

学术顾问: 马玉慧 李 粤 樊 磊

总 主 编: 梁森山 谢作如

副 总 主 编: 夏正仁 于方军

本 册 主 编: 谢作如 祁荣宾

主要编写人员: (按姓氏拼音排序)

陈建林 陈瑶瑶 傅海涛 管 庆 雷 鸣 梁光福
林森焱 刘啸宇 刘正云 陆雅楠 罗 亮 邱奕盛
申劲红 王怡婷 杨璐璐 应根球 于旭珩 俞 晓
张敬云 张天辉 郑 祥 周 鹏 周 琼 周源远

版权所有, 侵权必究。举报: 010-62782989, beiqinquan@tup.tsinghua.edu.cn。

图书在版编目 (CIP) 数据

信息技术·八年级下册 / 谢作如, 祁荣宾主编. —北京: 清华大学出版社, 2024.7 (2024.12 重印)

ISBN 978-7-302-66347-8

I. ①信… II. ①谢… ②祁… III. ①计算机课—初中—教材 IV. ① G634.671

中国国家版本馆 CIP 数据核字 (2024) 第 105936 号

责任编辑: 焦晨潇

封面设计: 王 静 薛玉斌 张思宇

责任校对: 赵琳爽

责任印制: 丛怀宇

出版发行: 清华大学出版社

网 址: <https://www.tup.com.cn>, <https://www.wqxuetang.com>

地 址: 北京清华大学学研大厦 A 座 邮 编: 100084

社 总 机: 010-83470000

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: 重庆升光电力印务有限公司

经 销: 全国新华书店

开 本: 185mm×260mm

印 张: 8.25

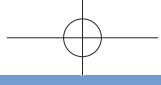
字 数: 142 千字

版 次: 2024 年 7 月第 1 版

印 次: 2024 年 12 月第 2 次印刷

定 价: 7.51 元

产品编号: 107193-02



清华大学出版社 前言

人类已进入全球化信息时代，信息科技作为现代科学技术领域的重要组成部分，对全球经济、社会和文化发展起着越来越重要的作用。信息科学所蕴含的思维方式改变了人们对虚拟世界和现实世界的认知方式，信息技术衍生出的虚拟世界重塑了人们沟通和交流的时空观念，深刻影响了人们的生活、学习和工作方式。信息素养已成为信息社会公民不可或缺的基本生存技能，无论从哪种意义上来说，信息素养的培育都势在必行。

党的二十大报告指出：全面贯彻党的教育方针，落实立德树人根本任务，培养德智体美劳全面发展的社会主义建设者和接班人。为落实新时代教育根本任务，满足社会发展对全体国民素质和人才培养的新要求，做好义务教育教材与高中新课标教材的衔接，我们编写了本套教材。

本套教材依据《义务教育课程方案（2022年版）》《义务教育信息科技课程标准（2022年版）》《普通高中信息技术课程标准（2017年版2020年修订）》的精神进行编写，全面落实党的育人方针，聚焦信息科技学科核心素养，借鉴了创客教育、STEAM教育、机器人教育多年的教学成果，积极创设真实的活动化、生活化、游戏化学习场景，以主题式项目学习组织知识、实验、活动与实践，旨在培养学生的创新思维和实践能力，提升学生的信息素养。

本套教材从培育与发展中小学生信息素养出发，采用多种策略适应中小学生的学习和认知特点，围绕“数据”“算法”“网络”“信息处理”“信息安全”“人工智能”六条逻辑主线，结合学生身边的事例、应用和真实情境，以项目式学习方式为主线，全面诠释了义务教育阶段学生应该了解和掌握的信息科技学科知识和应该具备的学科思维。

本册教材通过“体验活动”“实验活动”“实践活动”等多种学习活动，将学科知识、科学原理、问题解决方法、学科思维和对社会的影响系统地

融入其中。另外，教材创新性地选用了优质的国产自主知识产权平台、编程环境、开源硬件和配套资源等作为项目实践活动的支撑，从小处、细节入手，培养学生用国产、爱国产的文化自信。教材中所选项目和案例力求做到源于真实问题，在引领学生树立民族自豪感的同时，逐渐使他们养成负责任地使用信息科技解决实际问题的习惯，为今后的学习发展打下坚实的基础。

按照教育部印发的《中小学教材管理办法》等文件的要求，本套教材在具体编写过程中，特别强调中国在信息科技和信息化方面所取得的巨大成就，介绍信息科技在社会发展和解决重大问题中的核心作用，注重并强化信息科技学科的德育价值，强化学生做社会主义建设者和接班人的思想意识。

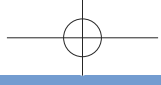
由于信息科技处在一个高速发展的过程中，中小学信息科技课程和教材也会随着教学实践与改革的深入而不断完善。鉴于编者水平有限，教材中难免存在不足之处，在此我们诚恳地希望广大师生给我们提出宝贵意见和建议，我们将及时予以修订。非常感谢每一位教材编写者所付出的心智与辛劳。

编 者

2024 年 2 月

目 录

第 1 单元 神机妙算：机器能预测	1
第 1 节 人工智能的起源与发展	4
第 2 节 机器学习初体验	11
第 3 节 机器学习大家族	19
第 4 节 用机器学习解决问题	29
单元小结	39
第 2 单元 洞明世事：机器能识别	41
第 1 节 神经网络与深度学习	44
第 2 节 卷积神经网络及其应用	53
第 3 节 用深度学习实现图像分类	65
单元小结	79
第 3 单元 妙笔生花：机器能创作	81
第 1 节 人工智能生成内容	84
第 2 节 图像生成模型	93
第 3 节 文本与图像的多模态模型	101
第 4 节 借助多模态模型进行创作	112
单元小结	123
附录	125



教材使用说明

本册教材供八年级第二学期使用，共分为 3 个单元，总计 16 学时。

本套教材设计了知识、活动和项目三条主线，这三条主线既彼此独立又相互融合。

(1) 知识主线主要包括以下几个方面。

【学习导引】结合学生生活实际或教学需求，简要介绍本单元将要学习的内容。

【正文】按照学科内在逻辑，系统性地阐述知识。

【拓展阅读】拓展延伸与本主题相关的信息科技知识，旨在拓宽学生的视野，提高学生探索未知的兴趣。

【概念解释】解释一些特定的专业术语，以帮助学生理解。

【知识回顾】以思维导图的形式呈现知识结构，帮助学生梳理学习内容。

(2) 活动主线主要包括以下几个方面，穿插在知识主线中。

【体验活动】需要让学生了解但是实施难度较高的活动。时间安排比较灵活，尽可能在课内完成，借助现有器材、设备、环境等进行体验，不需要过多的额外准备。

【实验活动】能够让学生探究的活动。相对正式，实验目标清晰体现学科核心素养或者关键知识点。

【实践活动】具有一定复杂度、综合性、实用性的技术操作活动，是知识内化之后的呈现，是创客式学习或 STEAM 跨学科学习的典型形式。

【问题讨论】提出开放性的问题，引导学生深入思考并开展讨论交流。

(3) 项目主线主要包括以下几个方面。

【项目情景】设置一个具体的项目情景，并在这个情景下提出问题。

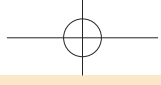
【项目方案】给出本单元项目中要用到的相关知识，并根据项目的目标与任务，对项目进行规划，给出主要的实施步骤，引导学生体验、动手操作或者进行探索实践，列出预期的实践成果。

【项目分工】引导学生根据任务内容进行分工，和与自己有共同想法的同学组成一组，并选出组长，列出每个人的具体分工。

【项目实施】引导学生结合项目内容进行项目实施。

【成果交流与评价】引导学生进行成果交流并学会分享，进行自评与他评。

在学习知识主线之前，先完成【项目情景】【项目方案】【项目分工】，【项目实施】穿插在知识主线中，【成果交流与评价】在单元最后。



第 1 单元

神机妙算：机器能预测

学习导引

提到人工智能，你的脑海中浮现的第一个画面是什么？是科幻电影里的机器人，是能战胜人类围棋高手的人工智能程序，还是能根据你的提示写出文章、画出图像的大模型？对于很多人来说，人工智能是一个熟悉而陌生的名词。一方面，在各种媒体上不断看到各种人工智能技术的最新进展，各类结合了人工智能技术的 Web 应用和手机 App 如雨后春笋般涌现；另一方面，我们却很难说清楚什么是人工智能，哪些表现称得上“智能”，人工智能和编写程序有什么区别。

科学家如何定义“智能”？如何让机器拥有类人的“智能”？本单元将从人工智能的起源和发展入手，结合最主流的机器学习技术，让同学们初步了解人工智能，感受人工智能在信息时代的重要作用；借助数据和算法，训练一个简单的模型，体验智能从无到有的全过程。

项目情景

小清是一个侦探迷。他在一些电视剧和电影中常常看到这样的场景：神探仅通过现场的脚印，就能初步推断出犯罪嫌疑人的身高。他查了一些资料后发现，原来人的身高与脚的长度、宽度和步伐长度有一定的关系。为了找出人的身高与这些因素的关系，他准备用人工智能中的机器学习解决这一问题，训练一个能推断人的身高的人工智能模型，但在探索实践过程中遇到了一些问题。



- (1) 如何使用机器学习训练一个能推断身高的模型?
- (2) 机器学习的算法有很多, 用哪一种算法能得到较好的效果?
- (3) 如何将这个人工智能模型部署成网页应用, 输入数据就能计算出结果?

.....

你是不是也很感兴趣? 快来和小清一起收集数据, 训练模型, 并实施这一方案吧!

项目方案

经过咨询与了解, 小青设计了以下方案。

知识学习	实施步骤	预期成果
(1) 人工智能的基础知识	(1) 掌握机器学习的基本流程	(1) 机器学习的基本流程 (PPT 格式)
(2) 机器学习的基本流程和常见算法	(2) 掌握借助 BaseML 完成模型训练与评估的基本方法	(2) 机器学习常见算法的认识 (PPT 格式)
(3) 数据集的收集、制作和划分	(3) 掌握从数据收集到数据集制作的简单方法	(3) 模型训练代码及部署代码 (代码、配套模型、运行视频)
(4) 模型的训练、评估和应用	(4) 利用开源工具库实现模型部署与应用	(4) 项目报告 (PDF 格式)
	(5) 撰写学习心得	

你对小青的项目方案有什么不同的看法或建议? 你准备如何设计项目方案? 请填写在下表中。

知识学习	实施步骤	预期成果

项目分工

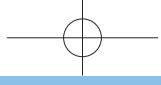
方案设计完成后，小清发现仅凭一己之力很难完成这个项目，于是邀请对此问题感兴趣的同学一起参与，并在项目方案中添加了以下表格。



姓名	角色	分工	任务
小清	组长	负责项目统筹、监督与管理	项目整体方案设计；项目实施过程的统筹、协调、监督、总结；项目文档撰写
同学甲	成员	负责进行数据集制作与文档撰写	收集数据，并制作数据集；撰写项目报告和相关文档
同学乙	成员	负责模型训练与评估	模型训练与评估的代码编写，训练一个满意的模型
同学丙	成员	负责模型应用与程序测试	设计模型应用的核心程序并测试程序，实现项目功能

你认为小清的项目组成员构成、分工和任务分配是否合理？请在下表中填写你的项目分工情况。

姓名	角色	分工	任务



第1节 人工智能的起源与发展

本节知识

- ◆ 人工智能的起源和发展历程
- ◆ 人工智能的研究学派和范式
- ◆ 新一代人工智能与人工智能模型

本节活动

- ◆ 体验人工智能应用

人工智能（artificial intelligence，AI）时代已经来临。语音输入、图像识别、自动驾驶、内容生成等人工智能应用，正以“润物细无声”的方式出现在我们的身边。而在十多年前，人工智能似乎离人们的日常生活还非常遥远。人工智能从陌生到熟悉的这段时间中，究竟发生了什么？在学习人工智能之前，我们需要先了解一下其发展历程。

一、人工智能的起源

和其他技术的起源一样，人类通过发明各种技术、制作各种工具改善自己的生产生活。比如，借助马车、汽车和飞机，人类的出行时间可以缩短；借助起重机、挖掘机，人类的工作可以更加省时省力等。数千年来，人类一直期望用复杂的机械实现人的智能。在古代的神话传说中，技艺高超的工匠可以制作人造人，并赋予其智能或意识，如希腊神话中有黄金机器人和皮格马利翁造人的故事。据《列子·汤问》记载，我国西周时期有“偃师造人”的故事。

人工智能的起源可以追溯到古希腊亚里士多德的“三段论”。“三段论”是一种形式化、机械化的论证方法，可以根据前提推断出结论。

例如：

贵州人是好客的；（大前提）

我是贵州人；（小前提）

我是好客的。（结论）

这种推断机制类似人的思考，是一种智能行为的表现。

“三段论”逐步发展为逻辑学，数学家用逻辑符号描述各种复杂的知识。于是，电子计算机出现后，新的思考也开始了：既然计算过程可以自动化，那么是不是就能实现“人工智能”了？这就是人工智能中一个重要的学术流派——符号主义（也称逻辑主义）的核心理念。

“机器会思考吗？”1950年，艾伦·图灵（Alan Turing）在论文《计算机与智能》开篇提出了这样的问题。图灵从各个角度讨论计算机这一人造的计算机器具备人类智能的可能性，并提出了著名的“图灵测试”。图灵还在论文中提出了“学习机器”的概念，认为可以通过编程将计算机模拟为儿童大脑，然后接受“正确的教育课程”，获得成人大脑。图灵的问题持续鼓励大量科学家投身于人工智能这一领域。

1956年夏天，约翰·麦卡锡（John McCarthy）、马文·明斯基（Marvin Minsky）等科学家在美国达特茅斯学院组织了一个研讨会，如图1.1.1所示。达特茅斯会议的研讨主题是让机器模拟智能的可能性，包含了人工智能领域面临的七个问题：自动计算机、编程语言、神经网络、计算规模理论、自我改进、抽象、随机性与创造性。达特茅斯会议的召开，标志着人工智能正式诞生，开启其漫长的征途。



图 1.1.1 达特茅斯会议与会者合影

问题讨论

什么是“智能”？在很多方面，“非智能”的机器已经远比人类聪明。比如，计算机程序可以计算很大数字的乘积，能一眼就辨认出二维码中的信息，但我们往往不会说它很聪明，而是说它算得“正确”。“智能”这个词，我们只用来描述人类特有的那些能力，比如辨认出熟悉的面孔，在交通高峰期的车流中穿行，精通某种乐器，能写出文笔流畅的文章等。那么，人的哪些表现才属于“智能”？请提出并讨论。

二、人工智能发展简史

达特茅斯会议之后，不同学科背景的学者对人工智能的实现提出了各自不同的观点，由此产生了不同的学术流派。这些学术流派中以符号主义、联结主义和行为主义为代表，此外还有贝叶斯学派、类推学派等。2019年，著名人工智能科学家杰弗里·辛顿（Geoffrey Hinton）对各种流派进行了分析，总结出人工智能发展的两种范式（可以理解为基本方式）——逻辑启发式（设计派）和生物启发式（学习派）。

设计派认为“智能”是人为设计出来的。这一思路来源于逻辑学，用一种“语言”将已有的知识进行精确编码，用各种运算符号表示各种规则，然后派生出新的知识。设计派的典型代表是符号主义学派，其核心工作是设计专家系统——一种基于特定的规则回答特定领域的问题的程序。而学习派认为“智能”因学习而来，用算法模拟人的大脑结构或者功能，机器能从大量的数据中找到“输入→输出”之间的映射关系。学习派的典型代表是联结主义学派，其重点工作就是用人工神经网络模拟人的大脑。

人工智能的发展并不是一帆风顺的，短短数十年几经波折。人工智能发展的早期，设计派的思路占了上风，但是很快就陷入了瓶颈，因为研究者发现，几乎没有办法对现有知识进行精确编码，常常“牵一发而动全身”，一个地方出现纰漏就要全部推倒重来。比如，专家系统需要人工定义规则，这项工作不仅费时费力，而且在语音识别、图像识别等自然输入的应用场合中难以实施。之后，学习派成为主流范式，从大量的数据中得到智能，这种方法也被称为数

据驱动的人工智能。因为学习的主体是机器，机器学习（machine learning）就成为人工智能研究方面最重要的技术领域。

拓展阅读

人工智能发展的两次低谷

人工智能概念诞生之后，迎来了第一个发展高峰期，快速涌现了许多相关成果，比如通过机器智能程序自动完成一些推理任务。可惜当时计算机的性能差，受限的内存容量和处理速度导致计算机程序无法解决较复杂的问题。再加上数据的严重缺失，计算机程序无法从数据中学习足够的知识，学习派无法施展手脚。因此，人工智能的研究进展逐渐减慢，科学家的预期迟迟无法实现，人们开始对人工智能感到失望。许多机构停止了对人工智能研究的资助。这是人工智能发展的第一个低谷期。

20世纪80年代，专家系统和人工神经网络的快速发展让人们对人工智能的热情再度高涨。然而，专家系统需要人工定义规则，这项工作不仅费时费力，而且在语音识别、图像识别等自然输入的应用场合中难以实施。随着个人计算机的出现，专家系统的功能很容易被个人计算机的通用软件所替代。而人工神经网络的研究受制于计算机的性能，于是人工智能进入第二个低谷期，大量人工智能公司倒闭。

三、新一代人工智能

2012年，得益于数据技术的发展、算法的突破和计算机性能的提升，人工智能开始突飞猛进。2013年，深度学习在图像识别和语音识别领域取得突破，标志着人工智能实现了感知智能。2016年，AlphaGo横空出世，它借助深度学习及先进搜索算法的强大威力，横扫围棋界，攻克了棋类运动中人类的智慧堡垒。2022年，ChatGPT的发布再次让世界为之瞩目，研究各种大模型应用成为人工智能的主流方向。智能时代已然来临，人工智能领域又一次掀起了新的浪潮。为区别之前的人工智能，科学家使用“新一代人工智能”或者“下

一代人工智能”进行表述。

1. 无处不在的人工智能应用

新一代人工智能的最大特点是不再停留在学术研究上，可以真正应用于生活，对各行各业产生巨大的影响。人们依靠智能导航出行，通过语音与机器互动，应用智能工具搜索知识信息，借助大模型生成文本和图像，已自觉或不自觉地处于人工智能的环境中。随着互联网的普及、物联网的渗透、大数据的涌现和信息社区的崛起，数据和信息在彼此融合的信息空间、物理空间和人类社会传播，新技术、新产业和新业态不断涌现，使人工智能迅速发展，在众多领域发挥着巨大的作用。

人工智能的能力越来越强大，已经从最初的判别式（discriminative，也称决策式）发展到生成式（generative，也称产生式）。也就是说，人工智能越来越像“人”，不仅能下棋、识别文字和图像，还能写诗作画，甚至能撰写文书等。

拓展阅读

我国发布《新一代人工智能发展规划》

2017年7月，国务院印发《新一代人工智能发展规划》，人工智能正式上升为国家战略。新一代人工智能发展规划将在大数据智能、群体智能、跨媒体智能、混合增强智能和自主无人系统等理论方面取得重大突破，推动人工智能的不断发展。《新一代人工智能发展规划》指出，发展人工智能是一项事关全局的复杂系统工程，要按照“构建一个体系、把握双重属性、坚持三位一体、强化四大支撑”进行布局，形成人工智能健康持续发展的战略路径。我国人工智能技术虽然起步较晚，但伴随着人工智能研究的热潮，产业化应用发展迅速，基础研究、技术及产业都进入了高速增长期。

2. 日益强大的人工智能模型

用通俗的话来表达，人工智能的研究目标就是让机器具备类似人的智能，而这一智能往往通过模型（model）实现。我们可以把人工智能模型看成一个

支持输入和输出的算法，类似于人的大脑，可以接收输入数据并生成相应的输出结果，如图 1.1.2 所示。输入数据可看作人的感官输入（如视觉、听觉等），模型接收这些数据并通过学习和分析处理它们。然后大脑会指挥身体做出特定的动作。人工智能模型根据输入数据的特征和模式，产生相应的输出结果，就像人脑“处理”眼睛看到的画面，然后生成相应的理解和反应一样。



图 1.1.2 人工智能模型和人脑对比图

随着数据、算法和算力方面的突破，人工智能模型的能力越来越强大，应用领域也越来越广泛。现在出现了一种新的云计算模式被称为模型即服务（model as a service, MaaS）。这种服务提供云端人工智能模型，用户不需要拥有自己的硬件设备或专业技能，也能使用这些模型设计各种应用或者完成某些工作。

拓展阅读

从模型到大模型

“模型”是人工智能最重要的概念之一。“模型”一词本来是一个数学概念，通常由数学公式、方程、图表、图形或其他数学概念组成，用来表示问题的不同方面或关系。人工智能模型则是指一种用来模拟和解决问题的方法或工具，基于数据和算法而构建，可以对信息进行处理、分析和预测。

每一个人工智能模型都有不同的特点和用途，用来解决不同的问题。人们常说的大模型（large model），泛指一些规模很大的模型。这些模型不仅文件的容量很大，而且运行这些模型也需要很高的算力，如 GPT-3（一个著名的大语言模型）的文件大小约为 700GB，每次前向传递的推理过程需要经历约 1750 亿个参数的共同运算，算力要求极高。而训练这样的模型还需要海量的数据，如 GPT-3 已学习的自然语言数据量约为 45TB。

体验活动

体验人工智能应用

训练和应用一个人工智能模型其实也很容易。通过浏览器就能在一些人工智能教学平台上训练一个简单的模型，或者通过浏览器就可以训练一个简单的模型，或者应用一个已经训练好的模型，如图 1.1.3 所示。请选择 1~4 个你感兴趣的模型应用，体验其功能并填写表 1.1.1。



图 1.1.3 AI 教学平台“AI 体验”栏目界面

表 1.1.1 人工智能应用的功能体验记录表

应用名称	功能简述	输入数据	输出结果



第2节 机器学习初体验

本节知识

- ◆ 机器学习的概念
- ◆ 传统编程和机器学习的区别
- ◆ 机器学习的基本过程

本节活动

- ◆ 用电子表格软件“学习”数据规律
- ◆ 训练温度转换模型

在人工智能发展的早期阶段，科学家尝试了各种方法，用逻辑符号推理模拟人脑思考，用人工总结规则的方式灌输知识，希望赋予机器“智能”，可惜困难重重，进展缓慢。在自然界无穷无尽的规律和人类数千年积累的知识前面，计算机引以为傲的运算速度和人工总结灌输知识的效率显得微不足道。兜兜转转，

科学家最终回到图灵提出的“学习机器”概念：能否实现人工智能的关键，很可能取决于“如何让机器拥有学习能力”。

一、从学习到机器学习

“学习”是一种伴随人类终生的普遍行为。它的含义可以是很广泛的，并不是一定非要在学校接受教育才能算是学习，也不一定必须读书、思考才能算是在学习。所有的对象，如果接受外界信息的刺激之后，能形成经验反应，并影响日后的行为，这个过程就可以被称为“学习”。就这种广义的学习定义而言，不仅局限于人类，几乎所有的生命体与生俱来都拥有着不同程度的学习能力。

1. 初识机器学习

正式探讨“机器学习”之前，我们需要先给它设定一个合理的期望与定位：

机器学习是人工智能中的一种重要技术，做出了许多令人惊叹不已的成果。不过，机器学习并非魔法，不能把机器“教育”成智慧机械生命。这里的“学习”是取其“从经验中自我改进”的含义。

1997 年，“机器学习之父”美国卡内基梅隆大学汤姆·米切尔（Tom Mitchell）教授提出：“机器学习是对能通过经验自动改进的计算机算法的研究。”台湾大学李宏毅博士则通俗地总结：机器学习就是找到一个函数（function），实现特定的功能。我们可以把“函数”理解为一段支持数据输入和输出的计算机程序。这个程序能够通过大量的数据训练，学习总结出输入数据（X）和输出数据（Y）之间的映射关系。

我们还可以用更加通俗的方式进行类比。假设家里养了一只名字为“AI”的小狗，你需要教会这只 AI 小狗分辨出主人和陌生人。显然，一开始这只小狗谁都不认识，随着你的不断调教，它就慢慢能分辨主人和陌生人了，说明 AI 小狗具备了“学习”能力，在学习过程中自动改进了分辨主人和陌生人的能力。这里的“调教”，在机器学习中被称作“训练”。所谓机器学习，不过是编写一段具备学习能力的程序（类似 AI 小狗），然后通过训练使其得到某种智能的方法。而“学习”的结果，就是具备了某种“智能”的模型。

拓展阅读

第一个机器学习程序——西洋跳棋

1952 年，IBM 的程序员阿瑟·塞缪尔（Arthur Samuel）在计算机上开发了一款西洋跳棋的程序。该程序并不是开发人员将下棋的方法通过算法编程的方式直接赋予计算机，而是通过算法赋予计算机一定的学习能力，在下棋过程中其可自行总结赢棋的方法或者策略。塞缪尔用该程序下了许多盘棋，并发现经过训练的程序比人下得更好。到 1959 年，这款程序打败了塞缪尔本人，并在 3 年之后打败其所在州的跳棋冠军，这是塞缪尔自己都做不到的。塞缪尔还创造了“机器学习”这一概念，并将其定义为“在没有明确编程指令的情况下赋予计算机学习能力的一个研究领域”。

2. 传统编程和机器学习的区别

显然，机器学习的算法也需要编程。但是机器学习的编程和传统编程不一

样。对于传统编程来说，利用设计好的函数或模块可以直接实现。比如，下面的程序能够将摄氏温度转换为华氏温度。

```
def celsius_to_fahrenheit(celsius):
    fahrenheit = celsius * 9/5 + 32
    return fahrenheit
n = int(input())
print(celsius_to_fahrenheit(n))
```

运行这段程序，输入“25”，这段代码会输出“77.0”（ $25 \times 9/5 + 32 = 77$ ）。这里的“ $fahrenheit = celsius * 9/5 + 32$ ”是程序事先给出的转换公式。如果有一段程序并不用事先给出这个公式，而是通过输入一组符合这个规律的数据（表 1.2.1），让程序通过“学习”这些数据的规律，“自动”得出这样的“公式”，这就是机器学习。

表 1.2.1 摄氏温度和华氏温度对应表

摄氏温度 / °C	华氏温度 / °F
0	32.0
25	77.0
27	80.6
29	84.2
100	212.0

从一组数据中总结出公式的做法，在数学上称为“回归分析”。“线性回归”则是回归分析中最简单的一种方法。实际上，机器学习的兴起，的确和数据统计的发展分不开。WPS、Excel、Numbers 等数据处理软件中就内置了类似功能。

如图 1.2.1 所示，我们绘制了一组数据的散点图，并给散点图增加了一条趋势线。绘制趋势线后，勾选“设置趋势线格式”的“趋势预测”中的“显示公式”和“显示 R 平方值”两个选项。这里的公式就是 Excel 通过数据“回归”推导出的规律，R 平方值则表示数据与公式的契合程度。

总结一下，传统编程是基于人为设定的数据规则实现相应的功能，而机器学习则是基于数据自动推导规则，这就是传统编程与机器学习的区别。

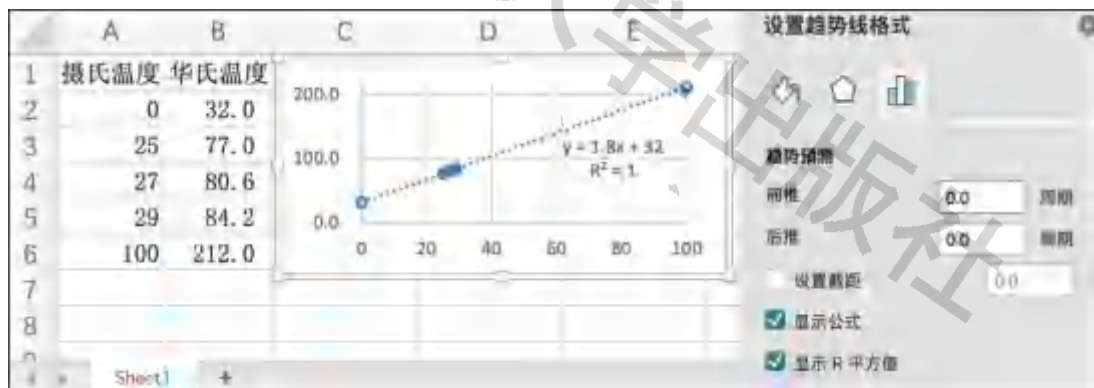


图 1.2.1 数据散点图与趋势线

拓展阅读

线性回归和R平方值

从图 1.2.1 中可以看出，如果把输入数据（摄氏温度）和输出数据（华氏温度）绘制在平面直角坐标系中，会看到一条直线。那么给出新的输入数据（ x 坐标）就能在直线上找到对应的输出数据（ y 坐标）。“线性回归”的核心思想是找到一条能够表示数据（一个或多个输入值和输出值）之间关系的直线，然后借助这条直线预测数据。这个表示数据关系的函数就是一个人工智能模型。那么，如何评价模型的准确度呢？数学家设计了一系列的衡量指标，其中最常用的指标叫做 R 平方值（R-squared）。

