

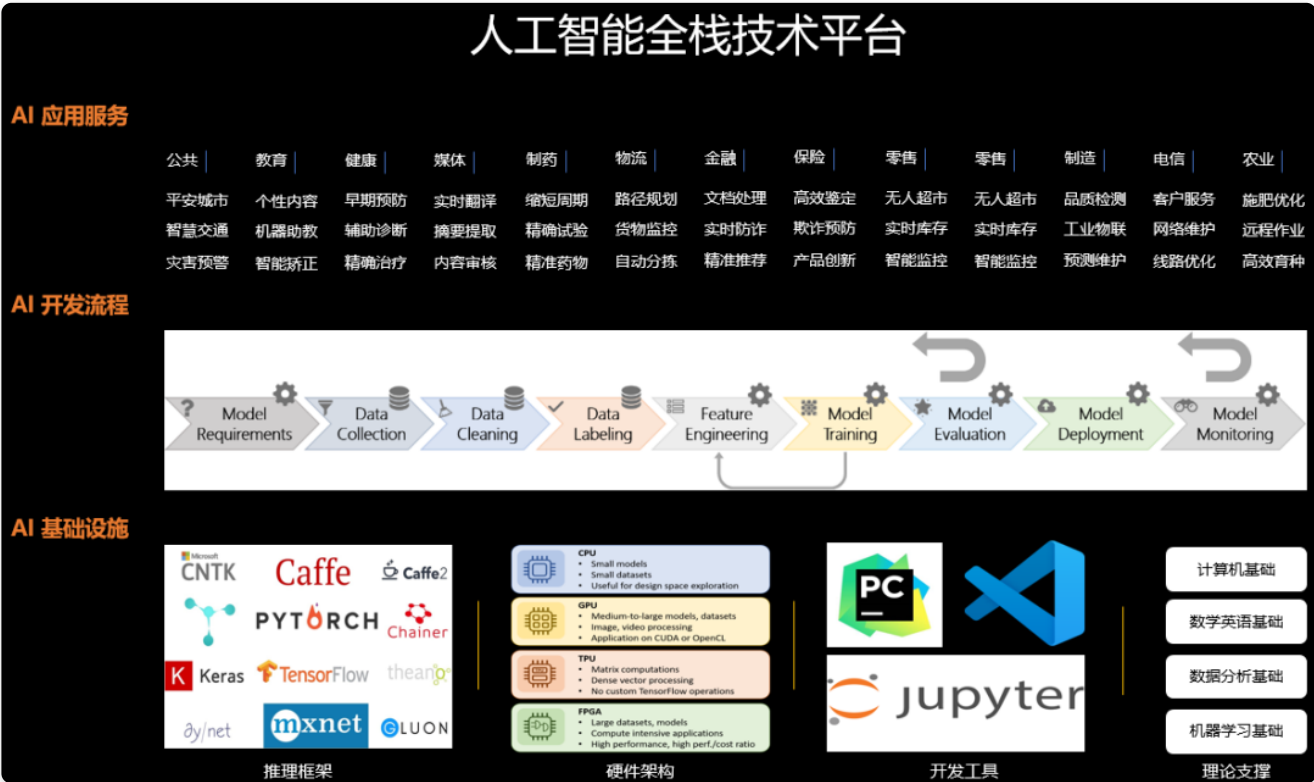
万字长文带你入门计算机视觉!

原创 派派星 CVHub 2021-06-07 22:21

收录于合集

#深度学习 90 #计算机视觉 95 #一文看尽系列 7

导读



CV技术栈

在这个AI大爆发的时代，仿佛只要与人工智能沾边的工作便显得高大上起来，与之相反的便是被大多数网友们痛批的一无是处的**生化环材**专业。虽然AI为我们的生活增添了许多便利，但是随着越来越多的人涌入这个圈子，也导致了整个AI圈迅速膨胀。**计算机视觉**，相对比自然语言处理这个领域来说入门门槛会稍微低些，而且图片视频相对于文字来说也更具有趣味性和易于理解。然而，这也会导致CV圈越来越卷，通过知乎相关热门问答便能体会到：

- 2016年——深度学习的春天是不是要来了？
- 2017年——人工智能是不是一个泡沫？
- 2018年——算法岗是否值得进入？
- 2019年——如何看待算法岗竞争激烈，供不应求？
- 2020年——如何看待算法岗一片红海，诸神黄昏？
- 2021年——如何看待算法岗灰飞烟灭？

CVHub有云，小卷怡情，大卷伤身，强卷灰飞烟灭。正所谓人在CV圈，哪有不被卷？正因为如此，我们如何才能在这个浮躁的环境中沉下心来，选定一条合适自己的学习路线，从千军万马中脱颖而出，荣登时代的C位呢？这便是本篇文章诞生的目的，本文旨在构建最完整的CV学习路线，帮助大家早日从“调参侠”的包袱中脱离出来，成为一名真正合格的算法工程师。

本文将从以下三大维度为大家逐一展开介绍。



软基础

软基础，可以理解为技能栈，它是支撑整个发展路线的核心。

数学与英语基础

相信对于任何一个理工科的学生而言，数学的重要性是不言而喻的。因此，掌握好数学基础，有助于我们更好地理解其它知识以及为后面的学习做铺垫。然而，数学领域所涉及到的知识可谓非常广泛，如果漫无目的的看，可能对于90%的人来说第一关就已经从入门到放弃了。下面，我们将给出计算机视觉领域所涉及到的的一些数学基础，对于需要重点掌握的知识我们均会提及一下，其余未涉及到的知识点可视个人时间精力和研究方向自行查阅相关资料：

微积分

《微积分》是现代数学的基础，也是支撑后续几门数学分支的重要基石。顾名思义，微积分里面包含的两个核心组件便是微分和积分。其中，与微分比较接近的概念便是**导数**，熟练的计算导数有助于我们后面推导整个梯度下降的过程。其次，掌握**上确界**和**下确界**的概念有助于我们理解诸如Wasserstein距离，它作为一种损失函数被应用于WGAN以解决训练不稳定的问题。此外，有涉及到GAN方向的同学也可以理解下**Lipschitz连续性的**概念，它对提高GAN训练的稳当性分析具有不可替代的作用。而对于基础的深度学习原理来说，我们更多的需要掌握的

概念是：**梯度、偏导数、链式求导法则、导数与函数的极值&单调性&凹凸性判断、定积分与不定积分**。最后，我们可以着重理解下**泰勒公式**，它是理解各种优化算法的核心。

