第1章

人工智能引论

人工智能是计算机学科的一个分支，旨在让机器模拟人类的智能行为，如学习、推理和决策。AI政务平台实现“一网通办”“最多跑一次”，彰显社会主义制度“执政为民”的本质特征；医疗领域通过AI辅助诊疗提高诊疗效率与准确率；工作与学习中的AI大模型应用极大地提升了效率；工业中的自动检测，农业中的病虫害、成熟度智能识别，都是AI的典型应用。

1.1 从人工智能的多场景应用到AIGC

工业质检：富士康公司引入AI视觉检测系统，通过高清摄像头捕捉产品图像，快速识别电路板上的细小裂痕或焊接不良，使得漏检率从5%降至0.1%，质检效率提升10倍。

农业质检：大型农场利用AI图像识别技术分析作物叶片，准确识别病虫害并给出防治建议，使得农药使用量减少20%，作物产量提升15%。

智能交通信号控制：杭州“城市大脑”系统通过实时监测交通流量，动态调整红绿灯时长。在试点区域，车辆等待时间减少30%，拥堵指数下降15%。

艺术与文案创作：如图1.1所示(见彩页)，你可以用AI工具生成一幅梵高风格的画作，或者让AI写一篇短篇故事。

图1.1 AI生成的艺术作品示例

近年来，大语言模型(Large Language Model，LLM，简称大模型)的快速发展极大地推动了AIGC的普及与应用。最初，大模型(如BERT)通过预训练和微调显著提升了自然语言处理任务的性能，奠定了现代大模型的基础。随后，GPT系列模型的推出将语言生成能

力推向新高度，其能够生成流畅且上下文相关的文本，广泛应用于对话系统和内容创作。之后，扩散模型的出现进一步拓展了 AIGC 的边界，使 AI 生成高质量图像和视频成为可能，例如 DALL-E 生成逼真的艺术作品。模型规模的扩大也带来了显著的性能提升，例如 ChatGPT 的问世展示了大模型在多任务处理和用户交互中的强大能力。与此同时，多模态模型的兴起使得 AI 能够同时处理文本、图像和音频，极大地丰富了生成内容的多样性。此外，开源生态的繁荣降低了技术门槛，研究者和开发者能够基于开源模型进行创新，加速了 AIGC 工具的普及。高效训练技术（如低秩适配（LoRA））进一步降低了模型开发和部署的成本，使中小型团队也能参与 AIGC 的开发。

从 AI 到 AIGC 的转变，体现了技术从“辅助决策”到“自主创作”的飞跃。AIGC 让普通人也能轻松参与内容创作，不需要专业技能即可生成高质量作品。这种能力正在改变人们的学习、工作和生活方式，为大学生提供了更多创新的可能性。人工智能已深入日常生活和各行各业，其生成内容的能力为个人和社会带来了前所未有的便利与创新。AIGC 的应用不仅限于技术领域，还广泛渗透到教育、职场和工业生产等场景中，为非专业人士提供了强大的工具。

1.2 人工智能助力校园学习、职场工作与工业生产

本节将从大学生活、未来职场和工业生产三个方面，详细探讨 AI 如何通过多样化的工具赋能学习、工作和生产，展示其对非专业人士的深远影响。

1.2.1 大学生活

AI 技术正在重新定义大学生的学习和生活体验，为非计算机专业的学生提供了高效、便捷的工具。智能学习平台通过分析学生的学习习惯和进度，生成个性化的学习路径。例如，Duolingo 利用 AI 算法根据用户的语言水平、学习速度和常见错误，动态调整练习内容，帮助学生更快掌握外语听说读写技能。类似地，Coursera 和 Khan Academy 的 AI 推荐系统能够根据学生的兴趣和薄弱点，推送定制化的课程和练习题，提升学习效率。这些工具让学生无须深入了解 AI 技术就能从中受益。

在学术写作方面，AI 工具显著降低了写作门槛。Grammarly 通过自然语言处理技术，实时检测学生的论文、邮件或社交媒体文本中的语法、拼写和风格问题，并提供改进建议，如图 1.2 所示。对于需要快速构思的场景，Jasper.ai 和 Writesonic 等 AIGC 工具可以生成论文大纲、创意写作初稿甚至诗歌，帮助学生克服“空白页恐惧”。例如，学生只需输入主题关键词，Jasper.ai 就能生成结构清晰的段落，供学生进一步修改和完善。这些工具不仅节

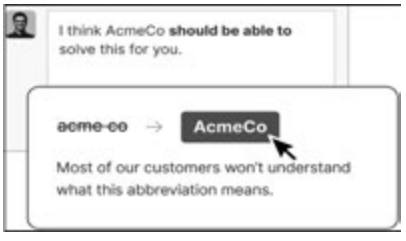

图 1.2 Grammarly 语法检查演示

省时间，还培养了学生的批判性思维，因为他们需要评估和优化 AI 生成的内容。

AI 还在学习辅助和考试准备中发挥重要作用。Quizlet 利用 AI 生成智能化的闪卡(flashcard)和模拟测试，根据学生的答题表现调整题目难度，帮助他们更高效地复习。Socratic(由 Google 开发)则允许学生拍照上传数学或科学题目，AI 会自动解析问题并提供详细的解题步骤，特别适合非理工科学生应对复杂课程。此外，AI 驱动的语音识别工具(如 Otter.ai)可以实时转录课堂讲座，生成可搜索的笔记，让学生更专注于课堂内容而无须手忙脚乱地记录。

在校园生活中，AI 聊天机器人提供了全天候的便捷服务。例如，xAI 开发的 Grok 能够回答关于课程安排、校园活动或图书馆资源的问题，减轻学生在信息查询上的负担。一些大学还引入了 AI 驱动的心理健康支持工具(如 Woebot)，通过对话分析学生的压力水平并提供情绪管理建议。这些工具以直观的方式融入学生的日常生活，让非专业人士也能轻松使用 AI，优化学习和社交体验。

1.2.2 未来职场

AI 正在重塑职场生态，成为提升工作效率和创造力的重要伙伴，尤其对非技术背景的职场新人而言，AI 工具显著降低了专业技能的门槛。在内容创作领域，Canva 利用 AIGC 技术，让用户可以通过简单的文字指令生成专业的海报、演示文稿和社交媒体图片。例如，市场营销人员只须输入品牌关键词和风格偏好，Canva 就能生成多种设计方案，省去烦琐的设计过程。同样，Copy.ai 和 Writesonic 能够快速生成营销文案、广告标语、产品描述甚至长篇博客文章，适合需要频繁产出内容的职业人士。这些工具通过直观的界面和预训练模型，让非专业用户也能创作出高质量内容。

在编程和数据分析领域，AI 工具为非技术人员打开了新世界的大门。GitHub Copilot 通过自然语言处理和代码生成技术，为程序员和非程序员提供代码补全和调试建议。例如，市场分析师可以用 Copilot 生成 Python 脚本，快速完成数据可视化任务，而无须深入学习编程语言。类似地，Tableau 的 AI 功能允许用户通过自然语言查询生成复杂的图表和仪表盘。例如输入“显示过去一年的销售额趋势”，即可自动生成折线图。这些工具让数据分析变得更加民主化，非技术背景的职场人士也能参与数据驱动的决策。