我们把通过模型得到的输出值和理想值之间的差距，叫做“误差”。R 平方值在 0 和 1 之间，若数据完全准确，则 R 平方值为 1；数值越接近 1，说明存在误差越小；反之，说明存在误差越大。R 平方值为 0 时，意味着模型与实际数据之间没有相关性。以射箭为例，R 平方值类似射中目标的准确程度。如果所有的箭都正中靶心，那么结果是完美的，R 平方值就是 1；如果箭都偏离得很远，就像是闭着眼睛射出的，那么 R 平方值接近于 0。



体验活动

用电子表格软件“学习”数据规律

当拿到“摄氏温度和华氏温度对应表”时，我们很难快速看出其对应转换公式，但是这种用数据推导规律的任务，却很适合用机器学习的方法解决。大部分电子表格软件集成了简单的数据统计功能，请用电子表格软件打开教材资源包中的“温度转换.xlsx”文件，为散点图添加趋势线，并记录公式和R平方值。

二、机器学习的基本流程

“不学《诗》，无以言。”再聪明、再智慧的大脑也需要不断学习，学成后，再遇到问题时，便会在之前的学习经验中寻找答案。机器学习也一样，需要在已知数据中学习这些数据蕴含的规律，从而建立起解决问题的模型，并且借助更多的数据自动修正、优化模型，最终利用模型解决问题。在已知数据中学习规律，叫做模型训练；将新的问题输入模型中得出结果，叫做模型推理。机器学习实际上分为两个阶段：首先是模型训练阶段，即“学习”；然后是模型推理阶段，即“应用”，如图 1.2.2 所示。

当模型训练好之后，应用这个模型解决问题，跟传统的编程就没有什么区别了。实际上，我们在学习 Python 编程时常常会导入一些内置了人工智能模型的库，如 OpenCV（计算机视觉库）、Pyttsx3（语音合成库）等，即使我们没有学过人工智能也能使用这些模型。对学习人工智能来说，重点是训练模型。训练模型的流程，就是机器学习的过程。如图 1.2.3 所示，典型的机器学习流程可以分为数据准备、模型搭建、模型训练与评估、模型应用等环节。

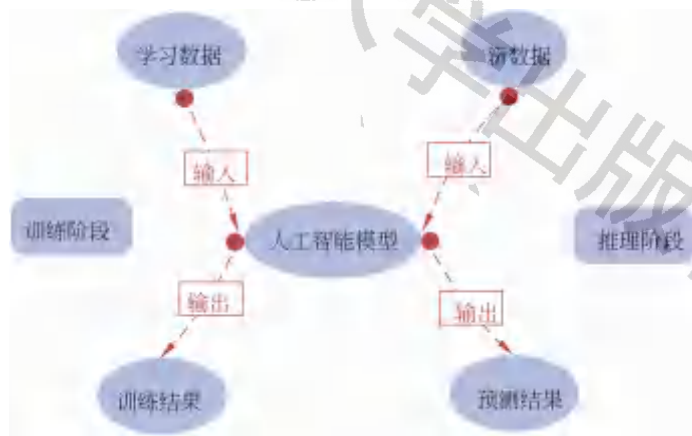


图 1.2.2 人工智能模型训练阶段和推理阶段

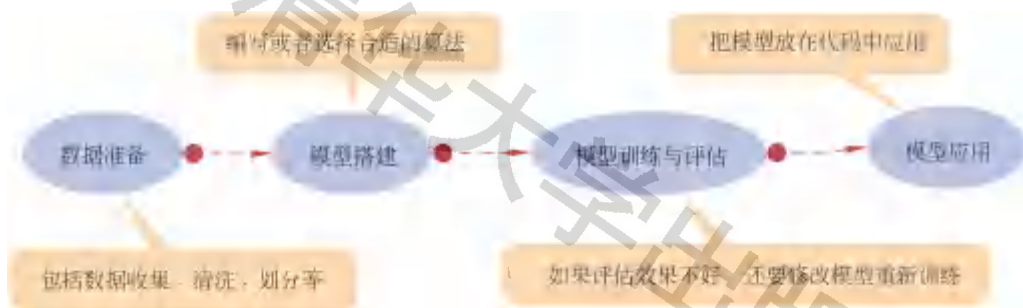


图 1.2.3 典型的机器学习流程

1. 数据准备

数据是描述客观事物或现象的符号记录，可以是数字、文字、图像、声音等形式。机器学习需要很多这样的数据集合，我们称为“数据集”。要训练怎样的模型，就要准备怎样的数据。例如要训练温度转换的模型，就要准备很多条类似“摄氏温度和华氏温度对应表”里的数据。

2. 模型搭建

搭建机器学习模型的核心工作是实现一个具有特定功能的算法。编写实现机器学习的算法程序难度较大。但好消息是 Python 有多个机器学习的库，这些库中内置了各种优秀的算法，只要根据需要选择合适的算法，就可以直接完成模型的搭建。

3. 模型训练与评估

对于训练好的模型，我们需要评估一下其推理能力，这就类似我们学习了某个课程后，还要做些测试题，看看掌握了多少。对线性回归任务来说，简单好用的评估指标之一就是 R 平方值。通过比对推断结果与实际结果的差异，可以计算出评估指标。如果推理效果不好，那么就需要重新检查数据和模型，再次训练。

4. 模型应用

如果训练出来的模型的评估表现不错，就可以保存模型了。保存的模型文件，可以直接导入使用或供其他程序使用，只要输入一组新数据，就能输出预测结果。



实践活动

训练温度转换模型

通过摄氏温度和华氏温度对应表，我们知道二者之间存在一定的对应关系。假设我们并不清楚它们之间的转换公式，能否通过训练线性回归模型找出二者之间的关系？请以小组为单位，参考资源包中的代码完成模型训练，实现“输入摄氏温度，输出华氏温度”的功能。主要实践内容包括：

- (1) 完善代码，输入数据集路径，进行模型训练和保存。
- (2) 将训练好的模型复制到模型应用代码对应的文件夹中，然后运行模型应用代码，输入摄氏温度，测试输出的华氏温度是否正确。
- (3) 进一步思考：如果要训练“华氏温度→摄氏温度”的模型，该如何实现？

拓展阅读

“从三到万”故事的启示

有一个家喻户晓的故事叫做“从三到万”。其大意是一个财主给儿子请了一位先生教写字，先写“一”“二”“三”，财主儿子学后说已会写字，财主便将先生辞退。一日，财主要给朋友写请帖，让儿子写。儿子从早到晚在纸上画满了横道也没写好，还说这朋友怎么偏偏姓“万”啊，画了一整天也才五百道，距离万还差得远。这个故事让人哭笑不得。

虽然这是一个编造的笑话，但从机器学习的角度看，我们能进行“科学”的解释——故事中的孩子用一万条横表示“万”字的问题根源就在于学习数据太少。在机器学习中，数据的质量非常重要，这里的“质量”指数据的准确性、完整性等。如果温度转换的训练数据有错误，那么训练出来的模型肯定也错误百出。另外，数据要有一定的规模，如果仅仅提供了几条数据，那么肯定训练不出好模型。

第3节 机器学习大家族

本节知识

- ◆ 机器学习典型的学习方式
- ◆ 分类任务和回归任务的区别
- ◆ 机器学习的开发工具：scikit-learn 和 BaseML
- ◆ 机器学习的评估与效果对比

本节活动

- ◆ 投石车落地距离预测的不同算法对比

经过数十年的发展，机器学习逐步成为人工智能最重要的研究方向之一。从第一代西洋跳棋程序到基于深度学习的大语言模型应用，从简单的数据判别到图文并茂的多媒体数据生成，机器学习已经演变成一个庞大的家族。机器学习在理论、工具和算法上不断创新，逐渐形成了许多分支。

一、机器学习的再理解

机器学习是人工智能的重要研究方向，其目标是让计算机能够从数据中学习并不断改进其性能，如那款西洋跳棋程序，随着对手越来越强大，程序的棋力也越来越好，最后打败了当地冠军。而通过“温度转换”的例子，我们已经体会到机器学习并不神秘，在特定算法的支持下，几行代码就能从数据中寻找规律。但是要挑战更多机器学习任务，尝试训练更多更优秀的模型，还需要进一步理解机器学习。

1. 学习方式：监督学习和无监督学习

机器学习的核心工作是寻找任意输入和输出的数据组合之间的数学关系。在“温度转换”的这个例子中，训练的数据为两列，其中一列用于输入，另一

列代表输出。在机器学习中，用于输入的数据叫做“特征”（feature），而代表输出的数据叫做“目标”（target），也叫做“标签”（label）。任务中的数据同时包含特征和目标，这类任务称为监督学习。监督学习如同教小孩识别各种物品，我们给他看不同物品的外观（特征）的同时，也要教给他不同外观对应的名称（目标）。数据集中的特征和目标要一一对照，这一过程称为数据标注。例如在“温度转换”这个例子中，就要依据摄氏温度这一特征明确对应的目标，即华氏温度，如图 1.3.1 所示。

	摄氏温度	华氏温度
1	-1	30.2
2	1	33.8
3	3	37.4
4	5	41.0
5	7	44.6
6	9	48.2
7	11	51.8
8	13	55.4
9	15	59.0
10	17	62.6
11	19	66.2

图 1.3.1 “温度转换”数据集范例

在机器学习中还有一类任务，给定的数据中没有相应的预测目标信息，也就是说数据集没有进行人为标注，这种学习方式称为“无监督学习”。无监督学习的主要目的是发掘数据间的联系。比如，给定一批数据，将其按照特点分成不同的类别。

拓展阅读

机器学习中的强化学习

机器学习涵盖多种类型，除了监督学习、无监督学习外，还有一种叫做强化学习。在强化学习中，智能体可以通过与环境进行交互，根据环境的反馈调整自己的行为。每次交互后，环境会给智能体一个奖励或惩罚，智能体需要通过尝试不同的行为获得最大的累积奖励。就像一个小孩在成长过程中需要不断尝试各种事情，如走路、说话、玩耍等。每次尝试后，他可能会得到表扬或批评，这样他就知道哪些行为是好的，哪些是不好的，从而逐渐学会适应环境。塞缪尔开发的西洋跳棋程序和人工智能早期三大学术流派中的行为主义，使用的就是强化学习的思想。

2. 学习任务：回归和分类

人工智能要解决的问题，我们称为任务。使用不同机器学习的学习方式已经能够解决很多任务。以监督学习为例，它主要解决两类任务：分类和回归。

这两类任务虽然都要提供标注过的数据集，但在应用场景上有显著的差异。

分类任务的目标是将输入的数据分配到预定义的类别中。这类任务要求算法不仅要理解数据的特征，还要能根据这些特征将数据归入特定的类别。分类可以是二分类（如判断西瓜甜不甜），也可以是多分类（如识别图像中的物体类别）。而回归任务是预测一个连续的数值，而非将数据分配到类别中。这类任务通常涉及预测数量，如房价、温度或销量等。以卖西瓜为例，分类任务中瓜农要判断西瓜好不好，回归任务则是要给不同外表的西瓜标上不同的价格。

3. 数据集的划分：训练集和验证集

在“温度转换”的实践活动中，一共使用了两个 CSV 格式的数据文件。其中用于模型训练的叫做训练集，用于模型评估的叫做验证集。一个完整的监督学习任务，数据集分为训练集（training set）、验证集（validation set）和测试集（testing set），以便进行模型训练、性能评估和测试。

数据集划分的主要目的是确保模型能够在未见过的数据上也有良好的表现，实现举一反三，这种能力也叫做“泛化”。因此，训练集、验证集和测试集的数据要保持独立，尤其不能将验证集和测试集的数据加入训练集中进行训练，不然训练出来的模型会评估出错，出现“过拟合”（类似学习中的“死记硬背，不会变通”）的情况，看起来得分很高，但遇到新的数据就表现很差。

问题讨论

为了确保数据集的有效划分，通常采用随机划分的方法。以小组为单位，讨论采用手动划分数据集的方式可能会带来哪些方面的弊端，以及还有哪些划分数据集的方法。

二、机器学习的开发工具

大部分编程语言都支持机器学习的开发。仅 Python 语言中就有好多机器学习库，如 scikit-learn、SciPy 和 BaseML 等。后来随着神经网络算法的兴起，

又出现了如 TensorFlow、PyTorch 和飞桨（PaddlePaddle）等库。这些工具的功能越来越强大，使用门槛却越来越低。

需要强调的是，无论使用何种工具，机器学习的基本流程都是不变的，都需要经历如图 1.2.3 所示环节，如数据准备、模型搭建、模型训练与评估和模型应用这几个必要的阶段。

1. 用 scikit-learn 训练模型

scikit-learn 是一个开源的机器学习库，支持监督学习和无监督学习。它还提供了用于模型拟合、数据预处理、模型选择、模型评估和各种其他实用程序的工具。

scikit-learn 内置了很多算法模型，我们可以很方便地调用。以搭建线性回归模型并训练为例，核心代码如下所示。

```
from sklearn.linear_model import LinearRegression
# 从库中导入线性回归模块

from sklearn.model_selection import train_test_split
# 从库中导入数据划分模块

model = LinearRegression()
# 实例化线性回归模型

x_train, x_valid, y_train, y_valid = train_test_split(x, y,
random_state=0)
# 将数据拆分为训练集和验证集

my_model = model.fit(x_train, y_train)
# 训练模型
```

2. 用 BaseML 训练模型

BaseML 是上海人工智能实验室开发的 XEdu 工具箱中的一款子工具，针对 scikit-learn 做了进一步的封装，使工具变得更加易用，只需几行代码就能实现机器学习的训练、评估或应用。“温度转换”模型的训练就可以使用 BaseML 编写，代码如下。

```
from BaseML import Regression as reg # 从库文件中导入回归任务模块
model = reg('LinearRegression')
# 实例化线性回归模型
model.load_tab_data('./data_train.csv')
# 载入训练数据
model.train()
# 训练模型
model.valid('./data_val.csv', metrics='r2')
# 载入验证数据并验证
model.save('./mymodel.pkl')
# 保存模型供应用
```


从以上代码可以看出，BaseML 库文件的导入只需要一行代码即可根据机器学习的任务类型导入相应的库。“Regression” 模块内置了回归任务的常见算法，“Classification” 模块则内置了分类任务的常见算法。“model = reg ('LinearRegression')” 表示指定的算法是“LinearRegression”（线性回归）。

使用“load_tab_data”，即可加载 CSV 格式的数据文件。这里要求数据文件每行一条记录（首行为表头，数据从第 2 行开始），输入数据（特征）列在前，输出数据（目标或标签）列在后，即最后一列为输出数据，其余列均为输入数据，以 CSV 格式存储。这种格式是最常见的监督学习数据集格式，如图 1.3.2 所示。

图 1.3.2 CSV 格式的数据集格式

BaseML 支持直接载入验证数据集进行模型评估，只需数据集格式及输入、输出列数与训练数据保持一致即可。这里会输出评估指标的计算结果（可以选择 R 平方值等评估指示）。BaseML 还可以进一步利用可视化功能，直观地了解模型的验证效果。核心代码如下所示。

```
model.valid('./data_val.csv',metrics='r2')
# 载入验证数据集，计算R平方值评估指标
model.metricplot()
# 模型验证效果可视化
```

运行结果如图 1.3.3 所示。

从下面的模型验证可视化效果图中可以看出，验证集已有的输出（ y ）为横坐标，通过模型推理得到的结果（ \hat{y} ）为纵坐标，两者构成的坐标点若落在灰色虚线上，则说明模型完全契合验证数据。通常实际构成的点没有落在灰色虚线上，而是围绕黑色虚线分布。两条虚线相差越大，说明模型效果越差。

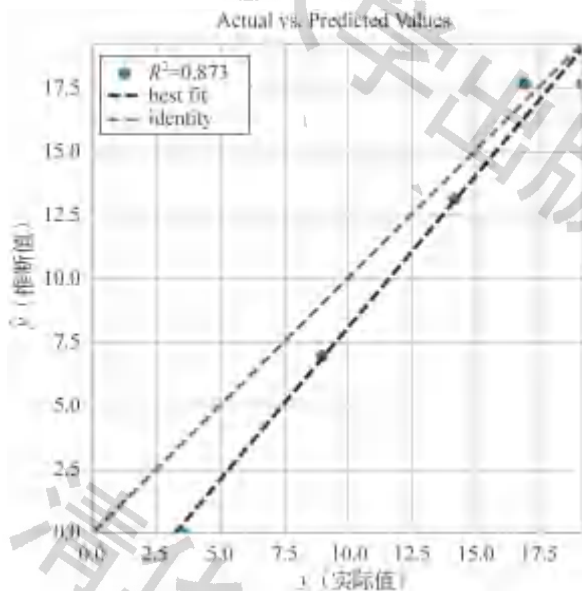


图 1.3.3 模型验证的可视化图表

训练好模型后，只要输入和训练数据相同类型的数据，便可以使用 inference 方法对新数据进行推理。如果想要在其他地方使用模型，只要执行 save 方法，即可保存这个模型。核心代码如下。

```
y = model.inference(300)    # 模型推理
print(y)
model.save('mymodel.pkl')    # 保存模型
```

BaseML 支持很多任务或算法，而基本语法是一致的。若要修改任务为分类，修改“Regression”为“Classification”即可；若要选择决策树算法（英文为 CART），只要将 reg('LinearRegression') 中的“LinearRegression”改为“CART”即可，如图 1.3.4 所示。

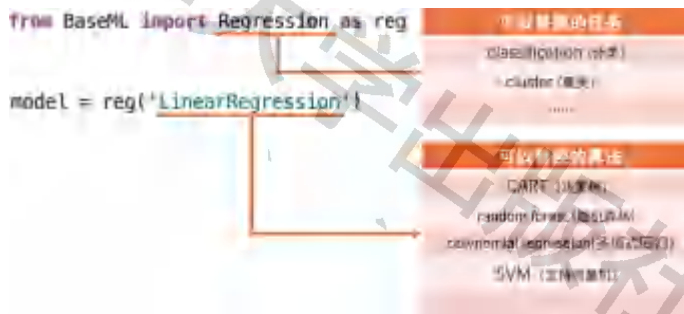


图 1.3.4 在 BaseML 中修改任务或算法

三、机器学习的算法家族

机器学习要想在复杂的数据中寻找出所蕴含的规律，需要使用合适的算法。机器学习的发展，其核心还是算法的发展。如图 1.3.5 所示，机器学习的算法发展史大致是从基于规则走向基于数据统计，然后走向深度学习，即用仿生模拟的方式模拟出人的大脑。基于深度神经网络的机器学习也称深度学习，我们将在第 2、3 单元继续学习。

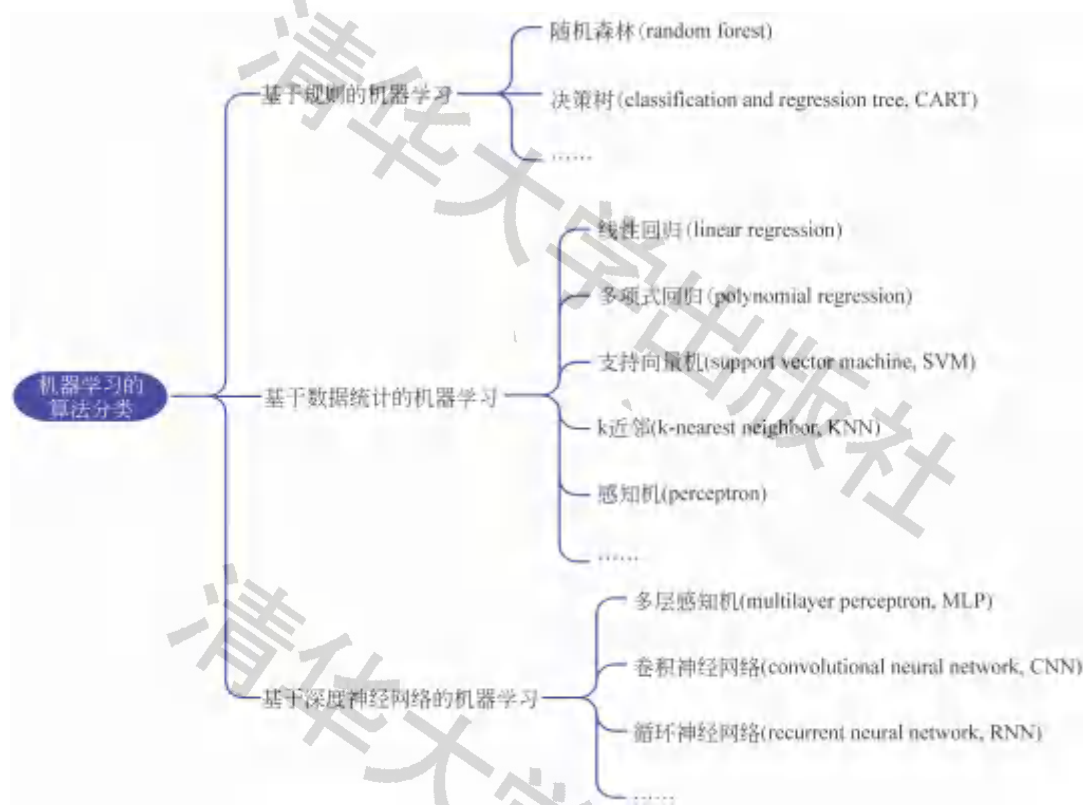


图 1.3.5 机器学习的算法分类

虽然图 1.3.5 中罗列的算法看起来已经很多了，但实际上这仅仅是机器学习算法家族中的一部分。不同的算法适合解决不同的任务。要用机器学习解决问题，首先要了解有哪些算法，并弄清楚这些算法擅长解决哪些问题。为帮助初学者了解这些算法的作用，BaseML 的文档中提供了一张“机器学习典型算法一览表”，节选如表 1.3.1 所示。

表 1.3.1 机器学习典型算法一览表（节选）

算法名称	BaseML中的算法名	适合任务	典型任务	算法解释
线性回归	LinearRegression	回归	适用于预测房价、销售额、贷款额度等	就像用直尺在散点图上画一条尽可能穿过所有点的直线，这条直线就能帮我们预测未来的值
多项式回归	Polynomial	回归	适用于预测房价、销售额、贷款额度等	就像是在一条直线上增加更多的弯曲，使这条线可以更好地贴合数据点。就像用橡皮筋在散点图上拉出一个曲线，这个曲线就能更好地帮助我们预测未来的值
支持向量机	SVM	分类/回归	适用于文本分类、图像识别等	想象你有两种颜色的方块，SVM 就是用一根棍子（在复杂情况下是一张弯曲的板），尽可能分开两种颜色的方块
决策树算法	CART	分类/回归	适用于客户分级、疾病诊断等	想象你在做一个选择（比如选择餐馆），你可能会根据一系列问题（离家近不近？价格怎么样？）决定。决策树算法就是通过一系列问题来达到决策的过程

拓展阅读

机器学习算法的参数调整

选择恰当的算法，就如同为特定任务挑选最合适的工具一样重要。但是在选择模型时，不仅要选择合适的算法，还要考虑模型参数的设置。面对某些比较复杂的任务，即使选择了合适的算法，但参数设置不合理，训练出来的模型效果也未必优秀。这个过程可以用烹饪一道菜类比，同样是做“苗家酸汤鱼”，哪怕使用的主材料和做法都一致，也会因厨师的刀工、烹饪火候的区别，烧出来的菜的味道各不一样。

简而言之，调整机器学习算法的参数，就像是为特定的问题找到最合适的解决方案。要综合考虑问题的性质、数据集的特征以及计算资源，我们才能找到最适合的算法和参数设置，以实现最优的模型表现。



实践活动

投石车落地距离预测的不同算法对比

投石车可以以指定角度向空中抛出石头，我们以恒定的力量，分别从不同角度抛出，测量石头落地点距投石车的距离，可以得到一个“角度-距离”对照表。请使用多种回归算法训练“投石落地距离”预测模型，对比不同算法的效果，并填写表 1.3.2。

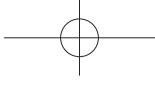
实验内容：以小组为单位，尝试编写代码，将不同算法对应的 R 平方值填入表 1.3.2 中，对不同算法的效果进行对比，并总结分析在该数据集上选择何种算法最合适。

实验准备：投石落地距离数据集。

表 1.3.2 投石模型训练效果对比及总结分析

算 法 名 称	R 平 方 值
线性回归 (LinearRegression)	
多项式回归 (Polynomial)	
支持向量机 (SVM)	
总结分析	

注：括号中的英文表示 BaseML 中的算法名。



项目实施

1. 了解了人工智能中最重要的研究方向——机器学习，进一步理解了机器学习的几大任务，你认为身高推断模型属于监督学习两类学习任务中的哪类任务？

2. 体验了机器学习的简单开发工具的使用和算法的选择，你准备选择哪些算法完成你的项目呢？

请在下面的横线上写出你的想法。

第4节 用机器学习解决问题

本节知识

- ◆ 数据集的收集和整理
- ◆ 机器学习的模型训练和评估
- ◆ 机器学习的模型应用

本节活动

- ◆ 训练回归模型推断身高
- ◆ 搭建一个身高推断系统

机器学习一般包括数据准备、模型搭建、模型训练与评估、模型应用等重要环节。接下来我们将详细探讨如何通过机器学习技术解决一个真实的问题，从问题分析开始，然后准备数据、选择算法、训练模型，最后进行评估和应用。通过这样的问题解决过程，我们可以感受机器学习的强大能力。

一、问题分析与数据准备

计算机不是万能的，人工智能也不是无所不能的。在尝试用机器学习解决问题的最初阶段，要先做问题分析与数据准备。先分析这一问题能不能用机器学习解决，属于哪一类问题，再来看能不能准备相应的数据。“巧妇难为无米之炊”，没有数据显然无法进行机器学习。

1. 问题分析

哪些问题适合用机器学习解决？一般来说，预测分析和模式识别（如识别图像、声音、文字）等看起来有规律的工作，都可以用机器学习的方法试一试。以“身高预测”为例，高个子的同学往往手的长度、鞋码、步伐间距都比较大，

显然这些因素之间有错综复杂的关联。如果拥有比较全面的身高和各种因素的对应表，那么就应该能训练出一个模型。这一任务就是经典的回归任务。至于各个因素之间的关系，机器可以通过特定的算法寻找。

查找相关资料发现，与身高有关联的因素很多，除了脚长，还有脚的宽度、步长、体重、性别等。脚长可直接由鞋码替代，身高、步长、脚宽均可测量，如图 1.4.1 所示。只要获取的数据足够多，就能创建出一个有效的数据集。



图 1.4.1 与身高有关联的因素

机器学习能不能从很多数据中寻找规律呢？答案是肯定的。前面使用的范例都比较简单，仅靠一个数据（特征）去预测另一个数据。机器学习的算法很强大，不仅可以找出多个数据和一个数据的关系，还能找出多个数据和多个数据之间的关系。

2. 数据收集

在上学期我们已经学习了很多数据收集的方法，包括通过在线调查和表单、物联网设备、自动化网络爬虫等进行收集。随着互联网的普及，使用在线调查和表单成为收集信息和数据的一种快速有效的方法。研究者可以设计调查问卷并分发给目标人群，参与者填写信息后直接通过网络提交，即可快速汇总各种信息，数据收集和整理在网络的支持下变得便捷而高效。这种方法特别适用于市场研究、消费者偏好分析和社会科学研究。一般而言，通过在线调查的方式收集身高数据最为方便。

问题讨论

制订一个详尽的计划将有助于确保数据收集过程的高效性和有效性，为构建一个准确的机器学习模型打下坚实的基础。请从数据多样性的角度出发，以小组为单位展开讨论：通过哪些途径来收集数据？应该收集哪些信息？如何设计调查问卷？

3. 数据整理

数据整理也称为数据清洗，涉及对原始数据进行错误识别、清理、修正和补全等工作，旨在为模型训练提供一个精确和完整的数据集。通常，初步采集的数据往往都会因为误操作等原因夹带一些有问题的数据。这些数据也称“脏数据”。如图 1.4.2 所示，这个通过问卷调查收集的数据集在身高（weight）、鞋码（foot size）等方面都存在“脏数据”，需要“整理”。数据整理的目的是保证数据的完整性、统一性和准确性。完整性涉及识别和处理数据集中的缺失值，统一性要求所有数据遵循相同的格式和标准，准确性涉及识别和校正数据中的错误或不合理的值。数据整理可能耗时较长，但它在整个机器学习项目流程中占据着不可替代的重要位置。

sex	weight	foot size	step	foot width	height
-1	52	36	59	6.5	155
1	70	41	65	8.5	170
-1	68	39	64	8.3	168
-1	56	38	61	7	160
1	62	4	70	8.5	185
1	55	43	61	6.8	160
-1	57	38	64	7.1	168
1	50	41	65	7.8	170
-1	110	39	61	7.6	160
-1	61	39	60	6.8	159
1	70	44	71	8.8	188
1	52	40	66	7.6	173
1	72	42	68	8.9	180
1	67	44	70	8.4	183
-1	52	37	64	6.7	168
1	70	42	63	7.6	167

