matlab** 就可以笑傲江湖的计算机视觉研究者，如今也纷纷需要开始补课学习 **Python** 和 **C++**了。
3. 充裕的 **GPU** 资源。深度学习的模型训练依赖于充裕的 **GPU** 资源，通过多机多卡的模型并行，可以有效的提高模型收敛速度，从而更快的完成算法验证和调参。一个专业从事深度学习的公司或实验室，拥有数十块到数百块的 **GPU** 资源已经是普遍现象
4. 创新的方法。以深度学习领域权威的 **ImageNet** 竞赛为例，从 2012 年深度学习技术在竞赛中夺魁到最后一届 2017 竞赛，方法创新始终是深度学习进步的核心动力。如果只是满足于多增加一点数据，把网络加深或调几个 **SGD** 的参数，是难以做出真正一流的成果的。

根据笔者的切身经历，方法创新确实能带来难以置信的结果。一次参加阿里巴巴组织的天池图像检索比赛，笔者提出的一点创新——使用标签有噪声数据的新损失函数结果竟极大地提高了深度模型的精度，还拿到了当年的冠军。

深度学习前沿

【前沿信息的来源】

实战中的技术进阶，必需要了解深度学习的最新进展。换句话说，就是刷论文：除了定期刷 **Arxiv**，刷代表性工作的 **Google Scholar** 的引用，关注 **ICCV**、**CVPR** 和 **ECCV** 等顶级会议之外，知乎的深度学习专栏和 **Reddit** 上时不时会有最新论文的讨论(或者精彩的吐槽)。

一些高质量的公众号，例如 **Valse** 前沿技术选介、深度学习大讲堂、**Paper Weekly** 等，也时常有深度学习前沿技术的推送，也都可以成为信息获取的来源。同时，关注学术界大佬 **LeCun** 和 **Bengio** 等人的 **Facebook/Quora** 主页，关注微博大号“爱可可爱生活”等人，也常有惊喜的发现。

【建议关注的重点】

- 新的网络结构。在以 **SGD** 为代表的深度学习优化方法没有根本性突破的情况下，修改网络结构是可以较快提升网络模型精度的方法。**2015** 年以来，以 **ResNet** 的各种改进为代表的各类新型网络结构如雨后春笋般涌现，其中代表性的有 **DenseNet**、**SENet**、**ShuffleNet** 等。
- 新的优化方法。纵观从 **1943** 年 **MCP** 模型到 **2017** 年间的人工神经网络发展史，优化方法始终是进步的灵魂。以误差反向传导(**BP**)和随机梯度下降(**SGD**)为代表的优化技术的突破，或是 **Sigmoid/ReLU** 之后全新一代激活函数的提出，都非常值得期待。笔者认为，近期的工作如《**Learning gradient descent by gradient descent**》以及 **SWISH** 激活函数，都很值得关注。但能否取得根本性的突破，也即完全替代当前的优化方法或 **ReLU** 激活函数，尚不可预测。
- 新的学习技术。深度强化学习和生成对抗网络(**GAN**)。最近几周刷屏的 **Alpha Zero** 再一次展示了深度强化学习的强大威力，完全不依赖于人类经验，在围棋项目上通过深度强化学习“左右互搏”所练就的棋力，已经远超过上一代秒杀一众人类高手的 **AlghaGo Master**。同样的，生成对抗网络及其各类变种也在不停地预告一个学习算法自我生成数据的时代的序幕。笔者所在的公司也正尝试将深度强化学习和 **GAN** 相结合，用于跨模态的训练数据的增广。
- 新的数据集。数据集是深度学习算法的练兵场，因此数据集的演化是深度学习技术进步的缩影。以人脸识别为例，后 **LFW** 时代，**MegaFace** 和 **Microsoft Celeb-1M** 数据集已接棒大规模人脸识别和数据标签噪声条件下的人脸识别。后 **ImageNet** 时代，**Visual Genome** 正试图建立一个包含了对象、属性、关系描述、问答对在内的视觉基因组。