线性代数与矩阵论

《线性代数》是工科生的必备基础。由于计算机现有的运算方式大都是基于并行计算的，无论是机器学习或者深度学习模型来说，数据都是以矩阵或者向量化的形式传入的。因此，我们首先要明白**线性空间、向量、标量、矩阵、范数、秩**的概念以及着重掌握矩阵和向量的基本运算方式，这对于我们后面理解一些网络模型结构，如自注意力机制是非常有必要的。此外，对于后面涉及到的机器学习算法学习，我们还应该理解**特征值、特征向量以及矩阵的正定**概念，同时也应该掌握**特征值分解**（主成分分析和线性判别分析等）以及**奇异值分解**（正太贝叶斯分类器及主题模型等）。矩阵论则是线性代数的进一步升华，所要掌握的重点知识为**矩阵分解、线性变换、相似矩阵、欧式空间、正交矩阵、对称矩阵、正定矩阵**。

概率论与信息论

《概率论与数理统计》是工科生数理基础的三把斧之一，也是支撑整个机器学习与深度学习原理的核心理论。概率论，本质上是用于描述对不确定性的度量，深度学习本质上也是通过网络进行学习，最终对输入图像计算出一个概率值。对于这门课，我们需要重点掌握的概念有：**随机变量、概率分布、条件概率、统计与假设检验、概率密度及质量函数、贝叶斯公式、期望、方差、协方差**。此外，我们还需要重点掌握一些常见的分布函数，如**0-1分布、二项分布（伯努利分布）、几何分布、高斯分布、指数分布、泊松分布**。最后，我们还应该重点理解**最小二乘法、最大似然估计、指数移动平均**等。而对于《信息论》来说，最需要掌握的一个知识点必定是**熵**。除了理解熵的概念以外，可以着重了解下几种不同的熵理论，如**条件熵、联合熵、相对熵、最大熵以及互信息**。

最优化方法

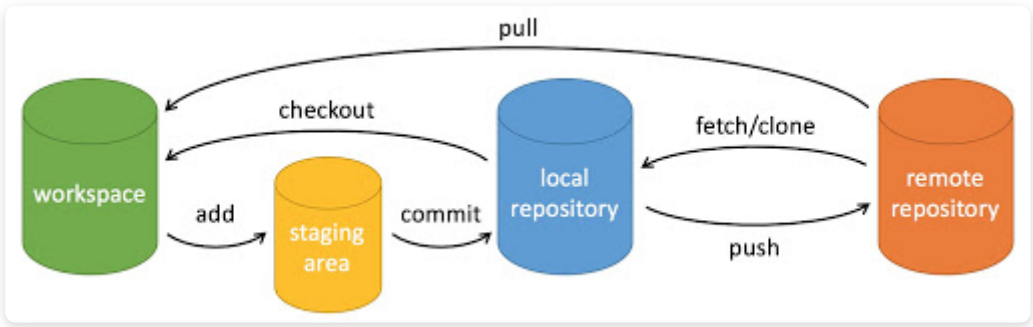
首先，我们可以简单的理解下什么是“最优化”？最优化，直白的说便是指在完成一个目标的时候，我们总是希望能够在**资源受限**的条件下以**最小的代价**来获取**最大的收益**。可以毫不夸张的说，最优化理论是整个机器学习中相对较难且非常重要的一个理论基础，同时也是面试机器学习算法岗或者在该领域继续进修所必不可少的一个门卡。在了解完基本概念的同时，我们也不觉的触及到了最优化问题中基本数学描述的三个基本要素，即**决策变量、目标函数以及约束条件**。其次，罗列下这方面所需要掌握的概念及原理：**凸集、凸函数、凸集分离定理、超平面和半空间**。最后，我们需要重点理解的便是**梯度下降法**，它是目前整个深度学习算法的核心。此外，对于一些诸如**拟牛顿法、阻尼牛顿法、随机梯度下降算法**等可留到后半部分学习。

英语

英语，虽然不是我们的母语，但是作为一名新时代的AI从业技术人员或者即将踏入硕博之路的人来说，这是一门必不可少的技能。首先，对于选择就业的同学来说，在日常的工作生活中，都会不可避免的会查阅大量的工具书和英文网站，这就需要我们具备一定的英语基础，才能够在工作中游刃有余，帮助我们快速的定位到关键点。其次，对于想往科研这条路的同学来说，英语更是不可或缺的必备技能。除了平常科研写作要经常用到之外，我们在申请海内外学校的时候还需要要求具有四六级证书（国内）或者雅思托福考试证明（境外），每年往往有许多发表不少科研成果的同学被卡在这里，着实令人惋惜。最后，掌握良好的英语沟通能力，无论对职场发展（去海外工作或者与国外客户合作）还是学术交流（会议）都是具有积极意义的。然而，语言这种东西不能一蹴而就，除了母语环境，我们能做的是每日学习，厚积薄发，后面方可应用自如。

计算机基础

Git基础



git各个分区的转换

Git作为一个版本控制软件，可以帮助我们更好地对我们的项目进行管理，也方便进行多人协同的工作。但是，估计大多数人最常用的还是git clone（误）。因此这个技能也常常会被大家所忽略但却又是进入工作之后说必须掌握的技能。

从逻辑上，git所管理的代码可以分为三个区，如上图所示分别为工作区(你正在编写的代码)，本地仓库（保存到本地仓库的代码）以及远程仓库（远程代码库），它们之间的转换所需要用到的命令如上图所示。比如 git commit（提交至本地仓库），git push（把本地代码推到远程库并进行合并），git pull（把远程代码下载下来并合并到工作区），git diff（比较工作区与暂存区的不同）。虽然平时不用的时候可能会忘掉这些命令，但是需要用的时候看看上图，基本上都能想起每个命令的意义。

Linux基础

大多数情况下，我们的训练环境都是运行在Linux平台上的，因此我们还是有必要掌握一些常用的Linux基础。下面列举一些实际开发工程中经常会用到的指令：

- 基本操作：mv(移动), rm(删除), cp(复制), mkdir(创建文件夹), touch(创建文件), scp(用于传送文件), cat(打印文件到命令行), grep(查找符合条件的字符串), ls(查看文件), chmod(修改权限), ln(创建软链接)
- 压缩/解压命令：zip/unzip和tar命令
- 系统监控相关：ps(查看进程), top(类似任务管理器), df(查看磁盘占用)
- 网络相关：ping(判断网络是否可达), ifconfig (显示网络设备状态)
- 显卡状态：nvidia-smi (需要安装), gpustat (需要安装, 这个显示更美观)
- 后台运行：nohup (关掉终端也不会停止运行), ctrl+z (切换到后台, 关闭终端会结束)

为了安全起见，一般来说不建议用rm来删除文件，而是建立一个垃圾回收文件夹。把平日里不想要的文件通过mv文件统一移动到该文件夹下，然后使用定时清理的脚本清空文件夹里面的内容，以避免“删库跑路”的尴尬。

算法基础

IT界有一句谚语——Talk is cheap, show me the code.虽然平时设计算法、训练模型以及算法部署过程中基本上都不会用到的数据结构相关的知识，奈何大部分互联网公司笔试面试就要撕代码，而且难度也在逐年提升（卷卷卷）。不仅如此，据身边许多人的反应，现在中小公司也学这一招了，所以手撕动态规划算法基本已经成为标配，因此，为了能够在求职的过程中尽可能不让Coding能力成为**减分项**，大家还是需要在这上面多花费点时间。