注：sex—性别（-1 表示女性，1 表示男性）。

图 1.4.2 收集的问卷数据

4. 数据集划分

在数据整理完成后，下一步是将数据集划分为训练集和验证集。划分数据

集可以手动完成,也可以通过编写代码自动完成。BaseDT 是一个数据处理工具。借助 BaseDT 不仅可以快速完成数据集的划分,提高效率,还可以根据需求定制数据划分的比例、选择特征列和标签列等,非常方便。同时,BaseDT 的自动划分脚本可以被保存、重复使用和共享,确保了数据划分过程的一致性。

BaseDT 提供的“split_tab_dataset()”函数能将一个表格文件一分为二。下面的代码可以将待拆分的 CSV 数据集按照 8:2 的比例划分为训练集、验证集。返回值是训练数据(tx)、训练标签(ty)、验证数据(val_x)、验证标签(val_y),并且会将训练集和验证集保存为 CSV 文件,分别命名为“原始文件名_train.csv”和“原始文件名_val.csv”。

```
from BaseDT.dataset import split_tab_dataset
# 指定待拆分的CSV数据集
path = './data/Height_data.csv'
# 指定特征数据列、标签列、拆分比重
tx,ty,val_x,val_y = split_tab_dataset(path,data_column=range
(0,3),label_column=3,train_val_ratio=0.8)
```

二、模型训练与评估

数据准备好之后,就进入模型训练阶段。选择合适的机器学习算法是关键,对于回归问题,常用的算法包括线性回归、随机森林回归和决策树回归等。使用训练集数据训练模型,并使用验证集数据进行模型评估,以评估模型的性能。

1. 模型搭建与训练

在第3节中,我们已经学习了机器学习回归任务的常见算法及工具的使用。下面以 XEdu 工具集的机器学习库 BaseML 为例,实现模型搭建与回归模型训练,载入身高数据集并实现身高推断。以下即为利用 BaseML 选择线性回归算法训练回归模型的核心代码。

```
from BaseML import Regression as reg # 导入回归模块
```

```

model = reg('LinearRegression')           # 搭建线性回归模型
model.load_tab_data('./data/Height_data_train.csv') # 载入训练集
model.train()                             # 开始训练

```

模型搭建涉及算法的选择，需指定一个算法，如线性回归，还可以选择随机森林回归、支持向量机回归等，必要时还需要完成相关参数的设置。

2. 模型评估

当机器学习模型准备就绪后，下一步是看看它在实际中表现如何。准备的验证集在模型评估时会起到至关重要的作用。使用 BaseML 可直接用一行代码完成验证集数据的评估并计算 R 平方值，实现代码如下。

```

model.valid('./data/Height_data_val.csv',metrics='r2')
# 载入验证集，计算R平方值评估指标
model.metricplot()
# 模型评估指标可视化

```

运行结果如下。

验证r2-score为：58.189889580821564%

如图 1.4.3 所示，R 平方值在 0.6 左右，虽然不是很完美，但是在很多实际情况下也算是一个不错的结果了。

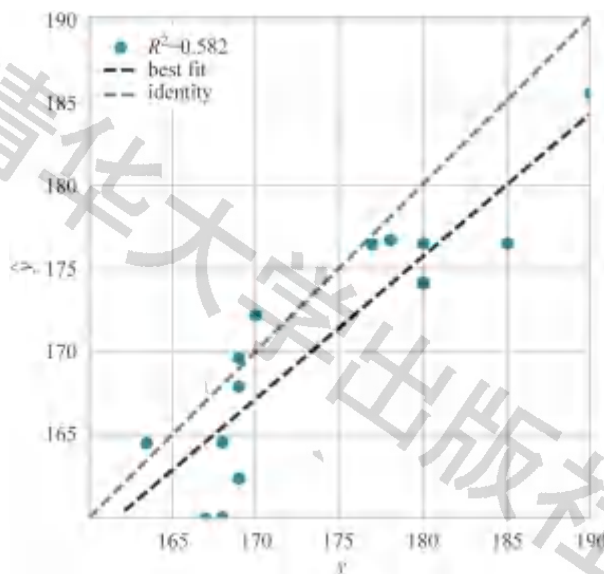


图 1.4.3 模型评估指标可视化

0

实践活动

训练回归模型推断身高

学习了 BaseML 训练模型的核心步骤和代码，现在请以小组为单位，使用小组准备的数据集，在此数据集基础上训练一个最好的身高推断模型并保存。核心实践内容包括：

- (1) 数据处理。对提供的数据进行整理和划分。
- (2) 完善训练代码。训练一个能预测身高的模型。
- (3) 选择算法并评估模型。尝试使用不同的回归算法训练身高推断模型，比较所训练的模型在对验证集进行模型评估时计算的 R 平方值，确认一个最好的模型。

三、模型优化和应用

机器学习虽然入门容易，但要训练出一个优秀的模型可不是一蹴而就的，还需要使用一些技术不断修改算法和调试参数。如果对模型预测的准确度有较高要求，就需要进一步优化模型。如图 1.4.4 所示，当训练的模型在评估阶段表现不够优秀，推理不够准确时，就需要从数据集和算法等多个角度进行检查并优化。一旦训练出一个不错的模型，就可以结合其他编程技术，将这个模型部署为一个程序或者部署到 Web 应用中，方便使用。



图 1.4.4 模型训练的一般过程

1. 模型优化

“学无止境”，训练模型也一样。如果训练的模型不够优秀，推理不够准确，那么就可以从数据集和算法两个角度进行检查并优化。

(1) 确保有一个高质量的数据集。什么是高质量的数据集？首先，尽量避免错误。训练数据中一旦混入了错误数据，那么训练出来的模型肯定效果很差，就好比你用错误的动作训练小动物，期望它能学会正确的表演。其次，数据量要大。如果仅仅提供了几条数据，那么肯定训练不出好模型。最后，数据量要尽可能扩大覆盖面。比如，仅仅用班级同学的数据训练身高推断模型，自然没有办法预测成人或者幼儿的身高。

(2) 选择合适的算法并将参数调到最优。算法的选择不仅要考虑任务的类型，还要综合考虑其他因素。比如，解决线性问题一般首选线性回归，但是数据集较小时，支持向量机可能是更好的选择；面对众多变量时，随机森林算法能够提供强大的处理能力；而如果想要深入理解变量之间的关系，那么决策树算法可能更加直观。对于大数据集和复杂问题，多层感知机表现出色，因为它能够很好地适应复杂的数据环境。但如果数据量有限，自适应增强回归可能是更好的选择。

2. 模型应用

模型应用是将训练好的模型部署到实际场景中。例如，集成到网站或移动应用中，让用户输入具体信息（如脚码大小和步伐长度）推断身高。模型的实际应用展示了机器学习技术解决实际问题的能力。

不同的人工智能模型有不同的推理框架，从模型训练到应用的一般流程如图 1.4.5 所示。一般来说，一个训练模型的工具也会自带推理功能，如在 BaseML 训练好模型并保存，下次使用时以同样的方式导入 BaseML 库并载入模型进行推理即可。还有一种方式是借助一些通用的模型推理库，如 XEdu 工

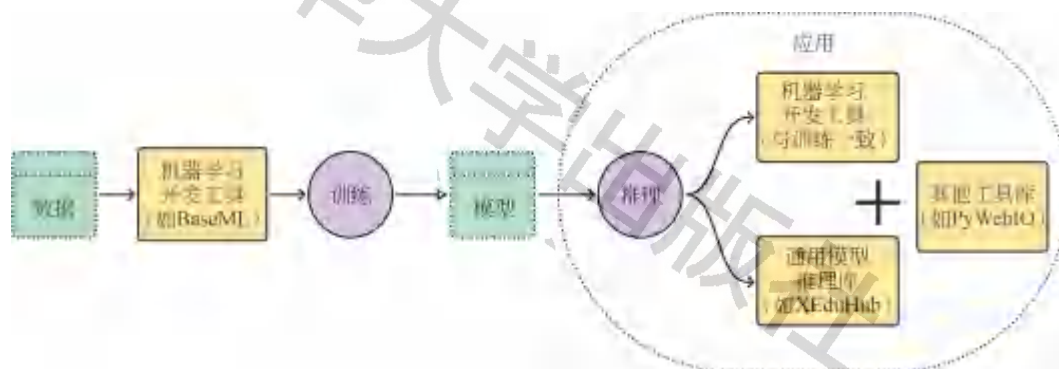


图 1.4.5 从模型训练到应用的一般流程

具中的 XEduHub 库，支持各种工具训练的模型完成模型推理，此类库的安装一般比机器学习开发工具简单很多。

下面的代码中，实现了用 BaseML 载入一个训练好的模型，再输入一组新数据，输出模型推理结果。

```
from BaseML import Regression as reg
model = reg('LinearRegression')    # 实例化模型
model.load('./model.pkl')          # 指定使用的pkl模型
data = [[53,38,7.8]]               # 指定测试数据，根据训练模型时使用的数据来定
y = model.inference(data)           # 进行模型推理
print(y)
```

借助 XEduHub 推理 BaseML 训练的模型的核心代码如下。

```
from XEdu.hub import Workflow as wf
baseml = wf(task='baseml',checkpoint='./checkpoints/model.
pkl')    # 指定使用的pkl模型
data = [[53,38,7.8]]    # 指定测试数据，根据训练模型时使用的数据来定
result= baseml.inference(data=data)    # 进行模型推理
print(result)
```

另外，要把自己训练好的模型放到网站上，让用户可以通过各类工具库（如 PyWebIO 和 Gradio）构建应用。只要把这些工具和推理代码结合起来，就能做出一个简单的智能系统。PyWebIO 的示例代码如下。

```
from pywebio.input import *
from pywebio.output import *
s = input('请输入你的名字: ')    # 文本输入
put_text('欢迎你,' + s)         # 输出文本
```

运行结果如图 1.4.6 所示，浏览器会自动打开一个本地的网址，出现以下界面，输入名字，单击“提交”后，会输出“欢迎你 + 名字”，短短四行代码就实现了一个简易的 Web 页面，只需修改相应的代码，就可以实现一个能推



图 1.4.6 PyWebIO 示例代码运行效果图

断身高的智能系统。

拓展阅读

机器学习模型推理库 XEduHub

XEduHub 是 XEdu 工具集的一款入门工具库，它像一个有各种螺丝刀、扳手、小刀等的工具箱，集成了许多机器学习领域的优质模型，可以直接利用它们完成不同的任务。此外，它还支持 ONNX 模型和 BaseML 训练的 pickle 模型的推理。

安装方法：在命令行中执行 `pip install XEdu-python` 或 `pip install xedu-python` 命令。

想要查看目前该库支持的所有任务，可以参照以下代码示例。

```
from XEdu.hub import Workflow as wf
wf.support_task() # 查看目前支持的任务
```



实践活动

搭建一个身高推断系统

已经学习了模型应用的相关知识，能否完成自己训练的模型的应用？请以小组为单位，参考资源包中提供的代码，利用 XEduHub 和 PyWebIO 搭建一个身高推断系统，实现效果参照图 1.4.7。



图 1.4.7 简易身高推断系统参照图

核心实践内容包括：

- (1) 完善代码，载入自己训练的模型并能完成推断结果的输出。
- (2) 增加个性化输入、输出交互设计。
- (3) 进一步思考：如果要制作一个功能更强大的身高推断系统，应该如何修改？

项目实施

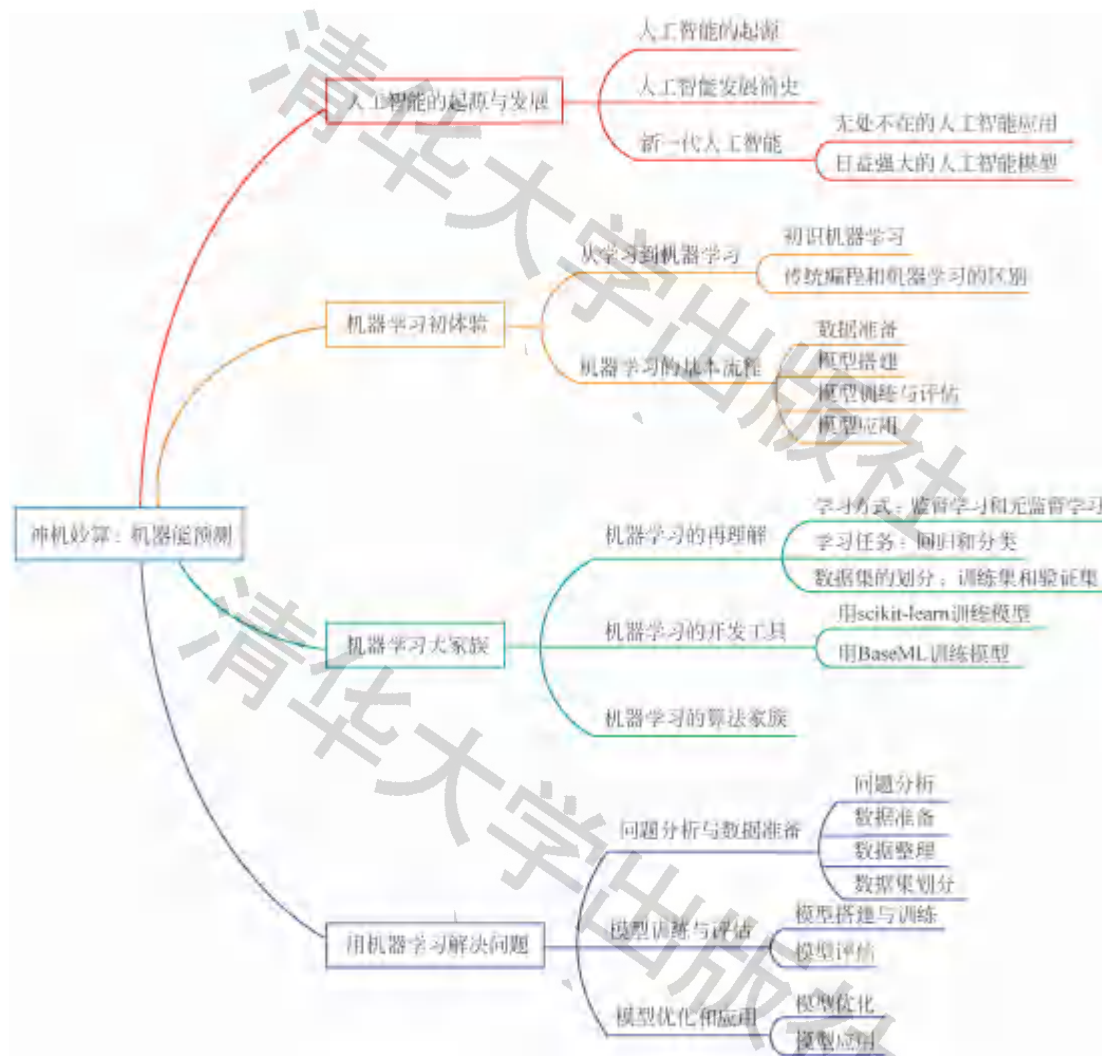
结合前面对数据准备、模型搭建、模型训练与评估、模型应用的学习，请你根据已拟定的项目方案，搭建一个简单的身高推断系统，完成问题深度分析，收集数据并完成数据集划分，选择合适的算法训练模型，为其搭建一个简易的网页应用，并填写表 1.4.1。

表 1.4.1 项目实施记录表

应用名称	简易身高推断系统
数据说明	收集方式： 涉及的变量： 划分比例： 数据量：
选择的算法	
模型应用工具	
程序代码	
实现效果	

单元小结

一、知识回顾



二、项目交流与评价

1. 参考本书附录“项目报告模板”撰写项目报告，并制作演示文稿。
2. 在课堂内展示自己的学习成果并分享经验，在下表中进行自评和他评。

项目成果评价表

评价维度	自评	他评
(1) 完整性 身高推断项目材料齐全，有数据收集规划、分工协作、项目实施记录表及最终成果（数据集、模型和展示系统）。	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般
(2) 实用性 成果内容具体，有真实的项目问题描述、有效的解决方案及相关工具等。	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般
(3) 规范性 项目报告规范，符合项目报告的一般格式要求，文字表述准确。	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般

3. 保存身高推断数据集与模型，整理源码等文档，并上传到校园网或者其他学习空间，与他人分享学习成果。



第 2 单元

洞明世事：机器能识别

学习导引

近年来，人工智能技术不断发展，深度学习作为一种基于人工神经网络的机器学习方法，在许多领域取得了突破性的进展，尤其在计算机视觉领域，深度学习显示出了显著的优势。2014 年，人工智能在人脸识别方面的准确率首次超过人眼；2015 年，微软亚洲研究院视觉计算组开发的计算机视觉系统，在 ImageNet 大规模视觉识别挑战赛（ImageNet Large Scale Visual Recognition Challenge，简称 ILSVRC）中首次超越人类进行对象识别分类的能力。之后，人工智能受到前所未有的重视，人工智能产业进入了快速发展阶段。

本单元将以“基于深度神经网络识别昆虫”为主题展开活动，让同学们体验图像识别的关键技术，感受深度学习技术的魅力，了解深度神经网络模型在图像识别领域的基本原理。

项目情景

小清同学的家在贵州省从江县的一个小镇上，这座山清水秀的小镇以种植各种农作物作为主要经济来源。近年来，小清发现农作物受到病虫害的影响，导致产量下降，但农村不少年轻人外出务工，除虫害劳动力不足，不由令人担忧。因此，小清希望利用无人机和深度学习技术识别多种昆虫，帮助家乡的农民监测农作物病虫害，减少经济损失，但在探索实践过程中遇到了一些问题。



- (1) 如何收集大量昆虫图像，整理出一份图像数据集？
- (2) 如何借助深度学习技术，训练一个深度神经网络模型，识别出不同的昆虫？
- (3) 如何将训练好的神经网络模型部署到实际应用场景中，完成一个智能应用？

你是不是也很感兴趣？让我们和小清一起了解计算机中的图像原理、认识神经网络技术、训练并部署神经网络模型帮助农民解决问题吧！

项目方案

经过咨询与了解，小青设计了以下方案：

知识学习	实施步骤	预期成果
(1) 认识神经网络和深度学习	(1) 搭建全连接神经网络，训练一个分类模型	(1) 对神经网络和深度学习的认识（PPT 格式）
(2) 认识卷积神经网络	(2) 用 EasyTrain 训练 LeNet 模型	(2) 一个昆虫数据集（ZIP 格式）
(3) 训练深度神经网络模型	(3) 收集与整理昆虫数据	(3) 模型训练代码并部署代码（代码、配套模型、运行视频）
(4) 评估与部署模型	(4) 用 MMEdu 训练一个昆虫识别模型，完成模型转换与部署	(4) 项目报告（PDF 格式）

你对小青的项目方案有什么不同的看法或建议？你准备如何设计项目方案？请填写在下表中。

知识学习	实施步骤	预期成果

项目分工

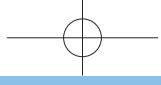
方案设计完成后，小青发现仅凭一己之力很难完成这个项目，于是邀请对此问题感兴趣的同学一起参与，并在项目方案中添加了以下表格。



姓名	角色	分 工	任 务
小青	组长	负责项目统筹、监督与管理	项目整体方案设计；项目实施过程的统筹、协调、监督、总结；项目文档撰写
同学甲	成员	负责进行数据集制作与文档撰写	数据采集，并制作数据集；撰写项目报告和相关文档
同学乙	成员	负责模型训练与评估	模型训练与评估的代码编写，训练一个满意的模型
同学丙	成员	负责模型部署与软硬件测试	设计模型部署的核心程序并测试程序，实现项目功能

你认为小青的项目组成员构成、分工和任务分配是否合理？请在下表中填写你的项目分工情况。

姓名	角色	分 工	任 务



第1节 神经网络与深度学习

本节知识

- ◆ 单层感知机模型的基本结构
- ◆ 多层感知机的基本原理和全连接神经网络的应用
- ◆ 深度学习的提出和优势
- ◆ 深度神经网络的开发框架和工具

本节活动

- ◆ 用全链接神经网络训练鸢尾花分类模型

线性回归模型在预测简单关系（如根据脚长预测身高）时表现良好，但在面对自然界的复杂关系（如看图识物、下围棋、自动驾驶等）时则显得力不从心。科学家为处理复杂的输入、输出关系进行了大量的尝试，设计了多种机器学习的算法，如非线性回归、支持向量机和神经网络等。其中神经网络

能够很好地表示复杂的物理变量关系。本节课，我们将从神经网络的起源出发，带领同学们逐步了解深度神经网络模型强大的预测能力。

一、人工神经网络的起源

20 世纪初期，科学家就已经知道人类的大脑有超过 800 亿个神经元。神经元的工作机制是当外部刺激达到一个阈值时，神经元会向下一个神经元发出信号。当时很难解释，为什么大量功能单一的神经元连在一起就能形成思维和智慧。1943 年，美国神经生理学家沃伦·麦卡洛克（Warren McCulloch）和数学家沃尔特·皮茨（Walter Pitts）合作，提出了“M-P 模型”，解释了如何通过神经元连接的网络进行逻辑运算，并提出了“人工神经网络”（artificial neural network）这一概念，如图 2.1.1 所示。

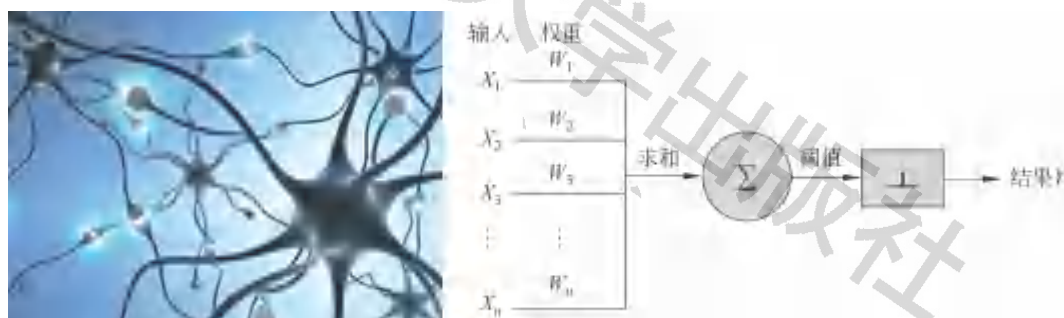


图 2.1.1 神经元和“M-P 模型”示意图

二、人工神经网络的发展

“M-P 模型”的提出在当时并没有引起人们的重视，直到 1957 年弗兰克·罗森布拉特（Frank Rosenblatt）设计了感知机（perceptron）模型，这才引发了一次人工智能领域的研究热潮。感知机模型是第一个具有学习能力的神经网络，罗森布拉特还造出了第一台硬件感知机“Mark-1”，它经过学习后能识别出英文字母，如图 2.1.2 所示。



图 2.1.2 罗森布拉特和感知机“Mark-1”

1. 单层感知机

罗森布拉特设计的感知机只有一层，也称单层感知机。单层感知机的成功引发了联结主义的兴起。但不久之后，人工智能奠基人之一的马文·明斯基

(Marvin Lee Minsky) 和麻省理工学院的西蒙·派珀特 (Seymour Papert) 从数学和逻辑上证明了单层感知机的重大局限——只能解决“线性可分”问题。明斯基还认为，虽然通过多层感知机可以解决线性不可分的问题，但连接数量太多会导致无法训练，研究两层乃至更多层的感知机是没有价值的。这一论断引发了连锁反应，给人工智能学科带来了沉重的打击。

拓展阅读

线性可分与线性不可分

线性可分与线性不可分是机器学习中划分数据集的两个术语。“线性可分”指数据集存在一个线性边界，使所有属于一个类的数据点都位于这个边界的一侧，而所有属于另一个类的数据点都位于另一侧。如图 2.1.3 (a) 所示，假设有两类数据点散布在二维平面上，一类是圆点，另一类是叉点。如果能找到一条直线，把两类数据分开，那么这两组数据点是线性可分的；相反，如果找不到这样的直线，那么该数据集被视为线性不可分。如图 2.1.3 (b) 所示，某些圆点被包围在由叉点形成的圈中，我们找不到一条单一的直线将它们完全分开，所以这个数据集是“线性不可分”的。

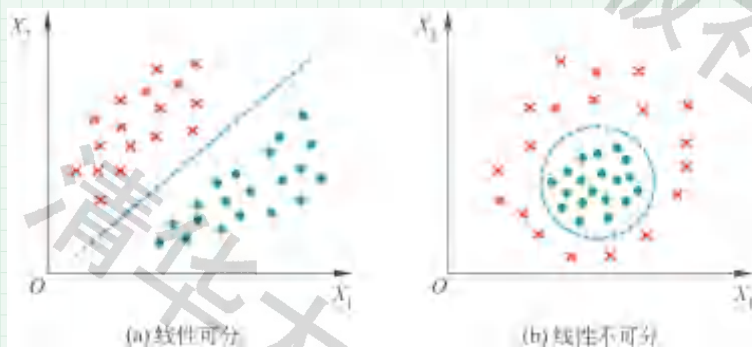


图 2.1.3 线性可分与线性不可分示意图

2. 多层感知机

因为明斯基对感知机的论断，神经网络被打入冷宫，所有“以机器模拟大脑结构”的研究被视作“异端”，只剩下少数人还在坚持。1986 年，杰弗里·辛

顿（Geoffrey Hinton）教授发现可以通过特定算法（误差反向传播）对多层神经网络进行训练，有效地解决更为复杂的非线性问题，人工神经网络逐步从单层走向了多层。

如图 2.1.4 所示，在单层感知机模型的基础上增加了多个隐藏层，形成了多层感知机（multilayer perceptron, MLP）模型，也称为多层神经网络。正如明斯基的论断，随着隐藏层的增加，神经网络的非线性表达能力将得到大大增强。

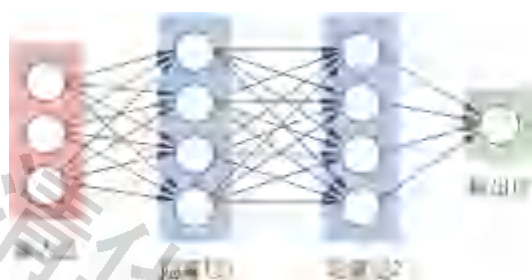


图 2.1.4 多层神经网络示意图

问题讨论

单层感知机增加若干隐藏层就组成了多层感知机，既然多层感知机能解决这么多问题，为什么不采用“多多益善”的思路，建构模型的时候尽可能多添加隐藏层？模型的隐藏层增多会导致哪些问题？

3. 全连接神经网络

全连接神经网络是一种由多层感知机构成的基本网络结构，可以处理更复杂的问题。例如身高推断模型，如果把“每天锻炼身体的时间”“饮食健康程度”“父母身高”等因素都加入数据集，那么多项式回归算法无法完成这个任务，但全连接神经网络依然能出色地完成任务。此外，全连接神经网络能同时适用于分类任务和回归任务，这就有点儿像“一招应万变”的必杀技。

机器学习中也有个算法叫做多层感知机，其实就是全连接神经网络。BaseNN（搭建简单神经网络的工具）和 BaseML 两个工具都提供了搭建多层神经网络模型的功能。以输入 4 个变量推断身高的回归模型训练为例，用这两个工具搭建全连接神经网络的参考代码如表 2.1.1 所示。

表 2.1.1 用两个工具搭建全连接神经网络

用 BaseML 搭建多层感知机算法	用 BaseNN 搭建全连接神经网络
<pre># 导入BaseML库 from BaseML import Regression as reg model = reg('MLP') # 搭建模型 model.set_para(hidden_ layer_size=(10,5))</pre>	<pre># 导入BaseNN库 from BaseNN import nn model = nn('reg') # 搭建模型 model.add(layer='Linear',size=(4, 10), activation='ReLU') model.add(layer='Linear',size=(10, 5), activation='ReLU') model.add(layer='Linear',size=(5, 1))</pre>

表 2.1.1 所示两段代码都能搭建一个输入维度为 4，输出维度为 1，隐藏层数量为 2 的全连接神经网络，如图 2.1.5 所示。使用这一网络，能够实现对多个输入变量进行训练，并得到可预测结果的模型。但要搭建更加复杂的神经网络，只能使用 BaseNN。若使用这个神经网络来做分类任务，则需要将“model = nn('reg')”修改为“model = nn('cls')”，并相应调整输出维度。其中“reg”指代回归任务，“cls”指代分类任务。