AI 也在招聘和培训中发挥了关键作用。Mya Systems 的智能聊天机器人通过分析简历和面试对话，自动筛选候选人，减少人为偏见并提高招聘效率。HireVue 的 AI 评估工具则通过分析求职者的语音和面部表情，提供更客观的面试评分，特别适用于大规模招聘场景。在员工培训方面，Articulate 360 利用 AI 生成互动式在线学习模块，模拟真实工作场景，帮助新员工快速掌握技能。例如，零售行业的员工可以通过与 AI 模拟的客户互动进行练习，学习如何处理投诉。Docebo 等学习管理系统则通过 AI 分析员工的学习进度，推荐个性化的培训内容，提升培训效果。

AI 工具还在项目管理和协作中展现了潜力。ClickUp 和 Asana 的 AI 功能可以自动生成任务优先级建议、会议摘要和进度报告，减轻团队的管理负担。Notion AI 则能将零散的笔记整理成结构化的文档，或根据用户需求生成会议议程。这些工具通过将烦琐任务自动化，让职场人士有更多时间专注于创意性和战略性工作，同时增强了非技术人员的竞争力。

1.2.3 工业生产

在工业生产领域，AI 工具通过优化设计、制造和维护流程，显著提高了效率和可持续性。Autodesk 的生成设计工具利用 AIGC 技术，根据工程师输入的材料、成本和性能参数，可以生成数百种产品原型。例如，汽车设计师可以通过该工具快速生成轻量化车身结构方案，缩短研发周期并降低成本。类似地，Runway ML 可以帮助工业设计师生成 3D 模型概念图或产品渲染图，为早期设计阶段提供灵感。这些工具直观易用，即使是非专业用户也能参与初步设计工作。

在生产线上，AI 视觉系统提升了质量控制的精度。Cognex 的 Deep Learning Vision 系统通过分析产品图像，检测微小的表面缺陷或装配错误，适用于电子产品和汽车零部件的制造。相比传统的人工检测，AI 系统不仅检测速度更快，还能持续学习以适应新产品。此外，IBM Maximo 的 AI 工具通过分析生产线数据，优化生产调度和资源分配，减少浪费。MindSphere（由 Siemens 公司开发）则利用 AI 生成工厂的数字孪生模型，实时模拟生产流程，帮助管理者优化布局和能耗。

预测性维护是 AI 在工业生产中的另一大应用。Uptake 和 GE Predix 通过分析设备传感器数据，预测机械故障的可能性并建议维护时机。例如，风力发电机的 AI 系统可以根据振动和温度数据，提前预警轴承磨损，延长设备寿命并减少停机损失。这些工具通过直观的仪表盘呈现分析结果，让非技术背景的工厂管理者也能轻松理解和操作。此外，AI 还在供应链管理中发挥作用，例如 Blue Yonder 的 AI 平台通过预测市场需求，优化库存和物流路线，降低运营成本。

AI 工具的应用让工业生产更加智能化和可持续。它们不仅提高了效率，还降低了技术门槛，使非专业人士也能参与到产品设计、质量控制和运营优化的流程中。AI 技术正在推动全球制造业向更高效、更环保的方向发展。

1.3 人工智能面对的伦理道德挑战

随着人工智能技术的迅速普及，尤其是 AIGC 的广泛应用，AI 伦理问题已成为社会关注的焦点。AI 不仅改变了我们的学习、工作和生活方式，还带来了复杂的道德和社会挑战。非计算机专业的大学生作为 AI 工具的直接用户，需要了解这些伦理问题，以负责任地使用技术并参与相关讨论。本节将探讨 AI 伦理的核心议题，包括数据隐私、算法偏见、内容真实性、劳动替代及社会影响，分析其对个人和社会的意义，并提出应对策略。

1.3.1 数据隐私与安全

AI 的强大功能依赖于海量数据，但数据收集和使用引发了严重的隐私担忧。许多 AIGC 工具，如语言模型和图像生成模型，需要从用户输入或公开数据中学习。例如，学生使用 AI 写作工具时，上传的文本可能被用于进一步训练模型。这些数据如果未经妥善保护，就可能被滥用或泄露，导致个人隐私受损。此外，AI 在教育和职场中的应用，如学习分析系统和招聘算法，常常涉及敏感信息，包括学生的成绩或求职者的个人信息。未经明确同意的数据收集可能违背用户的知情权。

为应对这一问题，数据隐私法规如欧盟颁布的《通用数据保护条例》(GDPR)要求企业明确告知用户数据使用目的，并提供删除数据的权利。大学生在使用 AI 工具时，应选择有透明隐私政策的平台，并避免上传敏感信息。例如，使用 Grok 等工具时，可以检查其隐私声明，了解数据如何被处理。培养数据隐私意识，不仅能保护个人权益，还能推动更负责任的 AI 开发。

1.3.2 算法偏见与公平性

AI 算法的决策可能因训练数据的偏差而产生不公平结果。例如，招聘类 AI 工具可能因历史数据中性别或种族的偏见，导致倾向于推荐某一类候选人，强化社会不平等。在 AIGC 领域，图像生成模型(如 Stable Diffusion)可能因训练数据以某些文化为主，导致生成的艺术作品缺乏多样性，或对特定群体产生刻板印象。学生在使用 AIGC 工具生成内容时，可能无意中传播这些偏见，例如生成带有文化误解的文本或图像。

解决算法偏见需要从数据和算法设计入手。开发者可以通过多元化训练数据和定期审计模型输出等方法减少偏见。大学生作为 AI 工具的用户，可以通过批判性思维评估生成内容的公平性，例如检查 AI 生成的文本是否包含歧视性语言。此外，支持公平 AI 的倡议，如要求算法透明度和公开测试结果，也有助于推动 AI 技术向更包容的方向发展。

1.3.3 内容真实性与版权

AIGC 的兴起模糊了真实与虚假的界限，带来了内容真实性和版权的伦理挑战。AI 生成的文本、图像或视频可能被用于传播虚假信息，例如深度伪造(deepfake)的视频或图片可能会误导公众，如图 1.3 所示。在学术环境中，学生过度依赖 AI 写作工具可能导致抄袭或学术不端，因为 AI 生成的内容可能基于现有作品，却难以追溯来源。此外，AIGC 作品的版权归属问题尚未明确，例如，AI 生成的艺术作品的版权应属于用户、开发者还是训练数据的原始作者？

图 1.3 AI 生成的虚假新闻图片

为应对这些挑战，技术公司正在开发内容溯源工具，例如为 AI 生成的内容添加数字水印，以便区分真伪。学术机构也制定了 AI 使用指南，鼓励学生在使用 AI 工具(如 ChatGPT)时注明出处，并将 AI 视为辅助而非替代。大学生在使用 AIGC 工具时，应确保生成内容的原创性，并在学术或创意作品中明确标注 AI 的贡献。支持清晰的版权法律框架，也有助于保护创作者的权益。

1.3.4 幻觉问题

AIGC 大模型在生成内容时可能会生成“无中生有”的结果，即所谓“幻觉问题”，这主要是由于训练数据、模型架构与训练目标的三重缺陷。首先，训练数据可能包含噪声、错误信息或过时内容，例如医学数据中未及时更新最新研究成果，导致模型复现错误结论；其次，模型架构与训练目标的局限性增加幻觉风险，概率驱动的解码策略可能选择概率高但错误的内容，例如虚构学术文献。幻觉问题的另一个关键原因是上下文理解能力不足。

