



基本理论 + 经典案例详解深度学习在图像处理中的应用

深度学习技术

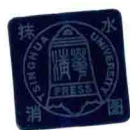
图像处理入门

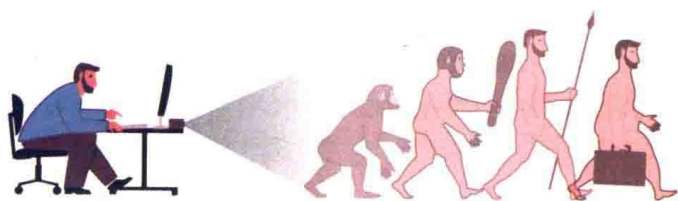
· 杨培文 胡博强 著 ·



理论部分通过具体的简易程序代码讲解取代大量的公式推导
实际应用部分大量讲解参数调整、模型优化的技巧
所有代码均可在云GPU服务器基于容器技术直接执行使用

清华大学出版社





· 杨培文 胡博强 著 ·

深度学习技术 图像处理入门



清华大学出版社
北京

内 容 简 介

本书从机器学习、图像处理的基本概念入手，逐步阐述深度学习图像处理技术的基本原理以及简单的实现。继而以几个实战案例来介绍如何使用深度学习方法，在数据分析竞赛中取得较高的排名。最后，通过一个实战案例，介绍如何将模型放入 iOS 程序，制作相应的人工智能手机App。

本书适用于对深度学习有兴趣、希望入门这一领域的理工科大学生、研究生，以及希望了解该领域基本原理的软件开发人员。此外，本书所有案例均提供了云环境上的代码，便于读者复现结果，并进行深入学习。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目（CIP）数据

深度学习技术图像处理入门/杨培文,胡博强著. —北京:清华大学出版社, 2018
ISBN 978-7-302-51102-1

I. ①深… II. ①杨… ②胡… III. ①图像处理软件 IV. ①TP391.413

中国版本图书馆CIP数据核字（2018）第195622号

责任编辑：王金柱

封面设计：王 翔

责任校对：闫秀华

责任印制：宋 林

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>, <http://www.wqbook.com>

地 址：北京清华大学学研大厦A座

邮 编：100084

社 总 机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者：北京天颖印刷有限公司

经 销：全国新华书店

开 本：190mm×260mm

印 张：16.75

字 数：422千字

版 次：2018年10月第1版

印 次：2018年10月第1次印刷

定 价：69.00元

产品编号：077101-01

序

深度学习是人工智能领域一种新兴的关键技术，其概念由多伦多大学 Hinton 等人于 2006 年提出，含多隐含层的人工神经网络感知器就是一种深度学习结构。深度学习具有很强的学科交叉性，涉及认知科学、脑科学、生物学、数学、运筹学、控制、计算机、通信、大数据、云计算、语言学乃至人文科学、社会科学等学科。近年来，深度学习无论是在基础研究还是工程应用方面都完成了很多创新性工作，例如有监督卷积神经网络、无监督深度置信网络、第一个战胜人类围棋世界冠军的人工智能程序 AlphaGo、人脸识别、语音识别、自然语言处理等。GPU、人工智能芯片等硬件技术的突破，让人们对于深度学习的实用化增添了信心。作为全球性的前沿研究问题，深度学习已经受到众多科学家、工程师的关注，其应用几乎涉及工程、金融、法律、文化、娱乐、艺术等社会生活的各个方面，多篇里程碑意义的论文在 Nature、Science 等国际顶尖学术刊物上横空出世，显示出深度学习技术旺盛的生命力和广阔的应用前景。

呈现在读者面前的《深度学习技术图像处理入门》一书，在兼顾介绍深度学习基础理论的同时，更侧重于记录作者本人在学习深度学习技术时的历程、感悟和经验，包括难免走过的弯路，作者也坦承地娓娓道来。通读全书，与同类书籍比较，我认为该书具有“亦师亦友、寓教于乐”的特点：入门部分介绍得通俗易懂，文风严谨亦不失活泼，让高深的理论不那么神秘、枯燥，使得读者更有阅读的乐趣和自信，易于沉浸其中，非常适合作为自学的教材和工具书；案例部分的代码，均基于实际的竞赛，偏重于应用，使得读者也更有兴趣参与这样的竞赛，以“小试牛刀”。