学习数据结构的方法除了看书，弄明白基础数据结构的原理外，最好的学习方法便是实战。关于这点大家可以以Leetcode为主，辅助《剑指offer》等面试宝典进行刷题。对于刚开始刷题的同学，建议按照专题来刷，这样能够比较好得掌握这一类题型的解决方法，更容易加深记忆。比如按照**数组、字符串、链表、栈、队列、哈希表、二叉树、堆、递归、深搜广搜、动态规划**等专题来刷。最后每样都要有所涉及，避免面试出现自己完全没坐做过的类型，几分钟都憋不出一行代码就有点尴尬了，而且给面试官的印象也不好。

好玩的事情那么多，谁又喜欢刷题呢？这里给大家一个刷题的思路，让大家更加高效地刷题：

- 看懂题目 (1-2min)
- 设计数据结构+算法，尽量回想一下同类型的题目 (1-2min) 【没有思路直接去看题解】
- 尝试动手将整份代码给敲出来，争取把给定的测试用例都通过 (3-5min)
- 最后剩下最苦逼的便是调bug了 (0- ∞ min)

注意：尽量不要在一道题目上花过多的时间，因为有些题目并不是你想就能想出比较好的解决方法。许多题目往往有一套**固定的解题模板**，而我们需要做的是应用好这套模板去解决问题。关于大家都会关心的问题是刷题的时候要用什么语言来刷比较合适？这里推荐大家优先用**Python**吧！相对于其它语言，特别是C++来说，Python的开发效率会显得非常有优势，重点是面试管也能接受。

编程基础

在计算机视觉领域中，前对于前期的模型设计、训练与验证，我们会基于Python语言进行快速的版本迭代，而对于后期模型的部署以及底层op的优化，考虑到实际的生产环境和运行效率则一般会基于C++来实现。

- **Why Python?** 作为一门高级语言，Python具备高度的可阅读性，这使得它能够更容易被初学者所接受和掌握，也更易于学习。此外，Python**强大的社区和丰富的第三方库、框架和扩展**也为开发人员带来了极大的便捷，这是它最大的优势！
- **How to learn Python?** 学习任何一门编程语言，包括Python，最好的方法便是实战。因此，第一步我们要了解**IDE**的概念，不同的IDE有不同的适用性，选择合适的开发工具有利于初学者更好的上手。第二步便是找到合适的教程来学习编写Python脚本，这里比较推荐的纸质书籍是《Python从入门到精通》这本书或者直接上在线网站**Python3 教程 | 菜鸟教程^[1]**。最后一步初学者只需再稍微了解下**解释器**的概念即可，解释器的作用是让计算机能够读懂你的代码脚本应执行相应的操作。
- **第三方库：**语法基础熟悉之后，作为进阶教程，我们可以进一步地学习一些常用的第三方库。比如用于数据分析的Numpy、Pandas；用于可视化的Matplotlib、Seaborn；用于机器学习的Scikit-learn；用于图像处理的Opencv、Scikit-image等。再学习的过程中建议搭配本文第3部分《工具链》中3.4节所列出的数据科学速查表进行有针对性的学习。
- **环境配置：**这里笔者强烈推荐大家将Miniconda 3作为我们的虚拟环境包管理。Anaconda是一个用于科学计算的Python发行版，其支持Linux, Mac, Windows系统，提供了包管理与环境管理的功能，可以很方便地解决多版本python并存、切换以及各种第三方包安装问题。而Miniconda则是Anaconda的一个轻量版，足够满足日常的开发需求。
- **集成开发环境：**即Integrated Development Environment, IDE，是用于提供程序开发环境的应用程序，一般包括代码编辑器、编译器、调试器和图形用户界面等工具。关于这方面的选择，也可参考第3部分《工具链》中3.2节所列出的常用开发工具首选。
- **再多唠两句：**Python功能其实远不止上面这些，它对于编写爬虫，搭建个人博客、网站，批量处理文件等都十分方便。对于在这方面有兴趣的同学也可自行探索。
- **Why C++?** 尽管Python具有十分强大的功能，然而比较致命的缺点是运行速度慢，相较于C++语言来说是非常慢的。当我们的AI模型训练好之后，下一步便是落地部署。与学术界更关心模型的精度不同，工业界更关心的是模型的部署性能。在实际的生产环境中，由于Python的可移植性和运行速度远不如C++，所以对于后期模型的部署以及底层op的优化都会首选C++。

- **如何部署?** C++部署一般有两种形式。第一种是利用C++从头搭建模型,实现模型训练和推理。第二种是前期先利用Python搭建网络进行训练,后期再将模型载入C++中实现推理。前者开发难度偏大,商用价值较高。而后者开发难度适中,且能满足大多数项目的需求。对于后者,部署C++有两种常用方式: **Pytorch** → **libtorch**和**Pytorch**→**onnx** → **tensorrt**。后期我们会在公众号的专栏进行专门的讲解,敬请关注。

数据分析与数字图像处理基础

数据处理与分析

深度学习项目的成功落地三要素:数据、算法和算力,数据有效地处理与分析是非常关键的一步。简单来说,数据处理与分析即是从大量的、杂乱无章的数据中抽取并分析出有价值的信息。

在近十几年的发展上来看,算法本身的创新是非常有限的,大都还是停留在修修补补的阶段。对于一名普通的计算机视觉算法工程师来说,工作中有80%的时间都是在处理数据,因为模型的改动需要考虑到底层算子的适应性,效果不一定好并且工作量会大大增加。因此,我们要学会根据实际应用场景针对性地分析业务需求,通常是基于已开源的大数据集作为pretrain,然后利用业务数据集来finetuning。所以重心还是要学会对业务数据进行**数据清洗**以及选择合适的**数据增强**方式来扩充有限的数据集。而对于一名数据竞赛者来说,更多的应该了解如何开始一个比赛。下面简单的罗列出数据分析比赛大体的几个步骤,具体的可以参考后面第3部分<**工具链**>中所罗列出来的竞赛网站自行寻找自己感兴趣的或方向相关的比赛去参加实战一波即可:

Exploratory Data Analysis: 导入数据 → 数据总览 → 变量相关性分析 → 缺失值及异常值处理 → 数据分布变换 → 特征分析(数值型和类别型);

Feature Engineering: 异常值处理(通过箱形图观察及长尾截断处理等) → 缺失值处理(分为随机性缺失和非随机性缺失,对应的处理方法分别为RF、SVD、Mean、Median、KNN、NN & Binary、GSimp、QRILC等) → 特征离散化 → 数据类型转换(one-hot编码、哈夫曼编码等) → 特征组合(将具有相关性的特征组合起来形成新的特征,以此来丰富特征的多样性,提高模型的泛化能力) → 数据降维(PCA);