应对大模型幻觉问题，需要从技术、模型优化与流程管控三个维度切入。在技术层面，可以通过数据清洗与增强提升输入质量，包括：剔除训练数据中的噪声和错误信息，引入多模态数据；采用检索增强生成(RAG)技术，在生成前强制调用权威数据库进行验证。在模型架构优化方面，可以通过惩罚预测不确定性高的词序列来降低幻觉出现的概率。在流程管控层面，需要建立全链条监控体系，在金融、法律等高风险领域，强制模型标注回答的置信度与数据来源。此外，跨行业协作构建可信数据生态至关重要。

1.3.5 劳动替代与社会影响

AI 的自动化能力引发了关于劳动替代的担忧。AIGC 工具在写作、设计和编程等领域的应用，可能减少对某些岗位的需求，如初级文案或平面设计师。对于大学生而言，这意味着未来的职场竞争可能更加激烈，需要掌握与 AI 协作的技能。然而，AI 也创造了新的职业机会，如 AI 模型训练师、数据标注员和伦理审核员，凸显了技术与人类合作的潜力。

应对劳动替代的关键在于教育和技能转型。大学生应主动学习 AI 相关的基础知识，例如如何使用 AIGC 工具优化工作流程，同时培养 AI 难以替代的软技能，如创造力、情感沟通和伦理判断。在政策层面，政府与企业可以通过再培训项目和终身教育计划，帮助劳动者适应 AI 驱动的职场变化。

1.3.6 社会与文化影响

AI 的普及还对社会和文化产生了深远影响。AIGC 工具使内容创作民主化，让非专业人士也能生成高质量的艺术作品。这种趋势丰富了文化表达，但也可能导致内容泛滥，降低原创作品的价值。此外，AI 生成的内容可能放大某些文化视角，削弱边缘群体的声音。例如，基于西方数据训练的模型可能在生成内容时忽视非西方文化背景。

为确保 AI 促进文化多样性，开发者需要使用更具代表性的训练数据，并邀请不同文化背景的用户参与模型测试。大学生在使用 AIGC 工具时，可以有意识地探索多元化的创作主题，例如用 Midjourney 生成反映本地文化的艺术作品。参与 AI 伦理的公共讨论，例如倡导透明的算法设计，也有助于塑造更具包容性的技术生态。

1.3.7 面向未来的 AI 伦理教育

面对 AI 伦理的复杂性，非计算机专业的大学生需要培养伦理意识和负责任的使用习惯。学校可以通过通识课程介绍 AI 伦理的基本概念，例如隐私保护和算法公平，帮助学生理解技术的双面性。学生自身也可以通过实践加深理解，例如在使用 AI 工具时记录其优缺点，或参与开源社区的 AI 伦理项目。此外，关注 AI 伦理的最新动态，如国际组织制定的

AI治理原则，有助于学生站在更广阔的视角看待技术的影响。

总之，AI伦理不仅是技术问题，也是社会和文化的共同责任。数据隐私、算法偏见、内容真实性、劳动替代和社会影响等议题，提醒我们需要在技术进步与伦理约束之间找到平衡。非计算机专业的大学生作为AI工具的活跃用户，通过理解和实践伦理原则，可以成为负责任的技术参与者，为构建公平、透明的AI未来贡献力量。

1.4 习题

一、单项选择题

1. 人工智能的概念最早在何时被正式提出？（　　）

A. 1940年代　　B. 1956年的达特茅斯会议

C. 1980年代　　D. 2000年代

2. AIGC的主要技术不包括以下哪项？（　　）

A. 生成对抗网络(GAN)　　B. 大语言模型(LLM)

C. 扩散模型(Diffusion Models)　　D. 传统数据库管理系统

3. 以下哪项是AI伦理中的核心问题之一？（　　）

A. 提高计算速度　　B. 数据隐私与安全

C. 优化图像分辨率　　D. 增加模型参数

二、简答题

1. 简述从人工智能(AI)到人工智能生成内容(AIGC)的技术演进过程，并说明生成对抗网络(GAN)在AIGC中的作用。提示：参考1.1节，结合AI历史和GAN的生成机制。

2. 在大学生活中，AI工具如何帮助非计算机专业的学生提升学习效率？举出至少两种具体工具并说明其功能。提示：参考1.2.1节，提及工具如Duolingo、Grammarly等。

3. 解释AI算法偏见可能导致的社会问题，并提出大学生在使用AIGC工具时可以采取的应对措施。提示：参考1.4节，关注算法偏见和公平性问题。

三、论述题

1. 结合1.3节内容，分析AI技术发展对未来职场的影响，讨论大学生应如何准备以适应AI驱动的职业环境。提示：考虑AI的自动化能力和新职业机会，强调技能学习和软技能。

2. 从AI伦理的角度，探讨AIGC工具在内容真实性和版权问题上的挑战，并提出至少两种解决方案。提示：参考1.4节，涉及深度伪造、学术诚信和法律框架。

四、实践题

1. 使用任意AIGC工具(如Jasper.ai、Canva或Midjourney)生成一篇短文或一幅图像，主题为“未来的大学生活”。记录生成过程，分析生成内容的优点和局限性，并讨论其是否可能涉及伦理问题(如偏见或版权)。提示：参考1.2和1.4节，注意工具使用和伦理评估。

2. 查找一篇关于AI伦理的新闻报道(例如涉及隐私或算法偏见的案例)，总结其核心问题，并结合1.4节内容，提出你认为合理的解决方案。提示：关注现实案例，结合数据隐私或公平性等议题。

第2章

人工智能通识基础

2.1 计算机视觉基础

AI视觉系统通过实时分析百万路摄像头数据，构建城市级公共安全神经网络。例如海康威视公司的“城市之心”平台，在杭州G20峰会期间实现0.3秒异常事件识别响应。技术应用需要严格遵循《个人信息保护法》第26条，采用差分隐私技术对人脸特征进行脱敏处理。

通过搭建包含伦理约束的YOLOv5目标检测模型，可以在代码中嵌入“最小必要”数据采集原则，践行“科技向善”的价值观，理解“保护公民合法权益”与“维护国家安全”的辩证统一，强化社会主义法治观念。

计算机视觉(Computer Vision，CV)是人工智能的一个核心分支，旨在使计算机能够像人类一样“看懂”并理解视觉信息，例如图像和视频，其目标是通过算法和模型处理、分析和解释视觉数据，赋予机器识别物体、场景或行为的能力。计算机视觉广泛应用于人脸识别、自动驾驶、医学影像分析等领域，对非计算机专业的大学生而言，掌握其基本理论有助于理解AI工具的原理和应用场景。

1. 图像表示与处理

计算机视觉的基础是数字图像的表示与处理。图像由像素(pixel)组成，每像素包含颜色或灰度信息。灰度图像的像素值通常在0(黑)到255(白)之间，而彩色图像则由多个通道组成，最常见的是红、绿、蓝(RGB)通道，每个通道的像素值表示该颜色的强度。例如，一张分辨率为1920×1080的RGB图像可以表示为一个1920×1080×3的矩阵。此外，HSV(色调、饱和度、亮度)等颜色空间常用于简化特定任务，如目标分割或颜色检测。

