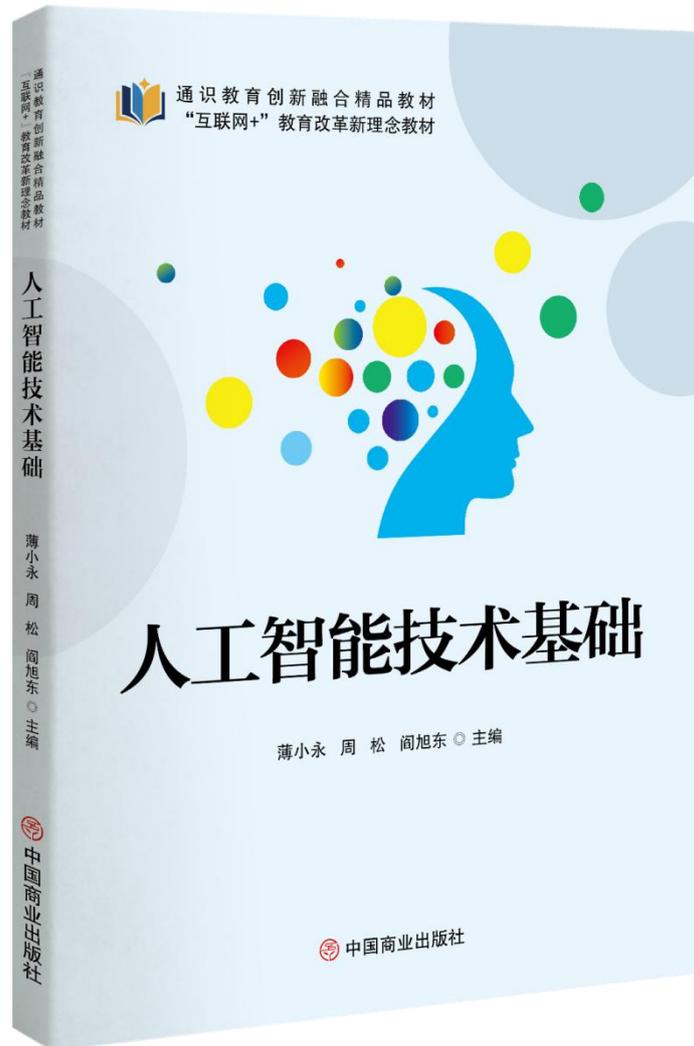


人工智能技术基础



类目：人工智能类
书名：人工智能技术基础
主编：薄小永 周松 阎旭东
出版社：中国商业出版社
开本：正 16 开
书号：978-7-5208-3645-6
使用层次：通用
出版时间：2025 年 10 月
定价：49.80 元
印刷方式：双色
是否有资源：有

责任编辑：滕 耘
封面设计：旗语书装



通识教育创新融合精品教材
“互联网+”教育改革新理念教材

“通识教育创新融合精品教材”
“互联网+”教育改革新理念教材

人工智能技术基础



人工智能技术基础

薄小永 周松 阎旭东◎主编



人工智能技术基础

薄小永 周松 阎旭东◎主编



学习资源库



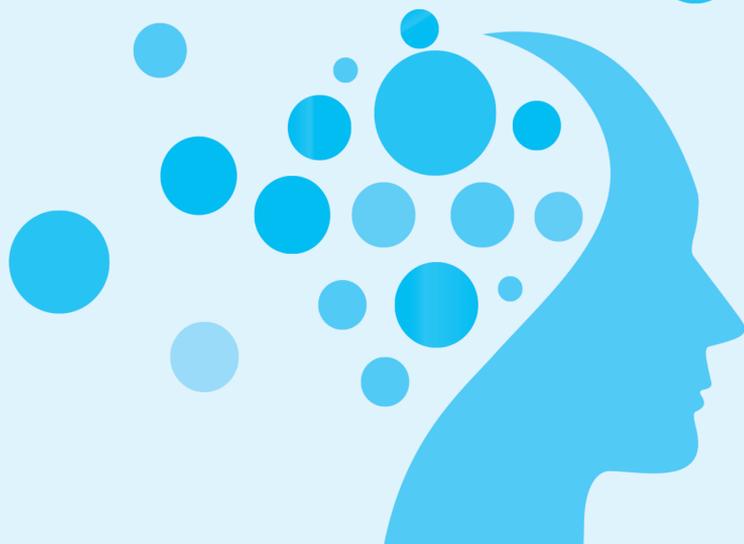
定价：49.80 元

中国商业出版社

中国商业出版社



通识教育创新融合精品教材
“互联网+”教育改革新理念教材



人工智能技术基础

薄小永 周 松 阎旭东 ◎ 主 编

徐智龙 马文路 ◎ 副主编

 中国商业出版社

图书在版编目 (CIP) 数据

人工智能技术基础 / 薄小永, 周松, 阎旭东主编.
北京: 中国商业出版社, 2025. 10. -- ISBN 978-7-5208-3645-6

I. TP18

中国国家版本馆 CIP 数据核字第 202571TM35 号

责任编辑: 滕 耘

中国商业出版社出版发行

(www.zgsybh.com 100053 北京广安门内报国寺1号)

总编室: 010-63180647 编辑室: 010-83118925

发行部: 010-83120835/8286

新华书店经销

唐山唐文印刷有限公司印刷

*

787毫米 × 1092毫米 16开 11.5印张 255千字

2025年10月第1版 2025年10月第1次印刷

定价: 49.80元

* * * *

(如有印装质量问题可更换)

前言

PREFACE

在信息化与数字化交织的今天，人工智能已不再是科幻电影中的想象，而是渗透到生活方方面面的现实力量。从智能音箱响应指令到智慧交通疏导车流，从医疗影像辅助诊断到金融风控精准预警，这项技术正以独特魅力重塑世界运转逻辑，成为引领时代变革的核心引擎。它不仅颠覆着我们的日常起居，更在产业升级、经济发展、社会治理等领域释放出不可估量的能量，推动人类社会迈向智能化新纪元。

正是基于这样的时代语境，《人工智能技术基础》应运而生。我们深知，面对蓬勃发展的人工智能浪潮，初学者往往渴望系统入门却苦于找不到合适路径。因此，本书致力于搭建一个全面且易懂的学习平台，让零门槛读者也能轻松迈入人工智能的大门。在编写过程中，我们始终坚守层次性与可读性原则：既避免用晦涩术语构筑壁垒，也拒绝为追求通俗而牺牲专业深度，通过由浅入深的内容编排，引导读者像剥洋葱般逐层探索技术内核。

作为基础入门级书籍，本书以六个项目构建起完整的知识框架。“人工智能概述”勾勒学科全貌，从概念起源到发展脉络厘清技术经纬；“人工智能技术”聚焦现实场景，展现技术在各行业的落地形态；“Python 程序设计基础”夯实实践根基，为后续学习提供工具支撑；“区块链与大数据”揭示技术生态，阐明人工智能背后的数据与信任基石；“人工智能算法”解析核心逻辑，拆解机器思考原理；“生成式人工智能模型”紧跟前沿，探讨新兴领域的创新应用。



为强化学习实效，教材特别设置九个特色模块。“项目导读”引出核心问题，“学习目标”和“德育目标”明确能力培养方向，“任务描述”与“任务分析”引导逻辑拆解，“任务实施”提供知识点，“分析与讨论”鼓励深度思考，“学以致用”设计实践场景，“巩固练习”检验学习成效。这些模块环环相扣，让内容既紧凑有序又逻辑清晰，形成“学思用”一体化的学习闭环。

本书定位兼具普适性与针对性，作为公共基础课程教材，能满足非计算机专业学生的系统学习需求；作为科普读物，可为人工智能爱好者提供自学参考。无论是课堂上的学生，还是职场中希望跨界的读者，都能在这里找到适合自己的入门路径。

编者

PREFACE



项目

<<< 人工智能概述 \ 1

- 任务一 人工智能的概念 \ 2
- 任务二 人工智能的发展历程 \ 5
- 任务三 人工智能应用探究 \ 10
- 任务四 人工智能的未来 \ 14

项目

<<< 人工智能技术 \ 19

- 任务一 物体识别 \ 20
- 任务二 人脸识别 \ 23
- 任务三 语音识别 \ 30
- 任务四 视频识别 \ 39

项目

<<< Python 程序设计基础 \ 54

- 任务一 Python 概述 \ 55
- 任务二 Python 语法基础 \ 63
- 任务三 程序流程控制 \ 69



任务四 Python 序列类型 \ 77

任务五 函数 \ 96

任务六 模块 \ 101

项目 四

<<< 区块链与大数据 \ 111

任务一 认识区块链 \ 112

任务二 大数据产业的理解 \ 118

任务三 大数据框架介绍 \ 120

项目 五

<<< 人工智能算法 \ 128

任务一 人工智能主流算法技术 \ 129

任务二 深度学习算法知识 \ 140

任务三 人工智能算法应用研究 \ 144

项目 六

<<< 生成式人工智能模型 \ 151

任务一 变分自编码器 \ 152

任务二 生成对抗网络 \ 155

任务三 流模型 \ 164

任务四 扩散模型 \ 169

参考文献 \ 178

项目一

人工智能概述



项目导读

本项目围绕“人工智能概述”展开，旨在帮助读者对人工智能形成整体认知。任务一聚焦人工智能的概念，对其相关的基本界定与范畴进行阐述，为学习提供基础认知。任务二梳理人工智能的发展历程，呈现其在不同阶段的发展脉络与重要进程。任务三探究人工智能的应用情况，涉及该技术在多个领域中的运用概况。任务四着眼于人工智能的未来，探讨其可能的发展走向及相关的广泛影响。通过本项目的学习，读者可对人工智能的整体面貌有较为全面的了解。

学习目标

知识目标

1. 掌握人工智能的核心概念，包括图灵测试、定义及强弱人工智能的区别。
2. 了解人工智能发展中的重要节点、学科诞生及发展历程。
3. 知晓人工智能在医疗、交通等多领域的典型应用场景。
4. 理解人工智能的近期目标及未来发展趋势。

能力目标

1. 能准确阐述人工智能的基本概念和不同类型特征。
2. 能梳理人工智能发展脉络，分析各阶段的特点。
3. 能结合案例分析人工智能在具体领域的应用价值。
4. 能对人工智能未来发展趋势进行初步研判。



素质目标

1. 培养对人工智能的学习兴趣和探索精神。
2. 提升对新技术的敏感度和适应科技发展的素养。
3. 树立辩证看待人工智能发展的意识。
4. 培养跨学科思维，提升分析问题的能力。

德育目标

认识人工智能发展的重要意义，树立科技报国理想，将个人发展与国家科技事业相联系。

任务一 人工智能的概念

任务描述

本任务聚焦人工智能的概念，包含三个核心板块。首先，围绕图灵测试与新图灵测试展开，探究早期判断机器是否具备智能的经典方法及当代衍生测试的差异；其次，明确人工智能的定义，从学科交叉视角解析其本质，涵盖模拟人类智能的技术与系统；最后，区分强人工智能与弱人工智能，对比两者在能力范围、目标设定上的根本区别，旨在通过系统梳理，帮助读者构建对人工智能概念的基础认知框架，理解不同维度下智能的判定与分类逻辑。