Modeling: 选择合适的模型(LightGBM & XGBoost等) → 交叉验证 → 模型调整(GridSearch & Bayes & 贪心调参等);

Ensemble: 一般选定3个左右的强模型+多个相关性较小的弱模型进行Voting or Stacking or Boosting。

数字图像处理

数字图像处理 (Digital Image Processing) 是通过计算机对图像进行去噪、增强、复原、分割、提取特征等处理的方法和技术。

尽管深度学习能实现很高的性能,但目前为止人们仍难以解释其内部工作的机制原理。因此,深度学习本质上还是一个黑盒子。了解数字图像的一些基本概念以及掌握基础的图像处理技能,有助于我们理解更深层次的内容。下面列举一些需要重点了解和掌握的知识点:

- **图像数字化:** 指的是将模拟(连续)信号的图像转换为数字(离散)信息的过程,主要包含**采样**和**量化**两个步骤。二维图像在计算机中通常是以矩阵的形式表示,这里需要了解图像的一些基本属性,如**图像格式**(BMP & JPEG & GIF & PNG的定义和区别)、**图像分辨率和通道数**(8位单通道二值图像 & 24位RGB通道彩色图像 & 32位RGBA通道 & 以及各种通道之间的互相转换)、**图像尺寸**(像素 & 图像宽度和高度)、**图像色彩颜色空间**(RGB & HSV & HSI & CMYK)、**图像插值方法**(最近邻插值、双线性插值、三线性插值)及**图像成像方式**(伽马射线 & X射线等)。
- **图像压缩:** 目的是减少图像中的冗余信息,以更高效的格式进行存储和传输数据。一般可分为有损压缩和无损压缩,这里我们只需要简单了解下有哪些经典的压缩算法即可。有兴趣的可以了解下**JPEG压缩算法**的原理和步骤(涉及离散余弦变换 & 量化 & YCbCr色彩空间)。
- **图像增强:** 指的是利用各种数学方法和变换手段来提高图像对比度与清晰度,使改善后的图像更适应于人的视觉特性或易于机器识别。简单来说,就是要突出感兴趣的特征,抑制不感兴趣的特征,从而改善图像质量。如强化图像中的高频分量,可使图像中物体轮廓清晰,细节更加明显;而强化图像中的低频分量可以减少图像中的噪声影响。这里我们需要重点掌握的是**灰度直方图**的概念,了解什么是**直方图**和**对比度**,掌握**直方图均衡化**、**限制对比度的自适应直方图均衡化**(CLAHE)以及伽马校正和仿射变换操作。
- **图像复原:** 指的是利用退化过程的先验知识,去恢复已被退化图像的本来面目。这是由于受采集设备和光照等外部环境的影响,图像在形成的过程中不可避免的会失真或者引入背景噪声干扰。因此,我们首先要知道**噪声的来源**(电子元器件发热或传输损失 & 成像系统的调制与缺陷 & 光照等外部环境因素的干扰)以及各种**噪声的来源和特点**(高斯噪声 & 脉冲噪声[椒盐噪声 | 胡椒噪声 | 盐粒噪声] & 泊松噪声 & 斑点噪声)。最后便是如何**去噪**,这里需要重点掌握一些常见的**线性和非线性滤波**算法(均值滤波 & 中值滤波 & 维纳滤波 & 卡尔曼滤波 & 高通滤波 & 低通滤波 & 高斯滤波 & 双边滤波 & 拉普拉斯滤波 & 卷积核 & Gabor滤波器)。需要注意的是,我们在学习的过程中不能停留在定义上,而是应该着重理解各种滤波背后的**工作原理及应用范围**,如低通滤波可用于消除噪声、高通滤波常用于提取边缘,又比如高斯滤波就是用来去除高斯噪声、

均值滤波和中值滤波有助于去除胡椒噪声、边滤波则能够在滤波的同时保证一定的边缘信息，但是计算复杂度较高。

- **图像基本运算**：指的是对图像执行一些基本的数学运算。这里涉及到的运算主要可分为**点运算**（线性 & 分段 & 非线性点运算）、**代数运算**（加法 & 减法运算）、**逻辑运算**以及最重要的**几何运算**（图像平移 & 旋转 & 翻转 & 镜像 & 缩放）
- **图像边缘检测**：这里仅需了解下有哪些边缘检测算子以及重点掌握一些常见的算子。如一阶微分算子（Sobel算子 & Roberts算子 & Prewitt算子）、二阶微分算子（Laplacian算子 & LOG算子）及Canny算子。
- **图像形态学操作**：指的是一系列处理图像形状特征的图像处理技术。这里需要着重掌握的有**腐蚀和膨胀**、**开运算与闭运算**、**形态学梯度**（用于保留边缘轮廓）、**白色和黑色顶帽变换**；此外，也可了解下细化、厚化、击中击不中变换、边界/孔洞/联通分量提取。
- **图像变换**：指的是将图像阵列从源域转换到目标域。了解几种常见的变换方式，如**傅里叶变换**、**离散余弦变换**。此外，可以重点学习下用于特征提取的**霍夫变换**，实战下如何利用该技术进行直线、圆和弧线等局部特征的提取。最后，再重点梳理下傅里叶变换与**小波变换**之间的区别和联系。
- **图像分割**：主要是基于灰度值的不连续和相似的性质将图像划分为前景区域和背景区域。对于不连续的灰度值，常用的方法是边缘检测。而对于相似的灰度，我们一般常用**阈值处理**（局部多阈值 & 全局阈值 & Otsu自适应阈值）、**区域生长**、**分水岭算法**等。
- **图像质量评价**：主要是对图像的某些特性进行分析研究，评估出图像的失真程度。这里需要重点掌握几个评价指标：**SSIM**（结构相似度）、**PSNR**（峰值信噪比）及**MSE**（均方误差）。

机器学习与深度学习基础

机器学习

机器学习是一门多领域交叉学科，涉及概率论、统计学、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。

- 了解并区分什么是监督学习、无监督学习、半监督学习、弱监督学习、多示例学习、迁移学习、元学习、强化学习、对比学习、少样本学习、零样本学习；
- 了解参数与超参数、数据拟合（欠拟合、过拟合、Under fit）、偏差与方差、训练集/验证集/测试集、生成模型与判别模型、奥卡姆剃刀/丑小鸭定理/没有免费午餐、样本统计（TP/TN/FP/FN）、交叉验证、参数搜索的概念；理解损失函数、梯度下降、正则化（L1 & L2）、数据降维（PCA & LDA）、数据归一化（Min-Max 标准化 & Z-score标准化）等原理及适用场景；
- 掌握机器学习的十大基础算法，即Linear Regression、Logistic Regression、LDA、LVQ、Naive Bayes、KNN、Random Forest、Decision Tree、SVM、Bagging&Boosting&AdaBoost及K-Means算法。梳理各种树模型（GBDT & XGBoost & RF）之间的原理和区别；