《深度学习技术图像处理入门》是一本能够帮助读者很快掌握和应用深度学习技术进行图像处理的工具书，对生物、医学、安防、机器人等众多图像处理领域的工程应用问题具有很大的参考价值。在我国高度重视人工智能技术的发展、将其上升为国家战略的时代背景下，相信该书的出版，对于推动我国深度学习技术的培育、推广、应用具有积极的作用。

中国农业大学博士生导师 陈建

2018 年 8 月 1 日

前言

回想 2017 年 4 月，当清华大学出版社的编辑找到杨培文和我，商量着写一本与深度学习相关的书时，我还是比较缺乏信心的。首先，自己本专业是基因组学，或者说是生物学，机器学习方面的知识都是自学的。其次，我根本就没有写过书，由我参与撰写，可能是班门弄斧，内容有误都是小事，万一写的内容给读者灌输了错误的观念、在大方向上误导了初学阶段的读者，实在是难辞其咎。

出版社方面同样了解我们的情况，跟我们说出版社这次想出一本面向非数学、计算机相关专业的书，希望语言更加通俗易懂，例子更贴近实际项目，让非专业出身的人看了以后，对机器学习、图像处理以及深度学习三者有一个最基本的认识。这里，我经常向生物、医学专业背景的人解释机器学习模型的原理，而培文则有多次数据分析竞赛名列前茅的经历，因此出版社希望我们两位尝试一下。

所以接下来编写书籍的过程中，我们的定位就是相比现在市面上主流的相关书籍，前几章写得更加通俗，把入门的门槛再降低一些；然后后面的章节基于参加数据分析竞赛的实战过程，把最终的目标再定高一些；最后我们的配套代码以及环境（<http://github.com/Jinglue/DL4Img>）要让初学者可以很容易地跑起来，把书籍的内容落在实际运用中。

我们希望这本书可以让非科班出身的读者快速了解深度学习的基本原理，将相关技术举一反三，运用在自己的课题、项目中。以我自己为例，在书籍编写完成后的审阅过程中，我仔细阅读了培文撰写的运用循环神经网络进行验证码识别这一章节（第 10 章）的内容，后来参加百度 AI 挑战赛时，最初的模型就是培文整理的配套代码，后来经过调整，最后取得了第二名的成绩。

最后一点，阅读本书，需要读者具有基本的 Python 编程基础，以及科学计算相关模块的了解。这部分内容本书并未涉及，但读者可以通过斯坦福大学 cs228 相关配套入门习题进行简单的了解，我们对此进行了汉化（<https://jizhi.ai/blog/post/cs228-py>）。

在此感谢景略集智的王文凯、柯希阳在书籍编写过程中提供的帮助。

胡博强

2018 年 7 月 18 日

目 录

第 1 章 搭建指定的开发环境	1
1.1 为什么要使用指定的开发环境	1
1.2 硬件准备	2
1.2.1 在亚马逊租用云 GPU 服务器	2
1.2.2 在腾讯云租用 GPU 服务器	4
1.2.3 在云服务器中开启搭载开发环境的 Docker 服务	8
1.3 软件准备	9
1.3.1 在 Ubuntu 16.04 下配置环境	9
1.3.2 在 CentOS 7 下配置环境	12
1.4 参考文献及网页链接	12
第 2 章 温故知新——机器学习基础知识	13
2.1 人工智能、机器学习与深度学习	13
2.2 训练一个传统的机器学习模型	15
2.2.1 第一步，观察数据	16
2.2.2 第二步，预览数据	17
2.3 数据挖掘与训练模型	29
2.3.1 第一步，准备数据	29
2.3.2 第二步，挖掘数据特征	31
2.3.3 第三步，使用模型	37
2.3.4 第四步，代码实战	44
2.4 参考文献及网页链接	49
第 3 章 数形结合——图像处理基础知识	50
3.1 读取图像文件进行基本操作	51
3.1.1 使用 python-opencv 读取图片	51