任务分析

本任务从基础概念切入，层层递进解析人工智能的核心要素。图灵测试与新图灵测试的对比，揭示了智能判定标准随技术发展的演变，体现从行为模仿到深层认知评估的转变；人工智能定义的阐释，需融合计算机科学、心理学等多学科视角，明确其“模拟、延伸、扩展人类智能”的核心特质；强人工智能与弱人工智能的区分，则是理解当前技术边界与未来发展方向的关键，弱人工智能的工具属性与强人工智能的自主意识目标形成鲜明对比，为后续学习人工智能技术应用与伦理问题奠定概念基础。

任务实施

人工智能（Artificial Intelligence，AI）是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的新兴技术科学，自图灵测试起逐步进入研究领域。



一、图灵测试与新图灵测试

1936年，哲学家阿尔弗雷德·艾耶尔对心灵哲学问题展开研究：我们究竟如何知晓其他人曾有过相同的体验。在《语言、真理与逻辑》这部著作中，艾耶尔阐述了“有意识的人类及无意识的机器之间的区别”。

1950年，艾伦·麦席森·图灵在其论文《计算机器与智能》里预言了创造出具备真正智能的机器的可能性。鉴于注意到“智能”这一概念难以精准定义，他提出了闻名遐迩的图灵测试：若一台机器能够（借助电传设备）在5分钟内回答由人类测试者提出的一系列问题，且超过30%的测试者无法判定被测试者是人还是机器，那么这台机器便通过了测试，可称这台机器具有智能。图灵测试是人工智能哲学领域的首个严肃提案，其中的30%是图灵对2000年时机器思考能力做出的一个预测。

图灵构想采用“问”与“答”的模式，也就是观察者通过打字机向两个测试对象进行通话，其中一个是人类，另一个是机器。要求观察者持续提出各类问题，以此来辨别回答者究竟是人是还是机器。

图灵指出，倘若机器在某些现实的条件下，能够极为出色地模仿人类回答问题，以至于提问者在相当长的一段时间内误将其当作人类而非机器，那么这台机器就可以被视作是能够思维的。

从表面来看，要让机器回答按照一定范围所提出的问题似乎并无太大困难，通过编制特定的程序便可实现。然而，要是提问者不遵循常规标准，那么编制回答的程序就会是一件极其艰难的事情。

美国科学家兼慈善家休·罗布纳为把图灵的设计转化为实际行动，在20世纪90年代初设立了人工智能年度比赛，比赛设有金、银、铜三个等级的奖项。

2014年6月8日，俄罗斯的一个团队开发了名为“尤金·古斯特曼”的人工智能聊天软件。在英国雷丁大学针对这一软件开展的测试中，它成功地让33%的人类对话参与者认为，与其聊天的对象是人类，而非计算机，进而成为有史以来首台通过图灵测试的计算机——这一事件被视作人工智能发展的一个里程碑式事件。

2015年11月，《科学》杂志封面刊登了一篇重磅研究：人工智能终于能像人类一样学习，并通过了图灵测试。测试的对象是一种人工智能系统，研究者分别展示它未见过的书写系统中的一个字符例子，并让它完成写出同样的字符、创造相似字符等任务。结果表明，这个系统能够迅速学会写陌生的文字，同时还能识别出非本质特征（也就是那些因书写造成的轻微变异），通过了图灵测试，这也是人工智能领域的一大进步。

1980年，约翰·塞尔在《心智、大脑和程序》一文中提及中文屋子思想实验，对图灵测试予以批评。

数十年来，研究人员始终运用图灵测试评估机器模仿人类思考的能力，然而这一针对人工智能的评判标准已沿用70年之久，研究者觉得应当更新换代，研发出新的评判标准，



推动人工智能研究朝着现代化方向更进一步。

新的图灵测试会涵盖更为复杂的挑战，如加拿大多伦多大学计算机科学家赫克托·莱韦斯克所提议的“威诺格拉德模式挑战”。该挑战要求人工智能回答有关语句理解的一些常识性问题。例如：“这个纪念品无法装在棕色手提箱内，因为它太大了。问：‘什么太大了？’回答 0 表示纪念品，回答 1 表示手提箱。”

加里·马库斯的建议是在图灵测试中增添对复杂资料的理解，包括视频、文本、照片和播客。例如，一个计算机程序可能会被要求“观看”一个电视节目或者网络视频，然后依据内容来回答问题，像是：“为什么电视剧《绝命毒师》中，老白打算甩开杰西？”

二、人工智能的定义

人工智能属于计算机科学的一个分支，是一门由自然科学、社会科学以及技术科学相互交叉所形成的边缘学科，其涉及的学科内容涵盖哲学和认知科学、数学、神经生理学、心理学、计算机科学、信息论、控制论、仿生学、社会结构学与科学发展观等诸多方面。

人工智能是对人的意识以及思维的信息过程所进行的模拟。人工智能并非人的智能，然而却能够像人一样进行思考，甚至在某些方面还有可能超越人的智能。人工智能旨在了解智能的本质，进而生产出一种能够以与人类智能相似的方式作出反应的智能机器。自其诞生之后，人工智能的理论和技术的越发成熟，应用领域也在不断拓展。可以预见，人工智能所带来的科技产品将会成为人类智慧的“容器”。所以，人工智能是一门极具挑战性的学科。

人工智能的定义可拆分为“人工”与“智能”两个部分。其中，“人工”指的是凭借人力或者人的智力能够制造出来的东西，“智能”则牵涉到意识、自我、思维之类的问题。然而实际上，人类对于自身智能的理解状况，以及对构成人类智能的必要元素的认知程度都较为有限，这就导致很难确切地定义什么是由“人工”制造出来的“智能”。对人工智能进行的研究，常常会关联到对人类自身智能的研究，并且其他有关动物或者人造系统的智能方面的研究，也普遍被看作和人工智能相关的研究课题。

尼尔逊教授给人工智能下过这样一个定义：“人工智能是关于知识的学科，也就是怎样对知识进行表示，以及怎样获取知识并运用知识的科学。”温斯顿教授则认为：“人工智能就是探究如何让计算机去从事以往只有人类才能完成的智能工作。”这些观点体现了人工智能学科的基本思想与基本内容。

自 20 世纪 70 年代起，人工智能被称作世界三大尖端技术（空间技术、能源技术、人工智能）之一，同时也被视为 21 世纪三大尖端技术（基因工程、纳米科学、人工智能）之一，原因在于近 30 年间人工智能得到了迅猛发展，在众多领域均获得了广泛应用，并取得了丰硕的成果。



三、强人工智能与弱人工智能

对于人的思维模拟，能够从两个方向展开。其一为结构模拟，即依照人脑的结构机制，制造出“类人脑”的机器；其二是功能模拟，也就是从人脑的功能过程着手进行模拟。现代电子计算机的产生，便是对人脑思维功能的模拟，是对人脑思维信息过程的一种模拟。

强人工智能又被称作多元智能。研究人员期望人工智能最终能够发展成多元智能，并超越大部分人类的能力。部分人觉得要达成上述目标，或许需要具备拟人化的特性，如人工意识或者人工大脑。上述问题被视作具有人工智能完整性：若要解决其中某一个具体问题，就必须将所有问题都解决掉。即便是一个简单且特定的任务，像机器翻译，也要求机器依据作者的论点（推理），知晓人们所谈论的内容（知识），忠实地展现作者的意图（情感计算）。所以，机器翻译被认为具备人工智能完整性。

强人工智能的观点主张，有可能制造出真正能够推理和解决问题的智能机器，并且这样的机器将会被视作有知觉、有自我意识的。强人工智能可以分为以下两类。

一是类人的人工智能，即机器的思考和推理方式如同人类的思维一样。

二是非类人的人工智能，即机器产生了与人类截然不同的知觉和意识，运用和人类完全不一样的推理方式。

弱人工智能的观点认为，不可能制造出能够真正推理和解决问题的智能机器，这些机器只不过看起来是智能的，但实际上并不真正拥有智能，也不会有自主意识。

当前，主流的研究活动主要集中在弱人工智能领域，并且已经取得了可观的成果，而强人工智能的研究则只是稍有进展。

分析与讨论

结合日常接触的例子（如智能音箱、聊天机器人等），说说它们是属于弱人工智能还是强人工智能，并说明判断依据。

任务二 人工智能的发展历程

任务描述

本任务围绕人工智能发展历程展开，包含四个要点：介绍推动智能理论的前驱学者及通用机器的早期构想，追溯人工智能学科正式诞生的标志性事件，梳理其发展历程中的关键阶段与重要突破，分析人工智能产生的社会必然性。旨在帮助读者系统了解人工智能从理论萌芽到学科确立、逐步演进的历史脉络，把握其与社会发展的关联。



任务分析

本任务以历史逻辑为核心。“大师与通用机器”是理论源头，展现早期思想奠基；“人工智能学科的诞生”标志着研究进入系统化阶段，具有里程碑意义；“人工智能的发展阶段”呈现技术与认知的起伏推进，反映领域探索的曲折；“人工智能的社会必然性”揭示其是科技积累与社会需求共同作用的结果，为理解当下发展提供历史视角与深层动因解析。

任务实施

繁重的科学研究和工程计算本需人脑承担，如今计算机不仅能完成此类计算，还能比人脑做得更快、更精准，所以人们已不再将这种计算视为“只有人类智能才能完成的复杂任务”。由此可见，复杂工作的定义会随着时代发展与技术进步而改变，人工智能的具体目标也随着时代变化而发展，它一方面不断取得新进展，另一方面又朝着更有意义、更具难度的新目标迈进。科学家已经制造出非常多的像汽车、火车、飞机这样的技术系统，它们模仿并拓展了人类身体器官的功能，然而到目前为止，人们只是知道人类大脑是由数十亿个神经细胞组成的器官，对它的了解还十分有限。

一、大师与通用机器

艾伦·麦席森·图灵出生于英国伦敦帕丁顿，毕业于普林斯顿大学，是数学家、逻辑学家，被誉为“计算机科学之父”“人工智能之父”，是计算机逻辑的奠基者，如图 1-1 所示。1950 年，图灵在其论文《计算机器与智能》中提出了著名的“图灵机”和“图灵测试”等重要概念。

约翰·冯·诺依曼出生于匈牙利，毕业于苏黎世联邦工业大学，是数学家，是现代计算机、博弈论、核武器和生化武器等领域内的科学全才，被后人称为“现代计算机之父”和“博弈论之父”，如图 1-2 所示。他在泛函分析、遍历理论、几何学、拓扑学和数值分析等众多数学领域及计算机学、量子力学和经济学中都有重大成就，也为世界第一颗原子弹和第一台电子计算机的研制作出了巨大贡献。



图 1-1 图灵



图 1-2 冯·诺依曼



电子计算机俗称电脑，简称计算机，是一种通用的信息处理机器，它能够执行可被详细描述的任何过程。用于描述解决特定问题的步骤序列被称作算法，算法能够转化为软件（程序），从而确定硬件（物理机）可以做什么以及做了什么。创建软件的这一过程被称为编程。