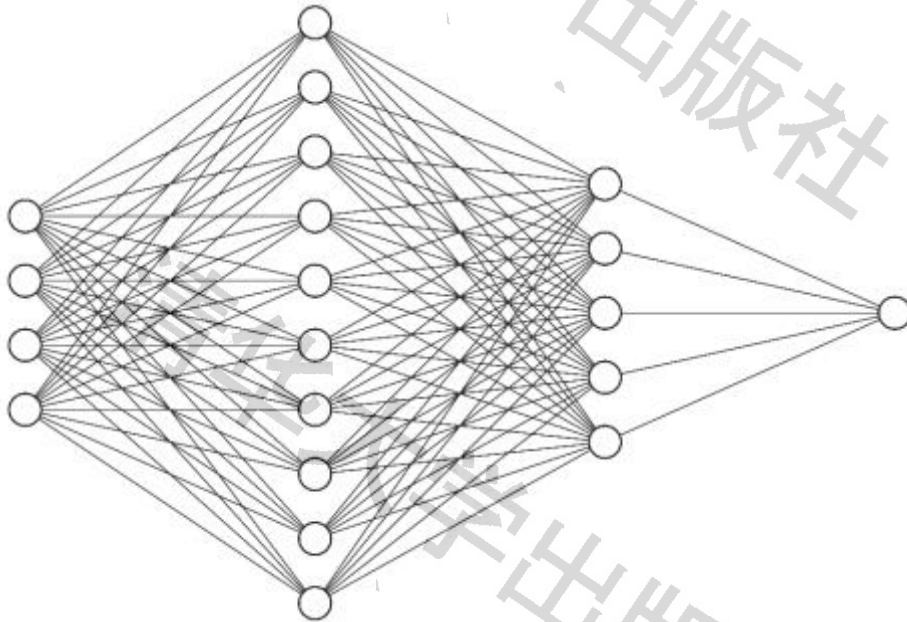


图 2.1.5 具有 2 个隐藏层的全连接神经网络

我们使用身高数据集训练这个神经网络时，会得到神经网络的各个神经元之间连接的“权重”。我们通常说的“模型”，实际上包含神经网络的结构和

训练得到的“权重”数据。为了更好地理解神经网络，下面我们结合图 2.1.5，介绍常见的专业名词。

- 神经元。图中的每一个圆圈表示一个神经元。
- 层。图中每一列就表示一层，除去输入层和输出层，这个图中有两个隐藏层。在 BaseNN 中，“model.add()”用来增加层，“layer='Linear'”表示线性层，后面我们还会学到更多的层。“size=(10, 5)”表示隐藏层的输入维度是 10，输出维度是 5。
- 参数。所有神经元之间的连接线就是参数，参数包含“权重”。模型搭建好后，参数是随机的，训练模型实际上是在“找”合适的参数。
- 激活函数。激活函数用来模拟神经元控制继续传递信号的方式，可以看成是一个 if 语句，若符合条件则向下一个神经元输出信号。常见的激活函数有 sigmoid、tanh、relu 和 softmax 等，对于分类模型来说，最后一层用的都是 softmax。



实践活动

用全连接神经网络训练鸢尾花分类模型

鸢尾花虽然常见但品种很多，只有资深的花农才能辨别。1936 年，统计学家、生物学家罗纳德·费希尔（Ronald Fisher）在加拿大加斯佩半岛上，测量了一批有三个种类的鸢尾花，形成了一个鸢尾花数据集。该数据集包含 150 个数据样本，分为 3 类（*Iris versicolor*、*Iris setosa* 和 *Iris virginica*），每类有 50 个数据，每个数据包含 4 个属性（花萼长度、花萼宽度、花瓣长度和花瓣宽度），如图 2.1.6 所示。



图 2.1.6 3 类鸢尾花

请以小组为单位，使用多层神经网络训练模型，实现对鸢尾花的分类。核心实践内容包括：

- (1) 使用 BaseNN 搭建多层神经网络。
- (2) 载入数据集，训练模型并测试。
- (3) 进一步思考：使用类似的网络结构训练身高预测模型，能否得到不错的准确度？

三、从浅层学习到深度学习

虽然辛顿等人提出的误差反向传播算法给神经网络研究注入了新的希望，但是没有直接促成联结主义研究的复兴。人们之所以不愿意研究具有更多隐藏层的神经网络，背后的重要原因之一依然是隐藏层数量增加会带来更高的训练难度和更大的训练数据需求。这种困局持续了 20 年才得到改变。

1. 深度学习的提出

2006 年，辛顿等人发表了论文《通过神经网络进行数据降维处理》。在该论文中，辛顿提出了“深度学习”的概念，第一次清晰地说明了深度学习相对于传统浅层学习（指隐藏层少的神经网络）的价值和优势。深度学习是指利用深度神经网络技术进行机器学习的一种过程，而深度神经网络是指拥有多个隐藏层的神经网络，如多层感知机。

深度学习的最大优势是什么？简而言之，神经网络可以在学习过程中逐层自主提取数据特征。我们已经知道，在传统的机器学习中选取特征是关键。比如训练身高预测模型时，选择兴趣爱好、皮肤颜色、头发长短等特征，再好的算法也将束手无策。因此，机器学习能否成功依赖于人工分析是否准确。如果要让机器分辨图像中的动物是猫或者狗呢？按照人工分析的老办法显然行不通，因为在图像中寻找特征（猫和狗的区别）非常困难。但是，深度神经网络可以在训练过程中自动提取图像、语音和文本中的有效特征，甚至效果比人工分析还要好。这一优势对于机器学习来说非常显著，如图 2.1.7 所示。

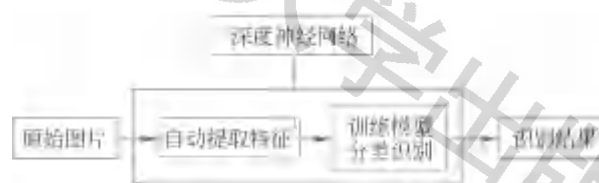


图 2.1.7 深度神经网络的优势

有了深度学习的理论基础之后，研究人员开始不断挖掘深度学习的价值。2012年多伦多大学开发的 AlexNet 网络模型在著名的 ImageNet 大规模视觉识别挑战赛中夺冠，仅仅 8 层就远超第二名，展现出了超越传统机器学习算法的性能。此后，基于深度学习的系统在图像分类上的错误率持续下降（见图 2.1.8），其识别能力已超越人类。各种深度神经网络结构也不断涌现（如卷积神经网络、循环神经网络、残差网络等），深度学习能解决的任务也越来越多，在模式识别、自动控制、生物、医学、经济等领域成功解决了大量难题。在通用并行计算平台（如 CUDA）的支持下，深度学习训练模型的速度也在加快。目前，深度学习已经成为人工智能最重要的研究方向。2014 年，生成对抗网络（GAN）的出现开启了生成式人工智能的新时代。

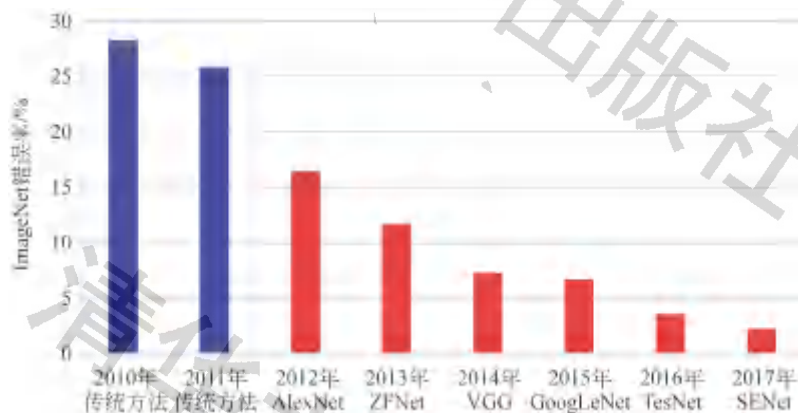


图 2.1.8 ImageNet 大规模视觉识别挑战赛历年成绩

问题讨论

深度学习的巨大优势背后，需要哪些技术的支持？请谈谈深度学习与大数据技术、机器高速运算能力的关系。

2. 深度学习的开发工具

很多编程工具都支持深度学习开发。由于具有开源和易用的特点，Python 语言成为了人工智能编程的首选语言。它拥有多个人工智能开发框架和工具包，如 TensorFlow、Keras、PaddlePaddle 和 PyTorch 等，如表 2.1.2 所示。

表 2.1.2 常见的深度学习开发框架和工具包

名 称	开发团队	发布时间	功 能 特 点
TensorFlow	谷歌	2015 年	是很多人进入人工智能领域第一个听到或者接触的开发框架
Keras	谷歌	2015 年	是在 TensorFlow 的基础上发展起来的，开发门槛更低，拥有大量的用户。可以把 Keras 看成 TensorFlow 的入门简化版本
PaddlePaddle	百度	2016 年	集核心框架、基础模型库、端到端开发套件、丰富的工具组件、用户社区于一体，是中国首个自主研发、功能丰富、开源开放的产业级深度学习平台
PyTorch	Facebook	2017 年	用于张量计算、自动微分和 GPU 加速，深受科研人员喜爱

TensorFlow 和 PyTorch 是迄今为止最受用户欢迎的两个人工智能开发框架，都拥有丰富的编程接口、广阔的用户群体，目前广泛用于学术研究和商业应用。Keras 简化了 TensorFlow 的开发门槛，更适合初学者。PyTorch 也支持 Keras，并且出现了 FastAI 以及 OpenMMLab 等工具，同样拥有大量用户。

2022 年 12 月，上海人工智能实验室浦育团队发布了开箱即用的深度学习开发工具——MMEDu，在中小学教育领域引起了广泛关注。随后，浦育团队又相继发布了 BaseML、BaseNN 和 XEduHub 等工具，与 MMEDu 合并为 XEdu，成为中小学生学习人工智能的必备工具之一。

第2节 卷积神经网络及其应用

本节知识

- ◆ 认识卷积神经网络
- ◆ 训练卷积神经网络
- ◆ 用训练好的网络模型进行推理

本节活动

- ◆ 字符画的制作
- ◆ 用 BaseNN “观察” LeNet 模型
- ◆ 训练一个昆虫分类模型

随着层数的加深，神经网络从数据中提取特征的能力也大大提升，但新的问题紧跟而来：相对于 CSV 数据等表格数据，图像和语音中的数据量要大很多。网络层数、图像分辨率等因素的增加，使所需要的参数急剧增长，不仅会导致模型训练时间变长，模型推理效率变低，

甚至还会导致模型无法正常训练。于是，科学家再次从生物神经学领域找到灵感，仿照人类的视觉处理机制提出了卷积神经网络，有效地解决了神经网络在计算机视觉领域的应用难题。

一、认识卷积神经网络

卷积神经网络(convolutional neural network, CNN)是一种特殊的神经网络，在深度学习中应用非常广泛，也是深度神经网络中最具影响力的一种模型。要理解卷积神经网络，首先要了解图像的数字化原理。

1. 图像的数字化原理

图像是由一个个像素点构成的。对于黑白图像来说，像素点的值只有两

种，即 0 和 1，分别代表黑、白两种颜色。而灰度图像中像素点的亮暗程度，则是通过 0~255 的数值控制，从而显示出不同的图案。将一张灰度图像压缩为 25×25 像素，再通过 Python 代码读取图像灰度值，结果如图 2.2.1 所示。

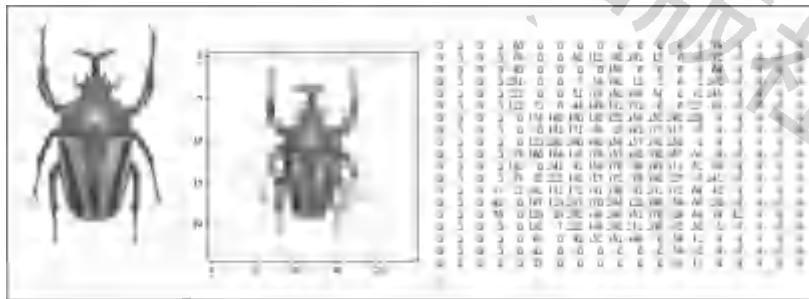


图 2.2.1 灰度图像及其灰度值

彩色图像的数字化原理与黑白图像类似，只不过是用 RGB（红、绿、蓝）三原色混合而得到各种色彩。如图 2.2.2 所示，用 Python 代码读取彩色图像，每一个像素点会出现由三个 0~255 的数值组成的列表，这三个值对应像素点的 R、G、B 三种颜色。图 2.2.2 中的 [255 255 255] 表示白色，[174 179 184] 表示一种带有蓝绿色调的浅灰色。

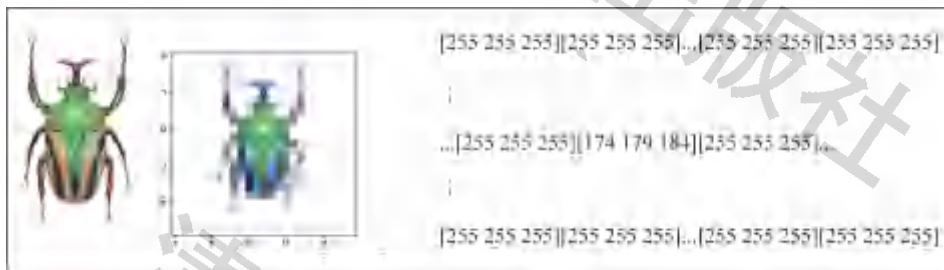


图 2.2.2 彩色图像及像素值



体验活动

字符画的制作

一张灰度图像的亮暗程度通过 0~255 的数值控制，小青编写了一个有趣的 Python 代码，根据数值大小输出不同的字符，从而得到了一幅字符画，如图 2.2.3 所示。请打开“字符画.jpynb”文件，体验字符画的制作。



Figure 1.10: A stylized drawing of a person's head and shoulders, facing right. The person has dark, wavy hair and is wearing a dark, high-collared garment. The drawing is composed of thick, black, expressive strokes. The background is white. The drawing is positioned on the left side of the page, with the right side of the page being blank.

图 2.2.3 有趣的字符画

2. 卷积神经网络的作用

神经网络用图像作为训练数据时，需要读取整幅图像作为神经网络模型的输入。以全连接神经网络为例，假如输入的是一幅 100×100 像素的灰度图像，那么输入层就有 100×100 个神经元，隐藏层任何一个神经元都将有 100×100 个参数需要训练，即使隐藏层和输入层的神经元数量一致，仅仅两层神经网络就有 $100 \times 100 \times 100 \times 100$ 个参数。这个计算量非常可怕。因此，神经网络层数不断增多，会出现“参数爆炸”的情况。如图 2.2.4 所示， x 表示输入元素，每一根连接线代表一个参数，图（b）比图（a）多了 2 个输入数据，参数却多了 10 个。如果输入元素由 5×5 像素（5 行 5 列 25 个像素）的图像变为 10×10 像

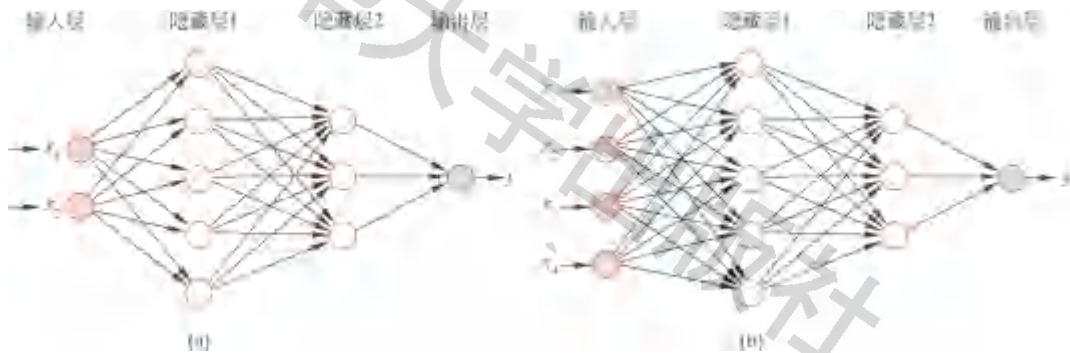


图 2.2.4 输入数据对参数的影响

素的图像，参数会迅速变为原来的很多倍。

科学家在脑神经科学领域找到了新的突破，他们参考动物视觉感受野的理论设计了卷积神经网络。卷积神经网络中有两种特殊的网络，分别为卷积层和池化层。卷积层用来提取图像的特征，池化层用来减少数据的运算量。对于计算机来说，图像不过是一个充满数字的表格，这种表格在数学上叫做数字矩阵，而卷积（convolutional）和池化（pooling）就是数字矩阵中两种重要的“计算”方法。如图 2.2.5 所示，与全连接神经网络相比，卷积神经网络的卷积层每个神经元仅与上一层的某个区域连接，这样参数量得以大大减少。

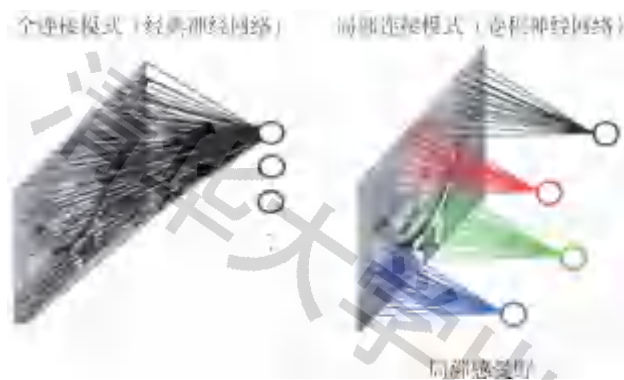


图 2.2.5 全连接神经网络和卷积神经网络的比较

卷积神经网络最早由图灵奖获得者杨立昆（Yann LeCun）教授于 1998 年提出，随后在各个方面被广泛应用。2012 年辛顿团队夺冠的 AlexNet 是卷积神经网络的代表作之一。经过多年的发展，除了图像处理之外，语音识别、自然语言处理等领域也都离不开卷积神经网络。

拓展阅读

从视觉感受野到卷积神经网络

1959 年，神经科学家大卫·休伯尔（David H. Hubel）和托斯坦·威泽尔（Torsten Wiesel）在哺乳动物视觉领域上有一个重大的发现：不管眼睛看到了什么东西，当图像进入大脑的视觉皮层时，神经元都会将它们拆分成一系列形状的组合。这一发现不仅获得了诺贝尔生理学或医学奖，同时也为人工智能研究提供了两大思路。一种是从局部连接到全局感知。

每个神经元只需要对视野的局部（即局部感受野）进行感知，然后将局部信息综合起来而得到全局信息。另一种是在视野（也就是在整张图像）上进行重复操作。用多个不同功能的神经元对图像进行多次检测。

卷积层提取图像特征的主要操作是用特定的卷积核与图像进行卷积运算。可以把卷积核看成一个过滤器，拿过滤器在图像上滑（检测）一遍，图像中与卷积核的相似度高的特征就会被记录下来，即用卷积运算对图片进行特征提取。如图 2.2.6 所示，我们用两个卷积核把图像的竖直边缘和水平边缘特征提取出来。

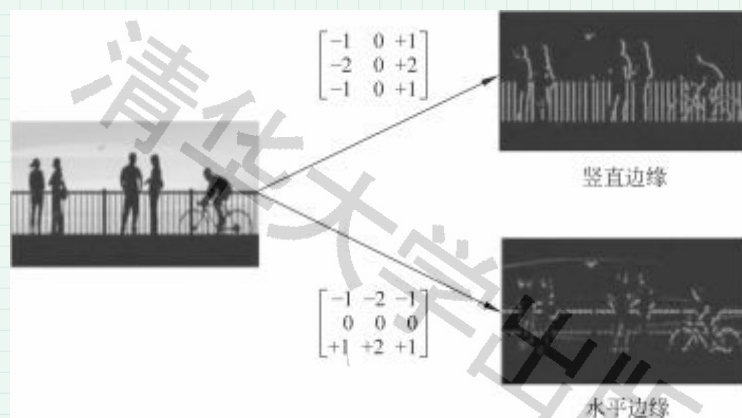


图 2.2.6 用卷积核提取图像特征

二、LeNet 模型的搭建

LeNet 模型由杨立昆提出，是第一个成功应用于数字识别问题的卷积神经网络，在经典的数字手写体（MNIST）数据集上，LeNet 模型可以达到约 99.2% 的正确率。LeNet 模型是卷积神经网络入门的模型，至今依然被广泛应用于一些简单的图像分类场景。

1. 认识 LeNet 模型的结构

LeNet 模型有多个版本，一般指 LeNet-5 模型。如图 2.2.7 所示，在每一

个卷积层的后面都有一个池化层，用来减少数据和参数的数量。很多人模仿 LeNet 模型，通过增加或者减少卷积层和池化层的数量，搭建解决特定任务的神经网络模型。如果数据集使用的是灰度图像，特征比较明显且图像尺寸较小，直接使用 LeNet 模型一般也能取得很不错的效果。

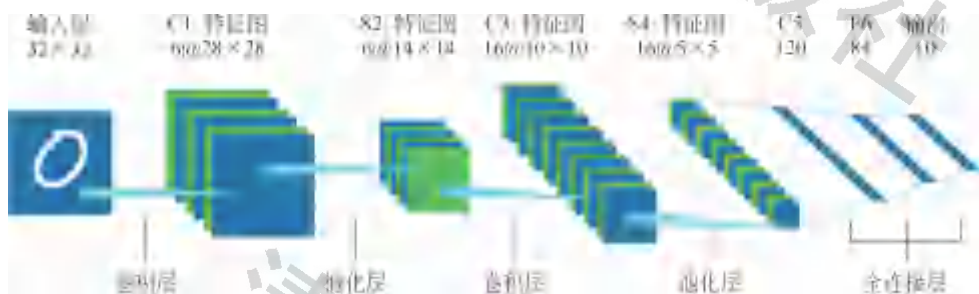


图 2.2.7 LeNet 模型的结构

LeNet-5 模型共有七层，输入的是 32×32 像素的灰度图像，经过一层卷积后，得到 6 个 28×28 的特征图；这些特征图经过池化后，变为 6 个 14×14 的特征图；特征图再依次经过卷积、池化、展平处理，得到 120 个 1×1 的特征图；最后经过两个全连接层后，实现 10 个手写数字分类。LeNet-5 模型的参数数量为 6 万多个，是一个效率非常高的模型。

2. 用 BaseNN 搭建 LeNet 模型

使用 BaseNN 可以搭建出一个 LeNet 模型且代码输出结果会呈现各层说明。

```
# 搭建LeNet网络
model.add('Conv2D', size=(1,6),kernel_size=(5,5),
activation='ReLU')
model.add('MaxPool', kernel_size=(2,2))
model.add('Conv2D', size=(6,16), kernel_size=(5,5),
activation='ReLU')
model.add('MaxPool', kernel_size=(2,2))
model.add('Linear', size=(16*5*5,120), activation='ReLU')
model.add('Linear', size=(120,84), activation='ReLU')
model.add('Linear', size=(84,10), activation='Softmax')
```

运行结果如下。

增加二维卷积层,输入维度:1,输出维度:6,kernel_size: (5, 5)
使用relu激活函数。
增加最大池化层,kernel_size: (2, 2)
增加二维卷积层,输入维度:6,输出维度:16,kernel_size: (5, 5)
使用relu激活函数。
增加最大池化层,kernel_size: (2, 2)
增加全连接层,输入维度:400,输出维度:120。
使用relu激活函数。
增加全连接层,输入维度:120,输出维度:84。
使用relu激活函数。
增加全连接层,输入维度:84,输出维度:10。
使用softmax激活函数。

上面的代码中,用“model.add()”添加神经网络层,其中 Conv2d 表示添加的是卷积层,MaxPool 表示添加的是最大池化层,“Linear”表示添加的是全连接层。七层网络结构中前几层使用“relu”激活函数,最后一层使用“softmax”激活函数。



体验活动

用BaseNN“观察”LeNet模型

BaseNN 内置函数 visual_feature() 可以查看图片在网络中传递时发生变化的过程,从而让我们更直观地了解各个层的作用,如图 2.2.8 所示。请运行教材资源包中的代码“观察 LeNet 模型.ipynb”,输入一张图片,“观察”经过每一层后的数据变化,以及卷积层提取的特征。

```
import cv2
from BaseNN import nn
model = nn('cls')
model.load('./mn_ckpt/basenn.pth') # 保存的已训练模型载入
path = './test_IMG/single_data.jpg' # 指定一张图片
img = cv2.imread(path,flags = 0) # 图片数据读取
model.visual_feature(img,inlimg = True) # 特征的可视化
```



图 2.2.8 特征图可视化

三、LeNet 模型的训练

深度学习是机器学习的一个重要分支，因此深度学习的模型训练流程和机器学习是一致的。相对来说，深度学习在训练数据的容量、模型搭建的难度、训练方法的复杂度和对算力的要求方面，都超过其他机器学习。“数据是燃料，模型是引擎，算力是加速器”，支持并行计算的 GPU 设备成为深度学习模型训练的基础设施。

拓展阅读

贵州着力打造全国算力保障基地

《国务院关于支持贵州在新时代西部大开发上闯新路的意见》（国发〔2022〕2号）明确提出：“加快推进‘东数西算’工程，布局建设主数据中心和备份数据中心，建设全国一体化算力网络国家枢纽节点，打造面向全国的算力保障基地。”2023年11月，贵州出台《关于促进全国一体化算力网络国家（贵州）枢纽节点建设的若干激励政策》，从支持算力中心建设、加快数据归集和数据流通交易、支持算力产业发展等9个方面提出了具体举措，着力打造面向全国的算力保障基地。

1. 常见的深度学习模型训练工具

深度学习的开发框架都可以用来训练深度学习模型，但难度较高。为了降低模型训练的技术门槛，一些企业和研究机构逐步推出了专用的深度学习模型训练工具或者平台，比如微软的 Azure ML、亚马逊的 AWS SageMaker、华为的 ModelArts、百度的 BML 等。这些平台支持开发者以低代码或者无代码（也称零代码）的形式训练模型，即不写代码或者写一点点代码，就能训练出深度学习的模型，如图 2.2.9 所示。



图 2.2.9 百度的 BML 平台界面

在用 BaseML 训练机器学习模型的过程中，我们可以看出训练模型的程序代码其实非常简单，甚至看不到分支结构和循环结构。因此，开发无代码训练深度学习模型的工具的难度并不高。一些面向青少年学习人工智能的平台也增加了无代码训练的插件，如 XEdu 系列工具中的 EasyTrain（见图 2.2.10）。



图 2.2.10 EasyTrain 任务选择界面

2. 用无代码工具训练 LeNet 模型

EasyTrain 是 XEdu 中的无代码训练插件，需要在 MMEdu 和 BaseNN 的环境下运行。借助 EasyTrain，不需要编写代码就可以训练一个 LeNet 模型，适用于人工智能入门初学者。

启动 EasyTrain 后，浏览器将自动打开一个本地网页。页面上方呈现一个进度条，根据进度流程的提示完成所有操作即可完成模型训练。LeNet 是一个图像分类模型，要先选择“分类任务”，再选择“LeNet”模型，完成“数据集选择”。然后开始设置参数，其中分类数量要和数据集保持一致，最后单击“生成代码”按钮，网页中将生成一段训练代码，如图 2.2.11 所示。

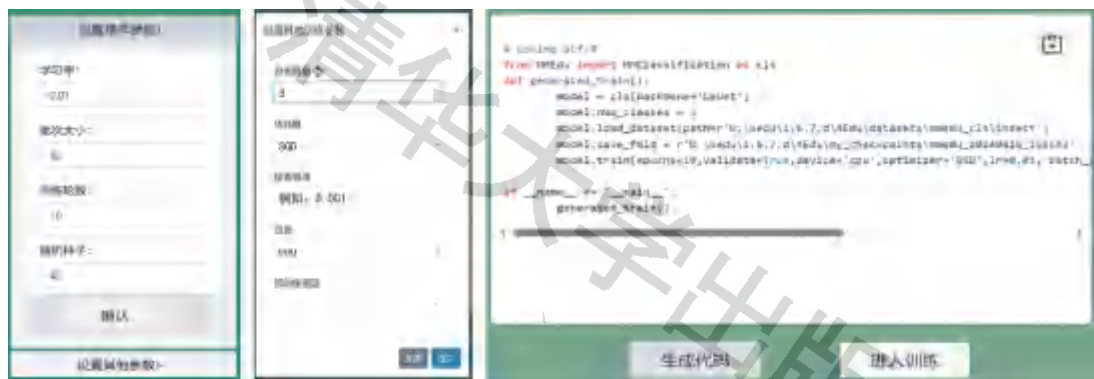


图 2.2.11 EasyTrain 参数设置界面

单击“进入训练”按钮，浏览器便会跳转至训练页面。页面中的“loading”表示模型正在训练中。如图 2.2.12 所示，训练过程中会可视化呈现两个图，分别是损失函数曲线（Loss Chart）和识别准确率曲线（Accuracy Chart），图中的横坐标为训练轮数，纵坐标为对应数值。训练结束后，页面上会显示模型保存的路径。