3.1.2	借助 python-opencv 进行不同编码格式的转换	52
3.1.3	借助 python-opencv 改变图片尺寸	53
3.2	用简单的矩阵操作处理图像	53
3.2.1	对图像进行复制与粘贴	53
3.2.2	把图像当成矩阵进行处理——二维码转换成矩阵	54
3.3	使用 OpenCV “抠图”——基于颜色通道以及形态特征	59
3.4	基于传统特征的传统图像分类方法	64
3.4.1	将图片简化为少数区域并计算每个区域轮廓特征的方向	66
3.4.2	将 HOG 变换运用在所有正负样本中	68
3.4.3	训练模型	70
3.4.4	将训练好的分类器运用在新的图片中	71
3.5	参考文献及网页链接	73
第 4 章	继往开来——使用深度神经网络框架	74
4.1	从逻辑回归说起	74
4.2	深度学习框架	76
4.3	基于反向传播算法的自动求导	77
4.4	简单的深度神经网络框架实现	80
4.4.1	数据结构部分	81
4.4.2	计算图部分	83
4.4.3	使用方法	85
4.4.4	训练模型	86
4.5	参考文献及网页链接	89
第 5 章	排列组合——深度神经网络框架的模型元件	90
5.1	常用层	92
5.1.1	Dense	92
5.1.2	Activation	92
5.1.3	Dropout	93
5.1.4	Flatten	94
5.2	卷积层	94
5.2.1	Conv2D	94
5.2.2	Cropping2D	101
5.2.3	ZeroPadding2D	101
5.3	池化层	102
5.3.1	MaxPooling2D	102

5.3.2	AveragePooling2D	102
5.3.3	GlobalAveragePooling2D	103
5.4	正则化层与过拟合	104
5.5	反卷积层	105
5.6	循环层	109
5.6.1	SimpleRNN	109
5.6.2	LSTM	109
5.6.3	GRU	110
5.7	参考文献及网页链接	110
第 6 章	少量多次——深度神经网络框架的输入处理	112
6.1	批量生成训练数据	113
6.2	数据增强	115
6.3	参考文献及网页链接	117
第 7 章	愚公移山——深度神经网络框架的模型训练	118
7.1	随机梯度下降	119
7.2	动量法	120
7.3	自适应学习率算法	121
7.4	实验案例	124
7.5	参考文献及网页链接	128
第 8 章	小试牛刀——使用深度神经网络进行 CIFAR-10 数据分类	129
8.1	上游部分——基于生成器的批量生成输入模块	131
8.2	核心部分——用各种零件搭建深度神经网络	131
8.3	下游部分——使用凸优化模块训练模型	132
8.4	参考文献及网页链接	133
第 9 章	见多识广——使用迁移学习提升准确率	134
9.1	猫狗大战 1.0——使用卷积神经网络直接进行训练	135
9.1.1	导入数据	135
9.1.2	可视化	137
9.1.3	分割训练集和验证集	138
9.1.4	搭建模型	140
9.1.5	模型训练	141
9.1.6	总结	142
9.2	猫狗大战 2.0——使用 ImageNet 数据集预训练模型	142

9.2.1	迁移学习	142
9.2.2	数据预处理	143
9.2.3	搭建模型	143
9.2.4	模型可视化	144
9.2.5	训练模型	145
9.2.6	提交到 kaggle 评估	146
9.3	猫狗大战 3.0——使用多种预训练模型组合提升表现	146
9.3.1	载入数据集	147
9.3.2	使用正确的预处理函数	147
9.3.3	搭建特征提取模型并导出特征	147
9.3.4	搭建并训练全连接分类器模型	148
9.3.5	在测试集上预测	149
9.4	融合模型	150
9.4.1	获取特征	150
9.4.2	数据持久化	151
9.4.3	构建模型	151
9.4.4	在测试集上预测	152
9.5	总结	153
9.6	参考文献及网页链接	154
第 10 章	看图识字——使用深度神经网络进行文字识别	155
10.1	使用卷积神经网络进行端到端学习	155
10.1.1	编写数据生成器	157
10.1.2	使用生成器	157
10.1.3	构建深度卷积神经网络	158
10.1.4	模型可视化	158
10.1.5	训练模型	160
10.1.6	计算模型总体准确率	161
10.1.7	测试模型	161
10.1.8	模型总结	162
10.2	使用循环神经网络改进模型	162
10.2.1	CTC Loss	163
10.2.2	模型结构	164
10.2.3	模型可视化	165
10.2.4	数据生成器	167
10.2.5	评估模型	168