世界上第一台通用电子数字计算机 ENIAC 于 1946 年诞生，如图 1-3 所示。中国的第一台电子计算机则在 1958 年问世。在 2020 年 11 月公布的全球 TOP 500 超级计算机榜单中，中国超级计算机的数量延续了上一届的强大优势，中国的神威·太湖之光超级计算机继续保持第四的名次。此次全球 TOP 500 超级计算机榜单中，中国共有 217 台上榜，美国以 113 台位居第二，日本则以 34 台排在第三。

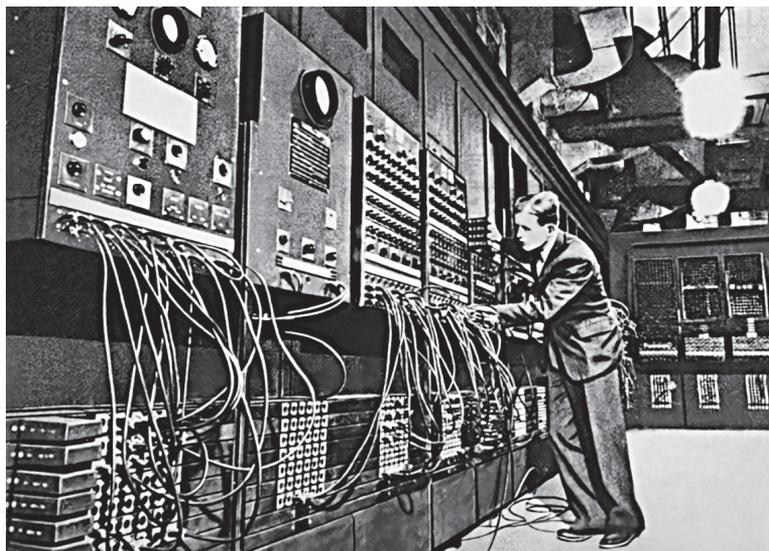


图 1-3 世界上第一台通用电子数字计算机 ENIAC

现代计算机被界定为“在可变更的程序的管控下，能够存储并处理信息的机器”。这里包含两个要点。

其一，计算机是用于处理信息的装置。也就是说，把信息存入计算机，计算机会把信息转化为新的、有价值的形式，接着通过显示或其他途径输出信息。

其二，计算机是在可改变的程序控制下运行的。计算机并非唯一的能够处理信息的机器。例如，当你使用计算器时，就是在输入信息（数字），对信息进行处理（如计算连续的总和），然后输出信息（如显示结果）。再如，油泵在给油箱加油时，会利用输入的每升汽油价格以及来自传感器的信号，读取汽油流入汽车油箱的速度。油泵将这些输入转化为加了多少升汽油以及应付多少钱的信息。不过，计算器或者油泵并非完整的计算机，它们只是内置了用于执行单个特定任务的嵌入式计算机。

二、人工智能学科的诞生

电子计算机的问世使信息存储与处理发生了革命性变革，计算机理论不断发展并催生



了计算机科学，最终推动人工智能诞生。尽管计算机提供了必要的技术基础，但直至 20 世纪 50 年代早期，人们才关注到人类智能与机器之间的关联，也就是反馈控制。一个大家熟知的例子是自动调温器，它会把收集到的房间温度和人们期望的温度进行比较并作出反应，通过调大或关小加热器来控制环境温度。这项研究的重要意义在于：所有智能活动皆为反馈机制的产物，而反馈机制是能够用机器模拟的。这一发现对早期人工智能的发展影响深远。

电子计算机的出现，使在技术上创造出机器智能成为可能，人类由此真正拥有了一个可模拟人类思维的工具。此后，无数科学家为实现这个目标而不懈努力。如今，全球几乎所有大学的计算机系都有人从事这门学科的研究与学习。

1956 年夏季，以麦卡锡、明斯基、罗切斯特和香农等为首的一批有远见卓识的年轻科学家在达特茅斯会议上聚会，共同研究和探讨用机器模拟智能的一系列有关问题，首次提出了“人工智能”这一术语，标志着人工智能这门新兴学科的正式诞生。

1997 年 5 月，IBM 公司研制的深蓝计算机战胜了国际象棋大师卡斯帕罗夫，这是人工智能技术的一次完美表现。

三、人工智能的发展阶段

人工智能六十余载的发展之路并非一帆风顺，其间充满了波折与挑战，其发展大致可划分为以下六个阶段，如图 1-4 所示。

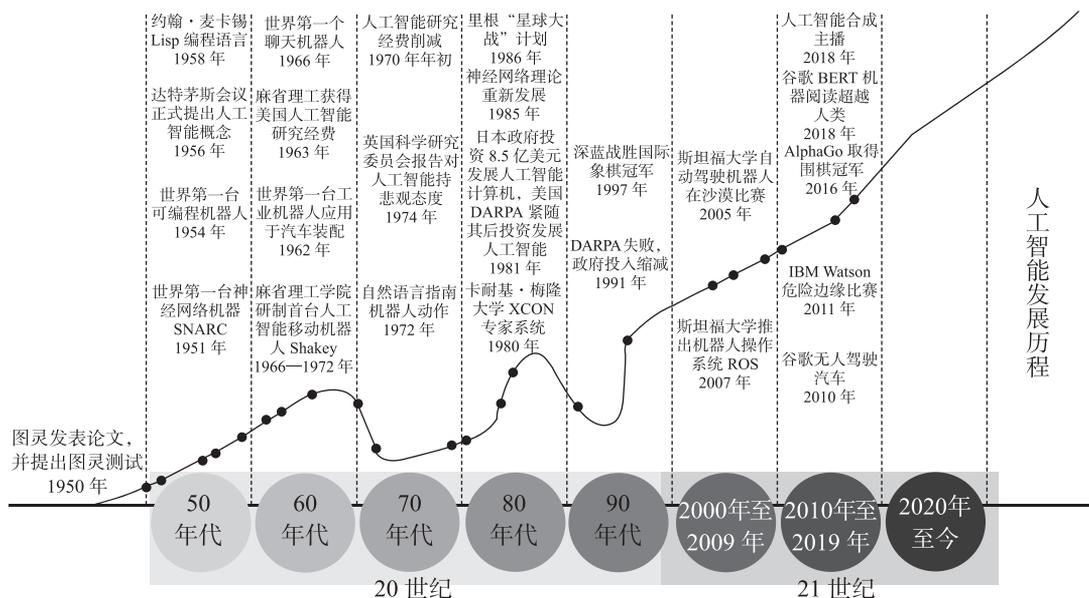


图 1-4 人工智能的发展阶段

(一) 起步发展期

1956 年至 20 世纪 60 年代初，在人工智能概念首次被提出后，相关研究迅速崭露头角，接连诞生了机器定理证明、跳棋程序、LISP (LISt Processing) 表处理语言等一系列



令人瞩目的成果，由此掀起了人工智能发展史上的第一个高潮。

（二）反思发展期

20 世纪 60 年代至 70 年代初，人工智能在发展初期取得的突破性进展，极大地拔高了人们对这一领域的期待。在这种高涨的预期下，研究者们开始着手挑战更复杂的任务，甚至提出了一些脱离实际的研发目标。然而，现实却未能如预期般推进，机器无法证明两个连续函数之和仍是连续函数、机器翻译频频出现荒诞笑话等接连不断的失败，以及预期目标的逐一落空，让人工智能的发展陷入了低谷。

（三）应用发展期

20 世纪 70 年代初至 80 年代中期，专家系统的出现成为人工智能发展的重要转折点。这类系统通过模拟人类专家的知识与经验来解决特定领域问题，实现了人工智能从理论研究迈向实际应用、从对一般推理策略的探讨转向专门知识运用的重大突破。凭借在医疗、化学、地质等多个领域的成功应用，专家系统有力推动人工智能进入了应用发展的新高潮。

（四）低迷发展期

20 世纪 80 年代中期至 90 年代中期。随着人工智能应用规模的不断扩大，专家系统存在的应用领域狭窄、缺乏常识性知识、知识获取困难、推理方法单一、缺乏分布式功能、难以与现有数据库兼容等问题逐渐暴露出来。

（五）稳步发展期

20 世纪 90 年代中期到 2010 年，网络技术，尤其是互联网技术的蓬勃发展，使得信息与数据的汇聚节奏持续加快。与此同时，互联网应用的日益普及，也为人工智能的创新研究注入了强劲动力，推动着人工智能技术在实用化的道路上不断前行。这一时期，1997 年 IBM 深蓝超级计算机击败国际象棋世界冠军卡斯帕罗夫，2008 年 IBM 提出“智慧地球”概念，这些事件都成为人工智能发展进程中具有标志性意义的里程碑。

（六）蓬勃发展期

自 2011 年起，互联网、云计算、物联网、大数据等信息技术不断革新，泛在感知数据的规模化积累与图形处理器（Graphics Processing Unit, GPU）等计算平台的迭代升级，为以深度神经网络为核心的人工智能技术注入了强劲动能，使其迎来飞速发展期。这一进程显著消弭了科学研究与实际应用之间的“技术鸿沟”，图像分类、语音识别、知识问答、人机对弈、无人驾驶等极具应用潜力的人工智能技术，成功实现了从“不能用、不好用”到“可以用”的关键跨越。由此，人工智能的发展正式步入爆发式增长的崭新阶段。

四、人工智能的社会必然性

总体上看，人工智能当前的发展具有“四新”特征：一是以深度学习为代表的人工智



能核心技术取得新突破，二是“智能+”模式的普适应用为经济社会发展注入新动能，三是人工智能成为世界各国竞相战略布局的新高地，四是人工智能的广泛应用给人类社会带来法律法规、道德伦理、社会治理等一系列新挑战。因此，这个机遇与挑战并存的新课题引起了全球范围内的广泛关注和高度重视。

虽然人工智能未来的创新发展还存在不确定性，但大家普遍认可其蓬勃兴起将带来新的社会文明，推动产业变革，深刻改变人们的生产生活方式，是一场影响深远的科技革命。人工智能技术的发展反映了生产力发展的要求，其产生具有必要性。

（一）人工智能是工具进化的结果

与以往的劳动工具相比，人工智能的进步体现在能够对大脑进行模拟，其技术已然超越了以往各类技术，有力推动了生产力的发展。此外，相较于之前的生产工具，人工智能不仅丰富了人的内心世界，还增强了人类的身体机能——它比以往的工具有了更多的肢体功能，因高度模仿人类技能而具备极强的拟人性，呈现出拟人装置的显著特征。

（二）人工智能响应生产力发展要求

人工智能的广泛应用催生了众多新兴行业，这些行业的发展速度与模式均超越了以往。在生产过程中，任何重大的科学技术创新都需要同步发展生产工具、设施、劳动者及生产管理方法，以此进一步提升生产力、拓展生产规模、提高人类改造客观世界的效率。作为一种辅助工具，人工智能协助人类重塑客观世界，以实现效率最大化，这一特性与生产力发展的要求相契合。而人工智能的迅猛发展，不仅解放了人类的智力与体力，还提升了管理效率和机器生产效率，拓宽了劳动者的实践领域，丰富了劳动改造的对象，进而推动了生产力的提高。