- 理解掌握常用的评价指标，如Accuracy、Recall (Sensitivity)、Precision、Dice (F1-score)、Jaccard、AUC曲线、P-R曲线、MIoU等。
- 学有余力的可以了解下期望最大化、隐马尔科夫模型及条件随机场等原理。

深度学习

深度学习严格意义上是属于机器学习的一个分支，旨在构建更深层的网络来进行表征学习。本文主要介绍与计算机视觉相关的一些概念及技术。

- 了解相关的一些**高频名词**和计算机视觉有哪些应用方向；
- **卷积结构**： Plain Conv, DW Conv, Ghost Conv, Deformer Conv, Octave Conv, Gropu Conv, 1×1 Conv, HetConv, Dilated Conv, SCCov, Pyramid Conv, Tined Conv, Dynamic Conv, Decoupled Dynamic Conv, Involution, etc.
- **激活函数**： Sigmoid, Softmax, ReLU, Tanh, PReLU, ELU, GELU, SELU, LeakeReLU, Soft Plus, ACON, etc.
- **损失函数**： Binary/Weighted/Balanced Cross-Entropy, Mixed Focal/Focal Loss/GFocal Loss, Dice Loss, IoU/CIoU/GIoU/DIoU/CDIoU/EIoU Loss/Focal-EIOU Loss, Tversky Loss, Triplet loss, etc.
- **池化相关**： Max pooling, Average pooling, GAP, Stochastic pooling, etc.
- **优化方法**： Adam、SGD、Adentum、Nesterov、AdaDelta、RMSprop, etc.
- **正则化技术**： Dropout, DropBlock, Label Smoothing, SelfNorm & CrossNorm, etc.
- **后处理技术**： Watershed algorithm、CRF、TTA、Overlap Prediction、NMS, etc.
- **归一化技术**： BN, LN, GN, IN, BGN, SwitchableN, SBN, SSN, FRN, EBN, KalmanN, DualN, etc.
- **学习率衰减**： ReduceLROnPlatea, MultiStepLR, ExponentialLR, CosineAnnealingLR, StepLR, etc.
- **注意力机制**： SE, SK, Shuffle Attention, Non-local, CBAM, GC, OCR, CBM, CBAM, BA2M, FCANet, Coordinate Attention, etc.
- **多尺度机制**： ASPP, SPP, Big-Little, Inception, SFA, ASC, DCFAN, etc.
- **特征可视化技术**： CAM, Grad-CAM, Grad-CAM++, Smooth Grad-CAM++, score-CAM, ss-CAM, Ablation CAM, etc.
- **数据增强**
 - 几何增强： Horizontal/Vertical flip, Rotation, Affine transformation, Translation, Cropping, Perspective transformation, Zoom, etc.
 - 色彩增强： Contrast, Brightness, Saturation, Color space conversion, Color jitter, Channel shuffling, Filling, Superimposed noise, etc.
 - 其它增强： Mixup, RandAugment, mosaic, dropout, cutout, cutmix, augmix, MoEx, RandErase, ObjectAug, InAugment, KeepAugment, Co-Mixup, ISDA, etc.

- **距离度量公式:** Manhattan Distance, Euclidean Distance, Chebyshev Distance, Minkowski Distance, Cosine Distance, Mahalanobis Distance, Hamming Distance, Edit Distance, Earth Mover's distance, etc.
- **Backbone:** LeNet-5, AlexNet, VGGNet, GoogleNet, ResNet, DenseNet, VoVNet, MoblieNet, ShuffleNet, Xception, queezeNet, RexNeXt, Res2Net, SENet, SKNet, DCNet, CSPNet, FBNet, EfficientNet, RegNet, ResNeSt, ReXNet, HaloNets, etc.
- **语义分割:** FCN, UNet, ENet, ThunderNet, RefineNet, SegNet, PSPNet, DeepLab, DenseASPP, OCRNet, HRNet, BiSeNet, etc.
- **目标检测:**
 - **Two-stage:** R-CNN & SPP & FastR-CNN & Faster R-CNN & Cascade-RCNN & Sparse R-CNN, etc.
 - **One-stage:** YOLO v1-v5 & PPYOLO & SSD & RetinaNet & RefineDet & YOLOR & YOLOF & YOLObite & NanoDet & OneNet, etc.
 - **Anchor-free:** CornetNet & Objects as Points & CenterNet v1-v2 & FCOS, etc.
- **实例分割:** MaskRCNN, PolarMask, PolarMask++, PointRend, BlendMask, ISTR, SOLO v1-v2, Sparse RCNN, A2Net, etc.
- **自监督学习:** SimCLR, SimSiam, BYOL, SwAV, MoCo v1-v3, OBoW, DINO, etc.
- **生成对抗网络:** GAN, DCGAN, Conditional GAN, InfoGAN, BigGAN, WGAN, StyleGAN, CycleGAN, Pix2Pix2, StackGAN, LSGAN, CGGAN, PD-GAN, etc.
- **重特征参数化:** ACNet v1-v2, DBBNet, RepVGG, ResRep, etc.
- **Transformer相关** ViT, DETR, METR, SETR, DeiT, TNT, CrossViT, Swin Transformer, LeViT, RVT, PVT, BoTNet, TrTr, MOTR, ISTR, TransGAN, Local-ViT, IPT, DeepViT, CoTr, CaiT, CeiT, PiT, ViViT, CvT, T2T-ViT, TransT, SiT, LV-ViT, MViT, PRTR, CoaT, Segmenter, etc.
- **多层感知机相关:** MLP, MLP-Mixer, ExternalAttention, RepMLP, ResMLP, gMLP, etc.

上面只列举了一小部分方向，对于CV其它领域的总结大盘点和一些训练技巧、常见踩坑点、模型部署等知识点，以及上面所涉及到的每一个知识点，我们均会以图文讲解的形式从原理、算法到具体的应用进行一一讲解和总结。

硬基础

硬基础，可以理解为求职申学的加分项，优秀的履历是一张去往金字塔顶端的入场卷。

学校 & 学历 & 年龄 & 性别

- **学校:** 实际上，大多数的在校生或多或少都具有名校情结。一个名牌学校的毕业生，除了学校品牌的价值，更多的体现在教学资源、硬件资源以及优质的校友资源等。当然，名校毕业的学生在享受以上资源的同时，也常常会倍感压力，这是由于社会对于此类群体的期许度过高。在我

