

第 1 单元 引论

支撑的课程目标

1. 能够基于智能信息处理的基本理论和技术，识别和理解数据处理与分析等相关问题的相关特性。

基本要求

1. 能够详述计算机视觉的定义、基本任务及解决方案，理解视觉信息处理问题的特性。
2. 能够详述自然语言处理的基本问题和发展趋势及解决方案，理解文本信息处理问题的特性。

教学重点与难点

- 重点： 计算机视觉的定义及解决方案。
难点： 深度学习对计算机视觉的影响。

教学过程设计

新课导入、知识讲授、教学目标达成考核、总结。

教学过程设计

本单元教学通过“互动、开放”的课堂形式，采用探究式学习、问题导入的教学方法，激发学生的学习兴趣，促成课程目标的达成。

教学学时

4 学时。

教师先阐述本单元的课程目标，再进入单元学习。

一、导入新课（10 分钟）

深度学习是一种基于神经网络的学习方法。和传统的机器学习方法相比，深度学习模型一般需要更丰富的数据、更强大的计算资源，同时也能达到更高的准确率。目前，深度学习方法被广泛应用于计算机视觉、自然语言处理、强化学习等领域。

二、知识讲授（160 分钟）

本单元要点：

* 计算机视觉

- 定义
- 基本任务
- 传统方法
- 仿生学和深度学习
- 现代深度学习

* 自然语言处理

- 基本问题
- 传统方法与神经网络方法
- 发展趋势

1 计算机视觉

1.1 计算机视觉的定义

计算机视觉是使用计算机及相关设备对生物视觉的一种模拟。它的主要任务是通过采集的图片或视频进行处理以获得相应场景的三维信息。计算机视觉是一门关于如何运用照相机和计算机来获取我们所需的，被拍摄对象的数据与信息的学问。形象地说，就是给计算机安装上眼睛（照相机）和大脑（算法），让计算机能够感知环境。

1.2 计算机视觉的基本任务

计算机视觉的基本任务包含图像处理、模式识别或图像识别、景物分析、图像理解等。除了图像处理和模式识别之外，它还包括空间形状的描述，几何建模以及认识过程。实现图像理解是计算机视觉的终极目标。

1.2.1 图像处理技术

图像处理技术可以把输入图像转换成具有所希望特性的另一幅图像。例如，可通过处理使输出图像有较高的信噪比，或通过增强处理突出图像的细节，以便于操作员的检验。在计算机视觉研究中经常利用图像处理技术进行预处理和特征抽取。

1.2.2 模式识别技术

模式识别技术根据从图像抽取的统计特性或结构信息，把图像分成预定的类别。例如，文字识别或指纹识别。在计算机视觉中模式识别技术经常用于对图像中的某些部分，例如分割区域的识别和分类。

1.2.3 图像理解技术

图像理解技术是对图像内容信息的理解。给定一幅图像，图像理解程序不仅描述图像本身，而且描述和解释图像所代表的景物，以便对图像代表的内容做出决定。

1.3 计算机视觉的传统方法

在深度学习算法出现之前，对于视觉算法来说，大致可以分为以下 5 个步骤：特征感知，图像预处理，特征提取，特征筛选，推理预测与识别。

传统算法：缺乏对特征的重视、图像特征提取需要人力、依赖特征算子

传统算法的成功或半成功的例子：指纹识别算法、基于 Haar 的人脸检测算法、基于 HoG 特征的物体检测

1.4 仿生学与深度学习

1.4.1 神经网络示例

输入：图片

输出：标签

1.4.2 视觉系统信息处理机制

发现人：David Hubel

时间：1981

意义：证明大脑的可视皮层是分级的。

视觉功能划分

抽象：把非常具体的形象的元素抽象出来形成有意义的概念。

迭代：把有意义的概念往上迭代，变成更加抽象，人可以感知到的抽象概念。

1.4.3 抽象与迭代的过程

案例说明：人脑感知气球、模拟识别人脸、识别摩托车

1.5 现代深度学习

1.5.1 卷积神经网络（CNN）

卷积神经网络是一种对人脑比较精准的模拟，它模拟了人脑识别图片时感知图片中的局部特征，之后将局部特征综合起来再得到整张图的全局信息的过程。其卷积层通常是堆叠的，低层的卷积层可以提取到图片的局部特征，高层的卷积能够从低层的卷积层中学到更复杂的特征，从而实现到图片的分类和识别。

卷积就是两个函数之间的相互关系在计算机视觉里面，可以把卷积当作一个抽象的过程，就是把小区域内的信息统计抽象出来。

1.5.2 现代深度学习的过程

用学习好的卷积和对图像进行扫描，然后每一个卷积和会生成一个扫描的响应图，我们叫 response map，或者叫 feature map。如果有多个卷积和，就有多个 feature map。也就说从一个最开始的输入图像（RGB 三个通道）可以得到 256 个通道的 feature map，因为有 256 个卷积和，每个卷积和代表一种统计抽象的方式。

池化操作。池化操作在统计上的概念更明确，就是一个对一个小区内求平均值或者求最大值的统计操作。带来的结果是，如果之前输入有两个通道的，或者 256 通道的卷积的响应 feature map，每一个 feature map 都经过一个求最大的一个池化层，会得到一个比原来 feature map 更小的 256 的 feature map。