分析与讨论

结合计算机从承担繁重科学计算到如今助力人工智能发展的过程，说说工具的进化如何体现了“复杂工作的定义随时代发展而改变”这一现象？

任务三 人工智能应用探究

任务描述

本任务探究人工智能的应用，涵盖医疗、交通、金融、物流、教育、家居、制造七大领域。分别介绍人工智能在各领域的具体应用场景与形式，展现其如何融入不同行业。旨在让读者了解人工智能应用的广度与深度，明晰其在实际生活和产业中的落地情况，构建对人工智能应用的整体认知。



任务分析

从各领域来看，医疗领域应用人工智能可提升诊断效率与精准度，但需解决数据隐私与算法可靠性问题；交通领域人工智能技术能改善交通拥堵、减少事故，却面临法规不完善与技术瓶颈；金融领域借助人工智能可增强风险管控能力，却需防范算法偏见与系统安全风险；物流领域人工智能应用能降低成本、提高效率，却受限于基础设施与数据共享难题；教育领域应用人工智能有助于实现因材施教，但仍需平衡技术与人文教育；家居领域应用人工智能可以提升生活便捷性，但存在标准不统一与用户隐私顾虑；制造领域应用人工智能可以推动产业升级，但面临传统企业转型困难与技术投入压力。整体而言，探究这些应用对推动各行业发展意义重大，同时需应对多方面挑战。

任务实施

人工智能的应用领域极为广泛，涵盖医疗、交通、金融、物流、教育、家居、制造等多个行业，甚至在玩具、游戏、音乐等日常休闲领域也有深入渗透。

一、医疗

人工智能的显著优势在于，它既能在极短的时间内查阅海量数据，又能精准定位关注领域。

近年来，巴罗神经学研究所与 IBM 公司的人工智能计算系统“沃森”携手，在肌萎缩侧索硬化（俗称渐冻症）研究领域取得了突破性进展。作为应用人工智能技术的计算系统，“沃森”可查阅数千项研究成果，并从中筛选出与渐冻症相关的基因，这为渐冻症研究者带来了新的思路，也为药物靶点的探索和渐冻症治疗方法的研发提供了便利。

在医疗领域，人工智能还有一项极具前景的应用——预测药物疗效。以往，癌症患者往往服用相同的药物，人们需通过检测患者的反应来了解药效；而借助人工智能，可依据数据预测出对特定患者最有效的药物，进而提供高度个性化的治疗方案，有效节省了宝贵的时间与金钱。

二、交通

在智慧交通领域引入人工智能，能大幅缓解交通拥堵状况。近年来，尽管无人驾驶车辆引发的交通事故屡屡成为新闻焦点，但人工智能在该领域的应用，最终将显著减少道路伤亡人数。斯坦福大学的一则报告显示，自动驾驶汽车（如图 1-5 所示）不仅能减少因交通事故导致的人员伤亡，还能改变人们的生活方式：在通勤途中，人们可以把更多的时间用于工作、娱乐；关于居住的地点，人们也可以有更多的选择。其报告指出，“自动驾驶汽车也将越发舒适，人们的认知负担越来越少，共享交通的发展，这些因素可能会影响人们对于生活地点的选择”。



图 1-5 自动驾驶汽车

三、金融

人工智能的诞生与发展，不仅让金融机构的服务更具主动性与智慧性，有效提升了金融服务的整体效率，还增强了金融机构的风险管控能力，为金融产业的创新发展注入了积极动能。在金融领域，人工智能的应用场景十分广泛，涵盖智能获客、身份识别、大数据风控、智能投顾、智能客服、金融云等多个方面，而金融业也是人工智能渗透最早、覆盖最全面的行业之一。展望未来，人工智能将持续推动金融行业的智能应用升级，进一步提升行业运行效率。

四、物流

物流行业借助智能搜索、推理规划、计算机视觉和智能机器人等技术，在对运输、仓储、配送、装卸等流程实施自动化改造后，已基本可实现无人操作。像基于大数据进行商品智能配送规划，就能对物流供给、需求匹配、物流资源等进行优化配置。不过就目前情况而言，物流行业的大部分人力仍集中在“最后一公里”的配送环节。

五、教育

借助图像识别技术，可实现机器批改试卷、识题、答题等功能；通过语音识别技术，能对发音进行纠正与改进；利用人工智能大语言模型，则能完成在线答疑、文章修改、程序编写等任务。人工智能与教育的融合，在一定程度上缓解了教育行业师资分布不均、费用偏高的问题，从工具层面为师生提供了更高效的教学与学习方式。

六、家居

智能家居主要是基于物联网技术，通过智能硬件和软件系统、云计算平台构成一套完



整的家居生态圈，如图 1-6 所示。用户能够远程控制各类设备，设备之间可实现互联互通，并且具备自我学习等能力，从而整体提升家居环境的安全性、节能性与便捷性。尤其值得一提的是，近年来，随着智能语音技术的不断发展，智能音箱成为一个爆发性增长点。智能音箱并不仅仅是一款音响产品，而是集内容服务、互联网服务及语音交互功能于一身的智能化产品，它不仅支持 Wi-Fi 连接，能提供音乐、有声读物等内容服务以及信息查询、网购等互联网服务，还可以与智能家居相连接，实现场景化的智能家居控制。



图 1-6 智能家居

七、制造

智能制造，是基于互联网的物联网架构实现的贯穿企业与社会的全流程制造形态，它将“智能工厂”“智能生产”“智能物流”进一步拓展至“智能消费”“智能服务”等全环节的智能化范畴。人工智能在制造业的应用主要体现在三个维度：其一是智能装备，包含自动识别设备、人机交互系统、工业机器人及数控机床等具体装置；其二是智能工厂，涉及智能设计、智能生产、智能管理和集成优化等核心内容；其三是智能服务，涵盖大规模个性化定制、远程运维与预测性维护等服务形态。尽管当下人工智能解决方案还无法彻底满足制造业的需求，但作为通用性技术，人工智能与制造业的融合已然成为大势所趋。

分析与讨论

结合人工智能在医疗或交通领域的应用案例，说说这些应用如何体现了人工智能“既能高效处理海量信息，又能提供个性化或精准化服务”的优势？



任务四 人工智能的未来

任务描述

本任务围绕“人工智能的未来”展开，包含近期发展目标与长远前景探究。从近期视角来看，其发展目标聚焦关键技术突破，如多模态统一表征、智能系统自主决策能力提升等，同时推动人工智能在医疗诊断、自动驾驶、智能制造等领域的深度应用，完善技术落地的配套机制。从长远视角来看，需研判人工智能在社会结构、生产模式、伦理规范等方面的深远影响，探索人机协同的可持续发展路径，明确其在未来人类社会中的角色定位。

任务分析

近期发展目标的推进，面临技术瓶颈（如通用人工智能突破难）、数据安全与隐私保护压力，以及跨领域协同不足等问题，但能为产业升级提供直接动力。从长远来看，人工智能的未来发展将重塑就业市场，可能引发伦理争议与监管挑战，却也孕育着生产力飞跃的机遇。探究这些内容，可为制定科学发展战略提供依据，助力平衡技术进步与社会稳定，具有重要的现实与前瞻意义。

任务实施

人工智能已经深度融入人们的生活，重塑了社会运行模式。面对这一科技浪潮，人们不禁思索：人工智能的下一步将迈向何方？又会给人类生活带来怎样颠覆性的改变？

一、近期发展目标

从近期视角来看，人工智能的发展聚焦于攻克以下关键难题。

（一）可信人工智能

自2022年生成式人工智能取得重大突破后，其在文本创作、绘画设计和图像生成等领域展现出惊人的能力，甚至在部分任务上超越了人类水平，这也让通用人工智能的实现看似触手可及。然而，繁荣背后暗藏隐患：由于训练数据的局限性，生成内容常出现虚假信息，从编造学术论文到被不法分子用于电信诈骗，此类问题引发了严重的社会信任危机。当前，构建可解释、可验证、安全可靠的人工智能体系，已成为行业亟待突破的瓶颈，这不仅关乎技术发展，更是人工智能实现长期健康发展的核心命题。



（二）在虚拟环境下训练人工智能

近年来，机器人硬件技术日新月异，高性价比的智能设备不断涌现，如仅需几千元就能购置的便携式高清摄像无人机。但令人疑惑的是，日常生活中机械助手的普及程度却远低于预期。其主要原因在于，传统机器人执行特定任务时，需通过大量试错编程训练，而现实环境中的训练不仅成本高昂，还存在诸多安全风险与不可控因素。

构建虚拟仿真环境成为破局的关键路径。在虚拟世界中，机器人可以模拟各种复杂场景，进行高强度、无风险的训练，随后将优化后的模型参数迁移至实体设备。这一技术已在多个领域取得显著成效：科技企业通过虚拟训练让机械臂掌握多样化的抓取技能；自动驾驶企业利用虚拟道路测试大幅缩短研发周期、降低测试成本；2022年后，结合大型语言模型与虚拟机器人训练的创新模式更是成为行业新趋势，实现了语言指令驱动下的复杂任务执行。随着虚拟训练技术的成熟，人工智能的落地速度将大幅提升，推动行业进入高速发展的快车道。

二、长远前景探究

将目光投向更长远的未来，强人工智能时代的到来充满想象空间。

当前，人工智能尚处于弱人工智能阶段，其能力局限于特定领域，如 AlphaGo 在围棋领域所向披靡，却无法理解自然语言；人脸识别系统能精准识别人脸，却无法洞察人类情感。而强人工智能则被定义为具备人类同等甚至超越人类的通用智能，不仅能处理复杂任务，还拥有自主思考、抽象推理、经验学习等高级认知能力。尽管强人工智能的研发面临诸多技术与伦理挑战，但近年来大模型技术的突破和生成式人工智能的快速迭代，已为其发展奠定了坚实的基础。

回顾人工智能近 80 年的发展历程，从理论萌芽到技术爆发，每一次突破都在加速行业演进。从弱人工智能迈向强人工智能，或许还需经历数年甚至数十年的攻坚，但一旦实现，人工智能将进入指数级增长的“奇点”阶段，超级人工智能时代也将随之而来。届时，人工智能将突破人类现有的认知边界，创造出前所未有的可能性。我们既是这场变革的见证者，也将成为塑造未来的参与者，共同探索人工智能与人类社会深度融合的无限可能。

展望未来，人工智能有望在多个关键领域掀起变革浪潮。例如，在教育领域，人工智能或将为每个学生量身定制个性化学习方案，智能导师不仅能精准识别知识薄弱点，还能依据学生的学习习惯与兴趣偏好，动态调整教学策略，真正实现因材施教，让优质教育资源跨越地域与经济的限制，惠及更多人群。在医疗健康方面，人工智能辅助诊断系统将更加精准高效，通过分析海量医疗数据，提前预测疾病风险，为医生提供决策支持，甚至在手术中实现机器人辅助的精细化操作，极大提升医疗服务的质量与可及性。在能源领域，人工智能可助力优化能源分配与管理，通过智能算法预测能源需求，协调不同能源来源的供应，加速清洁能源技术的研发与应用，为缓解全球能源危机、推动可持续发展贡献