使用 EasyTrain 训练模型虽然不需要编写代码，但也需要做一些准备工作，比如，准备一个符合要求的数据集，并保存在规定的位置（关于数据集的格式在下一节介绍）；再如，了解各种模型和超参数的作用。图 2.2.11 中的超参数设置是训练模型工作的重点，如轮次（epoch）、学习率（lr）等。轮次表示要训练多少轮，学习率用来控制模型在训练过程中的模型权重更新速度。第一次训练模型时可以直接使用默认值训练。随着对模型和超参数理解的加深，我们就能基于其他网络结构（如 MobileNet、ResNet 等）训练出更好的模型。

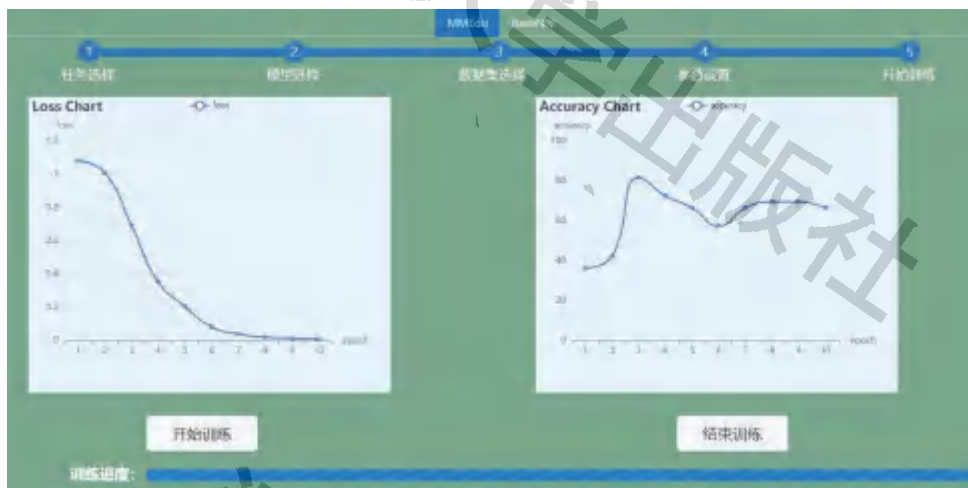


图 2.2.12 EasyTrain 训练页面

拓展阅读

自动化机器学习

自动化机器学习 (automated machine learning, AutoML) 的出现是为了简化机器学习的整个流程, 让非专业的用户也能够利用机器学习的技术解决实际问题。可以把 AutoML 想象成一台“自动洗衣机”, 训练模型就像我们把脏衣服放进洗衣机, 选择合适的洗涤程序, 按下“启动”按钮, 洗衣机就会自动完成清洗、漂洗、脱水等一系列复杂的过程。同样, 只要提供数据, AutoML 也会自动完成数据预处理、模型选择、超参数调整、训练、评估等一系列复杂的机器学习流程。即使是对机器学习一窍不通的人, 也能够轻松地利用机器学习解决实际问题。



实践活动

训练一个昆虫分类模型

借助 EasyTrain, 不需要编写代码就能训练各种经典的卷积神经网络模型, 用来解决各种图像分类的问题。请以项目组为单位, 使用 XEdu 内置

的昆虫数据集训练 LeNet 模型,并体验模型的推理准确度。核心实践内容包括:

(1) 在 EasyTrain 中选择分类任务,并选择 LeNet 模型和昆虫数据集,完成类别、数量等参数设置并进行训练。

(2) 生成代码并启动模型训练,观察准确率变化和训练时长。

(3) 使用模型转换得到的 ONNX 模型和代码,按照代码说明运行并输入测试集中的图片进行推理。

注:ONNX(open neural network exchange)是一种开放的深度学习模型交换格式。

项目实施

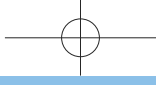
小清完成了卷积神经网络的训练,并使用训练的模型实现了推理。接下来,我们尝试训练一个昆虫识别模型吧!

针对昆虫识别模型的实现,你准备收集哪几类昆虫?用哪些渠道来收集?准备收集多少张?请在科学老师的指导下,收集数据。

昆虫类别	
收集渠道	
分工	
预计的数据规模	

EasyTrain 中还提供了很多图像分类的神经网络模型。请访问 XEdu 中的文档自主学习,记录这些模型的特点并思考你的昆虫分类模型将使用哪一种网络结构。

模型名称	适合解决哪些任务
LeNet	
MobileNet	
ResNet18	
ResNet50	



第3节

用深度学习实现图像分类

本节知识

- ◆ 准备图像数据集
- ◆ 训练图像分类模型
- ◆ 图像分类模型的应用

本节活动

- ◆ 整理一个 ImageNet 数据集
- ◆ 用 MMEdU 训练图像分类模型
- ◆ 用 Gradio 搭建模型展示应用

深度学习是机器学习最重要的组成部分。用深度学习解决问题，同样要参考机器学习的一般流程，从采集、整理和清洗数据开始，再搭建模型进行训练与评估，最后结合其他编程工具或者模块，以模型推理为核心功能，形成一个完整的人工智能项目。本节课，我们将用深度学习的方法实现图像分类并解决昆虫识别的问题。

一、准备图像数据集

合适的数据集是机器学习任务成功的关键，数据集的质量会直接影响模型的性能。针对图像分类模型训练，我们需要准备好一个图像数据集。

许多人关心的是“需要多少张图片？”以及“图片的尺寸应该是多大？”，答案取决于所期望的模型识别精度。一般来说，数据越丰富、越多样，训练出来的模型的表现就越好。只有数量还不够，还需要确保数据的多样性，这意味着需要考虑光线、拍摄角度、背景等变化条件。采集数据前还需要明确想要模型识别和学习的图像类型，考虑模型最终的应用场景及采集的图像数据与要推理的图像是否一致。例如，昆虫识别模型的应用场景是校园农场，最好也用摄像头去采集校园农场里的昆虫的图像。

1. 图像分类数据集的规范

不同的人工智能开发工具或框架对数据集格式有特定的要求，比较常用的用于图像分类的数据集格式是 ImageNet 格式。ImageNet 大规模视觉识别挑战赛是著名的计算机视觉竞赛，其机制是提供统一的数据集，让不同算法进行比较。ImageNet 提供的数据集拥有超过 1500 万张图片，约 2.2 万种类别，其格式逐步发展为一种通用的图像分类数据集标准。

ImageNet 格式的数据集一般包含三个文件夹和三个文本文件。如图 2.3.1 所示，不同类别图片按照文件夹分类，通过 training_set、val_set、test_set 区分训练集、验证集和测试集。文本文件 classes.txt 说明类别名称与序号的对应关系，val.txt 说明验证集图片路径与类别序号的对应关系，test.txt 说明测试集图片路径与类别序号的对应关系。

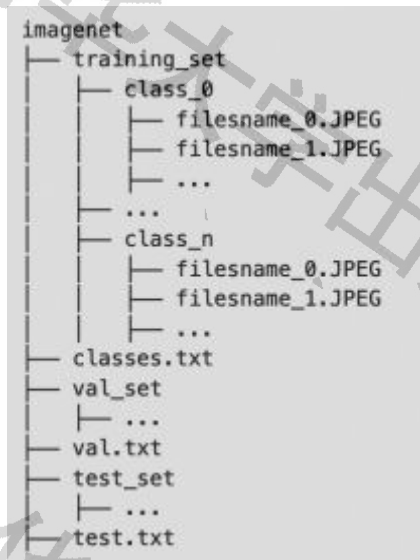


图 2.3.1 ImageNet 格式的数据集文件夹目录

2. 图像分类数据集的制作

第 1 单元已经使用过的 BaseDT 库不仅能完成 CSV 格式数据集的拆分，而且可以用于制作 ImageNet 格式的图像分类数据集。首先，将收集的数据进行初步整理，整理规范如图 2.3.2 所示，将所有图片按照类别存放（存放至以各类别名称命名的文件夹内），再将所有图片文件夹放入 images 文件夹，同时新建一个 classes.txt，按照次序逐行写上所有类别的名称，如图 2.3.3 所示。

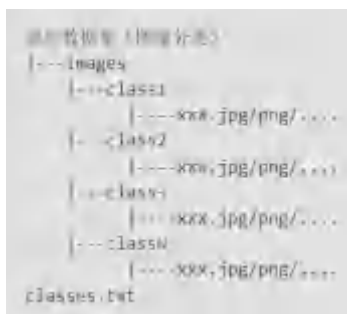


图 2.3.2 原始数据集整理规范

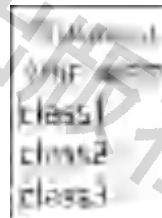


图 2.3.3 类别信息填写规范

然后，使用 BaseDT 对整理完毕的 images 文件夹进行格式转换，数据集格式转换的代码如下，需要在代码中指定生成数据集的路径、原始数据集的路径、原始数据集格式、划分比例（如不设置，则默认比例为训练集：测试集：验证集 = 7 : 1 : 2）。数据集划分代码运行结果如图 2.3.4 所示。

```
from BaseDT.dataset import DataSet
ds = DataSet('./dataset/insect')          # 指定生成数据集的路径
# 默认比例为train_ratio = 0.7, test_ratio = 0.1, val_ratio = 0.2
ds.make_dataset('./dataset/insect_data', src_format='IMAGENET',
train_ratio = 0.8, test_ratio = 0.1, val_ratio = 0.1)
# 指定原始数据集的路径，原始数据集格式选择IMAGENET
```

```
正在划分数据集，比例为 train:test:val = 0.8:0.1:0.1
分割中.....
分割完成
```

图 2.3.4 数据集划分代码运行结果

拓展阅读

其他数据标注工具

整理数据是深度学习中的核心工作之一。相对来说，图像分类任务比较简单，如果要做检测任务，那么还需要在图像上做各种标注。大部分人工智能开发平台都自带数据标注功能。如图 2.3.5 所示，用户可以在某人工智能平台完成数据集的数据标注工作。

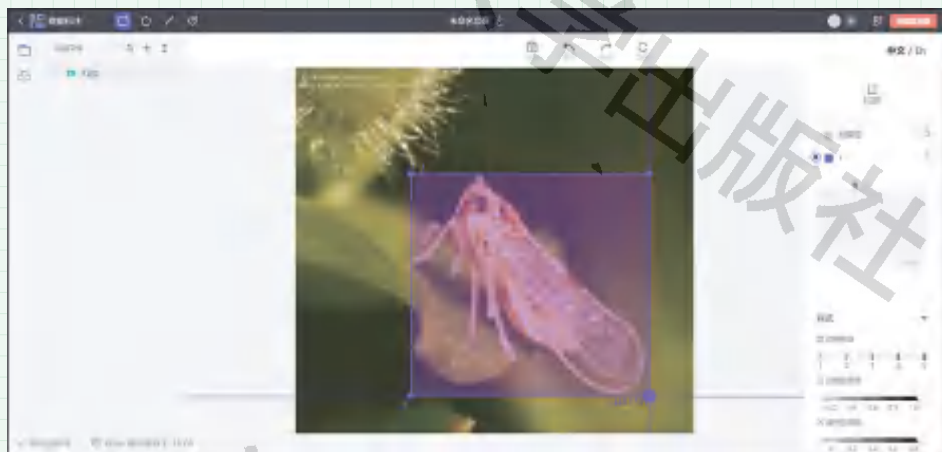


图 2.3.5 数据标注界面

数据标注工具能够帮助研究人员和开发者高效、准确地对数据进行分类、标记和整理。LabelImg、LabelMe 等都是开源的图像标注工具，支持图像中目标的标注与分类。用户可以在图形界面为图像用“拉框”的方式添加标签，操作比较方便。



实践活动

整理一个 ImageNet 数据集

“兵马未动，粮草先行。”没有好的数据集，就训练不出好的模型。MMEDu 自带的昆虫数据集的规模并不大，来源也比较单一。请使用各种方式对昆虫数据集进行补充，如在科学老师的指导下进行现场拍摄或者网络搜索。然后编写模型转换代码，制作一个 ImageNet 格式的昆虫数据集。核心实践内容包括：

- (1) 数据集扩充。对 MMEDu 自带的昆虫数据集进行扩充，增加类别或者增加现有类别的图像数量。
- (2) 数据集转换。完善代码，输入数据路径进行格式转换，并检查格式是否正确。
- (3) 进一步思考：如果因为误操作把数据分类弄错了，有什么办法可以快速修正？

二、图像分类模型的训练

MMEdu 是一个为青少年设计的计算机视觉算法包，其简化了神经网络模型搭建和训练参数，让初学者通过简洁的代码即可完成各种模型训练。借助 EasyTrain，使用者还可以用无代码的方式调用 MMEdu 训练模型。MMEdu 内置了大量优秀的算法（如 SOTA 模型等），支持图像分类和检测任务。

1. 图像分类 SOTA 模型

SOTA（state-of-the-art）是指在某个研究任务中目前表现最好、最先进的模型。不同的 SOTA 模型适合解决不同的问题，要根据具体的任务进行合理选择。比如，LeNet 模型在灰度数字手写体字符识别任务上表现非常优秀，也适用于简单的灰度图像分类任务。除此之外，MMEdu 在图像分类方面还内置了 MobileNet、ResNet 系列的模型。

（1）MobileNet

MobileNet 系列模型的设计初衷是“for mobile vision applications”（为移动设备的视觉应用而开发），即提出一个轻量化的卷积网络模型，在显著减少计算量和参数量的同时，保持较高的准确率以提升计算效率，能用在手机、机顶盒等算力较弱的移动终端（也称边缘设备）上。

MobileNet 最大的优点就是网络的轻量化。如果昆虫识别任务要采用移动终端部署，如行空板、树莓派和龙芯派等设备，那么 MobileNet 是一个不错的选择。

（2）ResNet

我们已经知道，随着网络层数的增加，获取的信息会更加丰富，但过多的网络层数可能导致网络性能下降，甚至难以训练。而何恺明、孙剑和汤晓鸥等提出的深度残差网络（deep residual network, ResNet），有效地解决了这个问题。ResNet 的提出是计算机视觉领域的一个里程碑事件，用 ResNet 训练出来的模型在当年 ImageNet 的图像分类和目标检测任务中都获得了冠军。

ResNet 有很多版本，常见的有 ResNet18、ResNet34、ResNet50、ResNet101、

ResNet152 等。迄今为止，ResNet 仍是难以替代的主流模型之一，被广泛应用于分类、检测、分割等领域。

拓展阅读

浦视OpenMMLab：一个最全面的计算机视觉算法包

MMEdu 源于国产人工智能视觉算法集成框架 OpenMMLab（浦视）。OpenMMLab 属于上海人工智能实验室的计算机视觉算法开源体系。OpenMMLab 是深度学习时代极具影响力的视觉算法开源项目，为学术和产业界提供一个可跨方向、结构精良、易复现的统一算法工具库。

OpenMMLab 已经累计开源了超过 30 个算法库，涵盖分类、检测、分割、视频理解等众多研究领域，拥有超过 300 种算法、2400 多个预训练模型。MMEdu 保留了 OpenMMLab 的各种参数，尤其是模型训练的所有常见参数。如果要训练出更加专业的模型，那么需要深入了解 OpenMMLab。

2. 图像分类模型的训练过程

MMEdu 的 MMClassification 模块的主要功能是对图像进行分类。以下是使用 MMEdu 训练图像分类模型的核心代码，其设计思路旨在保持代码的简洁性和高效性。用 EasyTrain 训练模型时，会自动生成类似的训练代码。

```
from MMEdu import MMClassification as cls      # 导入库
model = cls('LeNet')                          # 实例化模型
model.num_classes = 3                        # 配置基本信息（类别数量）
model.load_dataset(path='./dataset/insect')    # 指定数据集路径
model.save_fold = './my_model'               # 指定模型保存集路径
model.train(epochs=10, validate=True)         # 训练模型
```

以上代码体现了图像分类模型训练的五个主要步骤。

① 模型实例化：选择一个适合任务的 SOTA 模型作为模型的骨干网络（backbone）。

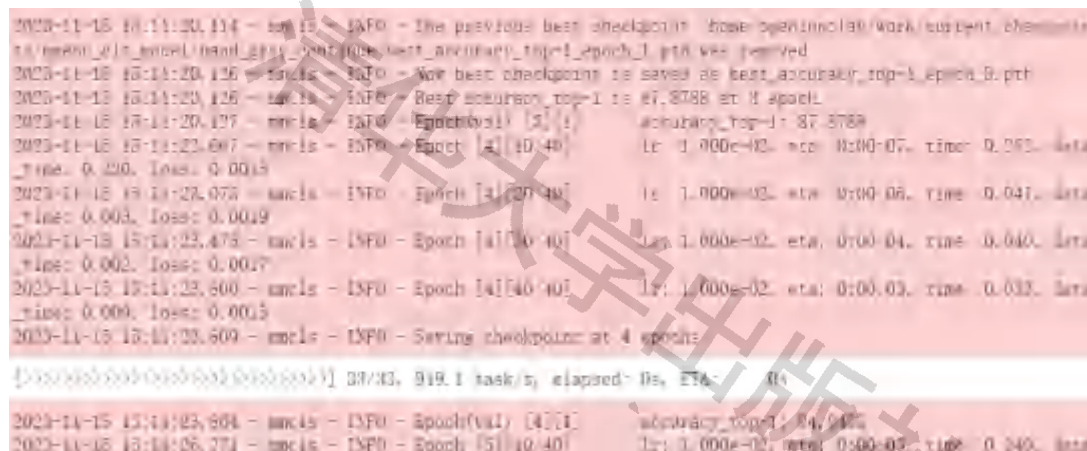
② 配置类别数量：根据实际数据集中的类别数量配置模型，确保模型输出与数据集相匹配。

③ 加载数据集：明确指定数据集的路径，让模型知道从哪里获取训练所需的数据。

④ 设置模型保存路径：指定一个路径来保存训练好的模型，便于后续的使用或部署。

⑤ 开始训练：指定训练的轮次（epoch）并选择是否在训练过程中进行验证，以确保模型的泛化能力。

运行结果如图 2.3.6 所示。



```

2023-11-15 15:11:20.114 - mmcls - INFO - The previous best checkpoint 'home/openmmlab/work/current_checkpoint/ckpt_model_best.pth' was removed
2023-11-15 15:11:20.125 - mmcls - INFO - Now best checkpoint is saved as test_accuracy_top-1_epoch-8.pth
2023-11-15 15:11:20.126 - mmcls - INFO - Best accuracy_top-1 is 0.000 at 1 epoch
2023-11-15 15:11:20.127 - mmcls - INFO - Epoch[val] [1/1] accuracy_top-1: 0.000
2023-11-15 15:11:23.697 - mmcls - INFO - Epoch [1/10/40] lr: 1.000e-02, eta: 0:00:07, time: 0.003, data_time: 0.000, loss: 0.0015
2023-11-15 15:11:23.675 - mmcls - INFO - Epoch [1/20/40] lr: 1.000e-02, eta: 0:00:06, time: 0.047, data_time: 0.003, loss: 0.0019
2023-11-15 15:11:23.475 - mmcls - INFO - Epoch [1/30/40] lr: 1.000e-02, eta: 0:00:04, time: 0.040, data_time: 0.002, loss: 0.0017
2023-11-15 15:11:23.600 - mmcls - INFO - Epoch [1/40/40] lr: 1.000e-02, eta: 0:00:03, time: 0.033, data_time: 0.000, loss: 0.0015
2023-11-15 15:11:23.609 - mmcls - INFO - Saving checkpoint at 4 epochs
[5.00000000000000000000000000000000] 08/03, 919.1 task/s, elapsed: 0s, ETA: 0s
2023-11-15 15:11:23.604 - mmcls - INFO - Epoch[val] [4/1] accuracy_top-1: 0.040
2023-11-15 15:11:26.271 - mmcls - INFO - Epoch [5/10/40] lr: 1.000e-02, eta: 0:00:03, time: 0.040, data
  
```

图 2.3.6 MMEdu 模型训练运行结果

从图 2.3.6 中可以看出，每轮训练结束后出现的 `accuracy_top-1` 表示模型在验证集上验证得到的准确率。MMEdu 训练模型的过程中不仅会保存每一轮的模型权重文件，还会将最佳准确率的模型进行重命名保存，并呈现在训练结果中。

如果对当前模型不够满意，MMEdu 支持加载之前训练过的模型继续训练，在 `train` 函数中增加 `checkpoint` 参数，并指定模型权重路径即可。下面的代码是继续训练的核心代码。

```

checkpoint = './latest.pth' # 指定使用的模型权重文件
model.train(epochs=50, validate=True, checkpoint=checkpoint)
# 进行再训练
  
```

0

实践活动

用MMEdU训练图像分类模型

已经学习了 MMEdU 训练模型的核心步骤和代码，请使用自带的昆虫数据集，也可以使用已经完成扩充与格式转换的 ImageNet 数据集，训练一个自己最满意的昆虫分类模型。

请以小组为单位，尝试使用不同的 SOTA 训练昆虫分类模型，留下一个最好的模型。核心实践内容包括：

- (1) 完善训练代码，输入数据集路径进行模型训练。
- (2) 选择不同的 SOTA 模型，并设置不同的参数，指定不同的保存路径，分别完成训练并在表 2.3.1 中记录最佳 accuracy_top-1。
- (3) 思考：面对无代码和有代码训练两种方式，你更喜欢哪一种？请说明理由。

表 2.3.1 活动记录表

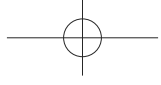
算法类型	LeNet	MobileNet	ResNet18	ResNet50
accuracy_top-1				

拓展阅读

迁移学习

通常，从头开始训练一个深度学习模型需要大量的数据和计算资源，既费时又费力。但如果将一个已经在类似任务上训练好的模型作为起点，那么就可以直接利用这个预训练模型已经学到的知识，包括特征提取、模式识别等能力。这就是迁移学习（transfer learning）的魅力所在。在 MMEdU 中选择之前的模型“继续学习”，其实就是迁移学习。

迁移学习允许我们利用一个任务中的知识加速另一个相关任务的学习过程。如果选择在猫狗分类模型的基础上训练牛羊分类模型，那么训练



速度也会快很多。这就像已经学会了骑自行车，现在想学习骑摩托车，由于两者在平衡和控制方面有相似之处，那么骑摩托车的学习过程会比从零开始学习要快得多。

三、图像分类模型的应用

完成了模型的训练，接下来便要思考模型应用的问题。比如，可以将模型应用在校园智慧农场里，对害虫进行监控。再如，可以将模型应用在某个生态区域，对昆虫类型进行分类并记录。不管是哪一种应用场景，都需要借助摄像头拍摄图像，调用模型对图像进行实时推理。

1. 深度学习模型的类型和推理工具

在深度学习领域中，几乎每一种深度学习开发框架都会有自己特定类型的模型，如 PyTorch 的模型格式为“.pth”，TensorFlow 的模型格式为“.h5”等。为了让深度学习模型更容易转化为应用软件，不同的企业和机构相继推出了它们自己的推理工具（也称推理框架、部署工具），如微软、亚马逊、Facebook 和 IBM 等公司推出了 ONNX Runtime，腾讯推出了 NCNN，Intel 推出了 OpenVino，英伟达推出了 TensorRT。

目前，ONNX Runtime 得到的企业支持较多，是一个应用较为广泛的推理框架。ONNX Runtime 支持的模型格式为 ONNX（open neural network exchange，即开放神经网络交换），旨在实现不同深度学习框架之间的模型互操作，便于模型的共享和部署。ONNX Runtime 支持多种硬件平台，包括 CPU 和 GPU。

2. 深度学习模型的转换

为提高模型推理的速度，很多硬件厂商设计了专用的推理加速芯片，如 TPU（张量处理器）、NPU（神经网络处理器）等，而不同的芯片会针对某些类型的模型进行优化。模型转换可以使开发者充分利用硬件资源，提升模型

性能，实现数据的快速推理。

例如，一个在 PyTorch 中训练的模型（PTH 格式）可以通过转换工具转换为 ONNX 格式，然后进一步部署到支持 ONNX Runtime 的环境中，实现高效的模型推理。

MMEdU 内置 convert 函数，其针对用 MMEdU 训练的模型，可直接完成一键式模型转换。下面的代码实现了将训练的 PTH 格式模型转换为 ONNX 格式。

```
from MMEdU import MMClassification as cls
model = cls(backbone='MobileNet')      # 实例化模型
checkpoint = './my_model/best.pth'      # 指定PTH模型
out_file='./my_model/best.onnx'         # 指定输出文件的路径
model.convert(checkpoint=checkpoint, out_file=out_file)
# 模型转换
```

3. 从模型推理到人工智能应用开发

当一个深度学习模型训练完成后，最终的任务是要结合其他编程工具，编写一个人工智能应用。一般来说，这些规模较小的模型运行在一些边缘设备（指性能较弱的移动端和嵌入式设备）上。以昆虫分类模型为例，一般要将模型和代码放在边缘设备上，然后安装在校园农场里，结合摄像头，其应用流程如图 2.3.7 所示。



图 2.3.7 图像分类模型应用的一般流程

（1）借助 XEduHub 实现 ONNX 模型推理

XEdu 工具集中的 XEduHub 也支持 ONNX 模型的推理，以下是 MMEdU 模型转换后推理的示例代码。

```

from XEdu.hub import Workflow as wf
mm = wf(task='mmedu',checkpoint='./my_model/best.onnx')
# 模型声明

image = './resources/test.jpg'
# 待推理图像路径
res,img = mm.inference(data=image,img_type='cv2')
# 模型推理
result = mm.format_output(lang="zh")
# 标准化推理结果
mm.show(img)
# 可视化结果图像

```

运行结果如图 2.3.8 所示。

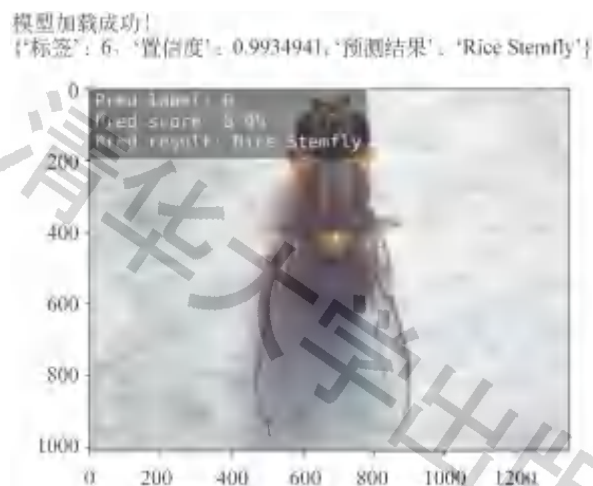


图 2.3.8 示例代码运行结果

(2) 用 Gradio 搭建人工智能应用

Gradio 是一个开源库，用于快速设计和部署机器学习模型的交互式界面。它提供了简单易用的接口，让用户能够轻松地创建和展示机器学习模型并与其交互。

下面是一段 Gradio 的示例代码，主要作用是创建一个用户界面，让用户可以上传一张图片，并直接显示这张图片而不对图片进行任何处理。

```

import gradio as gr
# 定义一个方法：输入一张图片，对图片不进行任何处理就输出
def display(input_img):
    return input_img
# 创建一个用户界面（包括一个接收用户输入图片的输入框和一个显示图片的输出框）
demo = gr.Interface(fn=display, inputs=gr.Image(type='filepath'),
    outputs='image')

```

```
# 启动用户界面
demo.launch()
```

在以上代码的基础上，结合模型推理代码即可搭建一个用户上传一张图片就显示推理结果的人工智能应用，如图 2.3.9 所示。



图 2.3.9 害虫识别小应用展示图

拓展阅读

结合OpenCV和pinpong开发AI科创作品

OpenCV (open source computer vision library) 是一个开源的计算机视觉和机器学习软件库。借助 OpenCV 可以轻松地完成图像和视频处理。而人工智能结合物联网技术，就形成了智联网应用。借助 pinpong 库，我们就能利用人工智能模型的推理结果驱动相应的硬件，比如发现害虫就保存照片并告知农民，发现偷食的小鸟就发出声音驱逐，发现杂草则启动激光设备“清除”等。下面的代码实现的功能是使用摄像头拍摄照片，若识别出照片中是害虫，则点亮警示灯 (D13 端口接 LED)。

```
# 使用摄像头拍摄照片，若识别出照片中是害虫，则点亮警示灯 (D13端口接LED)
import cv2
import time
from pinpong.board import Board,Pin

Board().begin() # 初始化，选择板型和端口号，不输入端口号则进行自动识别
led = Pin(Pin.D13, Pin.OUT) # 设置数字13引脚为输出模式，用于控制LED灯
cap = cv2.VideoCapture(0) # 选择0号摄像头，0表示默认摄像头
```

```

print('一秒钟后开始拍照.....')
time.sleep(1)
ret, frame = cap.read()          # 从摄像头中读取一帧图像
cv2.imshow('./my_photo.jpg', frame) # 展示图像到窗口中
cv2.waitKey(1000)                # 显示持续1秒(这里单位是毫秒)
cv2.destroyAllWindows()          # 关闭窗口
cv2.imwrite('./my_photo.jpg', frame) # 将图像保存到文件
print('成功保存 my_photo.jpg')
cap.release()

.....                            # 此处省略导入XEduHub完成模型推理的代码

# 判断推理结果, 若是害虫, 则点亮LED灯
if res['标签'] == 0:              # 标签0表示害虫
    led.write_digital(1)          # 点亮LED灯
    time.sleep(1)                 # 保持1秒
    led.write_digital(0)          # 熄灭LED灯

```

结合 OpenCV 和 pinpong 库, 可以创建出既能通过摄像头“看”到环境, 又能通过硬件反馈进行“动作”的 AI 科创作品。这种结合使项目不仅限于软件层面的交互, 还能与物理世界进行互动, 极大地拓宽了创作的可能性。



实践活动

用Gradio搭建模型展示应用

已经学习了图像模型应用的相关知识, 能否搭建一个自己训练的图像分类模型应用? 请以小组为单位, 参考资源包中提供的代码, 利用 XEduHub 和 Gradio 搭建模型展示应用。核心实践内容包括:

- (1) 完善代码, 载入自己训练的模型并实现用户输入图片、输出推理结果。
- (2) 增加个性化输入、输出交互设计。
- (3) 进一步思考: 如果要制作一个功能更强大的模型展示应用, 如何修改?