10.2.6	评估回调	169
10.2.7	训练模型	169
10.2.8	测试模型	171
10.2.9	再次评估模型	171
10.2.10	总结	173
10.3	识别四则混合运算验证码（初赛）	173
10.3.1	问题描述	174
10.3.2	数据集探索	174
10.3.3	模型结构	176
10.3.4	结果可视化	181
10.3.5	总结	182
10.4	识别四则混合运算验证码（决赛）	183
10.4.1	问题描述	183
10.4.2	数据集探索	184
10.4.3	数据预处理	186
10.4.4	模型结构	192
10.4.5	生成器	195
10.4.6	模型的训练	197
10.4.7	预测结果	198
10.4.8	模型结果融合	199
10.4.9	其他尝试	200
10.4.10	小结	202
10.5	参考文献及网页链接	203
第 11 章	见习医生——使用全卷积神经网络分割病理切片中的癌组织	205
11.1	任务描述	205
11.1.1	赛题设置	205
11.1.2	数据描述	206
11.1.3	数据标注	206
11.2	总体思路	206
11.3	构造模型	207
11.3.1	准备数据	208
11.3.2	构建模型	214
11.3.3	模型优化	217
11.4	程序执行	225
11.5	模型结果可视化	226

11.5.1	加载函数	226
11.5.2	选择验证集并编写预测函数	226
11.5.3	根据 tensorboard 可视化结果选择最好的模型	228
11.5.4	尝试逐步降低学习率	230
11.6	观察模型在验证集上的预测表现	231
11.7	参考文献及网页链接	234
第 12 章	知行合一——如何写一个深度学习 App	235
12.1	CAM 可视化	235
12.2	导出分类模型和 CAM 可视化模型	236
12.2.1	载入数据集	236
12.2.2	提取特征	237
12.2.3	搭建和训练分类器	237
12.2.4	搭建分类模型和 CAM 模型	238
12.2.5	可视化测试	239
12.2.6	保存模型	241
12.2.7	导出 mlmodel 模型文件	241
12.3	开始编写 App	242
12.3.1	创建工程	242
12.3.2	配置工程	244
12.3.3	测试工程	249
12.3.4	运行程序	249
12.4	使用深度学习模型	250
12.4.1	将模型导入到工程中	250
12.4.2	数据类型转换函数	250
12.4.3	实施 CAM 可视化	252
12.4.4	模型效果	254
12.5	参考文献及网页链接	255

第 1 章

搭建指定的开发环境

“工欲善其事，必先利其器”。在介绍本书的具体内容之前，我们首先需要进行硬件方面的准备，同时安装必要的软件，进而基于这个开发环境学习本书中的案例代码。

本书所用代码及本章内容对应的开发环境，读者可以在 <https://github.com/Jinglue/DL4Img> 具体查阅。

1.1 为什么要使用指定的开发环境

使用指定开发环境的主要目的是方便读者运行代码。深度学习环境主要基于 Linux 操作系统搭建，此过程需要有 Linux 相关的知识作为铺垫。然而，使用指定的开发环境，可以将这一部分大大简化，使读者直接运行代码，降低入门门槛，否则读者可能会遇到很多无从下手的麻烦，影响主要内容的学习。

首先，确定开发环境的版本。Python 的 Scikit-learn、Tensorflow、Keras 等机器学习、深度学习常用库，很多基本用法在不同版本之间都是不同的。另外，可能一两年后，相比此时 Tensorflow 的语法又有很大不同，到那时读者想使用本书代码的话，就会发现代码运行错误，造成很多困扰。还有这些库的安装过程中，有时并不是安装一个库的问题，同时涉及与系统库

的交互。笔者依稀记得很多年前第一次接触 Python，在大型机的个人用户安装 Scipy 库时，需要通过逐一手动安装系统库来解决包依赖问题，折腾了整整三天。

其次，本书的代码可以在该开发环境下运行。如果读者觉得本书的案例和自己学习工作中的案例比较接近，可以直接使用自己的数据代替书中的数据，或者使用本书的模型快速得到一个和自己工作相关的模型。

最后，本书基础部分（第 2~7 章）的内容，计算量并不大，可以在普通的个人电脑上部署安装，甚至可以打开电脑浏览器，然后直接在景略集智官网里单击与本书内容相关的博客文章，在云端运行程序。当然，本书的案例部分（第 8~11 章）的内容都需要相当的计算量，这就需要准备一台带 GPU 的电脑了，而这样的电脑通常并不便宜。为了方便读者学习，我们专门配置了和本书配套的开发环境，方便读者租用云服务器上手学习。有关电脑配置和云服务器租用方式，详情请见 1.2 节的内容。