力量。

此外，随着人工智能与物联网、5G 等前沿技术的深度融合，“硅基生命”的概念正逐步从科幻走向现实。以传感器、处理器等硬件为“躯体”，以各类智能技术为“神经中枢”的智能体，将凭借全时在线、海量并行推理等特性，以智能终端、智能网联汽车、智能机器人等多元形态融入社会生活，进一步重塑人们的交互方式与生活体验。

但与此同时，人工智能的飞速发展也带来了诸如隐私保护、数据安全、伦理道德等一系列严峻挑战，亟待全球各界协同合作，制定合理的政策法规与行业标准，确保技术发展始终服务于人类福祉。

分析与讨论

结合“可信人工智能”的需求，说说生成式人工智能在提供便利的同时，可能带来哪些需要警惕的问题，以及可以从哪些角度应对这些问题？

学以致用

智能音箱的家居控制应用

以小米智能音箱为例，介绍如何利用人工智能技术实现对家居设备的智能控制，体验 AI 在日常生活中的便捷应用。

步骤一：准备设备与连接

确保智能音箱已通电并完成初始化设置，同时将需要控制的智能家居设备（如智能台灯、智能空调等）接入同一 Wi-Fi 网络，并在对应的米家 App 中完成设备添加与绑定。

步骤二：唤醒智能音箱

对智能音箱说：“小爱同学。”此时音箱指示灯亮起，表明已成功唤醒，进入待命状态。

步骤三：发出控制指令

根据需求向智能音箱下达控制指令。例如：

控制灯光：“小爱同学，把客厅的灯打开”“小爱同学，将卧室台灯调为暖光”。

调节空调：“小爱同学，把客厅空调温度调到 26℃”“小爱同学，关闭主卧空调”。

步骤四：查看设备状态与管理

打开米家 App，在设备列表中可查看被控制设备的当前状态（如是否开启、温度设置等）。若需修改控制参数或取消指令，可在 App 中直接操作，也可再次通过语音指令告知智能音箱进行调整。

通过以上步骤，智能音箱借助语音识别、自然语言处理等人工智能技术，实现了对家居设备的便捷控制，体现了弱人工智能在特定场景下的工具属性，能够根据预设程序和数据处理用户指令，但不具备自主意识和通用思考能力，为日常生活带来了高效与便利。



巩固练习

一、单选题

1. 1950年,图灵在其论文《计算机器与智能》中提出了图灵测试,若一台机器在5分钟内回答人类测试者的问题,超过多少比例的测试者无法判定其是人还是机器,则通过测试?()

- A.20% B.30% C.40% D.50%

2. 人工智能学科正式诞生的标志性事件是()。

- A.1936年阿尔弗雷德·艾耶尔对心灵哲学问题的研究
B.1950年图灵提出图灵测试
C.1956年达特茅斯学会上首次提出“人工智能”术语
D.1997年IBM深蓝计算机战胜国际象棋大师卡斯帕罗夫

3. 下列属于强人工智能观点的是()。

- A. 不可能制造出真正有智能、有自主意识的机器
B. 机器只是看起来智能,实际上不具备真正智能
C. 有可能制造出能真正推理、解决问题且有知觉和自我意识的智能机器
D. 机器只能在特定领域完成预设任务,不具备通用智能

4. 在人工智能发展历程中,专家系统的出现推动其进入应用发展新高潮,这一阶段是()。

- A. 起步发展期(1956年至20世纪60年代初)
B. 反思发展期(20世纪60年代至70年代初)
C. 应用发展期(20世纪70年代初至80年代中期)
D. 稳步发展期(20世纪90年代中期到2010年)

5. 下列哪项不属于人工智能在医疗领域的应用?()

- A. 借助AI预测药物对特定病人的疗效
B. 利用AI进行智能配送规划
C. 通过AI系统筛选与渐冻症相关的基因
D. AI辅助诊断系统提升诊断精准度

二、填空题

1. 人工智能是研究、开发用于模拟、延伸和扩展人的智能的_____、_____、技术及应用系统的一门新兴技术科学。

2. 强人工智能可以分为两类:类人的人工智能和_____的人工智能,其中类人的人工智能是指机器的思考和推理方式_____人类的思维一样。

3. 1997年5月,IBM公司研制的_____计算机战胜了国际象棋大师卡斯帕罗夫,这是人工智能技术的一次重要表现;2020年11月全球TOP 500超级计算机榜单中,中国



共有_____台超级计算机上榜。

4. 智能家居主要是基于_____技术，通过智能硬件、软件系统、_____平台构成一套完整的家居生态圈。

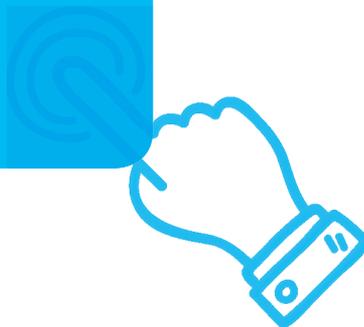
5. 从近期来看，人工智能的发展聚焦于攻克可信人工智能和_____环境下训练人工智能等关键难题；从长远来看，人工智能有望迈向_____时代。

三、简答题

结合人工智能在医疗领域的应用，谈谈其在提升医疗服务质量的同时，可能面临哪些挑战？

项目二

人工智能技术



项目导读

本项目围绕人工智能技术的关键应用领域展开，构建智能感知技术的认知体系。任务一从计算机视觉对图像的理解切入，阐释机器解析视觉信息的基础逻辑，并介绍相关核心任务。任务二聚焦人脸识别，涵盖技术内涵、发展历程及系统组成等维度，展现生物特征识别的独特价值。任务三转向语音识别，从音频认知出发，解析声音信号转化为可处理信息的技术路径。任务四则围绕视频认知、编解码技术及视频理解中的行为与场景识别展开，揭示动态视觉信息处理的关键机制。通过本项目的学习，读者可初步建立对人工智能感知技术的系统认知。

学习目标

知识目标

1. 了解物体识别、人脸识别等技术的基本概念及计算机视觉的任务。
2. 知晓人脸识别的发展历程和系统组成。
3. 理解音频认知及音频信号识别的基本过程。
4. 掌握视频认知、编解码技术及视频理解中的行为与场景识别相关知识。

能力目标

1. 能简述物体识别和人脸识别的核心内容。
2. 能说明音频信号识别的基本流程。
3. 能解释视频编解码技术及视频理解的关键环节。
4. 能初步分析人工智能感知技术在实际中的应用。



素质目标

1. 培养对人工智能感知技术的探究兴趣。
2. 提升对视觉、音频等信息处理技术的认知素养。
3. 树立严谨对待技术细节的科学态度。
4. 增强对人工智能技术应用的责任意识。

德育目标

认识人工智能技术发展的重要性，树立利用技术服务社会、推动进步的责任感。

任务一 物体识别

任务描述

本任务围绕物体识别展开，包含计算机视觉对图像的理解及计算机视觉的任务两部分。需探究计算机视觉如何解析图像中物体的形状、颜色、纹理等特征，以及物体所处场景的空间关系，从而理解图像的语义内容；同时明确计算机视觉在物体检测、分类、分割等具体任务的目标、常用算法及在智能安防、工业质检等场景的应用。

任务分析

本任务是计算机视觉领域的基础，为后续更复杂的物体识别任务奠定技术根基，能推动自动驾驶中障碍物识别、安防监控中异常物体检测等领域的发展。但面临图像模糊、物体遮挡、光照变化等导致的识别精度下降问题，且不同场景对算法适应性要求高，需不断优化算法模型以提高鲁棒性和泛化能力。

任务实施

安全检查是指对旅客及行李物品进行检查，查看其中是否携带枪支弹药、易燃易爆物品、管制刀具及其他可能威胁他人的危险物品。火车站安检处的物体识别如图 2-1 所示。为确保乘客的人身与财产安全，所有旅客均需接受安检，而在安检过程中进行必要的物体识别，也由此成为人工智能需要解决的课题。



图 2-1 火车站安检处的物体识别

一、计算机视觉对图像的理解

世间万物皆有其独特的形状与外观，这些特征通常能被人类的视觉所感知。而对于人工智能来说，计算机视觉技术的核心，就是让计算机或摄像头具备感知万物形状与外观的能力，进而完成对图像的处理。

从本质上讲，计算机视觉是一门研究如何让机器“看见”的科学，更具体地说，是用摄像机和电脑替代人眼，实现对目标的识别、跟踪与测量等功能。在人工智能领域，计算机视觉并非简单地停留在获取图像或进行裁剪、缩放、滤波等基础处理层面，而是要像人类一样去理解图像所承载的信息。

两个人过马路的场景如图 2-2 所示，人眼能轻松识别出老人、年轻人、斑马线、黑色提包、拐杖等元素，还能理解他（它）们之间的关联，如一个拎着黑色提包的年轻人正搀扶着拄拐杖的老人走过斑马线；甚至可以进一步推理，根据年轻人的小辫子推测其可能是女孩，或是从年轻人穿着的短袖判断当下或许是夏天。



图 2-2 人眼中的场景图



然而，在计算机的世界里，一切都只是数字，具体来说，是 0 到 255 之间的数字。那么，怎样才能让计算机拥有像人类一样理解图像的能力呢？

要实现计算机对图像的理解，首先得明确一点：计算机并非人类，它“观察”图像的方式与人类截然不同，更像是在进行“看图说话”。让计算机根据图片自动生成描述内容的技术，因其潜在的巨大应用价值而备受关注，如在人机交互、图像与语言理解等领域都有重要作用。这项技术既需要视觉系统识别图片中的物体，又需要语言系统对识别出的物体进行合理描述，因此面临着诸多复杂且极具挑战性的问题。

当计算机能够理解图像的语义后，再对图像进行分析以获取所需结果的技术，被称为图像处理，也叫影像处理。图像处理技术的核心内容包括图像压缩、增强、复原、匹配、描述和识别。

图像处理可分为三个层次。第一个层次是分类。分类是将图像转化为某一类别的结构化信息，即用预先设定好的类别或实例来描述图片内容。第二个层次是检测。检测聚焦于特定的物体目标，需要获取该目标的类别及其位置信息。如果说分类是对整张图片内容的概括描述，那么检测则是对图片前景与背景的深入理解，因此检测模型的输出通常是一个列表，列表中的每一项都会分别给出目标的类别和位置（坐标）。第三个层次是分割。分割是对图像进行像素级的精细描述，包括语义分割和实例分割。其中，语义分割要求将具有不同语义的图像部分区分开来；实例分割则需要精确描述目标的轮廓，相比检测框更为细致，适用于对理解程度要求较高的场景。

在完成图像处理之后，更深入的研究是图像理解。这不仅需要识别图中的物体，还需要获取一些描述性信息，对图像的某些特征进行判断和预测。之后，将这些标签和预测结果进行语义上的重组，就能形成一句完整的描述性语句。

二、计算机视觉的任务

通过计算机视觉对图像理解的过程，可以总结出计算机视觉任务的类型主要有以下几种。

（一）物体检测

物体检测作为视觉感知的起始环节，是计算机视觉领域里一个关键分支。其核心目标在于，通过框选的方式确定物体在图像中的具体位置，同时明确该物体的类别。这一技术的重点在于主动搜寻目标物体，不过有个明确的限制条件：所检测的物体必须具备固定的形状和清晰的轮廓。