图像处理涉及对图像进行变换和分析，以提取有用信息。常见的图像处理操作有以下几种。

- 滤波：通过卷积操作平滑图像或检测边缘。例如，高斯滤波用于降噪，Sobel算子用于边缘检测。
- 变换：傅里叶变换用于将图像从空间域转换到频域，便于分析周期性特征。
- 增强：直方图均衡化调整图像的对比度，使像素值分布更均匀，提升图像质量。这些基础操作不仅是计算机视觉任务的预处理步骤，也是理解高级算法的起点。

2. 特征提取

在深度学习兴起之前，计算机视觉依赖手工设计的特征提取方法来识别图像中的关键

信息。特征提取旨在从图像中抽取有意义的模式，如边缘、纹理或形状，以便后续的分类或检测。经典的特征提取算法包括以下几种。

- SIFT(尺度不变特征变换)：通过检测图像中的关键点并生成描述符，实现对物体旋转、缩放和光照变化的不变性，广泛应用于图像匹配和目标识别。
- HOG(方向梯度直方图)：通过计算图像局部区域的梯度方向分布，提取物体的形状特征，常用于行人检测。
- Haar-like 特征：用于快速检测图像中的特定模式，如人脸检测中的 Viola-Jones 算法。这些算法虽然有效，但需要领域专家设计，且在复杂场景下性能有限。

3. 机器学习在计算机视觉中的应用

机器学习为计算机视觉提供了强大的工具，通过训练模型从数据中学习模式。传统机器学习方法在特征提取后，使用分类器进行识别。当然，识别结果有时是错误的，例如，图 2.1 中将猫的图像识别为狗。常见的方法包括以下几种。

- 支持向量机(SVM)：通过寻找最大间隔超平面，将图像特征分类到不同类别，常用于图像分类和目标检测。
- 随机森林：由多个决策树组成，通过投票机制提高分类准确性，适用于图像分割和物体识别。

传统机器学习方法依赖手工特征，限制了模型的泛化能力。深度学习的出现，特别是它在卷积神经网络(CNN)中的应用，彻底改变了这一局面。

图 2.1　目标检测(引自参考文献[86])

4. 迁移学习与预训练模型

迁移学习是计算机视觉中的重要技术，允许在预训练模型上微调，以适应特定任务。预训练模型(如 ResNet、VGG 和 Inception)在 ImageNet 上训练，学习通用的图像特征，用户只需要在自己的小数据集上微调模型的最后几层，即可快速构建高性能的分类器或检测器。例如，学生可以使用预训练的 ResNet 模型，通过迁移学习识别校园植物，无须从头训练模型。

迁移学习极大地降低了计算机视觉的应用门槛，使非专业人士也能利用现有模型解决实际问题。许多开源工具和平台，如 TensorFlow Hub 和 Hugging Face Transformers，提供了丰富的预训练模型和直观的 API，方便用户部署。

5. 数学基础

计算机视觉的理论基础涉及以下多个数学领域。

- 线性代数：图像表示为矩阵，卷积操作可视为矩阵乘法。主成分分析(PCA)通过矩阵分解实现降维，常用于特征选择。
- 概率论：用于处理不确定性，例如在目标检测中模型输出物体的置信度。贝叶斯方法在图像分割和跟踪中也有应用。
- 优化算法：梯度下降法是模型训练的核心方法，通过迭代更新参数，最小化损失函数。Adam 等自适应优化器广泛使用在深度学习中，提升了收敛速度。

这些数学工具为计算机视觉算法的设计和优化提供了理论支撑，理解这些基础概念有助于更好地掌握 AI 工具的原理，但非科班学生作为 AI 基础使用者，并不要求掌握这部分知识。

6. 开源工具与框架

计算机视觉的普及得益于开源工具和框架的发展。OpenCV 是最流行的计算机视觉库，提供了图像处理、特征提取和目标检测等功能。TensorFlow 是一个强大的深度学习框架，支持 CNN 模型的训练和部署。PyTorch 以其动态计算图和灵活性著称，广泛用于学术研究和快速原型开发，已经成为事实上的开发标准。这些工具不仅为开发者提供了强大的编程接口，还通过图形界面和预训练模型，使非专业用户也能参与图像处理和生成任务。

学生可以使用 OpenCV 的 Python 接口进行图像滤波和边缘检测，或通过 Tensor-Flow 的 Keras API 快速构建 CNN 模型。这些工具的文档和社区支持丰富，降低了学习门槛。

计算机视觉通过数字图像处理、特征提取、机器学习和深度学习等技术，赋予机器理解和生成视觉信息的能力。从手工特征到 CNN 的自动学习，再到迁移学习和开源工具的普及，计算机视觉的发展历程展示了技术的飞跃和应用的大众化。非计算机专业的大学生可以通过理解这些基本理论，更好地利用 AI 工具，探索视觉技术在学习、创作和创新中的无限可能。

2.2 CNN 基础

CNN(Convolutional Neural Network，卷积神经网络)是计算机视觉中最强大的工具之一。它可以帮助计算机“看懂”图片，例如识别照片里的猫咪、检测街上的交通标志，甚至生成艺术画作。CNN 的设计灵感来自人类的视觉系统，能自动从图像中提取特征，如边缘、纹理和物体形状，而无须人工编写复杂的规则。许多经典的机器学习模型都是由 CNN 构成的，如图 2.2 所示。

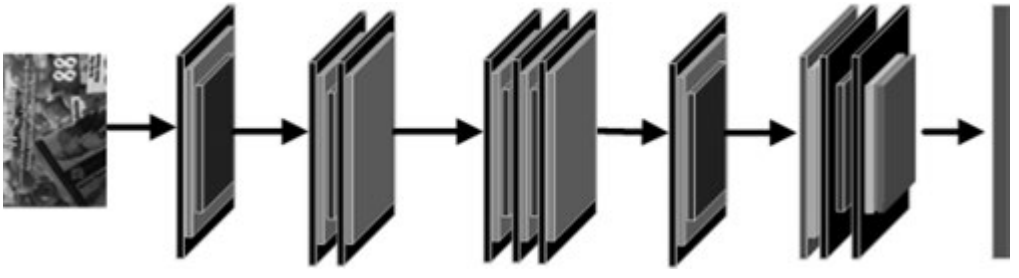

图 2.2 CNN 模型

2.2.1　CNN 的起源与基本原理

完整的 CNN 概念最早由 Yann LeCun 在 1998 年提出，当时是为了识别手写数字。它的工作方式模仿了人类大脑的视觉皮层：从简单的边缘检测开始，逐步理解更复杂的图案。例如，要识别一张猫的图片，CNN 会先找到线条和边缘，再识别毛发的纹理，最后判断出“这是猫”，这种分层方式让 CNN 在处理图像时非常高效。识别猫的 CNN 模型如图 2.3 所示(见彩页)。

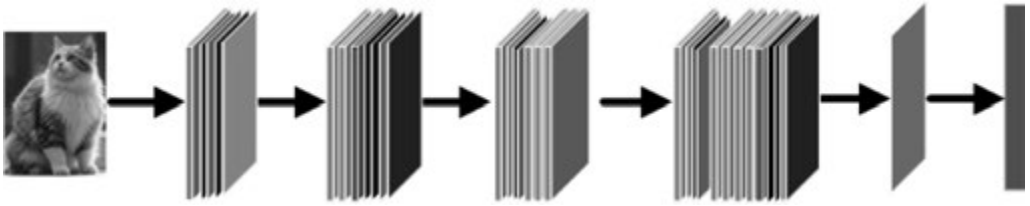

图 2.3　识别猫的 CNN 模型

2.2.2　CNN 的基本组成部分

CNN 主要由三部分组成：卷积层、池化层和全连接层。这三者就像一条流水线，各自负责不同任务，合作完成图像的分析。

1. 卷积层：特征提取的“侦探”