1.2 硬件准备

本书主要讨论使用深度学习技术进行图像处理分析，而深度学习技术在模型的训练阶段，需要使用除了 CPU 以外的硬件进行加速才能达到理想的训练速度，实现模型的快速收敛。这些硬件包括 TPU、FPGA、Intel Xeon Phi 处理器等，但目前阶段使用最广泛、开发最简单的还是 GPU，即英伟达（NVIDIA）公司的显卡。

之所以是 NVIDIA 公司的显卡，是因为现在主流深度学习框架大多基于 NVIDIA 的 CUDA 计算库，而 CUDA 计算库支持的硬件主要是自家产品。支持 CUDA 的硬件可以在 NVIDIA 官网中找到（<https://developer.nvidia.com/cuda-gpus>）。因此，想用自己电脑进行本书深度学习入门学习的用户，如果想借助 GPU 作为硬件帮助，基本要求是需要有一台近几年出品的 NVIDIA 显卡，其显存大小最好在 8GB 以上。

如果用户的电脑没有独立显卡或者显存大小不足，将无法训练本书最后几章中的案例。另外，电脑配置的是 AMD 显卡，是不是必须新买一台电脑呢？重新购置电脑当然也可以，但是毕竟成本比较高，用户同样可以考虑按小时收费，租用 GPU 云服务器。为了方便读者学习，我们为本书的内容在亚马逊云（aws）以及腾讯云（qcloud）上设计了专门的镜像，方便读者使用。

1.2.1 在亚马逊租用云 GPU 服务器

使用 aws us-east-1 节点，GPU 价格是 0.9 美元 / 小时，也可以使用实时竞价（Spot Instance）方式以获得更低的价格。使用方法如下：

（1）登录 aws 官网。进入 <https://console.aws.amazon.com/ec2/v2/home?region=us-east-1>，单击 Launch Instance，如图 1-1 所示。当然，也可以单击左侧的 Spot Instance 选项选择更低价的实时竞价机器，然后启动。

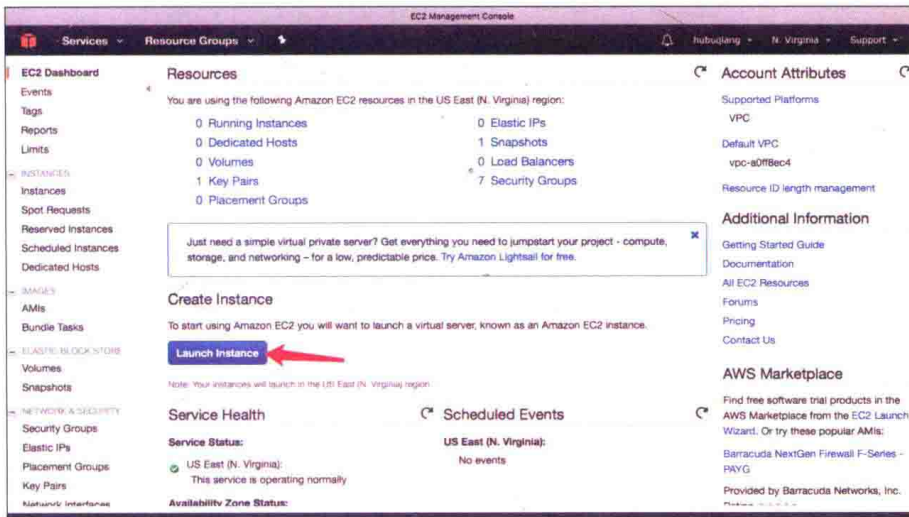


图 1-1 在 aws 上单击 Launch Instance

(2) 选择左侧的 Community AMIs 选项，然后在右侧的文本框中查找 NVIDIA Docker，再选择一个镜像，如图 1-2 所示。推荐使用 aws us-east-1 节点。