国，名校按大类来分，大致可分为985 | 211 | 双一流；如果想继续分，可分为C9 + 中下流985 + 211 + 其它；再细分还可以是清北+ 华五 + 国防七子 + 两电一邮 + 两外一法 + ...。当然，这更多的只是历史弥留的痕迹，相信随着时代的发展，会有越来越多好学校崛起，比如计算机专业较强的双非院校中发展势头最猛的深圳大学、南方科技大学、杭电南邮等以及苏州大学和上海大学等强劲211高校。对于大陆境外来说，常年排名靠前的有美国常春藤联盟+Stanford+CMU+MIT+UIUC，英国剑桥+牛津，瑞士苏黎世理工，新加坡国立大学+南洋理工大学，香港大学+香港中文大学+香港科技大学等。虽然，在我们求职或升学的面试过程中，好的学校通常会更容易获得HR的青睐，但这并不意味着其它学校名气不够好的同学就没有任何机会，对于学校出身不好的人也不用妄自菲薄，更多的应该把时间和精力花在提升自己的软技能上，而对于学校出身好的同学来说更应该珍惜这份机会，否则会适得其反。

- **学历：**对于互联网这个行业而言，对学历的要求可谓越来越高。特别是针对互联网大厂来说，每年大部分招的应届生有大部分都是研究生学历起步。特别是对于大厂AI Lab部门来说，现在基本都是海内外名校博士起步。当然，对于做开发的同学来说，只要把计算机基础打好以及Leetcode等算法题刷得快准狠，基本不会有太大的问题。而对于想从事CV算法的同学来说，更多的还是建议去修一个研究生学位。而对于想继续进修的同学来说，一定要在申请之前了解相关院校的要求，比如雅思托福成绩和就读学校门槛，并在截至日期到来之前将相关材料都准备好，一般来说，对于大多数人更多的还是建议国内读研国外读博。
- **年龄：**年龄这块其实是个很敏感的话题。可以发现在我国随着年龄越大，限制越多。最典型的莫过于某菊厂带头的“35岁优化”以及某福报厂所提倡的“为社会输送优质人才”等。对于高校老师而言，其实也会面临着严峻的考核要求，而且很多国家青年基金项目基本都要求35岁以下申请。对于想考公务员的同学同样也有本科35硕士40的年龄限制。因此，对于想进一步做出抉择的同学来说也需要考虑到未来的一个发展趋势和自身的竞争力，未雨绸缪。
- **性别：**俗话说男女有别，旧社会更有男尊女卑的说法。随着我国改革开放进入到新时代，人们对于这种观念已大大改善。尽管现如今职场上或多或少还存在着少部分的歧视女性，但对于大多数的公司而言，特别是对于互联网算法和开发部分而言，女生在团队里往往是最受照顾的，更多时候福利倾向原则反而是女士优先，所以这点大可放心。

业务理解 & 逻辑沟通能力

- **业务理解能力：**也可以理解为解决问题的能力。这个在工业界其实是比较注重的，特别是社招。对于大厂的大部分算法开发岗位来说，面试过程也经常会遇到这种问题，即给定一个场景，问候选人有什么相应解决方案，这类问题大都是开放题，没有标准答案。然而，这对于候选人来说其实是一个更大的挑战。遇到这种情况，我们首先要理解问题，清楚问题的定义是什么，这其实就是在工作中我们需要明确我们的客户需求是什么？只有明确了任务本身，下一步才是对问题进行分析，根据自身掌握的知识体系结合实际的应用场景给定一个最优的解决方案，最好是要能够落实到每一步应该做什么，尽可能地思考开发过程或应用过程中会出现哪些问题，应该如何解决。最后，才是根据最终所敲定的任务路线，应用已有的技术逐步地攻关解决。
- **逻辑沟通能力：**在我们的日常学习或者工作生活中，沟通是必不可少的。良好的沟通能力有助于团队之间更有效率的协作开发，强大的逻辑能力则更有助于个人解决问题。除了在找工作的过程中我们经常会遇到跟面试官交流或者跟HR谈薪外，在进行学术汇报和同行交流的时候也经

常需要我们有良好的逻辑沟通能力。这不仅能够提高一个人的自信，也能为我们的职场生涯增添筹码。

实习 & 工作经验

- **实习经验：**对于许多的在校大学生而言，接触实际的工程项目开发，熟悉公司的项目开发流程，掌握常用的开发工具，可能一个最好的办法就是去大厂实习。现如今大家都知道秋招，特别是春招的时候，竞争大厂CV算法工程师岗位往往是千军万马过独木桥。有一个很重要的原因是因为僧多粥少，这是由于每年大部分正式岗位的指标都留给实习生转正了。所以在这里奉劝大家一句，能去大厂实习的一定要去，不能出去实习的也要创造机实习，比如远程实习。实在不行的只能留在学校自己上github等相关网站上找一些开源项目学习。
- **工作经验：**正所谓学历只是一个敲门砖，它决定的是你能否进入面试。而工作经验，或者说项目经验，则是决定你在整个行业的天花板。一般来说，项目经验所丰富，职业生涯就能走的更远。同时，随着你工作年限的增长，业界有一个不成文的规则，即跳槽的时候一般薪水是会增长x%，有甚至直接翻n倍。除非你本来已经到达天花板，工资也足够高了。当然，如果能加上优质的学历和学校加持，你的升职加薪之旅也会加速。

物质资源 & 人脉资源化

- **物质资源：**这里可以暂且理解为硬件资源。众所周知，深度学习模型成功的三大要素便是数据、算法和算力。其中，算力是从根本上决定了我们是否能进行某个任务或课题的研发或研究。对于在校生成来说，首要关键的还是实验室的硬件基础设施，服务器资源是否能满足课题的需求，因为这同时也会关乎到你实验的进展。而对于进入工业界的朋友来说，如果遇到买不起足量服务器的公司千万不要去应聘该公司的算法工程师岗位。说一个身边真实的例子，有个别同学的导师要求该名同学用自己的破笔记本跑3D任务相关的实验以及某某公司为该公司旗下的算法部分分配每人一张卡。可想而知，这种开发和研究周期得有多长。大家也可以观察下，现在头部大厂如谷歌的最新研究基本都是基于多张V100甚至直接上TPU集群上训练。
- **人脉资源：**无论你今后是想从事工业界或者学术界，可以说人脉资源是你必不可少的东西。首先，简单地聊一下学术圈。学术圈这个圈子其实挺小的，包括大家平日里投的期刊，其实绝大多数都是不盲审的，这就意味着可操作性空间大。特别是对于某些期刊来说，很多时候就是一群自己人在圈子内“自嗨”，外人根本进不去（投稿基本是被拒的）。所以，这里更建议大家多去参加一些高级的学术会议，结识多点人脉，建立自己的小圈子，众人拾柴火焰高，唯有抱团取暖才是真谛。同样地，对于跟着大牛导师的同学也可以多利用导师手里边的资源为自己争取更好的机会。其次，对于工业界的朋友来说，平常也可以多参加点创投圈，结交多点大佬，不仅以后跳槽挖坑的机会更多，对于有志于创业的伙伴来说未尝不是一种积累。