项目实施

小清想要使用深度神经网络识别不同的昆虫，并将训练得到的模型部署到实际场景中。请你帮助他选择一种或多种人工智能应用开发的工具，结合前面的学习内容，实现这个任务。

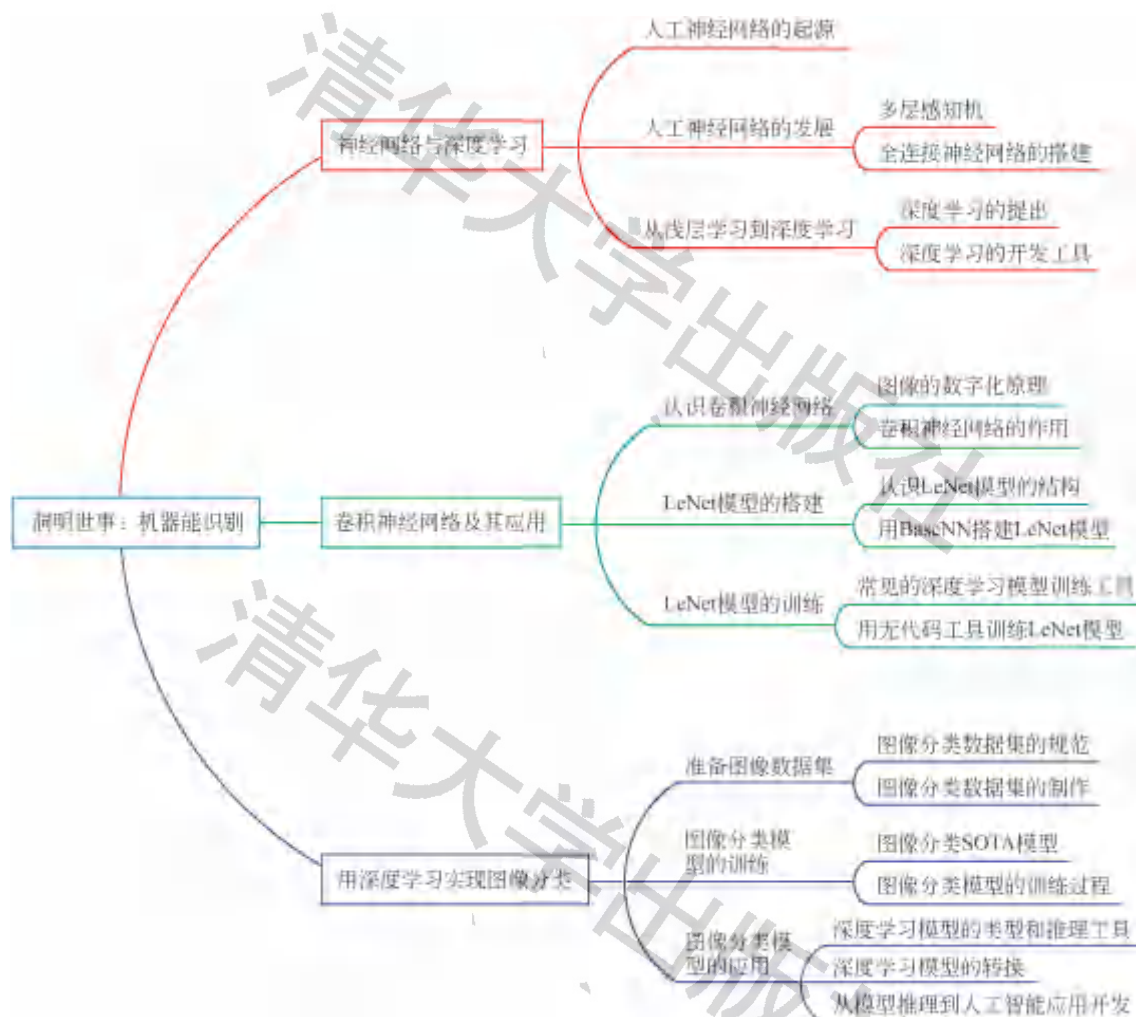
在你的昆虫识别应用中，你使用了哪些模型，计划用什么方式呈现模型效果呢？请在表 2.3.2 中填写实施过程。

表 2.3.2 项目实施记录表

使用的模型 1	模型名称： 训练轮次：
使用的模型 2	模型名称： 训练轮次：
训练的最佳模型	
模型的实际应用场景	
模型的输入来源	
模型的输出效果	
需采用的硬件清单	
使用的模型应用工具	
核心代码展示	

单元小结

一、知识回顾



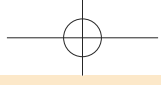
二、项目交流与评价

1. 参考本书附录“项目报告模板”撰写项目报告，并制作演示文稿。
2. 在课堂内展示自己的学习成果并分享经验，在下表中进行自评和他评。

项目成果评价表

评价维度	自 评	他 评
(1) 完整性 昆虫识别项目材料齐全，有数据收集规划、分工协作、项目实施记录表及最终成果（数据集、模型和展示系统）。	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般
(2) 实用性 成果内容具体，有真实的项目问题描述、有效的解决方案及相关工具等。	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般
(3) 规范性 项目报告规范，符合项目报告的一般格式要求，文字表述准确，项目实施记录表填写具体。	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般

3. 保存昆虫识别数据集与模型，整理源码等文档，并上传到校园网或者其他学习空间，与他人分享学习成果。



第 3 单元

妙笔生花：机器能创作

学习导引

随着时代的演进，文字、绘画等艺术形式已经成为记录人类历史、传承文化不可或缺的媒介。这些表达方式不仅是信息的载体，还是文化的创造者和传承者。它们承载着人类的思想与情感，同时也是宣传地方特色文化的重要手段。然而，长期以来，创作有价值的内容往往需要大量的积淀和人的参与。随着人工智能的快速发展，我们迎来了一个全新的时代，人工智能为内容创作带来了崭新的可能性。

人工智能是如何实现内容创作的呢？如何利用人工智能生成内容（artificial intelligence generated content, AIGC）工具助手生成符合需求的内容呢？本单元将以“贵州传统文化明信片创作”为主题，引导同学们深入探索 AIGC 的原理和机制，并学习如何利用 AIGC 工具助手生成符合特定主题的内容。通过亲身实践，我们将揭开人工智能在内容创作领域的神秘面纱，加深同学们对人工智能辅助创作的理解，为创作实践打下坚实基础。

项目情景

作为土生土长的贵州人，小青打算创作一些明信片，借助互联网弘扬自己家乡的传统文化。通过调研，小青准备使用人工智能助手辅助完成明信片的文案创作和图画创作，他找到了辅助生成文案的工具、辅助生成图像的工具以及



根据图像生成视频的工具。为了用好这些工具，小清做了很多尝试，但是在探索的过程中遇到了一些问题。

(1) 人工智能可以辅助生成哪些类型的内容？

(2) 人工智能助手可以辅助完成文案的编写、图像的生成和视频的生成，其背后的原理是什么？为什么这些工具助手可以生成特定的内容？

(3) 如何与可以生成内容的人工智能助手进行交互，并通过不断的调整，得到需要的内容？

.....

你是不是也很感兴趣？快来和小清一起学习如何利用人工智能生成助手辅助生成内容，探索 AIGC 背后的原理，并挑选与贵州传统文化相关的特定主体，借助人工智能生成助手完成主题明信片中文案的编写与图画的绘制。

项目方案

经过咨询与了解，小清设计了以下方案：

知识学习	实施步骤	预期成果
(1) 了解 AIGC 的基本概念与相关知识	(1) 了解 AIGC 的基本概念，掌握大语言模型助手的使用方法	(1) AIGC 的基本方法与其安全挑战的认知（PPT 格式）
(2) 理解机器生成图像的基本原理和常见算法	(2) 掌握机器生成图像的原理与训练扩散模型生成图像的基本方法	(2) 对 AIGC 常见算法的认识（PPT 格式）
(3) 探索文本与图像等多种模态数据之间的联系，了解根据文本生成图像与根据图像生成文本的基本原理	(3) 掌握借助多模态模型评估文生图匹配度的方法	(3) 能够合理编写提示语完成多种模态内容的生成与优化（提示语记录、作品）
(4) 运用 AIGC 工具生成文本内容与图像内容，掌握提示语的编写方法以生成符合需求的内容	(4) 利用开源的 AIGC 工具库搭建文生文、文生图作品的生成与优化模型	(4) 总结（PDF 格式）
	(5) 撰写作品生成的学习心得	

你对小清的项目方案有什么不同的看法或建议？你准备如何设计项目方案？请填写在下表中。

知识学习	实施步骤	预期成果

项目分工

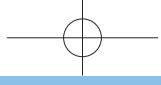
方案设计完成后，小青发现仅凭一己之力很难完成这个项目，于是邀请对此问题感兴趣的同学一起参与，并在项目方案中添加了以下表格。



姓名	角色	分 工	任 务
小青	组长	负责项目统筹、监督与管理	项目整体的选题与方案设计；作品生成过程的统筹、协调、监督、总结；撰写项目文档
同学甲	成员	负责大语言模型助手相关工具的调研与使用测试	完成大语言模型助手相关工具的调研，通过生成实际内容验证工具的使用方法与相关提示语编写的方法
同学乙	成员	负责文生图相关工具的调研与使用测试	完成文生图相关工具的调研，通过生成实际内容验证工具的使用方法与相关提示语编写的方法
同学丙	成员	负责记录和优化提示语，备份过程性作品	完成作品，生成主要的提示语记录，测试并优化提示语，完成最终作品的生成

你认为小青的项目组成员构成、分工和任务分配是否合理？请在下表中填写你的项目分工情况。

姓名	角色	分 工	任 务



第1节 人工智能生成内容

本节知识

- ◆ AIGC 的概念
- ◆ AIGC 生成工具与生成内容的类型
- ◆ AIGC 带来的问题与挑战

本节活动

- ◆ 使用大语言模型助手生成内容

“艾香飘飘舞龙舟，端午佳节情缱绻。香包红绳系家门，粽香袅袅传佳音。”诗文通过工整的格式、恰当的韵脚、细腻的语言勾勒出了端午佳节的热闹场景。这首七言绝句并不是人类创作的，而是由一个大语言模型助手生成的。AIGC 正在作为一种新的内容生产方式进入普通大众视野。

一、AIGC 概述

在过去，诸多文学、艺术作品是由人类费尽心力完成的，随着人工智能技术的发展，机器可以写文章、画图、生成动画，这就是人工智能生成内容。它具体是指使用人工智能模型生成文字、图片、音乐、视频等类型的内容。

1. AIGC 的发展历史

1950 年，图灵提出的“图灵测试”实际上就是在探讨机器是否能够“生成”与人类创作内容无差异的内容。

1957 年，莱杰伦·希勒（Lejaren Hiller）和伦纳德·艾萨克森（Leonard Isaacson）通过计算机程序完成了历史上第一支由计算机创作的音乐作品——

弦乐四重奏《依利亚克组曲》(Illiatic Suite)。

2017年,罗斯·古德温(Ross Goodwin)使用一台连接各种传感器的笔记本电脑,从纽约驾车前往新奥尔良。这台机器将旅途中感知到的一切以文字的形式输出,完成了世界上首部由人工智能创作的小说《The Road》。开头的一句话是“*It was nine seventeen in the morning, and the house was heavy*”(早上九点十七分,房子沉甸甸的)。

2018年,一幅由人工智能程序创作的画作《埃德蒙·贝拉米肖像》在佳士得拍卖行以43.25万美元的价格成交。作为首个出售的人工智能艺术品,它引发了关于人工智能在艺术和创造力中作用的讨论。

2022年11月30日,OpenAI实验室发布ChatGPT,短短5天时间,ChatGPT的全球注册用户就超过了100万。以ChatGPT为代表的AIGC工具开始被普通大众广泛关注。

图3.1.1展示了AIGC发展过程中的关键事件。



图 3.1.1 AIGC 发展过程中的关键事件

随着数据快速积累、算力性能提升和算法效力增强,人工智能不仅能够与人类进行互动,还可以进行写作、编曲、绘画、视频制作等创意工作。

2. 作品的潜在规律

图3.1.2中有三段描写“贵州的夏”的文字,阅读这三段文字,说一说哪段文字更符合人类创作的规律以及真实的情景。

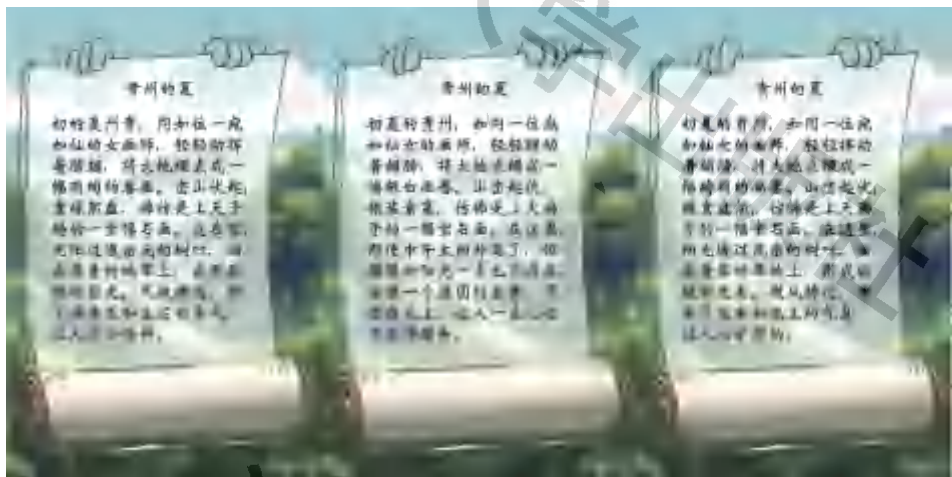


图 3.1.2 以“贵州的夏”为题生成的内容

显然，右侧的文字描述更符合真实的情景。左侧文字中的部分用词不符合中文词语的词序，比如“同如”“膀翅”“意绿然盎”等；中间部分的文字不符合贵州夏天的实际情景，银装素裹的景象在初夏的贵州几乎不会出现。不难看出，图 3.1.2 中的三段文字虽然都是由汉字组成的，但是并不是所有的内容都符合中文习惯。

图 3.1.3 展示了两幅绘画作品，仔细观察这两幅画，说一说其中哪幅画的意境更符合“贵州的夏”。

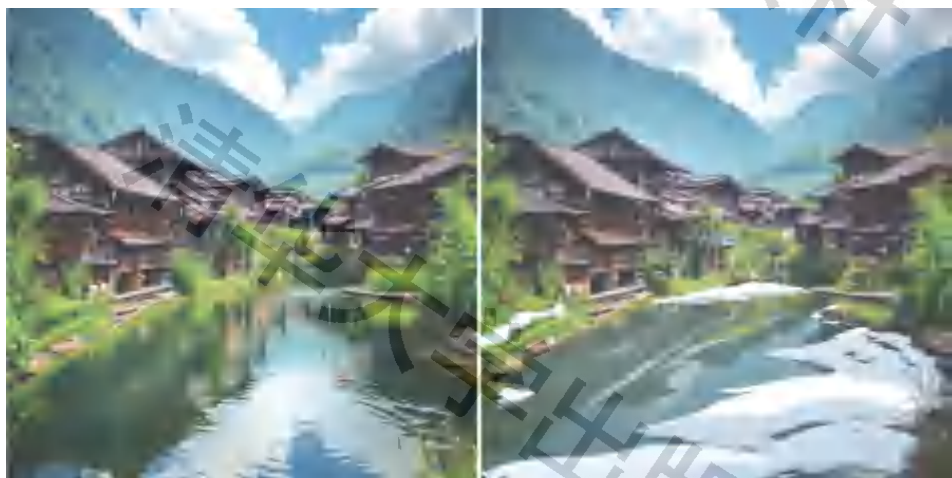
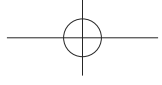


图 3.1.3 以“贵州的夏”为主题的画作

虽然右侧的画作也呈现了贵州省千户苗寨的风景，但是画作中的绿意盎然与水面的浮冰明显冲突，这不是贵州夏天常见的景象。



人类创作作品时，都会不自觉地遵循潜在的规律。比如，描述“贵州的夏”的文字作品，会使用“绿意盎然”“青翠”等辞藻；描绘“贵州的夏”的绘画作品，会以体现夏季特色的深绿为主，并搭配不同层次的绿色。以中文为例，即使不懂中文的外国人，也可以随意排列“汉字符号”形成一段文字，但是这样的文字组合可能并不符合中文潜在的规律。真正能够成为作品的创作，往往符合特定的规律，如规范的词语、合适的语法、美妙的意境等。即使是天马行空的画家，他的画作依旧有自己独特的风格。

二、使用 AIGC 工具生成内容

1. AIGC 的原理

人类成为优秀的作家或者画家要经过一系列复杂的学习过程。在文学领域，作家的成长往往从阅读经典文学作品开始，慢慢学习并积累丰富的语言和表达技巧。通过对语言文字和表达技巧背后规律的掌握和运用，他们开始通过创作短文、小说、诗歌等表达自己的思想和情感，并逐渐塑造出独特的文学风格。在绘画领域，画家一般要先通过学习经典的基础绘画技巧、色彩理论和构图等知识构建绘画基础，然后通过模仿大师作品和观察自然，逐渐培养对形式、色彩、光影等方面的敏感度，并开始创造具有自身风格的艺术作品。

因此，如果给机器准备大量作品作为数据进行学习，机器就可以找到人类作品的客观规律。这就是 AIGC 的原理。近年来，通过深度学习算法对大量作品数据进行学习，生成效果优良的模型开始大量涌现。

2. AIGC 生成工具与生成内容的类型

伴随生成效果优良的模型的出现，越来越多的 AIGC 工具被开发出来。当前，这些 AIGC 工具背后的生成模型包括大语言模型、文生图大模型、音乐生成模型、视频生成模型、代码生成模型等。应用 AIGC 工具，人类用户可以以自然语言的形式实现内容生成。

自然语言就是人类用来日常沟通和表达思想的语言，如汉语、英语等。人类用户编写的这种自然语言描述被称为提示语，如通过提示语“贵州省有哪些具有民族特色的节日？”向机器提问，得到的内容如图 3.1.4 所示。

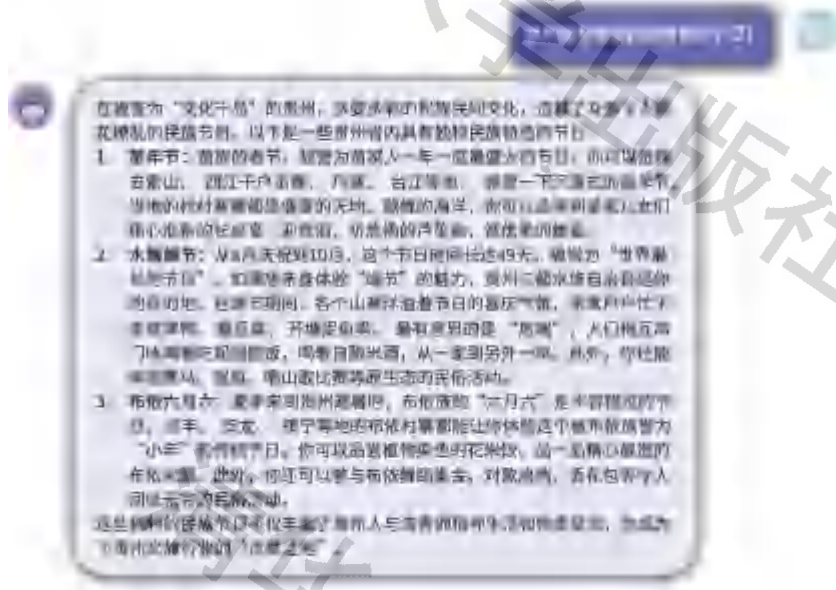


图 3.1.4 机器根据提示语“贵州省有哪些民族特色的节日？”生成的内容

AIGC 工具可以根据提示语的数据类型以及生成内容的数据类型进行分类，如表 3.1.1 所示。

表 3.1.1 AIGC 工具类型

工具类别	工具示例	输入的数据类型	生成的数据类型
大语言模型助手	天工 AI、商量、ChatGPT	文本	文本
文生图工具	天工 AI、秒画、DALL·E3	文本	图像
图、文生图工具	天工 AI、秒画、Muse Draw	文本、图像	图像
图、文生视频工具	可灵、Sora	文本、图像	视频
.....

借助 AIGC 工具可以生成不同类型的内容，包括但不限于表 3.1.2 中列出的内容形式。根据文本生成散文、根据文本生成程序、根据文本生成图像的效果如图 3.1.5 所示。

表 3.1.2 AIGC 工具生成的内容形式

生成的数据类型	生成的内容形式
文本	散文、小说、诗歌等文学作品； 文章、新闻稿等日常文本； 程序代码片段
图像	各种风格的图像，如漫画风格、山水画风格、凡·高风格

续表

生成的数据类型	生成的内容形式
音频	音乐，包括曲调、旋律及和声； 语音合成，将文本转为自然语音
视频	动画、电影短剧、广告宣传片
3D 交互内容	3D 场景，包括环境、物体和互动元素； 虚拟现实（VR）或增强现实（AR）体验的内容

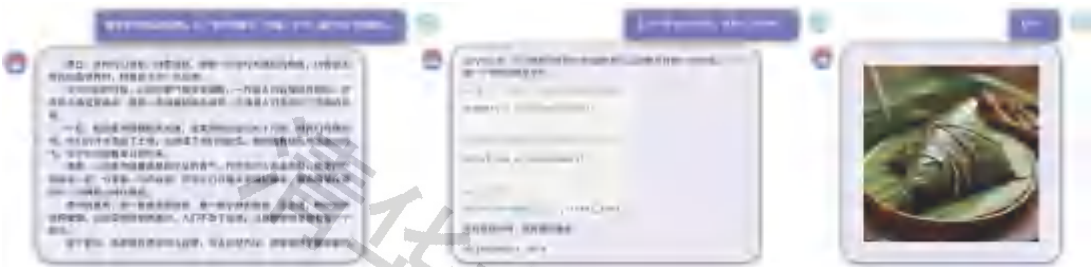


图 3.1.5 AIGC 生成内容示意图

体验活动

使用大语言模型助手生成内容

编写提示语与大语言模型助手进行对话，大语言模型助手能够理解提示语并以自然语言的形式进行反馈。任意选择一款大语言模型助手进行体验，输入下列提示语，观察根据提示语生成的内容，并将生成的内容记录在表 3.1.3 中。说一说生成的内容是否符合期望。

表 3.1.3 大语言模型助手工具体验记录表

提示语	生成的内容
我的家乡是 ××，请详细介绍 ×× 的特色文化与传统节日。	
小清是我的同学，请问他家在哪里？	
讲一下武松拳打镇关西的故事。	

三、AIGC 带来的问题与挑战

科学技术是一把双刃剑，既能为人类社会带来变革，也会带来潜在的风险，人工智能技术的崛起尤为明显。随着人工智能技术的普及和深入应用，一系列伦理、隐私和安全问题浮出水面。

1. AIGC 工具的使用边界

如果向大语言模型助手询问一些涉及他人隐私的问题，机器会如何回答呢？图 3.1.6 展示了大语言模型助手在这些边界问题上的答案。AIGC 工具背后有一些必须遵守的规则，如不会辅助生成威胁他人安全（如财产安全、隐私安全等）的内容。但是，当前 AIGC 工具可能依旧存在一些无法预知的问题，因此，其背后的模型还在不断优化学习中逐渐成长。

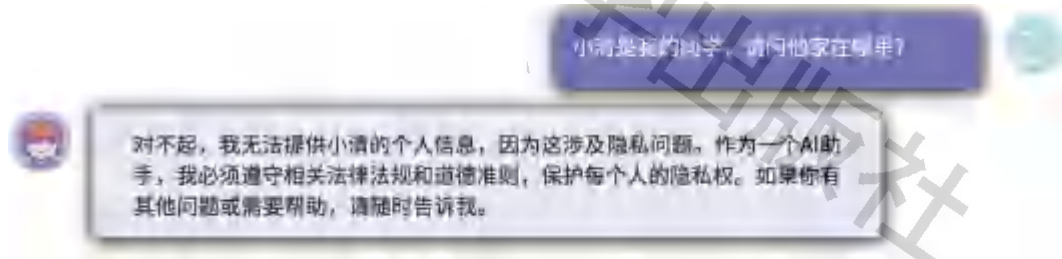


图 3.1.6 超越 AIGC 工具使用边界的使用示例

2. AIGC 带来的著作权问题

如图 3.1.7 所示，图中展示了两幅画作，左边是著名荷兰后印象派画家凡·高于 1889 年 6 月绘制的作品《星月夜》，右边是由人工智能根据提示语生成的图像。

问题讨论

相较于凡·高的著名作品《星月夜》，图 3.1.7 右边的作品是否具有创新性？右边的作品是否侵犯左边作品的著作权？请提出观点并讨论。



图 3.1.7 凡·高作品《星月夜》与人工智能生成的凡·高风格的画

3. 科学地使用 AIGC

应用 AIGC 技术，可以让创作更加高效，但是与此同时，AIGC 的广泛使用也引发了一系列安全、伦理等问题，因此使用 AIGC 时需要注意以下几个方面。

- 真实性问题：AIGC 生成的内容可能是虚构的，使用时需要判断生成内容的可信度。
- 道德责任：使用 AIGC 生成内容时，要注意语言的选择和表达，确保提示语的内容不冒犯或歧视他人；对生成的内容负责，不利用技术传播虚假信息、恶意言论或侵犯他人权益的内容。
- 隐私问题：注意保护个人数据与个人隐私信息，不要将过于私密的信息输入 AIGC 中，以防隐私泄露。
- 知识产权：尊重他人的知识产权，避免 AIGC 生成的内容侵犯他人的著作权。
- 创造性表达：过度使用 AIGC 可能会降低自身的创造力，因此要注意保持创造性表达，不只是复制、粘贴生成的内容，要加入自己的思考和独特见解。

与此同时，还需要制定管理办法、使用规范、法律法规约束技术的使用。作为使用者，即使在相关使用规则尚不健全时，也应该树立正确的伦理观念，自觉遵守使用边界，确保技术的应用不会产生负面影响。

拓展阅读

《生成式人工智能服务管理暂行办法》

2022—2023 年是生成式人工智能快速发展的两年。各类基于大模型的生成式人工智能应用层出不穷，极大地改变着人们的学习与生活方式。为了明确生成式人工智能服务提供者在内容生产、数据保护、隐私安全等方面的法定责任及法律依据，确立人工智能产品的安全评估规定及管理办法，2023 年 7 月 13 日，国家互联网信息办公室等联合发布了《生成式人工智能服务管理暂行办法》。该文件针对利用生成式人工智能技术向中华人民共和国境内公众提供生成文本、图片、音频、视频等内容服务，规定提供和使用生成式人工智能服务，应当遵守法律、行政法规，尊重社会公德和伦理道德。