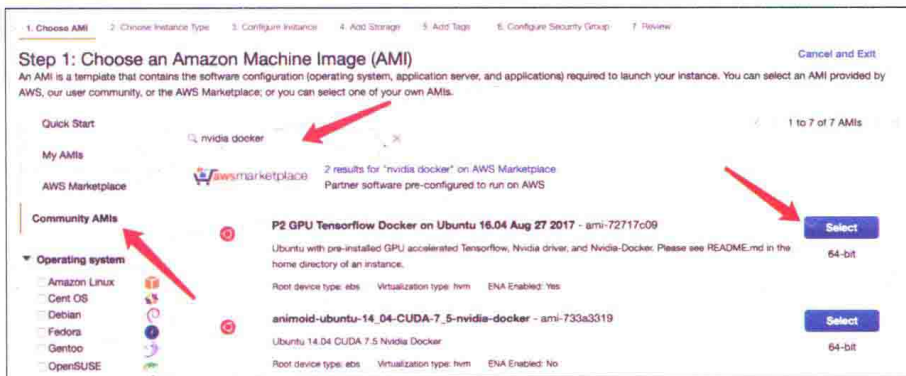


图 1-2 查找预装 NVIDIA Docker 的镜像

(3) 选择 p2.xlarge 节点，单击 Review and Launch 按钮启动服务，如图 1-3 所示。

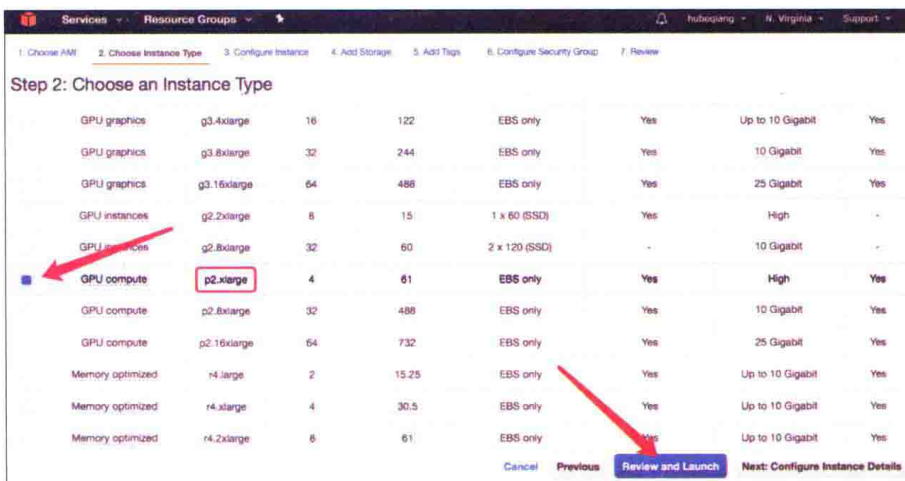


图 1-3 选择 p2.xlarge 节点

(4) 接下来选择公钥，如图 1-4 所示。如果还没有公钥就进行创建。创建公钥的过程中注意预留 22、8888、6006 这三个端口，供后续使用。短期使用时，可以选择开启所有端口。

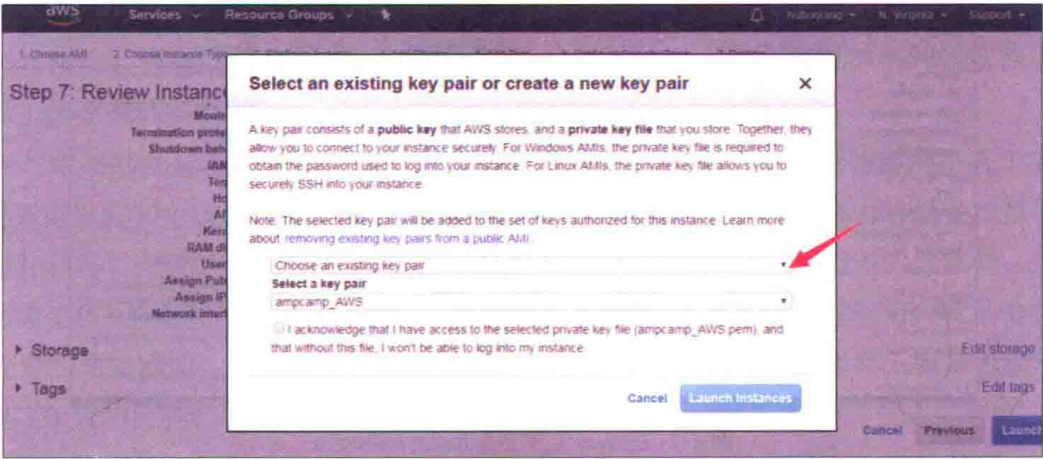


图 1-4 选择公钥

(5) 在 PuTTY (Windows) 或者终端 (Mac/Linux) 中，使用密钥启动。这里我们以 PuTTY 为例，需要填写公网地址，如图 1-5 所示。然后需要指定刚才登录时使用的公钥，如图 1-6 所示。