工具链

工具链，主要是向大家介绍一下日常工作学习生活中可以帮助自身提升工作效率的一些在线网站、实用工具、资讯媒介以及干货资料分享。

在线网站

- 数据竞赛
 - **Kaggle**^[2]: 全球最大的机器学习/深度学习竞赛平台
 - **DataCastle数据竞赛**^[3]: 全国专业的大数据培训和竞赛平台
 - **天池**^[4]: 由阿里巴巴举办, 面向全球科研工作者的算法竞赛平台
 - **KDD CUP**^[5]: 国际知识发现和数据挖掘竞赛
 - **Grand Challenge**^[6]: 比较全面地收集了医学图像相关的竞赛、数据集
 - **Biendata**^[7]: 国内很大的数据竞赛网站和社区
- 算法刷题
 - **Leetcode**^[8]: 题型多题目多且经典, 在国内绝对是头部的刷题网站了
 - **牛客网**^[9]: 求职找工作神站, 笔试\面试\实习\求职等各类资源集大成者
- 求职就业
 - **Boss直聘**^[10]: 找工作/实习, 打工是不可能打工的, 这辈子都不可能打工的
 - **牛客网**^[11]: 求职找工作神站, 笔试\面试\实习\求职等各类资源集大成者
 - **超级简历**^[12]: 在线制作属于你的个人简历
- 学术相关
 - **Google学术**^[13]: 论文搜索神器, 活用高级搜索功能
 - **arXiv**^[14]: 新鲜出炉的论文, 当然质量也是参差不齐了
 - **SCI-Hub**^[15]: 找论文必备
 - **dblp**^[16]: 是计算机领域内对研究的成果以作者为核心的一个计算机类英文文献的集成数据库系统, 往往可以比谷歌学术更快的搜到文献的引用
 - **Letpub**^[17]: 查看期刊的IF, 中科院分区等信息
 - **Connectedpapers**^[18]: 根据一篇论文找出其最相关的论文的图谱, 帮你快速发掘出论文创新点背后的其他论文!
 - **Papers With Code**^[19]: 帮助机器学习爱好者跟踪最新发布的论文及源代码, 快速了解最前沿的技术进展
 - **Overleaf**^[20]: 在线的论文latex编辑器, IEEE都推荐的编辑器~
 - **aideadlin.es**^[21]: AI会议deadline
- 新闻资讯
 - **知乎**^[22]: 蟹邀, 人在美国, 刚下飞机, CV水很深, 利益相关, 匿了。
 - **Quora**^[23]: 国外版知乎, 也有很多圈内大佬在这里谈笑风生
 - **Medium**^[24]: 优质的外文博客网站
 - **CSRankings: Computer Science Rankings**^[25]: 计算机各领域的高校/研究所排名
- 其它
 - **RGB配色表**^[26]: 在线RGB配色表, 可用于快速获取目标颜色的RGB值

- **Bilibili**^[27]: 这里面个个都是人才, 说话又好听, 我敲喜欢这里的
- **Github**^[28]: 全球最大的同性交友网站, 世界上资源最丰富的代码仓库
- **Graviti**^[29]: 专注于解决AI开发中的数据痛点, 从海量公开数据集社区 (Open Datasets) 到专业数据管理SaaS (TensorBay)
- **Learn Git Branching**^[30]: 用玩游戏的形式学会Git, 带领初学者快速的入门
- **jsoneditoronline**^[31]: 是一个简单、灵活、可视化在线的JSON编辑器, 支持差异化对比, 可查看、编辑和格式化JSON数据

实用工具

• 常用开发工具

- **Pycharm**: Python项目开发神器, 集成调试、语法高亮、Project管理、代码跳转、智能提示、自动完成、单元测试、版本控制等强大功能;
- **VSCode**: 便携式代码编辑器, 优点是短小精悍, 除了可以作为一款轻量级的编辑器, 也可作为中小型项目开发的首选工具, 除此之外还可以当作Latex编译器;
- **MobaXterm**: 是一款可以在Windows系统下远程SSH连接到Linux服务器的神器;
- **Jupyter Notebook**: 是一个开源的Web应用程序, 允许用户创建和共享包含代码、方程式、可视化和文本的文档。它的用途包括: 数据清理和转换、数值模拟、统计建模、数据可视化、机器学习等等。

• 论文写作工具

- **Mendeley**: 科研文献管理软件, 可以给文献加注释和分类, 同时也可以把参考文献按照指定格式进行导出;
- **EndNote**: 主要是用于Word写期刊论文时可以很方便的管理参考文献的自动引用;
- **Mathpix Snipping Tool**: Latex公式离线截屏工具, 只需通过软件进行截屏复制Latex代码即可编译运行;
- **Typora**: 轻量级的markdown编辑器, 同时附带许多有用的插件可供使用;

• 日常办公工具

- **Snipaste**: 是一款免费, 完全没有广告的截图工具, 具有强大的桌面贴图功能;
- **火柴**: 桌面搜索引擎, 可以快速帮你找到需要的文件, 同时附带许多实用的小工具;
- **V2ray, clashX, ssr**: fq必备软件, 领略世界文化;
- **向日葵 & Teamviewer**: 远程连接工具;
- **Adobe acrobat**: 功能全面的pdf阅读器;
- **IrfanView**: 轻量级图像浏览器, 可以实现图片编辑、压缩等功能
- **石墨文档**: 稳定、免费的在线文档, 记笔记、团队协作都很方便;
- **OneNote**: Windows官方记事本软件, 功能强大, 界面清新, 非常适合个人。不过由于服务器在国外, 所以多终端同步问题目前来说体验不是很好;

- **Picasa 3:** 谷歌开发的照片查看器，可用于快速的浏览照片，无广告，然而很早前已停止开发，不过不影响我们的日常使用；
- 坚果云/蓝奏云：不限速的云盘，免费版一个月的流量用来不同设备之间同步文件足够了
- **谷歌浏览器插件：**
 - 沙拉查词：在线划词实时翻译；
 - **Grammarly:** 在线语法自动纠正插件；
 - **Tempermonkey:** 俗称油猴，有丰富的js脚本，让你的浏览器更加强大；
 - **Octree:** 可以让github网页端的文件以树形结构显示；
 - **Adblock:** 广告拦截神器，拒绝牛皮癣，还你一个干爽的面页；

论文资讯

各大CV相关的顶会顶刊

- CCF A
 - CVPR (CV最高顶会), ICCV, AAAI, ICML, NIPS, ACMM
- CCF B
 - ECCV, WACV, ICME
- CCF C
 - BMVC, ACCV, ICPR
- JCR I 区
 - TPAMI (CV顶级期刊), TIP, IJCV
- JCR II 区
 - Pattern Recognition, IEEE Access (保毕业神器)
- JCR III 区
 - CVIU, IVC, PRL
- JCR IV 区
 - IET-CVI, IJRAI, Machine Vision and Applications
- AI+医疗交叉领域顶级会议和期刊
 - 会议: MICCAI, IPMI, ISBI, EMBC
 - 期刊: TMI, MIA, TBME
- **AI自媒体微信公众号平台：**
 - 机器之心：专业的人工智能媒体和产业服务平台；
 - 新智元：重点关注人工智能、机器人等前沿领域发展；

