

# 人工智能之多模态

6 班 陈静雯

## 一、图像理解

### （一）概述

将图像结构化为某一类别的信息，赋予每个像素类别意义，具体用到两种算法，一种是准确度较高的 two-stage 检测算法，产生候选区域并对其分类，另一种是 one-stage 检测算法，他不需要 region proposal 阶段，直接产生物体的类别概率和位置坐标，主要有 YOLO 和 SSD。

### （二）应用及其存在的问题

#### 1. 医学图像分析

传统的医学图像分析需要大量的时间和精力，并且这种人工针对特定场景设计的特征通用性欠佳，充分性和精细性也有待提高。而以深度神经网络为代表的深度学习方法能够从医学图像中自动进行特征学习，现被广泛应用。但是仍然存在一定的问题，比如在复杂场景下模型识别能力较弱，特征数据稀疏，数据获取困难等等。

对此，梁爽提出了基于注意力双分支网络的半监督学习方法和模型，以及卷积神经网络框架等<sup>[1]</sup>，能够在一定程度上更加充分与客观的反映医学图像中包含的丰富信息。

#### 2. 数字化病理

基于人工智能技术对数字病理图像进行特征提取和定量分析，并转为高保真、高通量的可挖掘、可分析的数据，在肿瘤早期诊断、分

级及构建后期模型等方面表现出独特的优势，相较于传统病理诊断模式的主观性和不准确性的弊端，数字化病理一定程度上解决了医疗资源低于分布不均匀的问题，极大提高了诊断的工作效率和准确性。

通过机器学习，可以从肿瘤病理图像中提取与患者预后相关的信息，构建预测模型，能够满足当下对常规组织病理学评估中高通量精度诊断的需求。但是目前的病理图像分析还是有它的局限性，如病理图像质量有待提高，缺少高质量的数据集，以及对人工智能诊断病情存在的道德伦理和法律上的问题。<sup>[2]</sup>

## 二、语音识别

### （一）原理

将复杂的声波分解为一个个组成部分，然后将每个频带（从低到高）的能量相加，为该音频创建一个有排序的识别码，将能量数字绘成图表，再对每个 20 毫秒的音频片段重复这一过程，最终得到一个频谱图。通过神经网络处理整个音频，根据其拥有的数据库，对这些基于发音的预测进行打分，找到最接近的转录文本，识别出想要的字符。

### （二）应用

#### 1. 声纹识别

每个人的发音控制器官不同，发音频率也不同，因此每个人的声纹都具有独特性、唯一性、不可复制性以及可识性。声纹识别逐渐成为主流的生物特征识别技术，该技术通过对声音的多语言信号特征进行分析，从中提出关键性特征，与数据库中的样本进行对比，实现对

个体声音特征进行针对性识别，从而实现身份识别。<sup>[3]</sup>

## 2. 海洋动物声信号检测

基于仿生应用与生态文明建设等多种需求，海洋哺乳动物声学快速发展，而人工智能在其声信号检测识别研究领域的应用刚刚起步。段德鑫提出了一种基于图像处理和随机森林的鲸豚类海洋动物回声定位信号检测方法，以及基于卷积神经网络提出的一种海洋哺乳动物声信号的识别算法，在低信噪比条件下有更高的召回率和检测效果<sup>[4]</sup>。但是从海量原始声音数据中检测识别出所需要的动物声信号仍是制约海洋哺乳动物声学研究的一道主要障碍。

## 三、文本表示

### （一）概述

大致有以下三种方法，一是基于向量空间模型的方法，比如潜在语义分析，通过奇异值分解的方法对原始矩阵降维，得到文档向量和词项向量；二是基于主题模型的方法，从概率生成模型角度来实现文本的表示，文档向量和词项向量的每一个维度都是一个“主题”；三是基于神经网络的方法，包括基于词向量合成、RNN 和注意力机制等等，而词向量是这些模型的基础。

### （二）应用

1. 对话生成

2. 机器翻译

## 四、参考文献

[1]梁爽. 基于深度神经网络的医学图像特征学习与分析[D]. 北京科

技大学, 2022. DOI:10.26945/d.cnki.gbjku.2022.000220.

[2] 李晓燕, 李晨, 陈灏源, 张勇, 张敬东. 人工智能在结直肠癌数字化病理图像分析中的应用[J]. 协和医学杂志, 2022, 13(04):542-548.

[3] 白曦龙. 人工智能技术在智能声纹识别系统中的应用[J]. 电声技术, 2021, 45(04):12-14+18. DOI:10.16311/j.audioe.2021.04.004.

[4] 段德鑫. 人工智能在海洋哺乳动物声信号检测识别研究中的应用[D]. 自然资源部第一海洋研究所, 2022. DOI:10.27058/d.cnki.gohyy.2022.000009.