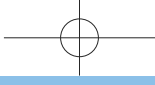
项目实施

小清期望通过与“大语言模型助手”交流选出一种合适的贵州传统文化或者节日，并以此作为明信片的主题，然后根据该主题生成一些文字作品作为明信片的文案。

请设计提示语，完成明信片主题的选择和初始文案的创作。将创作过程记录在表 3.1.4 中。

表 3.1.4 明信片主题及文案创作记录表

确定主题的提示语		确定的主题
主题相关文案生成的提示语与文案		
序号	文案的提示语	生成的文案
1		
2		



第2节 图像生成模型

本节知识

- ◆ 图像生成模型的原理
- ◆ 扩散模型的起源
- ◆ 扩散模型的原理

本节活动

- ◆ 图像生成模型的训练

随着 AIGC 的爆火，人们不仅可以通过与机器聊天生成文案、创作诗歌，还能利用其进行大量的艺术创作。目前，人工智能生成的图像已经变得越来越逼真和自然，越来越具有艺术创意和独特性，这些都得益于深度神经网络的快速发展，推动了众多图像生成大模型的出现，使模型的生成性能不断突破瓶颈。

一、图像生成模型的原理

图像是一种特殊类型的数据，均匀分布在图像空间中，如图 3.2.1 所示，图中的正方形区域代表图像所处的空间，其中每个点代表一张图像。图中“紫

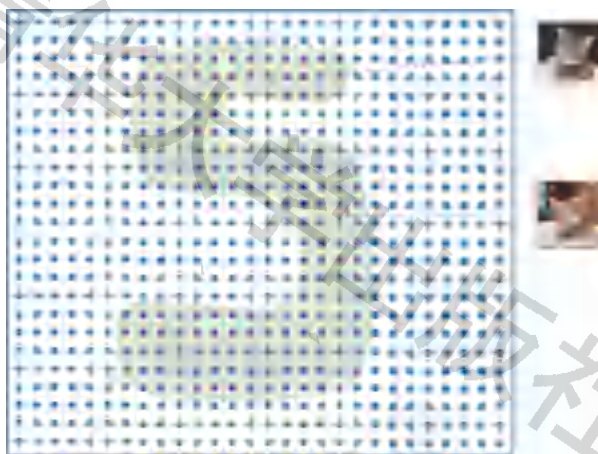


图 3.2.1 图像生成模型示意图

色”的点代表宠物猫的图片，宠物猫的图片在图像空间中呈现特殊的分布规律，见图 3.2.1 中“浅绿色阴影区域”。

假设，现在希望机器能够自动生成宠物猫的图片，可以让机器利用大量宠物猫的图片数据集进行训练。训练过程中，机器会从提供的数据集（图 3.2.1 中的紫色点）中学习宠物猫数据的真实分布规律，这个分布可以理解为宠物的毛发、种类、姿态等特征的统计规律。训练结束后，将得到一个模型，这个模型可以生成接近真实的宠物猫的图片（图 3.2.1 中的红色点）。得到的这个模型就是图像生成模型。具体来说，图像生成模型是一种深度神经网络模型，可以通过对大量图像数据的学习，找到图像数据的分布规律，然后按照规律，从图像数据空间中抽取样本，生成新的图像。生成模型学习的分布规律越接近图像的真实分布，则生成的图像质量越好。因为图像数据分布规律的复杂性，无法用经验规则或是显性的公式表达，只能通过大量数据的训练学习，才能实现图像生成模型的生成功能。目前，使用最为广泛的图像生成模型是生成对抗网络和扩散模型。

二、扩散模型的起源

扩散模型的起源可以追溯到热力学中的扩散过程，就是物质由高浓度的地方向低浓度的地方逐渐移动，最终形成一个各个方向均衡的状态。就像滴入水中的蓝墨水一样（见图 3.2.2），随着蓝墨水在水中扩散，水逐渐变成了淡蓝色，墨水分子均匀地分布在水中。这个状态非常简单，易于描述，而墨水入水时的状态却是千姿百态的，很难描述和刻画。假设扩散过程的每一步都可逆，只要扩散的“步子”足够小，就可以从当前的均匀分布状态，逆向推断出墨水最初入水的状态。用于图像生成的扩散模型就是受这一原理的启发而提出的。

以生成宠物猫的图片为例，直接给出猫的分布规律几乎是不可能的。想象一下，每次往清晰的猫的图片中加入噪声，随着时间的变化，加入的噪声越来越明显，图像会由清晰到模糊，最后完全变成雪花点的随机状态，如图 3.2.3 所示。这种随机状态就类似于墨水的均匀分布是易于描述的。然后训练一个模型学习逐渐加入的噪声，它就能逆向去除噪声，生成新的图像。



图 3.2.2 蓝墨水滴入水中的扩散过程

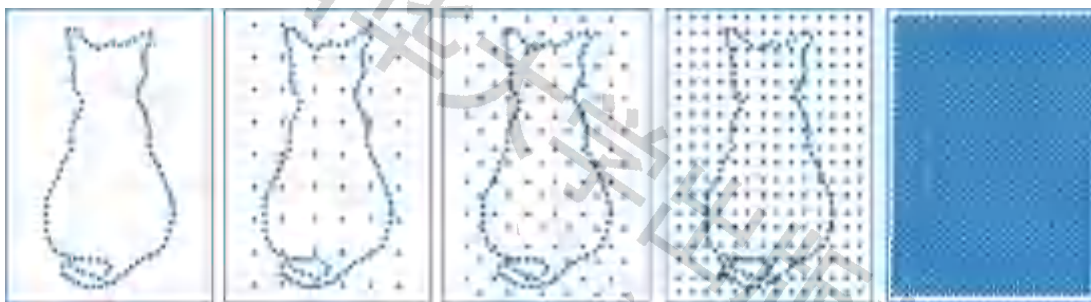


图 3.2.3 向猫的图像中不断加入噪声

三、扩散模型的原理

扩散模型就是一个先不断破坏（添加噪声），再逐步重建（去除噪声）的迭代生成的过程。

1. 扩散模型的正向过程

正向加噪过程如图 3.2.4 所示，图中 x_0 代表输入的真实图像，给真实图像 x_0 混入噪声会生成图像 x_1 ，经过第 t 步加噪后可以得到图像 x_t ，不断加噪直至第 T 步，图像会变成一幅没有任何含义的纯噪声图像 x_T 。 T 是预先定义的总的加噪步数，可以设置为 500、1000 等。 T 值越大，消耗的算力越多。



图 3.2.4 正向加噪过程

在正向加噪扩散的过程中，从前到后每一步加的噪声都是不同的。开始时，清晰的原图上只需要稍微加点噪声，就能明显看出混入了噪点。随着加噪步数的增加，为了让每次图像都有显著的变化，噪声会加得越来越多。

2. 扩散模型的训练

为了能够根据噪声图像生成新图像，需要训练一个神经网络预测正向所加的噪声。如图 3.2.5 所示，神经网络的输入有两项，分别是含有噪声的图像和当前所在的加噪步数 t 。

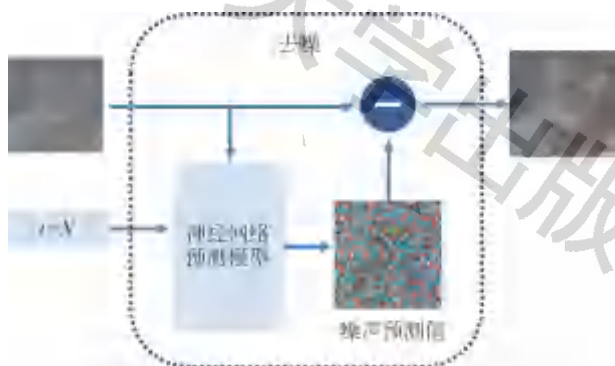


图 3.2.5 神经网络预测噪声（预测值）

图 3.2.5 中输入的噪声图像就是通过正向加噪到第 t 步获得的图像。神经网络期望的输出就是正向第 t 步时加入的噪声，如图 3.2.6 所示。扩散模型训练的目标就是对 $[1, T]$ 范围之间的任意步数的噪声图像，都能预测出其加入的噪声，从而恢复出上一时刻的图像，直至预测出第 0 时刻的图像，也就是生成新的图像。



图 3.2.6 正向不断加入噪声（真值）

3. 扩散模型的反向过程

经过训练后，神经网络可以预测每一步加入图像的噪声，然后从图像中去除噪声，逐渐生成全新的图像，如图 3.2.7 所示。



图 3.2.7 反向去噪生成过程

训练后的扩散模型学到了训练数据集的特征分布规律，但并不是记住了数据集中的图像再进行复制生成，因此它会生成与数据集特征相似的全新图像。

拓展阅读

生成对抗网络

生成对抗网络 (generative adversarial network, GAN) 是一类非常经典的生成模型。不同于扩散模型加噪去噪的生成原理，生成对抗网络是借助生成器和判别器间相互博弈的方式使生成器逐渐生成逼真的图像。

一个生成对抗网络包含两个基础网络：生成器 (generator, 简称为 G, 也被称为生成网络) 与判别器 (discriminator, 简称为 D, 也被称为判别网络)，两者既相互协作又相互对抗 (见图 3.2.8)。其中，生成器用于生成新数据，其生成数据的基础往往是一组噪声或者随机数，这些噪声或随机数经过生成网络，变换为生成数据 (比如图像)，生成器的目标是生成尽量真实的数据，最好能够以假乱真，而判别器用于判断生成的数据和真实数据哪个才是真的，其目标是让自己的判断准确性越来越高。

当生成器生成的数据越来越逼真时，判别器为维持准确性，就必须向判别能力越来越强的方向迭代。当判别器越来越强大时，生成器为了降低判别器的准确性，就必须生成越来越真的数据。这样不断地循环，直到生成器可以生成足够真实的数据，以至于判别器无法分辨真假。



图 3.2.8 一种简单的生成对抗网络示意图

4. 扩散模型的生成示例

本册中，我们借助手写数字图像数据集 MNIST，可以训练一个扩散模型，用于生成手写数字图像。MNIST 数据集的图像示例如图 3.2.9 所示，该数据集共有 70000 张图像，其中训练集 60000 张，测试集 10000 张。所有图像都是 28×28 像素的灰度图像，每张图像包含一个手写数字。

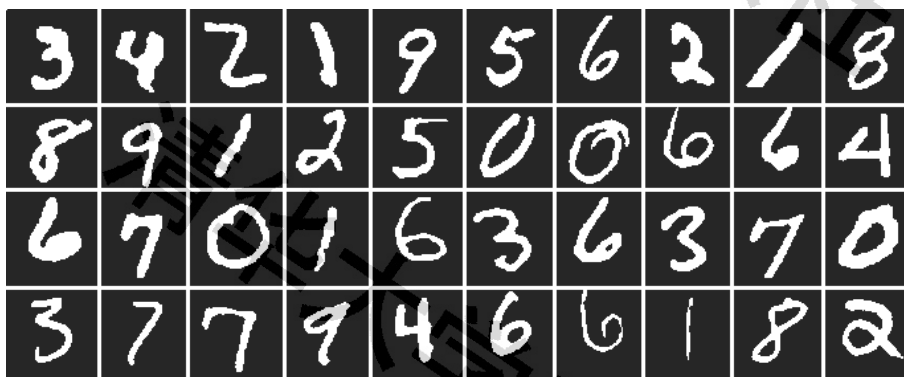


图 3.2.9 MNIST 数据集的图像示例

训练生成手写数字扩散模型的过程如下：

- ① 声明深度神经网络模型；
- ② 加载数据集；
- ③ 定义模型结构为扩散模型，指定模型训练的优化器和参数；

- ④ 进行正向加噪，训练扩散模型；
- ⑤ 应用扩散模型反向去噪，实现图像生成。

对应上述过程的关键程序如下。

```
# 导入依赖库
from BaseNN import nn
# 声明模型
model = nn()
# 加载数据集
model.load_img_data('/data/MELLBZ/mnist/training_set', batch_size=64)
# 定义模型结构：扩散模型
model.add('diffusion_model',img_size=28,timestep=500)
# 指定优化器（可省略）
model.add(optimizer='SGD')
# 正向加噪过程
model.noisy('/data/MELLBZ/mnist/training_set/3/0.jpg', timestep=500)
# 训练模型
model.train(epochs=10)
# 反向去噪过程
generated_imgs = model.inference(num=64, return_all_timesteps=True)
```

扩散模型训练时，正向加噪过程的效果如图 3.2.10 所示，可以看出，随着不断加噪，数字图像“3”变得越来越模糊。经过 500 次加噪后，最后变成了随机噪声。

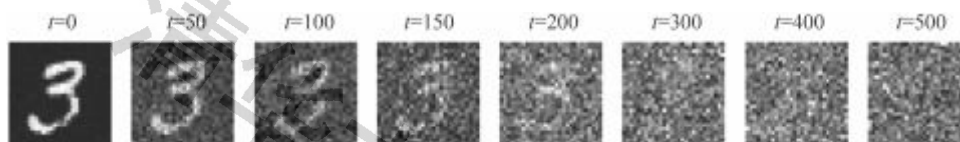


图 3.2.10 数字“3”的正向加噪过程

扩散模型生成时，反向去噪过程的效果如图 3.2.11 所示，由初始的噪声图像不断地去噪，最后生成一个清晰的数字图像“2”。

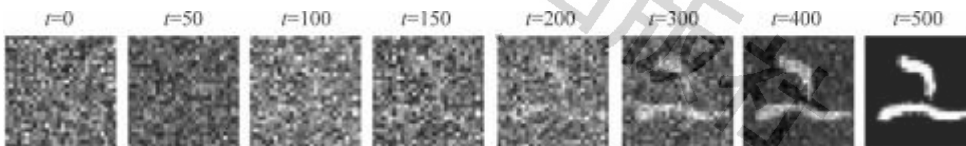


图 3.2.11 数字“2”的反向去噪过程

问题讨论

能够生成手写数字图像的扩散模型可以生成特定的数字吗？如果能，为什么？如果不能，有什么方法可以生成指定的数字？

实验活动




图像生成模型的训练

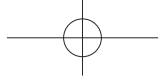
收集感兴趣的某类图像数据，或者直接选用手写数字图像数据集 MNIST，根据扩散模型的训练过程，完成扩散模型的训练，并生成新图像。观察生成的图像与训练集中真实的图像在图像特征一致性、清晰度、多样性等方面的差异，并进行评价。

实验内容：实现扩散模型的正向加噪过程和反向生成过程，并参考表 3.2.1 中正向加噪过程示例，将起始时刻、中间时刻及结束时刻的图像记录在表中。

实验准备：图像数据集。

表 3.2.1 图像生成效果展示及对比分析

过程	展示起始时刻的图像	展示中间时刻的若干张图像	展示结束时刻的图像
正向加噪过程示例			
正向加噪过程			
反向生成过程			
评价生成的图像与训练集中真实的图像在图像特征一致性、清晰度、多样性等方面的差异			



第3节

文本与图像的多模态模型

本节知识

- ◆ CLIP 多模态模型
- ◆ 根据文本生成图像

本节活动

- ◆ 依据文本生成不同类型的图像

随着生成式人工智能的不断发展，各种文生图的模型和工具应运而生，只要动动手指，输入一段文字，人工智能就能创作出符合文字指令的各种画作，让普通人也可以变身绘画魔法师。多模态模型技术功不可没，它

能将文本和图像两个模态信息映射到统一的空间中，建立文本和图像之间的语义联系，从而实现文本和图像的多模态数据的交互和生成。

一、CLIP 多模态模型

人类生活在一个由多种信息构成的世界中，人类可以借助眼、耳、口、鼻、皮肤等获取外界信息，这些信息包括语音、文字、图像、视频等多种类型的数据，这些不同的数据类型可以理解为模态，不同模态的数据有着不同的表达方式和表征空间。当研究的问题需要同时处理两种或多种模态信息时，我们将其称为多模态问题。

基于文本图像对比的预训练模型（contrastive language-image pre-training，后文简称 CLIP 模型）是多模态领域的经典之作，是一种可以同时处理文本和图像的基础模型。它能够将不同模态的原始数据映射到统一或相似的语义空间，实现不同模态信号间的相互理解，并能够基于此实现不同模态数据间的转化与生成。

1. CLIP 模型的数据集

CLIP 模型与以往的图像分类模型不同，它并没有使用大规模的带有标注的图像数据集，而是利用互联网上未经人工标注的图像-文本数据对进行训练。如图 3.3.1 所示，第一张图像是一只小猫，与其配对的文字是“A cute cat”……总共有 4 亿个图像-文本数据对被用于 CLIP 模型的训练，可见数据集规模相当大。

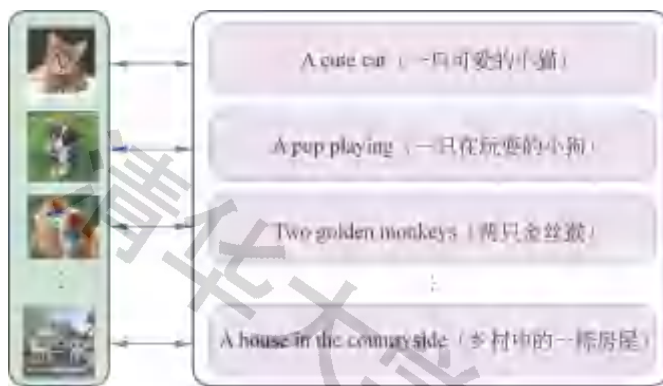


图 3.3.1 多组图像-文本数据对示意图

2. CLIP 模型的正负样本

因为图像-文本数据对规模很大，所以需要分批次输入 CLIP 模型中进行训练。假设每一个训练批次的大小为 N ，经过文本编码器和图像编码器后，会得到 N 个文本特征 $T_1, T_2, T_3, \dots, T_N$ 和 N 个图像特征 $I_1, I_2, I_3, \dots, I_N$ 。

这 N 个图像-文本特征对可以形成如图 3.3.2 右下方所示的矩形关系。图像和文本描述的内容是匹配的，沿着对角线方向的一系列图像-文本数据对就是正样本（用蓝色底纹标出的部分）。比如，猫的图像特征 I_1 和文本“A cute cat”的特征 T_1 是一一对应的关系，称为一个正样本；狗的图像特征 I_2 和文本“A pup playing”的特征 T_2 也是一一对应的关系，也是一个正样本，以此类推。

在这个矩形关系里，所有不是对角线上的数据对都是负样本。比如，一张猫的图像，用文本“A pup playing”描述，肯定是错误的，这种关系就称为负样本。

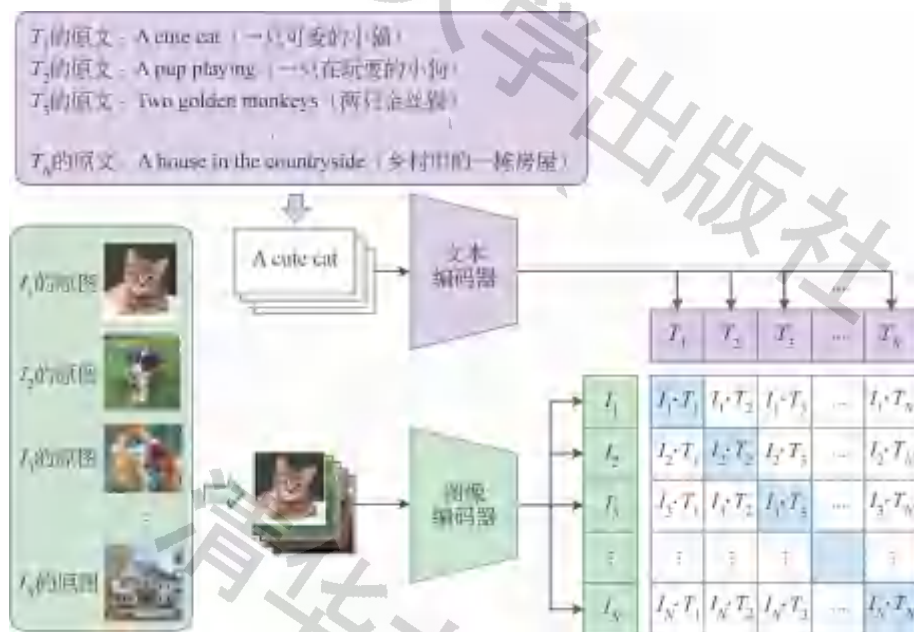


图 3.3.2 CLIP 模型的训练过程示意图

问题讨论

在一个训练批次大小为 N 的图像-文本数据对中，总共有多少个正样本？有多少个负样本？有了这样的图像-文本数据对，能否学习到图像和文本之间的语义联系？这种数据集相比起以往的手工标注图像的类别有什么优势？

3. CLIP 模型的对比学习

在图像-文本矩形关系中，可以通过“对比学习”，找到图像和文本的匹配关系。具体做法如下：计算矩形里每个格子中对应图像和文本的相似度。由于对角线上的图像-文本是成对的关系，训练的时候使其对角线上的相似度尽可能大，非对角线上的图像-文本的相似度尽可能小。这样，对于每一张图像，希望找到和它最相似的文本；而对于每一段文字，也希望找到和它最相似的图像，双向优化，最终得到图像和文本在多模态空间的特征表达，即一个图像编码器，一个文本编码器。

4. CLIP 模型的推理

在 CLIP 模型的推理阶段，文本的输入内容是可以指定的，比如可以是一系列类别信息的文本：“plane”“car”“dog”“bird”等，如图 3.3.3 所示。

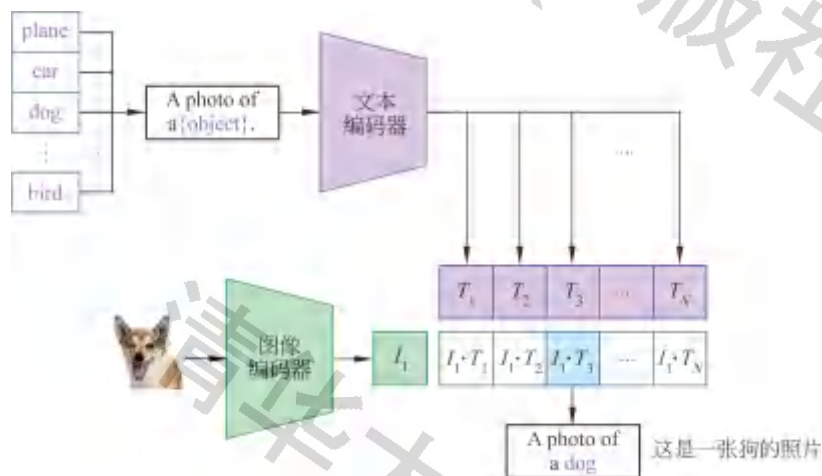


图 3.3.3 CLIP 模型的推理过程示意图

CLIP 模型通过文本编码器提取所有文本的特征。当给 CLIP 模型输入一张图像时，它先利用图像编码器得到图像的特征，不管这张图像它是否“见”过；接着，会将图像特征与前面一系列的文本特征做相似性比较；最后，把与图像特征最相似的文本特征所对应的句子输出，从而完成针对输入图像的分类任务，句子里描述的内容就是图像所包含的内容。

这种图像和文本的相似性比较不是顺次进行的，而是批次并行推理。

拓展阅读

CLIP 模型的性能评价

以图像分类任务为例，对比监督学习和 CLIP 模型。

在以往采用监督学习的图像分类任务中，需要准备带有标签的图像数据集进行训练。当遇到新的类别时，需要重新训练模型。比如已经训练了一个猫狗分类器，但是如果还想要区分鸟时，这个已有的猫狗分类器就不能用了，需要再训练一个包含猫、狗、鸟的分类器。

CLIP 模型用于新的类别进行图像分类时，无须重新训练。因为 CLIP 模型训练时，采用了 4 亿个图像-文本数据对，这个数据量几乎覆盖了所有可能的事物类别，它的监督信号完全来自配对的文本信息，不需要单独标注，就能通过相似性比较，找到文本和图像中匹配类别信息，完成图像的分类。

但是 CLIP 模型也有局限性，比如，它不适合用在计算物体数量及细粒度任务的分类上，如汽车型号、花卉种类等。另外，因为 CLIP 模型训练成本非常昂贵，所以如果现有 CLIP 模型对于待解决的任务起不到作用，那么训练自己的 CLIP 模型将会付出非常高的成本。

5. CLIP 模型的推理过程示例

首先，实例化图像和文本的嵌入模型；然后，对图像和文本进行编码，并计算两者的相似度。代码如下。

```
# 导入依赖库
from XEdu.hub import Workflow as wf
from XEdu.utils import get_similarity
# 实例化图像嵌入模型
img_emb = wf(task='embedding_image')
# 实例化文本嵌入模型
text_emb = wf(task='embedding_text')
# 示例输入
images = ['image_data/footprint.png', 'image_data/map.png']
texts = ['This is a tiger's footprint', 'this is a map']
# 对图像进行编码
image_embeddings = img_emb.inference(data=images)
# 对文本进行编码
text_embeddings = text_emb.inference(data=texts)
# 计算相似度
similarity = get_similarity(image_embeddings, text_embeddings,
                             method='cosine')
```

图像与文本相似度计算结果如图 3.3.4 所示，数值越大，说明其对应的行列上的图像-文本数据对相似度越高。从图 3.3.4 中可以看出，对角线上的数值相对其所在的行和列都是最大值，说明对角线上的图像和文本是成对的正样

本，而非对角线上的是负样本。



图 3.3.4 图像与文本的相似度结果

计算图像和 CIFAR100 数据集（一个经典的图像分类数据集，包含 100 个不同类别的图像）中类别的相似性，并显示前 5 个最为相似的类别。其中图像是任意的非 CIFAR100 中的图像。结果显示 CLIP 模型具有非常强的推理能力。比如，CIFAR100 数据集中没有“地图”这个类别，但是能将地图上显示的山脉、平原等地形识别出来，如图 3.3.5 所示。

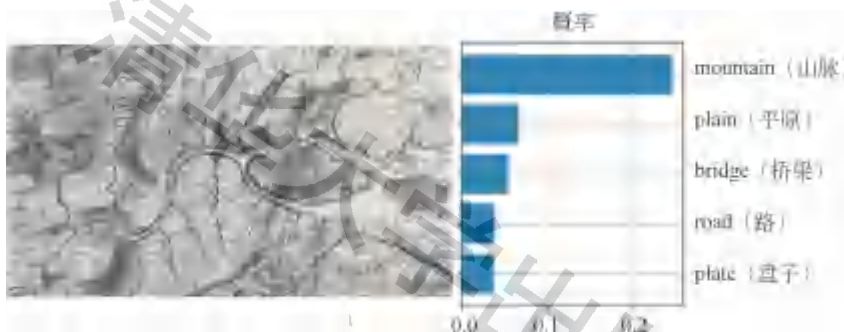


图 3.3.5 地图的识别结果

同样的，CIFAR100 数据集中没有“脚印”这个类别，但图 3.3.6 中的脚印在老虎、熊等动物的图像中会出现类似元素的内容，因此会被识别为这些

动物。

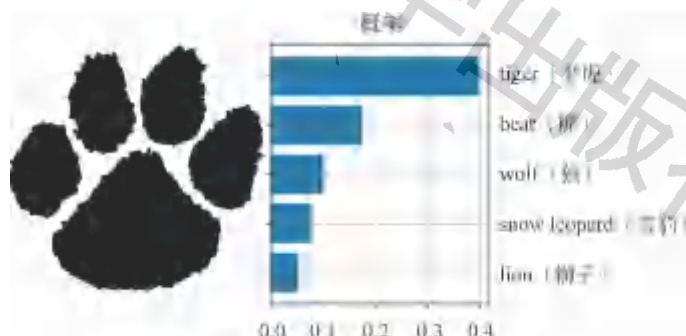


图 3.3.6 脚印的识别结果

再比如，CIFAR100 数据集中没有“酸汤鱼”这个类别，而图 3.3.7 中的鱼已经看不出其原貌，则盛鱼的碗会被识别为结果。



图 3.3.7 “酸汤鱼”的识别结果

二、根据文本生成图像

建立了文本和图像的多模态模型之后，就可以实现涉及文本和图像的各种跨模态任务，比如根据文本生成图像、图文检索、看图说话、图像编辑等，极大地延伸了生成式人工智能应用的广度。