卷积层是 CNN 的核心，专门从图像中找出重要特征。可以把它想象成一个“侦探”，用“放大镜”(称为“卷积核”)扫描图像，寻找线索。

卷积层的工作原理如下：卷积核在图像上滑动，计算每个区域的特征值，生成“特征图”。例如，一个卷积核可能找出边缘，另一个可能关注颜色变化。就像用手机滤镜处理照片，不同滤镜突出不同细节，卷积层是多个滤镜一起工作，捕捉图像的各种特征。

卷积的数学公式如下：

$$\text{特征图} = (I * K)_{i,j} = \sum_{m} \sum_{n} I_{i+m,j+n} K_{m,n}$$

其中，I 是输入图像，K 是卷积核。

卷积操作示意图如图 2.4 所示。

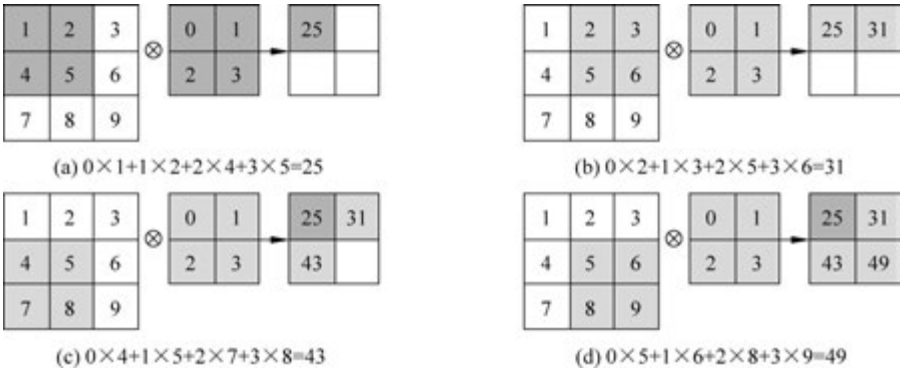

图 2.4　卷积操作示意图

2. 池化层：信息压缩的“整理员”

池化层负责把特征图变小，保留关键信息，同时减少计算量，就像整理行李，只带走最重要的东西。常见方法如下。

- 平均池化：计算区域平均值，平滑细节。
- 最大池化：挑出每个区域的最大值，突出主要特征。

池化的作用是让模型关注整体特征，减少无关细节的影响。池化操作示意图如图 2.5 所示。

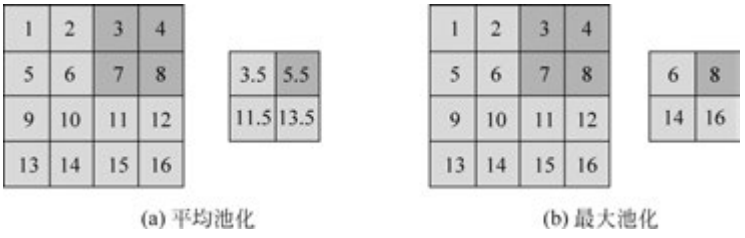

图 2.5 平均池化与最大池化操作示意图

3. 全连接层：决策的“指挥官”

全连接层是 CNN 的最后一步，把之前提取的特征整合起来，做出判断。例如，它会根据边缘和纹理，判断图像里是“猫”还是“狗”。这一层就像大脑，把所有线索汇总，输出最终结果。

2.2.3 经典 CNN 模型

- LeNet(1998 年)：第一个实用 CNN，用于手写数字识别。
- AlexNet(2012 年)：在图像分类比赛中获胜，使深度学习广为人知。
- ResNet(2015 年)：引入“残差连接”，使网络更深、更强，如图 2.6 所示。

这些模型就像从简单的手工工具进化到精密仪器，推动了图像处理的进步。

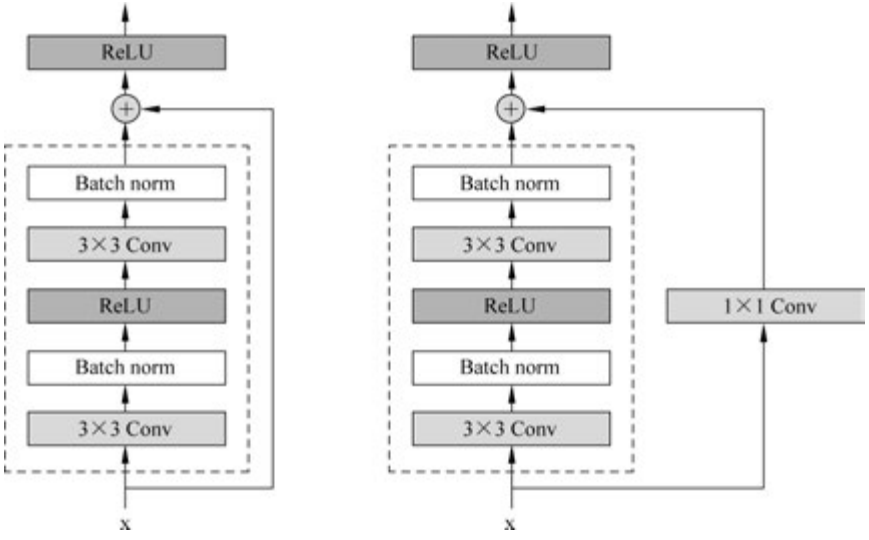

图 2.6 ResNet 的残差块结构(引自参考文献[62])

2.2.4　CNN 的应用与意义

CNN 不仅能识别图像，还能生成图像。比如，生成对抗网络（GAN）利用 CNN 创造逼真的艺术作品，像 DALL-E 这样的工具就是例子。它让普通人也能用 AI 创作，成为 AIGC 的重要技术。

商汤科技公司自主研发的 SenseParrots 深度学习框架，在卷积层设计上突破 GPU 并行计算瓶颈，实现了千亿级参数训练。青年学子通过搭建包含国产算子库的 CNN 模型，在池化层引入“自适应步长”算法（受大华股份的视觉压缩技术启发），在全连接层采用华为昇腾芯片的稀疏化方案。代码实践时标注每层结构中的国产技术突破点，例如激活函数融合中国科学技术大学类脑实验室的仿生设计，彰显“把论文写在祖国大地上”的科研担当。

2.3　目标检测基础

目标检测（object detection）是计算机视觉的核心任务之一，旨在识别图像中特定类别的视觉对象实例并提供其位置信息。由于各类物体有不同的外观、形状和姿态，加上成像时光照、遮挡等因素的干扰，目标检测一直是计算机视觉领域最具有挑战性的问题之一。