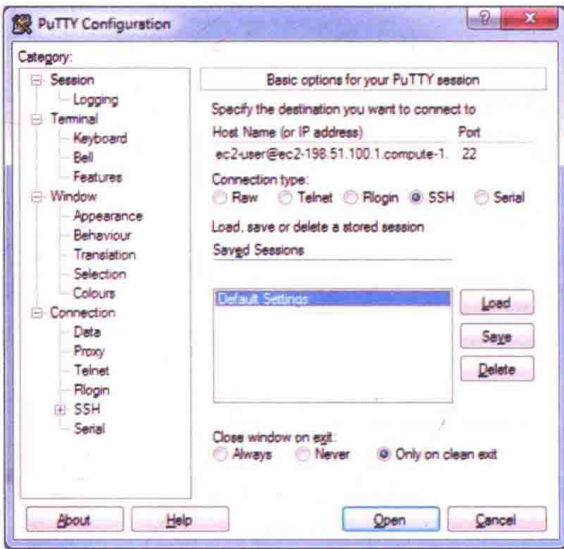


图 1-5 输入公网地址



图 1-6 指定登录时使用的公钥

(6) 此时单击 Open 按钮，即可登录 aws 服务器。如果遇到登录问题，可以在 aws 上问答 (http://docs.aws.amazon.com/zh_cn/AWSEC2/latest/UserGuide/putty.html) 中查询具体原因。

1.2.2 在腾讯云租用 GPU 服务器

亚马逊云是国外的机器，大家实际使用时可能网络连接速度并不快。这种情况下，大家可以考虑腾讯云、阿里云的按小时计费。这里以租用腾讯云为例。

(1) 进入网址 <https://cloud.tencent.com/product/gpu>，单击“立即选购”按钮，如图 1-7 所示。



图 1-7 选购 GPU 云服务器

(2) 选择“按量计费”，然后指定机型，这里 GPU 单卡足够，如图 1-8 所示。



图 1-8 选择计费模式和机型

(3) 接下来选择 GPU 镜像，单击“从服务市场选择”链接，如图 1-9 所示。

(4) 在弹出的窗口中选择腾讯云官方的 GPU 镜像，如图 1-10 所示。注意，这里显示的官方镜像 CUDA 是 7.5 版本，读者使用该镜像时需要下载并安装 8.0 版本以上的 CUDA。



图 1-9 选择镜像

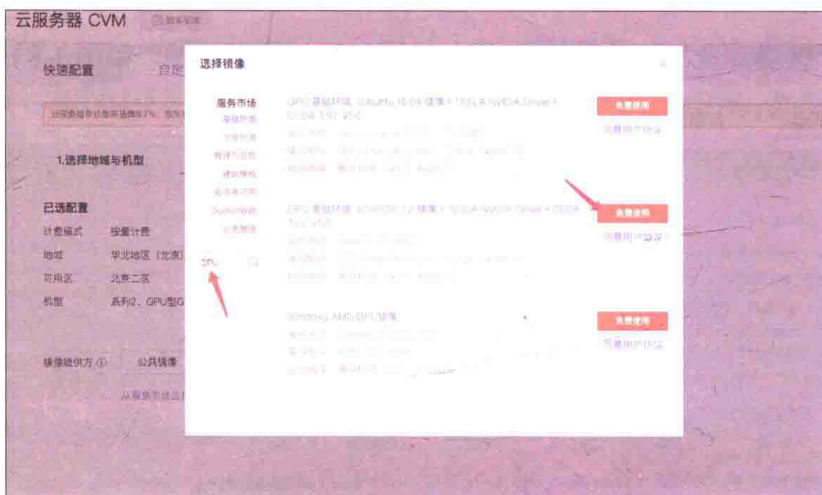


图 1-10 选择腾讯云官方镜像

(5) 选择完毕，单击“下一步：选择存储与网络”按钮，如图 1-11 所示。



图 1-11 单击“下一步：选择存储与网络”