LeNet 网络：LeNet 网络输入图像是大小为 32×32 像素的灰度图，第一层经过了一组卷积核，生成了 6 个 28×28 的 feature map，然后经过一个池化层，得到 6 个 14×14 的 feature map，然后再经过一个卷积层，生成了 16 个 10×10 的卷积层，再经过池化层生成 16 个 5×5 的 feature map。这 16 个大小为 5×5 的 feature

map 再经过 3 个全连接层，即可得到最后的输出结果。输出就是标签空间的输出。

1.5.3 现代深度学习

AlexNet 网络：

AlexNet 是基于 LeNet 的改进，它可以被看作 LeNet 的放大版。AlexNet 的输入是一个大小为 224X224 像素的图片，输入图像在经过若干个卷积层和若干个池化层后，最后经过两个全连接层泛化特征，得到最后的预测结果。

AlexNet 的特征：在第一层，都是一些填充的块状物和边界等特征；中间的层开始学习一些纹理特征；而在接近分类器的高层，则可以明显看到的物体形状的特征。最后的一层，即分类层，不同物体的主要特征已经被完全提取出来。

2. 自然语言处理

自然语言，区别于计算机所使用的机器语言和程序语言，是指人类用于日常交流的语言。而自然语言处理的目的却是要让计算机来理解和处理人类的语言。自然语言的一个句子背后往往包含着不直接表述出来的常识和逻辑。这使得计算机在试图处理自然语言的时候不能从字面上获取所有的信息。因此自然语言处理的难度更大，它的发展与应用相比于计算机视觉也往往呈现出滞后的情况。

为了将深度学习引入自然语言处理领域，研究者尝试了许多方法来表示和处理自然语言的表层信息（如词向量、更高层次、带上下文信息的特征表示等），也尝试过许多方法来结合常识与直接感知（如知识图谱、多模态信息等）。这些研究都富有成果，其中的许多都已应用于现实中，甚至用于社会管理、商业、军事的目的。

2.1 自然语言处理的基本问题

自然语言处理主要研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法，其主要任务包括：

语言建模：计算一个句子在一个语言中出现的概率。

中文分词：将中文句子恰当地切分为单个的词。

句法分析：通过明确句子内两个或多个词的关系来了解整个句子的结构。最终句法分析的结果是一棵句法树。

情感分类：给出一个句子，判断这个句子表达的情感。

机器翻译：最常见的是把源语言的一个句子翻译成目标语言的一个句子，最终预测出来的整个目标语言句子必须与给定的源语言句子具有完全相同的含义。

阅读理解：有许多形式。有时候是输入一个段落，一个问题，生成一个回答，或者在原文中标定一个范围作为回答，有时候是输出一个分类。

2.2 传统方法与神经网络方法的比较

人工参与程度：人的参与程度越来越低，但系统的效果越来越好。这是合乎直觉的，因为人对于世界的认识和建模总是片面的、有局限性的。如果可以将自然语言处理系统的构建自动化，将其基于对世界的观测点（即数据集），所建立的模型和方法一定会比人类的认知更加符合真实的世界。数据量：随着自然语言处理系统中人工参与的程度越来越低，系统的细节就需要更多的信息来决定，这些信息只能来自于更多的数据。

可解释性：人工参与程度的降低带来的另一个问题是模型的可解释性越来越低。在理想状况下，如果系统非常有效，人们根本不需要关心黑盒系统的内部构造，但事实是自然语言处理系统的状态离完美还有相当的差距。

2.3 自然语言处理的发展趋势

从传统方法和神经网络方法的对比中，可以看出自然语言处理的模型和系统构建是向着越来越自动化、模型越来越通用的趋势发展的。一开始，人们试图减少和去除人类专家知识的参与。因此就有了大量的网络参数、复杂的架构设计，这些都是通过在概率模型中提供潜在变量，使得模型具有捕捉和表达复杂规则的能力。之后，人们觉得每一次为新的自然语言处理任务设计一个新的模型架构并从头训练的过程过于烦琐，于是试图开发利用这些任务底层所共享的语言特征。在这一背景下，迁移学习逐渐发展。现在人们希望神经网络的架构都可以不需要设计，而是根据具体的任务和数据来搜索得到。这一新兴领域方兴未艾，可以预见随着研究的深入，自然语言处理的自动化程度一定会得到极大提高。

三、教学目标考核（20分钟）

讨论：

1. 计算机视觉处理任务的流程是什么？为什么需要深度学习？

2. 自然语言处理的基本任务是什么？深度学习对其有什么影响？

四. 总结 (10 分钟)

简要介绍了深度学习的应用领域。卷积神经网络可以模拟人类处理视觉信息的方式提取图像特征，极大地推动了计算机视觉领域的发展。自然语言处理是典型的时序信息分析问题，其主要应用包括句法分析、情感分类、机器翻译等。强化学习强调智能体与环境的交互与决策，具有广泛的应用价值。通过引入深度学习，模型的函数拟合能力得到了显著的提升，从而可以应用到一系列高层任务中。