回顾 20 年的发展，早期的目标检测算法主要基于手工设计的特征。Viola Jones 检测器于 2001 年提出，首次实现了无约束下的人脸实时检测。其核心技术包括“积分图”“特征选择”和“检测级联”，显著提升了检测速度。2005 年，HOG（方向梯度直方图）检测器通过计算方向梯度直方图来描述特征，在行人检测方面表现出色。HOG 检测器通过多尺度检测窗口和对比度归一化来提高特征的不变性和非线性。2008 年，DPM（Deformable Part-based Model）检测器问世。作为 HOG 检测器的延续，DPM 检测器采用多尺度特征与 latent SVM，遵循“分而治之”的检测哲学，其中训练可以简单地被视为学习一种适当的对象分解方法，而推理可以被视为对对象不同部分的检测集合。例如，检测“汽车”的问题可以分解为检测其车窗、车身和车轮。DMP 检测器作为 VOC-07、VOC-08 和 VOC-09 检测挑战赛的冠军，为后来的诸多检测方法提供了灵感。

随着深度学习的兴起，CNN 备受关注。既然 CNN 能够学习到图像的鲁棒性和高层特征，为什么不将其用于目标检测呢？自此目标检测的速度得到前所未有的提升。基于深度学习的目标检测算法主要分为两类：两阶段（two stage）和单阶段（one stage）。两阶段先选定区域，再用 CNN 进行分类；单阶段则直接使用 CNN 进行特征提取和分类。

1. 两阶段检测器

2014 年 RCNN（Regions with CNN）发布，RCNN 是两阶段检测器，通过选择性搜索生成候选区域建议，然后使用 CNN 提取候选区域的特征，最后用线性 SVM 进行分类和候选框精修。虽然 RCNN 取得显著进展，但其缺点也显而易见：对大量重叠区域（单张图像中

有超过 2000 个框)进行冗余特征计算,导致检测速度极慢(使用 GPU 时每张图像的处理需要 14 秒)。同年晚些时候发布的 SPPNet(Spatial Pyramid Pooling Networks)解决了这个问题。SPPNet 引入了空间金字塔池化层,使得 CNN 可以处理任意大小的图像或区域,避免了对图像的重新缩放,显著提高了检测速度。但仍有一些不足:第一,训练仍然是多阶段的;第二,SPPNet 只对其全连接层进行微调,而忽略了之前的所有层。2015 年,Fast RCNN 被提出,其引入了区域建议网络(RPN),实现了近乎实时的深度学习检测器,解决了 RCNN 中区域建议的瓶颈问题。2017 年 FPN(Feature Pyramid Networks,特征金字塔网络)被提出。在 FPN 之前,大多数基于深度学习的检测器只在网络的顶层进行检测,虽然 CNN 较深层的特征有利于分类识别,但不利于对象的定位。为此,FPN 中开发了具有横向连接的自顶向下体系结构,用于在所有级别构建高级语义,如图 2.7 所示。由于 CNN 通过它的正向传播,自然形成了一个特征金字塔,FPN 在检测各种尺度的目标方面显示出了巨大的进步。

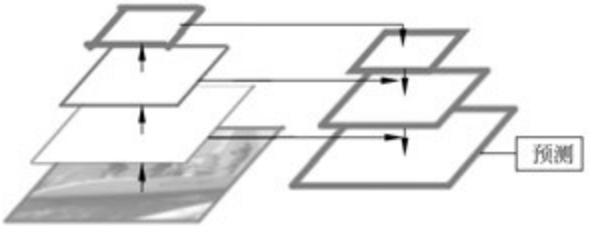

图 2.7　FPN 示意图(引自参考文献[79])

2. 单阶段检测器

YOLO(You Only Look Once)的第一个版本于 2015 年提出,是第一个基于深度学习的单阶段检测器。其通过将整个图像输入一个单一的神经网络来预测边界框和概率,其速度极快,但与两阶段检测器相比,在小目标检测的定位精度上有所下降。同年 SSD(Single Shot Multibox Detector)被提出,其引入了多参考和多分辨率检测技术,显著提高了单阶段检测器对小目标的检测精度。为了研究单阶段检测器的精确度如何赶上两阶段检测器,RetinaNet 于 2017 年被提出,其作者提出“焦点损失”(focal loss)函数来解决密集检测器中前景和背景类别不平衡的问题,为了证明该函数的有效性提出了一个简单模型 RetinaNet,使单阶段检测器在保持高速度的同时实现了与两阶段检测器相当的精度。2018 年 CornerNet 被提出,摒弃了传统的基于锚框的检测范式,将目标检测视为关键点(边界框的角点)预测问题,通过额外的嵌入信息对角点进行分组和重新组合以形成边界框。2019 年 CenterNet 被提出,同样基于关键点检测范式,但将物体视为一个点(物体中心),并基于该中心点回归物体的所有属性,消除了复杂的后处理步骤,如基于组的关键点分配和 NMS。后来随着 Transformer 的火热,2020 年 DETR 被提出,其将目标检测视为集合预测问题,并使用 Transformer 构建了一个端到端的检测网络,开启了不需要锚框或锚点即可检测物体的新纪元。目标检测的精确度变化如图 2.8 所示。

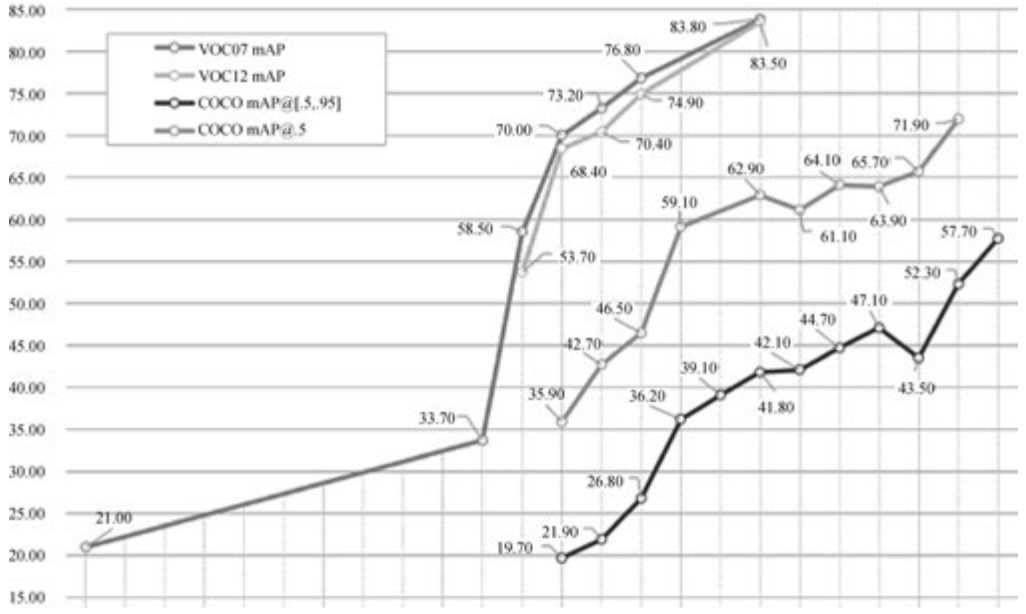

图 2.8　目标检测的精确度提升(引自参考文献[75])

2.4　自然语言处理基础

自然语言处理(Natural Language Processing,NLP)是计算机科学和人工智能的一个子领域,它利用机器学习使计算机能够理解人类语言并与之交流。通过将计算语言学、基于规则的人类语言建模与统计建模、机器学习和深度学习相结合,NLP 使计算机和数字设备能够识别、理解与生成文本和语音。