1. 根据文本生成图像的过程

输入文字“一只在街上奔跑的柯基狗”，生成模型很快就会生成一张图像，如图 3.3.8 所示，画面主角是一只柯基狗，背景是街道，姿态是在奔跑。可以发

现, 图像中的内容和文本的描述非常匹配, 文本涉及的关键词都能满足。



图 3.3.8 文生图过程示意图

采用同样的文本指令时, 生成模型会源源不断地输出不同的图像, 且都符合文本描述, 如图 3.3.9 所示。生成模型就好比是不同的创作者, 听到相同的指令, 但是会产生不同的创作表达。在创作的过程中, 生成模型不但要理解指令中每一个词语所表达的含义和其对应的影像, 还要补全文本中没有涉及的可能的想象, 如街道两侧是高楼林立的居民区, 还是红红火火的商业街, 或索性是虚拟化街道背景? 柯基狗奔跑的姿态是双蹄腾空, 还是左右腿交替奔跑? 这些信息在文本中都没有明确表达, 但是图像在生成的那一刻就要明确。

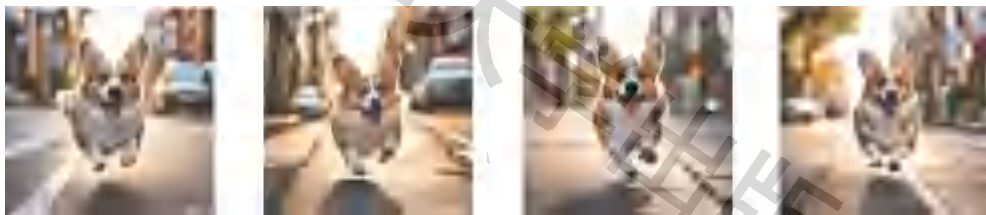


图 3.3.9 根据提示语“一只在街上奔跑的柯基狗”生成的系列图像

生成模型生成的这种图像既符合事物规律又带有随机性, 可以理解为从一个分布空间中随机抽取样本。文本“奔跑的柯基狗”对应很多不同的可能状态, 它们都符合这个文本描述, 在生成的时候, 生成模型会随机从这个文本描述对应的图像分布空间中抽取一个样本。

拓展阅读

常见的文本生成图像工具

2022 年 AIGC 爆发式发展, 同年 8 月, 在美国科罗拉多州举办的新兴数字艺术家竞赛中, 参赛者提交的 AIGC 绘画作品《太空歌剧院》获得了此次比赛“数字艺术/数字修饰照片”类别一等奖。参赛者没有绘画基础, 通过 AI 绘图软件 Midjourney 耗时 80 个小时创作出该作品, 如图 3.3.10 所示。

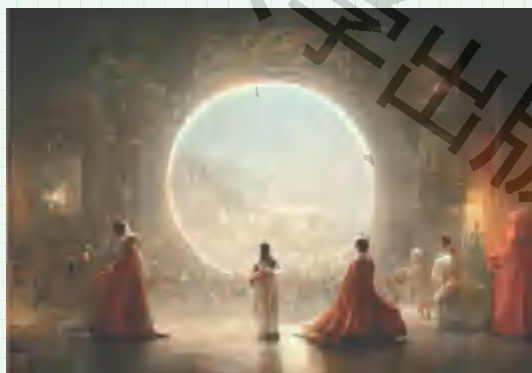


图 3.3.10 Midjourney 创作的画作《太空歌剧院》

目前，常见的文本生成图像工具有 MagicMaker、秒画、Midjourney、Stability AI、Imagen 和 DALL·E 系列等。秒画是由商汤科技自主研发的 AI 绘画平台，它基于一个文本生成图像大模型，依托商汤 AI 大装置的强大算力集群进行训练，其模型生成的图像质量更高、细节更丰富、风格更多样。

2. 编写提示语生成图像

目前，文本生成图像的模型功能非常强大，通过输入文字描述（提示语）就能得到想要的图像。可以按照一定的结构给出提示语，提示语一般包括画面主题、主题描述词、风格修饰词、画面质感增强词，每个提示语中间用逗号间隔。表 3.3.1 中展示了部分根据提示语生成的图像作品。

表 3.3.1 根据提示语生成的图像

提示语	粉色兔子，上班族，皮克斯动画风格	宣传海报，端午节，划龙舟，一群少年	一只戴着珍珠耳环的鹦鹉，维米尔风格，高画质，高清
图像			

3. 评估文本生成图像的匹配度

评估文本生成图像是一项非常有挑战性的工作，不仅要评估生成图像的真实性和质量，还要评估与文本描述的相关性。在具体评价的时候需要兼顾多项考虑，此处，借助 CLIP 模型评估一下图像的生成与其文本描述的相关性。如图 3.3.11 所示，标有红色下画线的图像-文本数据对，是依据该文本描述所创作的图像。左边第一张图是依据第一个提示语“一只在街上奔跑的柯基狗”所生成的。



图 3.3.11 文本生成图像的匹配性评价

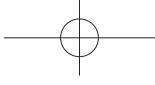
图 3.3.11 显示了图像与提示语之间的相似关系。有三张柯基狗的图像和前两个提示语很接近，但是仔细分辨，还是能看出明显差异。左边第一张图像相比后面两张图像，明显地表达出提示语“在街道上”和“奔跑”的信息，所以第一张图像与第一个提示语相似性分值会更高一些。第五张划龙舟的图像与后两个提示语较为接近，但与提示语“宣传海报，端午节，划龙舟，一群少年”的相似性分值更高，可能在端午节的场景下带有龙头的这种船会出现更多。通过文本与图像相似性的分值可以看出，依据文本生成的图像是比较真实、细致的。



实践活动

依据文本生成不同类型的图像

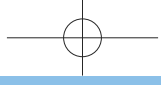
选用合适的文本生成图像工具，根据不同的图像类型，设计合适的提



示语，生成相应的图像。然后从生成的图像中挑选最符合期待的两张图像，并从图像生成的真实性、多样性、创意、意境、美感、图文匹配度等不同角度进行评价，填写在表 3.3.2 中。

表 3.3.2 根据提示语生成不同类型图像并作出评价

图像类型	提示语	生成的图像	评价
风景类			
人物类			
动物 / 植物类			
未来世界			



第4节

借助多模态模型进行创作

本节知识

- ◆ 多模态生成的创作过程
- ◆ 用于文本内容生成的提示语优化指南
- ◆ 用于图像生成的提示语优化指南

本节活动

- ◆ 使用多模态创作工具进行创意制作

人工智能生成工具如同魔术师，不仅能施展神奇的力量，将提示语瞬间变幻为鲜活生动的文本佳作，还能将天马行空的文字描述一键转化为栩栩如生的视觉场景。在本节中，我们将学习使用奇妙的多模态创作生成工具，通过输入多模态的提示语，生成用于制作电子明信片的宝藏素材。试着让我们的灵感与技艺碰撞出绚烂的火花吧！

一、多模态生成的创作过程

假如想生成一张“糯香满溢，端午寄思”的端午节明信片背景图，通过输入提示语“篮子里有若干粽子，整齐排列，粽叶青翠欲滴”，选择一种风格类型（如动画）和合适的生成比例（如4:3），即可生成一幅基本符合我们预想的图像，如图3.4.1所示。

想让画面中的粽子呈现更多细节，或者修改图像的风格，但是不改变整体画面的构图，该怎么办？如果通过修改提示语重新生成图像，新生成的图像具有一定的随机性，画面的整体构图很难维持，图生图可以解决这个问题。



图 3.4.1 “文生图”生成的粽子图

1. 图生图模型

图生图模型可以在参考图的基础上，根据提示语的描述生成新的图像。图 3.4.2 就是将图 3.4.1 作为参考图，并根据新的提示语“刚出炉的粽子，粽子上还有小水珠”以及新的风格类型“油画”所生成的。新生成的图像，整体构图与图 3.4.1 区别不大，但是呈现出了更多的细节，如新的绘图风格以及粽子上小水珠等。



图 3.4.2 “图生图”生成的粽子图

上述过程就是图生图工具的基本使用方法。在图生图的工具中，输入的内容包含两种模态的数据，分别是作为参考的参考图和指导模型生成新图像的提

示语。通俗来说，机器将理解文本提示语，并以提示语中的内容作为要求去修改参考图。利用“文本提示语 + 参考图”的方式进行图像创作，一定程度上可以减少图像生成的随机性，从而得到更符合期望的图像。

拓展阅读

不同的风格类型是怎么来的？

每种风格类型背后都对应一个基础模型。每个基础模型都是采用深度学习算法基于大量数据学习得到的大型神经网络模型，它能够根据给定提示语的描述，生成高质量的图像或完成复杂的任务。为了帮助理解，可以把各类基础模型想象成智能画家，它们都经过了大量的训练，掌握了许多绘画技巧和风格。不同的是，这些智能画家擅长的绘画风格各不相同。有的擅长写实，有的擅长素描，还有的擅长油画。如果想要生成油画风格的图像，那么应选择对应的基础模型，这样得到的艺术作品才会更符合所期望的风格。

2. 图像的局部编辑

假如想改变已经生成的图像中的某个物体或某个局部区域，该怎么办？图像的局部编辑功能可以解决这个问题。图像的局部编辑也是多模态生成中的一个重要功能，它能根据用户选定的区域以及输入的文本提示语，同时考虑原图像的内容信息，将原始图像选定区域修改为文本提示语中描述的效果。可以使用鼠标进行点选或者框选选定待修改区域或对象，选中后，编写文本提示语，描述被选中区域或对象想要生成的图像即可。利用这个功能，可以实现去除、添加、更改图像局部内容以及改变图像的某些细节等效果。

例如，图 3.4.3 展示了去除图像局部内容的操作。原图像中存在多余的绳结，选中多余的部分，然后使用图像编辑中的“智能清除”功能消除，即可将画面中的选中区域消除。

如果想让画面中的某个粽子呈现被剥开的状态，可以使用“点选”选中这个粽子，并输入文本提示“被剥开的粽子，里面是白色的糯米和红枣”，如图 3.4.4 所示。模型输出的效果如图 3.4.5 所示。



图 3.4.3 利用“智能清除”消除冗余元素

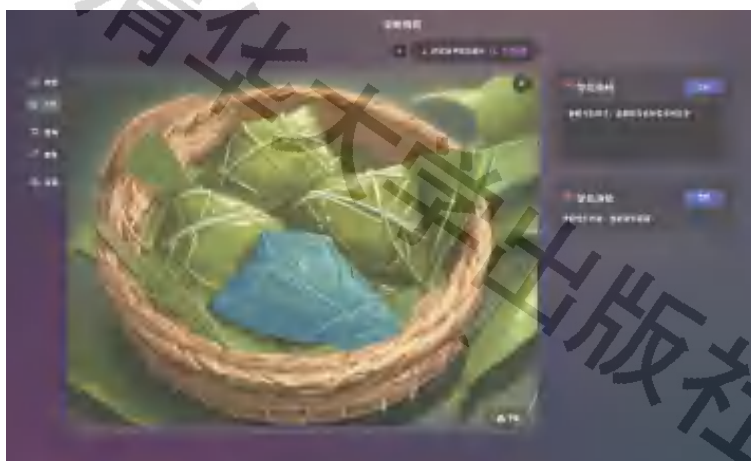


图 3.4.4 “智能编辑”过程图



图 3.4.5 重新编辑选中对象生成的效果图

刚才所选中的被粽叶包裹好的粽子变成了被剥开的状态，并露出了糯米和馅料。这里用到的原理就是在原图贴上蒙版，并编写提示语描述想要修改的效果。怎么理解蒙版？蒙版是一种非破坏性的编辑工具，核心作用在于通过一种灵活的方式帮助用户实现局部编辑，并且这种编辑是可以随时修改和撤销的。

拓展阅读

根据图像和提示语生成动图或视频

图像变成动图或者视频后，其表现力明显会增强。该如何让生成的图像动起来？视频生成工具可以解决这个问题。视频生成工具可以根据给定提示语、图像或视频中的单模态数据或者多模态数据，生成符合期望的视频内容。按照生成时输入的数据划分，视频生成工具可以分为文生视频、文图生视频、视频生视频等多种形式。

图 3.4.6 是一个文图生视频的工具示例，此图利用图 3.4.5 作为参考图，并编写期望生成视频效果的提示语“从画面右侧至左侧，阳光逐渐照亮画面的过程，生动展现光影交错的瞬间”，选择风格模型“动画”和动效强度参数，最终能够得到一个动态效果的视频。其中，动效强度数值越大，生成的效果越夸张，与原图的相似性越低；反之，数值越小，与原图相似性越高，动效幅度越小。



图 3.4.6 文图生成动态视频过程图



实践活动

使用多模态创作工具进行创意制作

小清已经使用“大语言模型助手”明确了明信片的主题，并生成了一些文本内容。

请试着利用多模态生成工具为该明信片的创作带来更多可能，结合具体需求为明信片生成合适的配图，并根据提供的示例，在表 3.4.1 中记录具体操作步骤。

表 3.4.1 使用多模态创作工具进行创意制作记录表

步骤	功能	输 入	输 出
示例	文生图	提示语：篮子里有若干粽子，整齐排列，粽叶青翠欲滴 风格类型：动画 生成比例：4:3	
1			
2			
...			

二、高质量提示语的编写指南

在多模态生成任务中,输入模型的提示语是用于引导模型生成特定类型内容的一组指令及信息。这些提示语包括文本描述、参考图像等多种模态,实际生成时,提示语可以是单一模态的,也可以是多模态混合的。例如,借助大语言模型助手文生文时,提示语就是文本这种单一模态;使用图像局部编辑工具时,提示语包含文本描述和参考图像,这就属于多模态混合。

不难发现,在使用大部分生成式人工智能工具进行创作时,基本上都需要编写文本类的提示语。因此,文本提示语编写的质量,是保证生成作品效果的关键。

1. 明信片文本提示语编写指南

大语言模型聊天助手能够回答任何问题,但是有时候它给出的“答案”并不是你期望的内容。实际上,大语言模型聊天助手的回答往往基于某个上下文背景,因此,上下文提示是使用此类工具时的入门技巧。上下文提示是指在提出问题前,向大模型提供背景或者情景信息,甚至为大语言模型助手分配合适的角色,引导模型根据预设的“身份”提供更相关、更有针对性的输出内容。例如,想给明信片配上与端午节相关的文案,可以将模型设定为“文创用品设计师”,从而激发它的创造力和专业性,如图 3.4.7 所示。

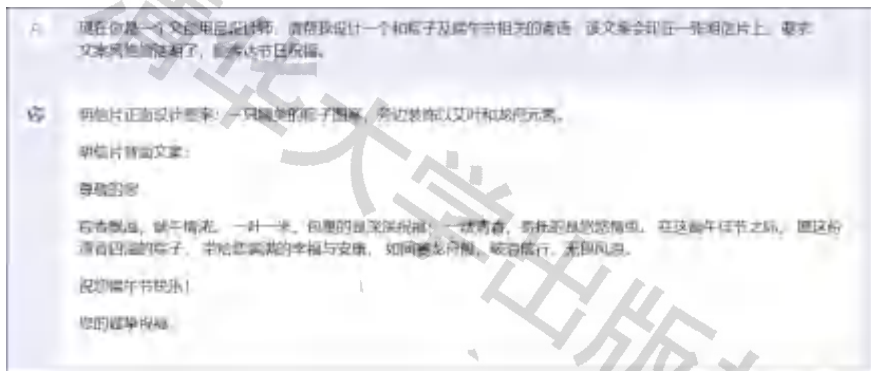


图 3.4.7 提出清晰且明确的指令

多轮对话是使用大语言模型助手的另一个技巧。在与大语言模型助手进行

对话时，很难一步到位生成令我们满意的内容。大语言模型助手具备较强的上下文理解能力，可以通过多轮对话的方法获得更全面的答案。比如，提供具体示例，供大语言模型助手模仿。如果想要类似于“无论咸甜，阖家团圆才是最好的馅”这样的文案，就可以告诉模型参照样例进行仿写，模型的回答如图 3.4.8 所示。

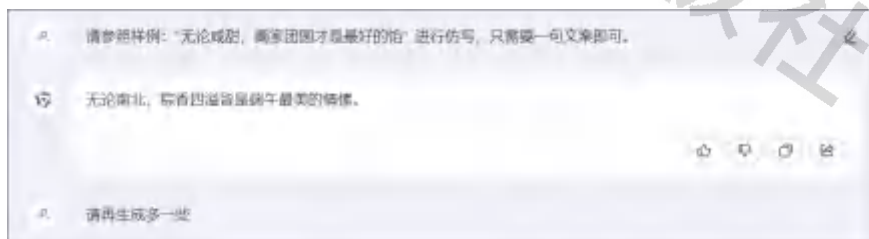


图 3.4.8 提供参考示例

表 3.4.2 给出了与大语言模型助手进行多轮对话的详细记录。每轮对话重新生成的内容可能或多或少会出现字数过多、侧重点偏离、生成的示例数量不够等问题。因而，在接收到模型的回答后，需要仔细检查是否符合要求。可以在下一轮的提问中加入明确的关键字、词或短语，对上一轮生成的内容提出更加明确的要求，以确保模型能够正确理解问题。

表 3.4.2 与大语言模型助手进行多轮对话记录表

对话轮次	提示语示例	提示语的构成	生成的文本
1	现在你是一名文创用品设计师，请帮我设计一个和粽子及端午节相关的寄语，该文案会印在一张明信片上，要求文案风格简洁明了，能传达节日祝福	提供详细的背景信息	明信片正面设计图案：一只精美的粽子图案，旁边装饰以艾叶和龙舟元素。 明信片背面文案： 尊敬的您， 粽香飘溢，端午情浓。一叶一米，包裹的是深深祝福；一缕清香，寄托的是悠悠情思。在这端午佳节之际，愿这份清香四溢的粽子，带给您满满的幸福与安康，如同赛龙舟般，破浪前行，无畏风浪。 祝您端午节快乐！ 您的诚挚祝福
2	请参照样例：“无论咸甜，阖家团圆才是最好的馅”，进行仿写，只需要一句文案即可	提供示例，进行模仿	无论南北，粽香四溢皆是端午最美的情愫

续表

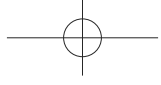
对话轮次	提示语示例	提示语的构成	生成的文本
3	请再生成多一些	明确的关键词/字：再、多	不论酸甜苦辣，共享天伦才是生活最醇的滋味。无论悲喜交织，围炉共话总是冬季最暖的画面。不管东西南北，归家团聚永远是春节最深的期盼。无问朝夕寒暑，手足相伴永远是最真的幸福旋律。不拘山海之隔，中秋赏月始终是亲情最浓的瞬间
4	文案内容需要围绕粽子，请重新生成	侧重关键词角度	粽叶飘香，丝丝缠绕的不只是五谷杂粮，更是千年流传的文化韵味。 一捧糯米、几片粽叶，包裹的是对端午佳节深深的情思与怀念。 每一只紧实饱满的粽子，都寄托着对屈子英魂无尽的追忆和崇敬之情。 不论甜咸南北，每一口粽子都是炎炎夏日里的一份清凉祝福与团圆期盼。 温馨手作的粽子，软糯之间流淌的是家庭和睦与节日欢聚的独特味道
5	文案字数需要限制在 50 字以内，请重新生成	明确的关键词：限制、50 字以内	粽叶裹糯米，端午情思浓。 甜咸皆佳品，一粽寄乡愁。 粽香飘满溢，屈子魂永存。 手作团圆味，糯心蕴深情

除了上述提问技巧以外，和大语言模型助手交流时还要留意以下几点。首先，尽量避免提出涉及主观判断的问题，因为模型不具备个人情感和价值立场，无法给出类似“你认为哪种方案更好”的主观评价。例如，在需要探讨某种观点时，可以明确询问“有哪些关于该话题的不同看法或观点”。其次，在与模型互动时，务必确保问题及潜在回答不触碰道德与法律底线。例如，不要尝试引导模型提供破解密码、侵犯他人隐私等违法违规信息。在合法、合乎道德规范的前提下使用大语言模型助手，使其在解答疑惑、辅助决策等方面发挥积极正面的作用。

2. 明信片图像提示语编写指南

在生成图像这类视觉内容时，提示语的精心设计与构建至关重要。和大语言模型提示语的不同之处在于文生图的提示语更关注视觉元素的具体呈现与艺术效果。

表 3.4.3 给出了三个不同细腻度级别的提示语示例。对比不同层级细腻度



的提示语所生成图像的效果，不难发现，提示语的语义信息越丰富，生成的图像在视觉表现力和细节还原度上的差别越明显。如表 3.4.3 所示，根据提示语“端午划龙舟”生成的图像，相比后两张图像，缺少了光照、环境及艺术风格等多层次的细节呈现，而风格化的图像往往能够传达特定的情感或氛围，合适的构图和独特的视角可以提升图像的视觉表现力。丰富的语义信息能够指导模型更精确地捕捉并理解用户的创作意图。例如，在提示词中给模型提供主体对象（形状、颜色、纹理、光照）和画面的空间布局等细节描述。此外，通过使用特定引导词可以有效地指导模型生成特定的艺术风格、情绪氛围等特征，如“中国山水画风格”或“庄重且祥和的节日氛围”等。

表 3.4.3 不同细腻度的提示语生成图像效果记录表

序号	提示语示例	提示语的构成	生成的图像
1	端午划龙舟	仅传达了基本的主题和场景	
2	华美的龙舟停泊在波光粼粼的江畔，在和煦的阳光下熠熠生辉，映衬出端午节庄重且祥和的节日氛围	提供了更多的主体细节、环境描述以及艺术风格指导	
3	一幅高清的艺术画，生动再现端午节清晨的江畔水景，碧波荡漾着龙舟竞渡的河道，天空被初升朝阳穿透薄雾洒下的光辉晕染成一片诗意盎然的霞光。在金色晨曦中，数艘龙舟蓄势待发或正在奋力划行，激起阵阵涟漪。强调光线对环境和活动的细腻渲染效果，采用类似中国山水画风格，确保画面色彩层次丰富且和谐，充分捕捉端午节热烈而庄重的气氛以及龙舟竞赛的动态美感。	不仅包含了时间、地点、光照情况、动态元素、自然现象等复杂内容，还指定了具体的艺术技法 and 美学追求。画面描述增加了空间描述，能够呈现丰富的层次	

同时，为了使模型能够更好地理解并构建出一个完整、协调且符合用户意图的视觉场景，要确保提示语之间具备逻辑关联性。例如，如果希望生成一幅包含“端午划龙舟”主题的图像，提示语不应局限于孤立的元素描述“端午划龙舟”，还应将与这个主题相关的元素有机地联系起来，如“华美的龙舟停泊在波光粼粼的江畔，在和煦的阳光下熠熠生辉，映衬出端午节庄重且祥和的节日氛围”。这样，模型就能够通过分析这些关联性强的提示语，理解整个画面的空间布局、时间情景及情感基调，并基于此创作出既符合描述又富有艺术感的高质量图像作品。

问题讨论

究竟怎样的提示语才能生成优质的图像？提示语中应该包含哪些内容？

试着以“贵州山水好风景”为主题，撰写一段提示语生成一幅可以充分展现“贵州地貌特色”的高质量图像。思考一下，提示语中必须包含哪些内容？对比一下各小组生成图像的效果和提供的提示语，思考并讨论：优质的作品究竟好在哪里？提示语中应该包含哪些内容才能生成优质的画作？

项目实施

制作并整合生成的素材，完成明信片的制作。

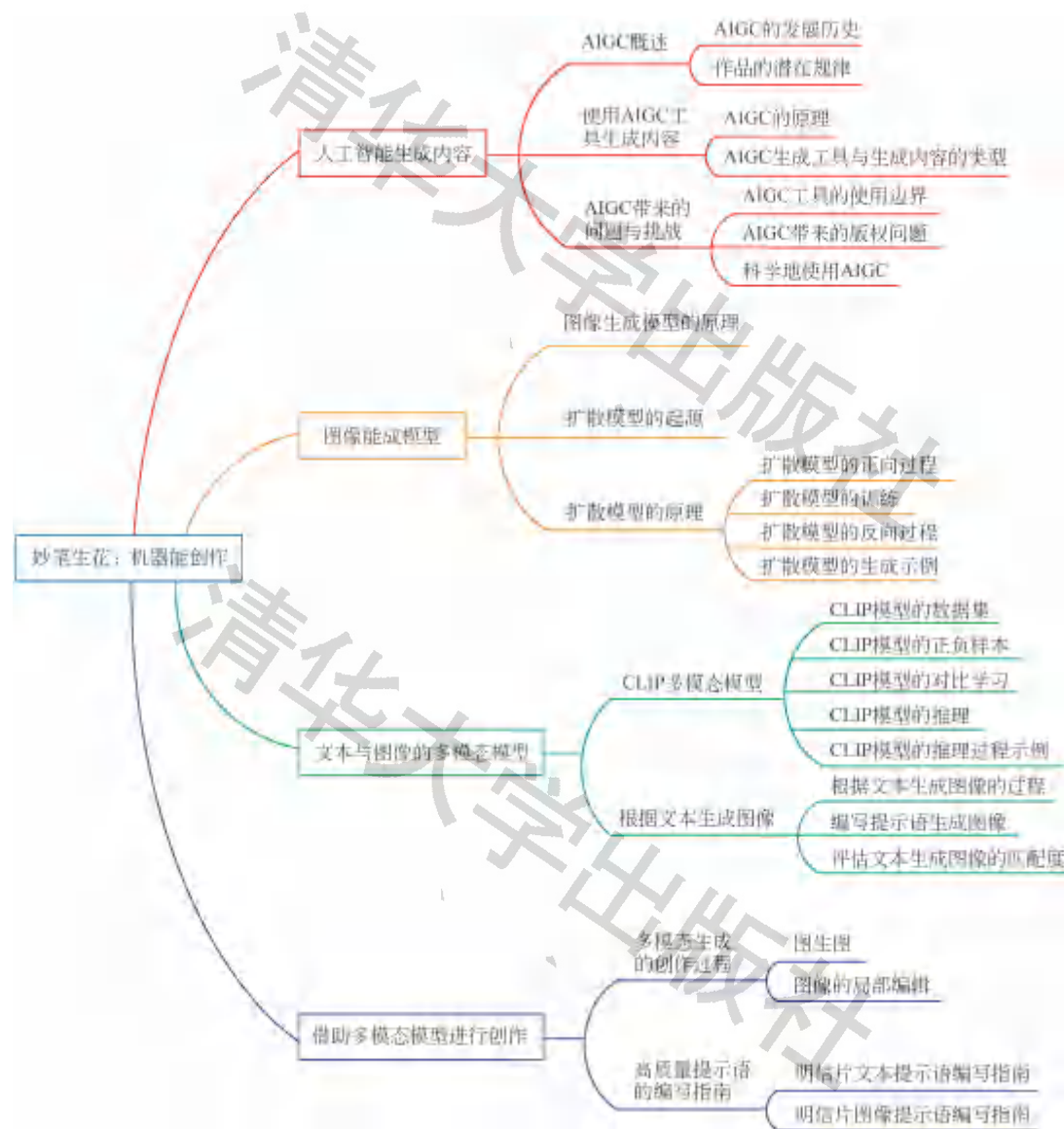
请试着围绕自己所选择的明信片主题，根据文本和图像提示语的编写指南，编写简洁而富有细节的文本提示语，综合运用大语言模型及多模态生成创作工具，创作出富有情感表达和视觉美感的个性化的素材。选择合适的图像与文本编辑工具，整合所有素材完成电子明信片的制作，并进行交流与分享。

明信片正面：

明信片背面：

单元小结

一、知识回顾



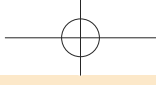
二、项目交流与评价

1. 参考本书附录“项目报告模板”撰写项目报告，并制作演示文稿。
2. 在课堂内展示自己的学习成果并分享经验，在下表中进行自评和他评。

项目成果评价表

评价维度	自评	他评
(1) 完整性 项目材料齐全，有作品构思、提示语设计、分工协作、提示语优化记录表（生成过程的提示语记录准确）及最终成果（明信片作品包含文生文、文生图两部分内容）。	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般
(2) 艺术性 作品内容具备一定的艺术性，文字部分阅读朗朗上口，图画部分具备一定的审美，作品与主题意境一致。	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般
(3) 规范性 项目报告规范，文生文、文生图提示语规范，符合基本要求。	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般	<input type="checkbox"/> 优秀 <input type="checkbox"/> 良好 <input type="checkbox"/> 一般

3. 保存作品创作过程中的提示语调整记录与最终作品，整理提示语、阶段性内容、最终作品等文档，并上传到校园网或者其他学习空间，与他人分享学习成果。



附录

项目报告模板

(此处填项目主题名称)

组员简介：姓名、组内职务与分工

项目问题：问题背景及需求

方案设计：知识学习、成果规划及成果呈现等

实践过程：实施过程的记录整理

成果汇总：活动成果整理

经验总结：实施活动的心得体会

撰写人：(签名)

日期： 年 月 日