（二）物体识别

物体识别，指的是判断一组图像数据中是否存在某个特定物体、特定图像特征或特定运动状态的技术。不过就目前而言，这一技术仍主要局限于高效解决特定目标的识别问题，如人脸识别、印刷或手写文件识别、车牌识别等场景均是其典型应用。



（三）图像分类

图像分类，是指判断单张图像中是否包含某种特定物体的技术。对图像进行特征描述，是物体分类研究的核心内容。从本质上来说，图像分类问题就是为输入的图像分配相应标签，这也是计算机视觉领域的核心问题之一。

（四）物体定位

物体定位，是指借助计算机视觉技术，精准确定图像中某一目标物体所处具体位置的过程。

（五）图像分割

图像分割，是为图像中的每个像素添加标签的过程。通过这一过程，带有相同标签的像素会具备某种共有的视觉特性。

分析与讨论

结合火车站安检的场景，说说计算机视觉在安检过程中需要完成哪些具体任务（如物体检测、识别等），并说明这些任务如何配合以实现危险物品的有效排查。

任务二 人脸识别

任务描述

本任务聚焦人脸识别，涵盖其技术内涵、发展历程及系统组成。要理解人脸识别基于人脸特征进行身份确认的原理，以及其非接触式、高效性等技术特点；梳理从早期基于几何特征的算法，到基于深度学习的卷积神经网络算法的发展历程，明晰技术迭代脉络；剖析人脸识别系统中数据获取、图像预处理、特征提取、匹配比对等模块的功能及协同工作机制。

任务分析

本任务在安防、金融等领域意义重大，大幅提升了身份验证的效率与便捷性。但存在人脸信息泄露的隐私安全风险，且在不同光照、姿态、表情变化下，识别稳定性有待加强，同时需在技术广泛应用与用户隐私保护、伦理规范之间找到平衡，避免技术滥用。



任务实施

一、人脸识别的技术内涵

人脸识别是依据人脸特征自动识别身份的生物识别技术，也叫面貌、面孔、面部识别等，通俗讲就是“刷脸”，是基于光学人脸图像的身份识别与验证的简称。其流程为：用摄像机或摄像头采集含人脸的图像，自动检测、跟踪、锁定人脸后，对图像进行后续操作。

例如，人脸可作钥匙开关安全锁；智能手机的“人脸识别”提升了隐私安全性。某些无人旅馆由机器人接待，接待处可刷脸，人脸也是进房的验证钥匙，如图 2-3 所示。



图 2-3 刷脸进房间

在人脸识别技术里，每张脸都具备独有的特征，这些特征体现为不同的维度。就算是看起来很相似的脸，在某些维度上也必然存在差异。由于机器只能通过解读数字来“认知”人脸，所以它需要先把特定的人脸转化成数字形式。而用来表示一张脸的、包含多维度信息的数字组合，就是特征向量，这个向量由按特定顺序排列的各类数字组成。

通过特征向量，给定一张人脸图片便可以用数值标注出不同特征，将其转化为特征向量，见表 2-1。

表 2-1 人脸特征向量表

| 脸部长度 (cm) | 脸部宽度 (cm) | 平均肤色 (RGB) | 嘴唇宽度 (cm) | 鼻子长度 (cm) |
|-----------|-----------|-----------------|-----------|-----------|
| 24.2 | 16.1 | (255, 224, 189) | 5.4 | 7.4 |

表 2-1 中的特征向量在 Python 中可以用元素的形式表示为 [24.2, 16.1, (255, 224, 189), 5.4, 7.4]。当然，这些只是人脸中的部分特征，还可以增加其他的特征，如头发的颜色、瞳孔的颜色、是否戴眼镜等。

有了特征向量，就能对图片进行解码，也就是将图片转换成数字形式，转换后得到的便是对应人脸的特征向量。当两张面部图片的特征向量极为相似，即两个特征向量数字之间的“距离”很小时，它们就可能属于同一张人脸，这正是人脸识别的原理。

机器学习在这一过程中主要完成两项工作：一是特征提取，在实际应用中，需要提取的特征值数量极多，手动罗列这些数值难度很大，而机器学习算法能够自动标注特征并提取出大量数值；二是实现匹配比对，在获取特征向量后，机器学习算法可以将图片与语料库中的特征向量进行匹配比对。

二、人脸识别的发展历程

对人脸识别的研究可以分为三个阶段。



（一）初级阶段

20 世纪 50 年代到 80 年代，人脸识别还只是被当作一般性的模式识别问题来研究。这一阶段主要围绕人脸的几何结构特征展开，研究重点集中在对剪影的分析上，不仅重要成果寥寥无几，也基本没有投入实际应用。

（二）快速发展阶段

20 世纪 90 年代，人脸识别技术进入快速发展阶段，其间出现了不少商业化运作的人脸识别系统，像知名的 Visionics（现名为 Identix）旗下的 Facelt 系统便在其中。从技术方案来看，2D 人脸图像线性子空间判别分析、统计表观模型以及统计模式识别方法，是这一阶段的主流技术。

（三）百家争鸣阶段

20 世纪 90 年代末至今，研究进一步深入，研究者更关注真实条件的人脸识别问题，主要集中在以下四个方面。

一是提出不同人脸空间模型，含线性判别分析等线性建模、Kernel 方法等非线性建模及 3D 人脸识别。

二是分析影响因素，涉及光照、姿态、表情等。

三是利用新特征表示，如局部描述和深度学习。

四是运用新数据源，如视频、素描、近红外图像的人脸识别。

目前，人脸识别技术在我国广泛应用于移动支付、交通枢纽安全、企业门禁考勤等场景。随着技术的提升，其应用将进一步扩大。未来，伴随人工智能等新一代信息技术在商业应用的提速，作为更有效的身份识别与验证手段，人脸识别技术的价值仍有巨大挖掘空间。

三、人脸识别系统组成

自动人脸识别可分为人脸检测、面部特征点定位、特征提取与分类器设计四个步骤，依据这一流程特点，便能明确人脸识别系统的组成，如图 2-4 所示。下面有针对性地对人脸识别系统中的每个部分进行分析。

（一）数据获取

每一张不同的人脸图像都可以通过摄像头、照片等渠道获取，

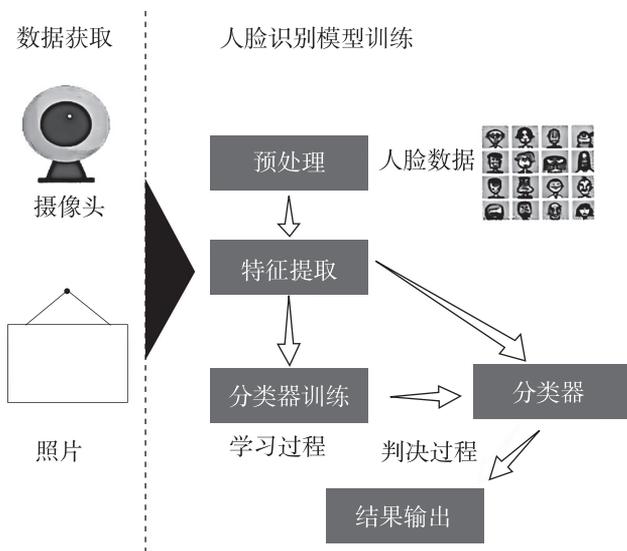


图 2-4 人脸识别系统的组成



可以获取不同的方位角度、不同的表情等。决定数据获取部分的主要因素有以下几个。

1. 图像大小

图像大小能在一定程度上体现人脸与摄像头的距离：人脸过小，说明距离过远，会对识别效果产生影响；人脸过大，则表示距离过近，会影响识别速度。一般来说，摄像头设定的最小识别人脸像素为 60×60 或 100×100 。

2. 图像分辨率

图像分辨率也是影响图像大小的因素之一，而图像大小又综合了分辨率，因此分辨率也会对摄像头的识别距离产生影响。通常，图像分辨率越低，识别难度就越大。就目前而言，7K 摄像头能看清人脸的最远距离为 20 米。

3. 光照环境

光照环境可能让图片出现过曝或过暗的问题，这会对人脸识别效果造成影响。针对这种情况，既可以借助摄像头自带的补光、滤光功能来平衡光照带来的影响，也能通过算法模型对图像光线进行优化。

4. 模糊程度

人脸相对摄像头移动时，往往会产生运动模糊，而模糊程度会对人脸识别效果造成影响。对于这一问题，既可借助部分摄像头的抗模糊功能来处理，也能通过算法模型进行优化。

5. 遮挡程度

在实际场景中，很多人脸都会被帽子、眼镜、口罩等物品遮挡，这部分数据有别于面部无遮挡、边缘清晰的图像，需要根据算法要求决定是否去除。

6. 采集角度

在很多场合中，除非是有组织地拍摄或不断调整角度，否则往往很难抓拍到人的正脸。对于算法模型而言，训练集应该包含左右侧人脸、上下侧人脸的数据。

（二）图像预处理

基于人脸检测结果对图像进行预处理，是特征提取的过程。由于获取的原始图像会受到各种外在条件限制和随机因素干扰，通常无法直接使用，因此必须在图像处理阶段对其进行灰度校正、噪声过滤等预处理操作。

一般的图像预处理可能包括如图 2-5 所示的步骤。

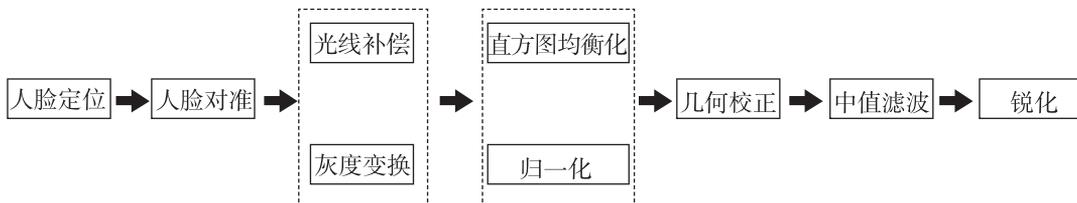


图 2-5 图像预处理的步骤

在图像预处理中，人脸定位是指机器对输入图片进行处理，找出人脸位置（如用框标



注出来)。不过，输入图片中人脸的大小并不固定，需要对原图进行多次缩放才能确定合适尺寸，这可能导致同一张人脸被多次识别，而我们只需保留最准确的结果。因此，会从这些可能的人脸定位结果中选取概率值最大的，以此实现人脸对准。

当光照条件改变时，人脸的颜色和灰度值会发生相应变化，这就需要通过行之有效的自适应亮度补偿算法来进行精确的人脸颜色补偿。

图像直方图是反映图像像素分布的统计表：横坐标代表图像像素的种类（可以是灰度的，也可以是彩色的）；纵坐标则表示每一种颜色值在图像中的像素总数，或占有所有像素个数的百分比。由于图像由像素构成，所以能反映像素分布的直方图，往往是图像中一个重要的特征。