2.4.1　基本理论

NLP 将计算语言学与机器学习算法和深度学习相结合。计算语言学利用数据科学来分析语言和语音。

NLP 的发展分为以下几个阶段。

1. 基于规则的 NLP

最早的 NLP 应用是简单的"if-then"决策树,需要预编程规则。它们只能根据特定的提示提供答案,例如最初版本的 Moviefone 就具有最基本的自然语言生成(NLG)功能。由于基于规则的 NLP 不具备机器学习或人工智能功能,因此其功能非常有限,而且无法扩展。

2. 基于统计的 NLP

后来发展起来的统计 NLP 可以自动提取、分类与标注文本和语音数据的元素,并为这些元素的每个可能含义分配统计可能性。这依赖于机器学习,可对语言学进行复杂的细分,如语音部分标记。统计 NLP 引入了将语言元素(如单词和语法规则)映射到向量表示的基本技术,这样就可以使用数学(统计)方法(包括回归或马尔可夫模型)对语言进行建模,这为 NLP 的早期发展提供了依据。

3. 深度学习 NLP

近年来，深度学习模型已成为 NLP 的主流模式，通过使用大量原始的非结构化数据（包括文本和语音），其准确性不断提高。深度学习可被视为统计 NLP 的进一步发展，不同之处在于它使用了神经网络模型。模型有以下几个子类别。

（1）序列到序列（seq2seq）模型：基于递归神经网络（RNN），主要用于机器翻译，将一个领域（如德语）的短语转换成另一个领域（如英语）的短语。

（2）Transformer 模型：利用语言的标记化（标记词或子词的位置）和自我关注（捕捉依赖和关系）来计算不同语言部分之间的关系。Transformer 模型可以通过在海量文本数据库上使用自监督学习进行高效训练。Transformer 模型的一个里程碑是谷歌公司的转换器双向编码器表示法（BERT），它成为了并一直是谷歌搜索引擎工作的基础。

（3）自回归模型：这类转换器模型经过专门训练，可以预测序列中的下一个单词，这代表着文本生成能力的巨大飞跃。自回归 LLM 的例子包括 GPT、Llama、Claude 和开源的 Mistral。

2.4.2 NLP 的工作方式

NLP 的工作原理是结合各种计算技术，以机器可以处理的方式分析、理解和生成人类语言。以下是典型的 NLP 流程及其步骤。

1. 文本预处理

文本预处理将原始文本转换成机器更容易理解的格式，为分析做好准备。如图 2.9 所示，首先通过将所有字符转换为小写，对文本进行小写标准化处理，确保像 University 这样的单词得到处理。然后是标记化，即将文本分割成单词、句子或短语等较小的单元，这有助于将复杂的文本分解为易于管理的部分。去停顿词是另一个常见步骤，在这一步骤中，back 和 to 等常用词会被过滤掉，因为它们不会给文本增加重要意义。词干提取或词形还原将词汇还原为其词根形式（例如 went 变为 go），这使得通过将同一词汇的不同形式进行分组来分析语言变得更加容易。此外，文本清理功能还能去除标点符号、特殊字符和数字等可能干扰分析的无用元素。经过预处理后，文本变得干净和标准化，机器学习模型可以对其进行有效解读。

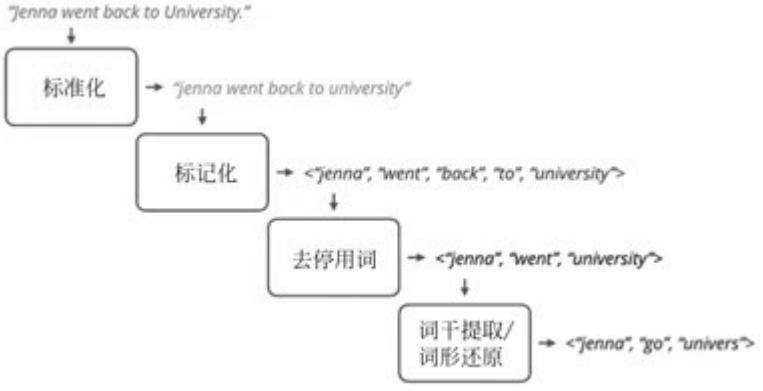

图 2.9　文本预处理（引自参考文献[123]）

2. 特征提取

特征提取是将原始文本转换为机器可分析和解释的数值表示的过程。这一过程通过使用相关技术，如词袋模型(bag of words)和词频-逆文档频率(TF-IDF)，将文本转换为结构化数据，从而量化文档中单词的出现频率及其重要性。更高级的方法有词嵌入技术等，如Word2Vec或GloVe，这些技术将单词表示为连续空间中的密集向量，从而捕捉单词之间的语义关系。上下文嵌入进一步提升了这一能力，通过考虑词汇出现的上下文环境，实现更丰富、更细腻的表示。

3. 文本分析

文本分析通过各种计算技术从文本数据中解释和提取有意义的信息。这一过程包括语音部分标记(POS)和命名实体识别(NER)等任务，前者用于识别单词的语法作用，后者用于检测名称、地点和日期等特定实体。依赖性解析分析词与词之间的语法关系，以了解句子结构；而情感分析则确定文本的情感基调，评估它是积极的、消极的还是中性的；主题建模可识别文本或整个文档语料库中的潜在主题或话题。NLU(自然语言理解)是NLP的一个子集，侧重于分析句子背后的含义。NLU使软件能够在不同的句子中找到相似的含义，或处理具有不同含义的单词。通过这些技术，文本分析可将非结构化文本转化为深刻见解。

4. 模型训练

经过处理的数据随后被用于训练机器学习模型，这些模型会学习数据中的模式和关系。在训练过程中，模型会调整参数，以尽量减少误差，提高性能。模型经过训练后，可用于对未见过的新数据进行预测或生成输出结果。通过评估、验证和微调，NLP建模的有效性不断得到完善，从而提高实际应用中的准确性和相关性。

在整个上述过程中，不同的软件环境都非常有用。例如，自然语言工具包(NLTK)是一套用Python编程语言编写的英语库和程序，它支持文本分类、词干化、标记、解析和语义推理功能；TensorFlow是一个免费、开源的机器学习和人工智能软件库，可用于训练NLP应用的模型。有兴趣学习此类工具的读者可以在网络上找到相关教程。

2.4.3　NLP的实际应用

LLM的交流技巧到图像生成模型理解请求的能力，NLP相关研究帮助促成了生成式人工智能时代的到来。对许多人来说，NLP已经是日常生活的一部分，它为搜索引擎提供动力，用口语命令提示客户服务聊天机器人，还应用于语音操作GPS系统，以及智能设备上的问题解答数字助理(如亚马逊的Alexa、苹果的Siri和小米的小爱同学)。

科大讯飞公司研发的“一带一路”多语种翻译系统支持60种语言互译，在中欧班列物流单据处理中实现98%的准确率。这种基于AI技术的语言互译系统，不仅推动了“一带一路”国家和地区的经济腾飞，而且增强了人们对于小语种的语义理解能力，能够自动识别并尊重当地的习俗用语，促进了各国人民的文化交流，也提升了“讲好中国故事”的传播能力。