- 量子位：追踪人工智能新趋势，报道科技行业新突破；
- **CVHub**：专注于发展成为计算机视觉领域的全栈平台，涵盖与CV相关的技术交流分享、学术论文写作指导、求职升学经验分享、实用干货教程分享以及最新的CV资讯速递；
- AI科技大本营：CSDN旗下的官方公众号，提供人工智能领域热点报道；
- OpenCV团队：致力于OpenCV开发、维护和推广工作；
- AMiner：AI赋能科技情报挖掘；

干货资料

• 在线教程

- **Fast AI**^[32]：重代码而非数学，不少初学者凭借Fast.ai课程所学技能称霸Kaggle比赛！
- 吴恩达**机器学习**^[33]、**深度学习**^[34]实战课程：最受欢迎的入门级课程，吴恩达老师课程最大的特点是通俗易懂；
- 李宏毅**机器学习**^[35]、**深度学习**^[36]实战课程：成为像宝可梦训练师那样的AI工程师，李宏毅老师课程最大的特点是诙谐幽默为主；
- 李飞飞**CS231**^[37]计算机视觉公开课：是计算机视觉领域入门的必看经典课程之一；
- 斯坦福**CS224n**^[38]深度学习自然语言处理课程：CV的同学可以稍微了解一下；
- **Pytorch官网**^[39]：这是学习Pytorch深度学习框架最好的课程资料，包括熟悉Pytorch的各类API的定义和功能以及一些快速入门的教程均可在官网上找到，质量也是杠杠的！

• 参考书籍

- 《python从入门到实战》：掌握Python语言的入门级必备书籍；
- 《统计机器学习》第2版：李航博士的力作，涵盖机器学习必备入门知识点；
- 《西瓜书》：由周志华教授主写，适合本科生看的中文机器学习书籍，关于书中所涉及到的公式推导，建议搭配《南瓜书》进行辅助；
- 《动手学深度学习》：这本书是由李沐大神写的，先前之有MXNet版本，后面也逐渐加入了Pytorch和Tensorflow版本；
- 《百面机器学习》&《百面深度学习》：由葫芦团队编写，机器学习、计算机视觉算法工程师等相关岗位的面试宝典；
- 《数字图像处理》：作者是日本的学者冈萨雷斯，是传统数字图像处理入门的必看书籍，这部分内容建议根据上面第1部分总结的重点有针对性的去了解即可；
- 《OpenCV 轻松入门：面向 Python》：这本书主要是用于辅助你学习opencv的API接口和验证理论实战用的，如果不是刚需也可以去github找一些实战项目亦可；
- 《深度学习》：也称为花书，被评为深度学习领域的圣经，可以帮助我们快速构建知识体系。当然，不少读者反映书中内容大都晦涩难懂，建议学有余力的同学可以去看看；
- 《剑指offer》：总共有 80 道典型的编程面试题，系统整理基础知识、代码质量、解题思路、优化效率和综合能力这 5 个面试要点，是非常适合新手面试刷的第一本题

库；

- 《神经网络与深度学习》：邱锡鹏深度学习经典教材，免费pdf下载可关注微信公众号CVHub，后台发送关键字“0607”获取；
- 《计算机视觉：算法与应用第二版》：计算机视觉经典教材；
- 《C++ Primer》：C++入门必备教材。

文章比较长，准备了PDF版本方便保存与阅读，关注微信公众号**CVHub**，后台发送关键字“**cvhub**”获取~

关注

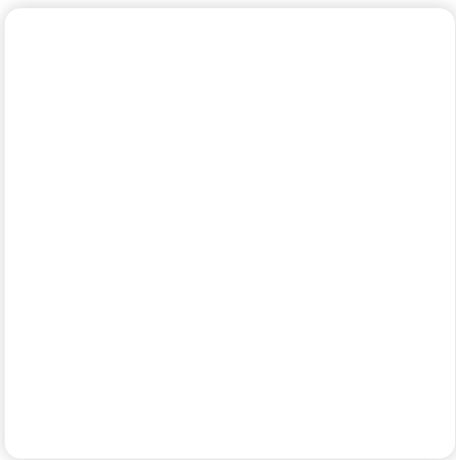


CVHub

门头沟学院AI视觉实验室御用公众号 | 学术 | 科研 | 就业

127篇原创内容

公众号



长按二维码添加CVHub个人微信加入学术交流群

References

- [1] <https://www.runoob.com/python3/python3-tutorial.html>
- [2] <https://www.kaggle.com/>
- [3] <https://js.dclab.run/v2/index.html>
- [4] <https://tianchi.aliyun.com>
- [5] <http://www.kdd.org>
- [6] <https://grand-challenge.org>
- [7] <https://www.biendata.xyz>
- [8] <https://leetcode-cn.com>
- [9] <https://www.nowcoder.com/>
- [10] <https://www.zhipin.com>
- [11] <https://www.nowcoder.com>
- [12] <https://www.wondercv.com>
- [13] <https://scholar.google.com.hk>
- [14] <https://arxiv.org>
- [15] <http://ww25.sci-hub.tw>
- [16] <https://dblp.uni-trier.de>
- [17] <https://www.letpub.com.cn>
- [18] <https://www.connectedpapers.com>
- [19] <https://paperswithcode.com>
- [20] <https://www.overleaf.com>
- [21] <https://aideadlin.es>
- [22] <https://www.zhihu.com>
- [23] <https://www.quora.com>
- [24] <https://medium.com>
- [25] <http://csrankings.org>
- [26] <http://www.wahart.com.hk/rgb.htm>
- [27] <https://www.bilibili.com>
- [28] <https://github.com>
- [29] <https://www.graviti.cn>
- [30] <https://learngitbranching.js.org>
- [31] <https://jsoneditoronline.org>
- [32] <https://course.fast.ai>
- [33] <https://study.163.com/course/courseLearn.htm?courseId=1004570029>
- [34] https://mooc.study.163.com/university/deeplearning_ai
- [35] <https://www.bilibili.com/video/BV1JE411g7XF>
- [36] <https://www.bilibili.com/video/av94519857>
- [37] <https://www.bilibili.com/video/av77752864/>
- [38] <http://web.stanford.edu/class/cs224n/index.html>
- [39] <https://pytorch.org>

收录于合集 #深度学习 90

上一篇

下一篇

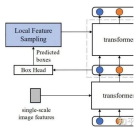
【卷积系列】CondConv | 跳脱出通道的限制，从Batch维度进行卷！

一文看尽深度学习中的各种注意力机制

喜欢此内容的人还喜欢

ECCV | 数据高效的Transformer目标检测器

计算机视觉研究院



ArcGIS中如何将两个面图层范围调整/磨合一致？

GIS信息化技术



Fred说KUKA系列篇--KUKA电池组

Madrigal Robot