图像灰度直方图是指一张数字图像的灰度级范围为 0~255，其直方图定义为：

$$h(g_k)=n_k \quad (2-1)$$

其中， g_k 代表第 k 个灰度级（如 255）， n_k 则是该灰度级的像素个数。

归一化直方图的定义是：第 k 个灰度级的像素数量与所有灰度级像素数量的总和之比，也就是该灰度级出现的概率。

当输入一张图像并将其转化为直方图时，效果如图 2-6 所示。

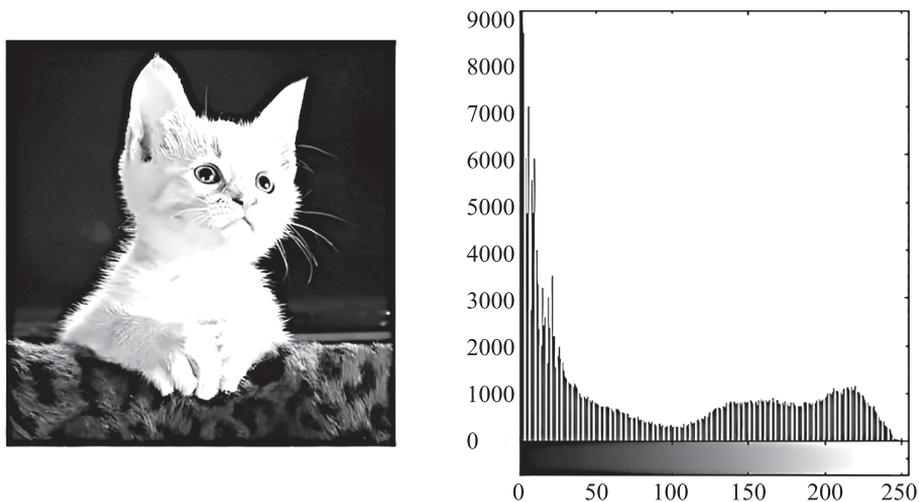


图 2-6 图像及其直方图效果

几何校正、中值滤波和锐化都是针对图片的平滑操作，以消除噪声。

（三）人脸识别的主要算法

在人脸识别领域，描述人脸特征的数据量较为庞大。为提取人脸的主要特征数据、提升识别系统的运行效率，对特征数据进行降维是必不可少的操作，而这里所讨论的人脸识别主要算法，正是针对人脸特征数据降维的。

1. 特征脸

把训练集中每个人的脸拉成一列，再组合成大矩阵 A 。若人脸图像为 $m \times m$ 大小，大



矩阵 A 的维度便是 $m \times m \times N$ (N 为训练集中人脸数量)。先对 N 个人脸在对应维度求平均, 得到“平均脸”; 再用大矩阵 A 中的 N 张图像分别减去“平均脸”, 得到新矩阵 B ; 随后计算新矩阵 B 的协方差矩阵, 进而求出该协方差矩阵的特征值和特征向量, 这些特征向量即“特征脸”。

其实这种算法就是一种从主成分分析中导出的人脸识别和描述技术, 叫主成分分析 (Principal Component Analysis, PCA) 算法。PCA 算法决定了特征脸方法的主要思路是将输入的人脸图像看作一个个矩阵, 通过在人脸空间中建立一组正交向量, 并选择最重要的正交向量作为“主成分”来描述原来的人脸空间。

2. 渔夫脸

Belhumer 将 Fisher 判别准则成功应用于人脸分类, 提出了基于线性判别分析的 Fisherface (渔夫脸) 方法。该方法旨在找到一种特征组合方式, 以实现最大的类间离散度和最小的类内离散度, 其核心思路很简单: 在低维表示下, 让相同类别紧密聚集、不同类别尽量远离。

具体而言, Fisherface 方法先通过 PCA 算法对图像表观特征进行降维, 再采用线性判别分析 (Linear Discriminant Analysis, LDA) 对降维后的主成分进行变换, 从而达到“类间离散度尽量大、类内离散度尽量小”的目标。线性判别分析的要求是将数据投影到低维度上, 投影后需使同一类别数据的投影点尽可能接近, 不同类别数据的类别中心之间距离尽可能大。

3. 拉普拉斯脸

拉普拉斯脸的核心是保持人脸空间的局部几何结构, 它通过对人脸空间进行描述以及对人脸图像进行降维, 能够在光照、角度、表情等非理想条件多变的情况下, 更准确地完成人脸的识别与预测。

拉普拉斯脸可以将人脸图像数据映射到二维空间中, 在这个空间里, 人脸图像的角度和表情呈现连续变化的状态。由于它考虑了人脸流形的局部几何结构, 因此降维结果能很好地保留人脸图像数据中原本的自然聚类。从本质上来说, 拉普拉斯脸是对人脸流形上拉普拉斯—贝尔特拉米算子特征函数的最优线性近似。

(四) 人脸识别的主要特征点

人脸识别算法的各类理论, 核心都是围绕人脸特征数据的降维展开。具体到人脸检测环节, 主要是通过识别能体现人脸主要特征的特征点, 再依据这些特征点完成人脸的对齐校准。2014 年, Vahid Kazemi 与 Josephine Sullivan 提出了一项方法: 从人脸的关键部位选取 68 个固定位置的特征点, 只需对系统进行相应训练, 就能在任意人脸上精准定位这 68 个点。在不少人脸识别框架中, 这 68 个特征点也被称为 Landmarks, 其具体分布如图 2-7 所示。

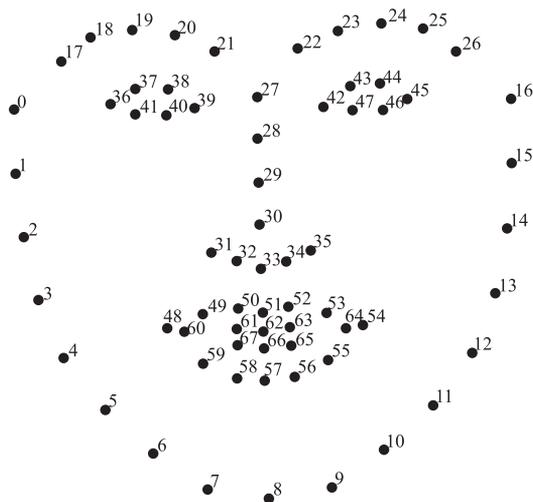


图 2-7 脸部的 68 个特征点

有了这 68 个特征点，就可以对人脸进行校正了，主要是通过仿射变换将原来比较歪的脸摆正，尽量消除误差。这里的仿射变换主要还是进行一些旋转、放大、缩小或轻微的变形，而不是夸张的扭曲。

（五）人脸检测和人脸识别的技术指标

在人脸检测过程中，存在一些关键性指标，具体如下。

1. 检测率

检测率指的是识别正确的人脸数量与图中所有人脸数量的比值。例如，摄像头抓拍了 100 张包含人脸的图像，系统识别出 90 张人脸数据，其中识别正确的有 85 张，其余为干扰项，此时检测率便基于这 85 张正确识别的人脸与图中所有 100 张人脸计算得出。

2. 误检率

误检率是识别错误的人脸数量与识别出来的人脸数量的比值。例如，摄像头抓拍 100 张人脸，系统识别出 90 张，但其中有 15 张是将树木、草植或路标等误识为人脸，误检率就由这 15 张错误识别的人脸与 90 张识别出的人脸计算得出。

3. 漏检率

漏检率为未识别出来的人脸数量与图中所有人脸数量的比值。例如，摄像头抓拍的 100 张图像中，实际有 95 张包含人脸数据，系统识别出 90 张，有 5 张未被识别，漏检率即根据这 5 张未识别的人脸与 95 张实际人脸计算得出。

4. 速度

速度是从图像采集完成到人脸检测完成所花费的时间。完成人脸检测，识别出包含人脸的图像后，还涉及人脸识别中的两个关键指标：精确率和召回率。以 100 张样本图片为例，其中共有 60 张正样本，在相似度为 0.9 的情况下，系统识别出 20 张图片，其中 19



张为正样本。精确率的计算公式为识别为正确的样本数除以识别出来的样本数，即 $19/20$ 。召回率的计算公式为识别为正确的样本数除以所有样本中正确的数量，即 $19/60$ 。虽然 0.9 阈值下的精确率高达 $19/20$ ，但正确输出的数量仅占有所有正样本的 $19/60$ ，这就很容易出现漏识情况。

分析与讨论

在实际生活中，人脸识别技术可能面临哪些潜在的问题或挑战？

任务三 语音识别

任务描述

本任务围绕语音识别，包含对音频的认知和音频信号识别过程。需了解音频的频率、振幅等物理特性，以及梅尔频率倒谱系数等声学特征，明确这些特征与语音内容的关联；探究音频信号从采集、预处理、特征提取到通过隐马尔可夫模型、深度学习模型等识别转化为文本的完整过程及各环节关键技术要点。

任务分析

语音识别推动了智能交互方式的革新，广泛应用于智能音箱语音控制、手机语音助手输入、客服系统自动应答等场景，提升了人机交互效率。但受方言、口音、背景噪声等因素影响较大，导致识别准确率下降，且多语种识别时，不同语言的语音特性差异对算法适配性要求高，是当前技术需突破的难点，未来技术优化空间较大。

任务实施

一、对音频的认知

人们所听到的声音一般具有音调、响度和音色三个属性。音调指的是声音的高低，由频率这一特征值决定，频率是每秒经过某一给定点的声波数量，其测量单位为赫兹，且频率越高，音调就越高。响度是人们主观上感觉到的声音大小，也就是通常所说的音量，它由振幅和人离声源的距离决定，振幅越大，人与声源的距离越近，响度就越大。音色又称音品，声音会因发声物体材料的不同特性而呈现出不同特征。虽然音色本身是抽象的，但可以通过波形直观地表现出来，不同的音色对应着不同的波形，如图 2-8 所示，钢琴和长

笛便具有不同的音色。



图 2-8 钢琴和长笛对应不同的波形

与机器进行语音交流，让机器明白你在说什么，这是机器的听觉系统的主要任务。语音识别系统也是这样的原理，就是让机器通过识别和理解过程把语音信号转变为相应的文本或命令。一个完整的语音识别系统通常包括信息处理和特征提取、声学模型、语言模型和解码搜索四个模块，如图 2-9 所示。

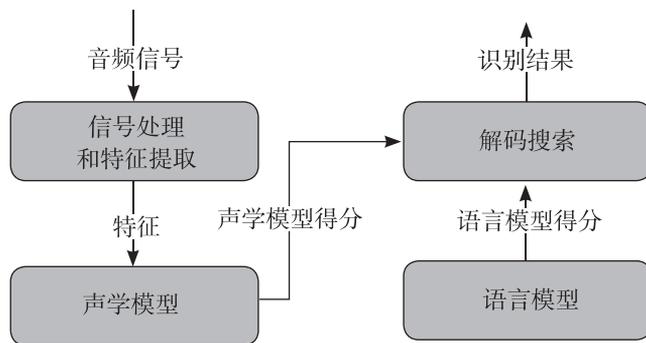


图 2-9 语音识别系统的构成

信号处理和特征提取可看作音频数据的预处理环节，由于实际使用的语音片段往往带有一定噪声，因此在正式进入声学模型前，需通过消除噪声、信道增强等预处理技术，先将信号从时域转换到频域，再为后续的声学模型提取有效的特征向量。语言模型借助自然语言处理得出语言模型得分，最后在解码搜索阶段，会综合声学模型得分与语言模型得分，将得分最高的词序列作为最终识别结果，这就是语音识别的一般原理。