2.5　习题

一、单项选择题

1. CNN的核心组成部分不包括以下选项中的(　　)。

A. 卷积层　　B. 池化层　　C. 循环层　　D. 全连接层

2. 自然语言处理(NLP)的文本预处理中,“Jenna went back to University”经过去停用词处理后应保留哪些词?(提示:参考 2.4.2 小节介绍的处理方法)(　　)

A. Jenna,went,university　　B. Jenna,back,university

C. Jenna,went,back　　D. back,to,university

3. 目标检测技术中,单阶段检测器(如 YOLO)的核心特点是(　　)。

A. 先生成候选区域再分类　　B. 直接预测边界框和类别概率

C. 依赖两阶段级联网络　　D. 仅适用于小尺度目标

二、简答题

1. ResNet 的残差块解决了 CNN 的什么问题?提示:参考 2.1.1 小节。

2. 列举 NLP 发展的三个阶段及其代表性技术。提示:参考 2.4.1 小节。

三、论述题

1. 目标检测技术从传统方法到深度学习的演进中,两阶段与单阶段检测器的核心差异是什么?结合 RCNN 和 YOLO 分析其优缺点。提示:参考 2.3.1 小节对目标检测技术发展的介绍。

2. 计算机视觉中,迁移学习如何降低技术应用门槛?举例说明其实际价值。提示:参考 2.1.1 小节。

四、实践题

1. 假设有一个 3×3 灰度图像矩阵 $\begin{bmatrix}1 & 2 & 3\\5 & 6 & 7\\9 & 10 & 11\end{bmatrix}$,使用 2×2 最大池化(步长 2)处理。提示:参考 2.2.2 小节。

2. 为句子“ChatGPT revolutionized AI interactions”设计预处理流程,需包含小写化、分词、去停用词、词干化四步,并写出每步骤的结果。提示:参考 2.4.2 小节及图 2.9。

第3章

有趣的AIGC综合案例

3.1 基本图像生成与风格迁移

3.1.1 Stable Diffusion 的工作机制

Stable Diffusion 能够生成高质量图像的生成模型。其工作原理基于扩散(diffusion)过程,扩散过程则是通过逐步添加噪声到图像数据,并在训练过程中学习去噪的逆过程。如图 3.1 所示,Stable Diffusion 利用了扩散过程和逆扩散过程来生成图像。

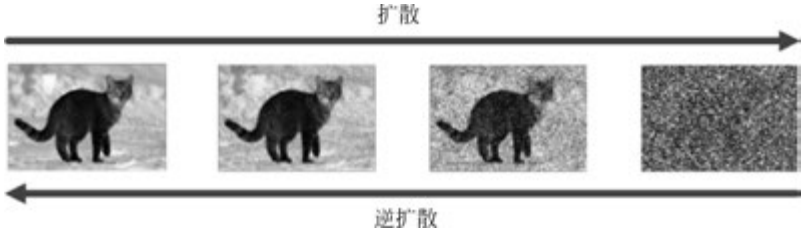

图 3.1 Diffusion 训练过程

扩散过程(Forward Process):在这个过程中,图像数据逐步被添加噪声,直至变得完全随机。这一步的目的是将复杂的数据分布转换为高斯分布。

逆扩散过程(Reverse Process):在逆扩散过程中,模型学习从噪声数据中逐步去除噪声,最终生成清晰的图像。这一步通过训练去噪自动编码器(denoising autoencoders)来实现。

简单来说,Diffusion 模型的训练包括两个主要步骤:首先,构建和训练去噪模型;其次,通过逆扩散过程生成图像。其优势在于能够生成高分辨率、细节丰富的图像,同时具有较高的稳定性和鲁棒性。对 Diffusion 进行大致介绍后,再对 Stable Diffusion 各个部分进行简单讲解。

如图 3.2 所示(见彩页)为 Stable Diffusion 生成图片的过程,该过程主要包括以下几个步骤。

1. 提示词处理

(1) 在将提示词(prompt)输入模型之前,必须对其进行预处理,因为模型无法直接解析文本数据。

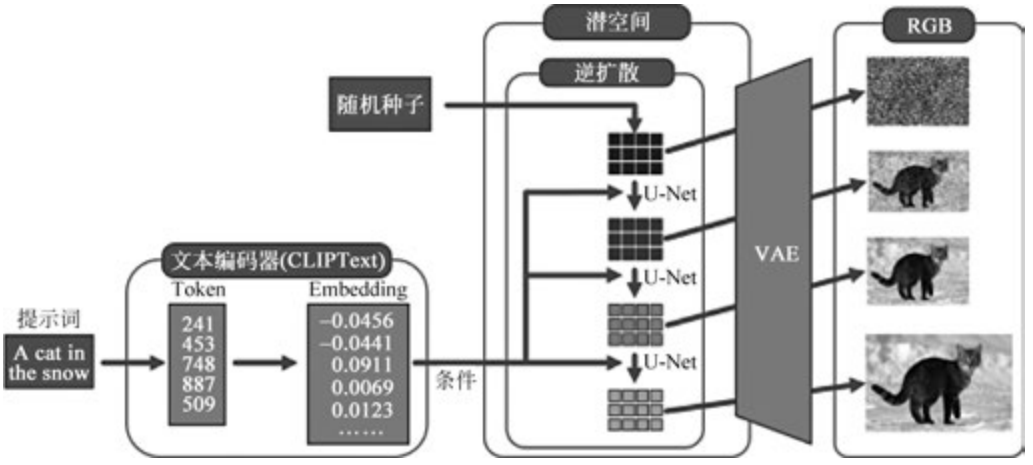

图 3.2 Stable Diffusion 过程

(2) 使用对比语言-图像预训练(Contrastive Language-Image Pretraining,CLIP)模型将提示词转换为标记(token),即一系列数字。一个词语可能被转换成多个标记。

(3) 将标记进一步转换为 Embedding,即特定维度的向量表示。

(4) 通过文本转换器(text transformer)将嵌入转换为模型的条件输入。

2. 噪声图生成

(1) 种子(seed)用于生成噪声图像。给出固定的种子和分辨率,生成的噪声图像是确定的。

(2) 模型以噪声图像为基础生成最终图像。需要注意的是,噪声图像并非实际的图片,而是潜空间中的表示,可理解为压缩后的噪声图像。

3. 图像生成

(1) 图像生成过程涉及去噪和从潜空间表示转换为实际图像。

(2) 去噪过程是图像生成的核心,也是 Stable Diffusion 的关键步骤。该过程涉及复杂的数学原理,以下为其概要:

- 模型基于噪声图像的内容,并结合提示词逐步修改图像。
- 修改过程是逐步进行的,早期修改幅度较大,后期修改幅度逐步减小。此过程使用 U-Net 模型。

(3) 去噪完成后,得到的结果图像仍在潜空间中,需要转换为可视化的实际图像。此步骤使用变分自编码器(Variational Autoencoder,VAE)完成。

(4) 值得注意的是,Stable Diffusion 模型自带的 VAE 模型效果可能不够理想,可以选择替换为其他 VAE 模型以获得更好的效果。

Stable Diffusion 在多个领域展示了广泛的应用前景,如广告设计、游戏设计、教育应用、室内设计、建筑设计、风景园林等,它生成的游戏素材如图 3.3 所示。这些只是 Stable Diffusion 的部分应用场景,其实际应用范畴可能远超这些,其强大的图像生成能力使其可以在几乎所有需要图像内容的领域发挥重要作用。