二、音频信号识别过程

任何自动语音识别系统的首要步骤都是提取特征，而梅尔频率倒谱系数（Mel-Frequency Cepstral Coefficients, MFCCs）便是在自动语音识别与说话者识别中得到广泛应用的特征。

（一）梅尔频率倒谱技术

梅尔频率倒谱是一种语音特征提取技术，其目的是从人发出的音频中排除噪声和情感的干扰，提取出特征值以便进行进一步分析。例如，在车辆行驶时，人与人的交谈会受到音乐声、空调声等环境因素的影响，进而降低语音识别的准确率，而梅尔频率倒谱技术就能应对这类问题。



要理解梅尔频率倒谱的原理，首先需要从语音信号的描述说起。对于语音信号而言，如何描述它十分重要，不同的描述方式能反映出不同的信息。通常，波形图（也叫声谱图）的运用能更好地帮助人们理解语音信号。

在对声谱图的研究中，峰值部分代表着语音的主要频率成分，这些峰值被称为共振峰。共振峰往往承载着声音的辨识属性，如同每个人的身份证标识一般，具有特殊的重要性，通过这些峰值可以识别不同的声音，具体如图 2-10 所示。

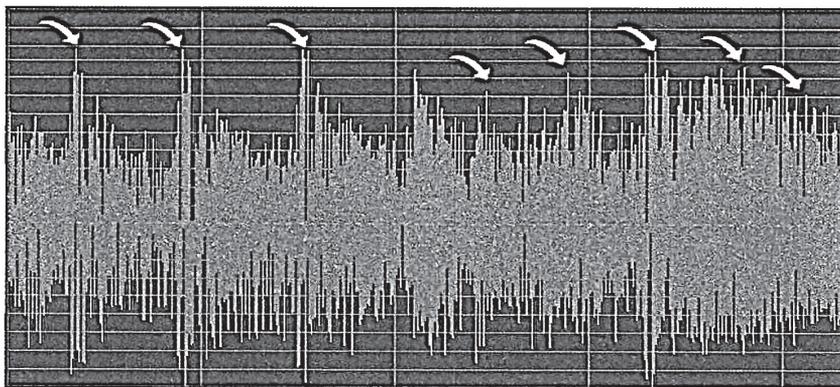


图 2-10 声谱图峰值表示

图 2-10 中所示的波峰并不是一成不变的，重要的波峰需要提取出来的不仅仅是共振峰的位置，还需要提取其转变过程。这些共振峰点形成的平滑曲线可以更好地体现其转变过程，这条平滑曲线称为包络，如图 2-11 所示。

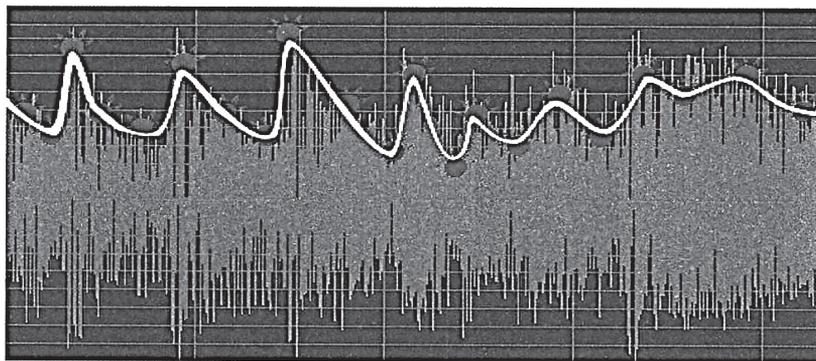


图 2-11 声谱图的包络曲线

由图 2-11 可知，这样的包络曲线其实也忽略了某些细节，所以将原始的声谱分为两部分：包络和频谱细节。可以把 $X[k]$ 比作某一部分的频谱， $H[k]$ 代表某一部分的包络， $E[k]$ 代表这部分包络的频谱细节。 $H[k]$ 与 $E[k]$ 的乘积可以得到 $X[k]$ ，两边取对数，就可以把乘法变成加法的问题。所得公式如下：

$$\log X[k] = \log H[k] + \log E[k] \quad (2-2)$$

依据公式 (2-2)，包络的提取成为解决问题的关键所在。由于包络呈现为一条平滑曲线，这一问题属于线性变化范畴，而线性变化问题可借助傅里叶变换来处理。傅里叶变换



作为一种线性积分变换，其核心思想是“任何连续周期信号都能由一组适宜的正弦曲线组合而成”。原始声谱虽类似三角形，看起来正弦曲线难以组合出带有棱角的信号，但实际上，利用正弦曲线能够对其进行极为精准的逼近，这种逼近程度可达到两种表示方法在能量上毫无差别的地步。不同幅度的正弦波能够无限逼近三角形，如图 2-12 所示。

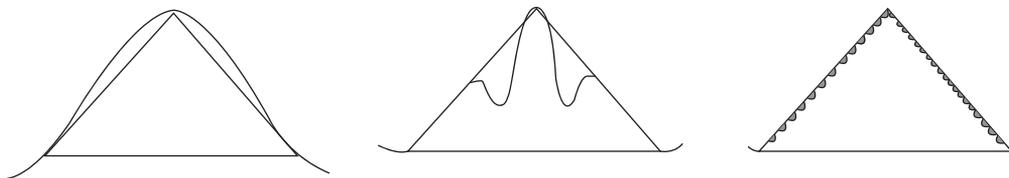


图 2-12 不同幅度的正弦波无限逼近三角形

由图 2-12 可知，随着正弦波数量逐渐增多，它们最终会叠加成一个标准的三角形。任何波形都可以用正弦波叠加起来，前提是正弦波要足够多。

傅里叶变换的公式如下：

$$F(\omega) = F[f(t)] = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \quad (2-3)$$

其中， $F(\omega)$ 叫作 $f(t)$ 的象函数， $f(t)$ 叫作 $F(\omega)$ 的象原函数。 $F(\omega)$ 是 $f(t)$ 的象， $f(t)$ 是 $F(\omega)$ 的原象。

借助傅里叶变换，我们能从一段语音的声谱图中得到其频谱包络。不过，心理学研究发现，人类对声音音调的感知并非呈线性关系，人类的听觉感知只会聚焦于某些特定区域，而非覆盖整个频谱包络。考虑到人类听觉感知实验中存在仅关注特定频率分量这一特性，研究者依据相关心理学实验，设计出了一组功能类似耳蜗的滤波器组，即梅尔滤波器组。这种滤波器组能够允许特定频率的信号通过，同时直接过滤掉那些无须感知的频率信号。将普通频率转换到梅尔频率的公式如下：

$$\text{Mel}(f) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \quad (2-4)$$

公式 (2-4) 对应的图像如图 2-13 所示。

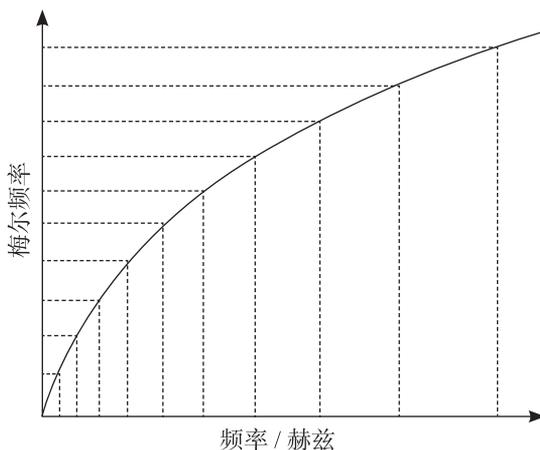


图 2-13 Mel 频率的图像表示

由图 2-13 可以看到，不统一的频率可以转化为统一的频率，也就是统一的滤波器组。

将线性频谱映射到基于听觉感知的梅尔非线性频谱，再转换到倒谱上，所得到的就是梅尔频率倒谱系数。这里所说的倒谱，是指对信号的傅里叶变换结果进行对数运算后，再进行逆傅里叶变换所得到的谱，其具体计算过程如图 2-14 所示。

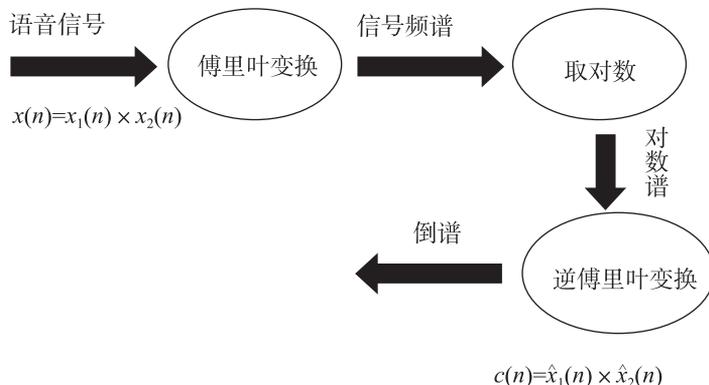


图 2-14 语音信号经傅里叶变换到倒谱的过程

梅尔频率倒谱系数的特征提取过程是先对语音进行预加重、分帧和加窗，再对每个短时分析窗通过快速傅里叶变换（Fast Fourier Transform, FFT）得到对应频谱，接着将该频谱通过梅尔滤波器组得到梅尔频谱，最后在梅尔频谱上进行倒谱分析以获得梅尔频率倒谱系数。这一过程概括表示如图 2-15 所示。

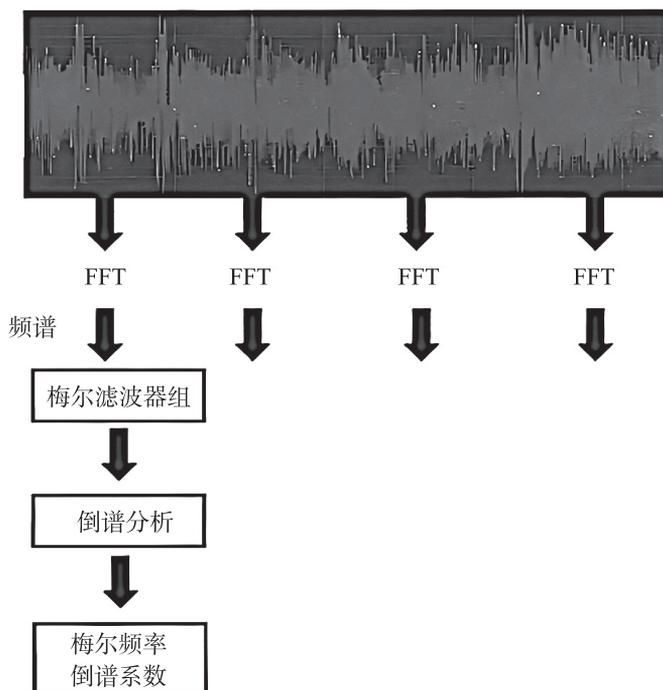


图 2-15 梅尔频率倒谱系数的计算过程

（二）隐马尔可夫模型

隐马尔可夫模型（Hidden Markov Model, HMM）是一种统计模型，广泛应用在语音