



AI座谈会分享

从输入输出看AI的演变

黄恩待 副研究员

智能医学与生物医学工程研究院

2024-12-19



CONTENTS

- 一、AI输入输出的变化
 - 二、大模型的突破
 - 三、AI未来重要方向
- 

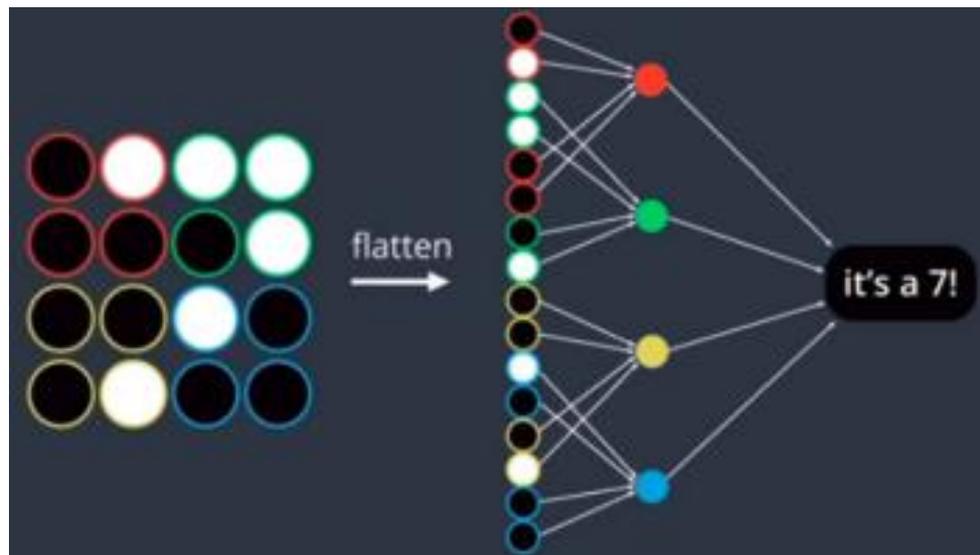
01

AI输入输出的变化

一、输入输出的变化

机器学习

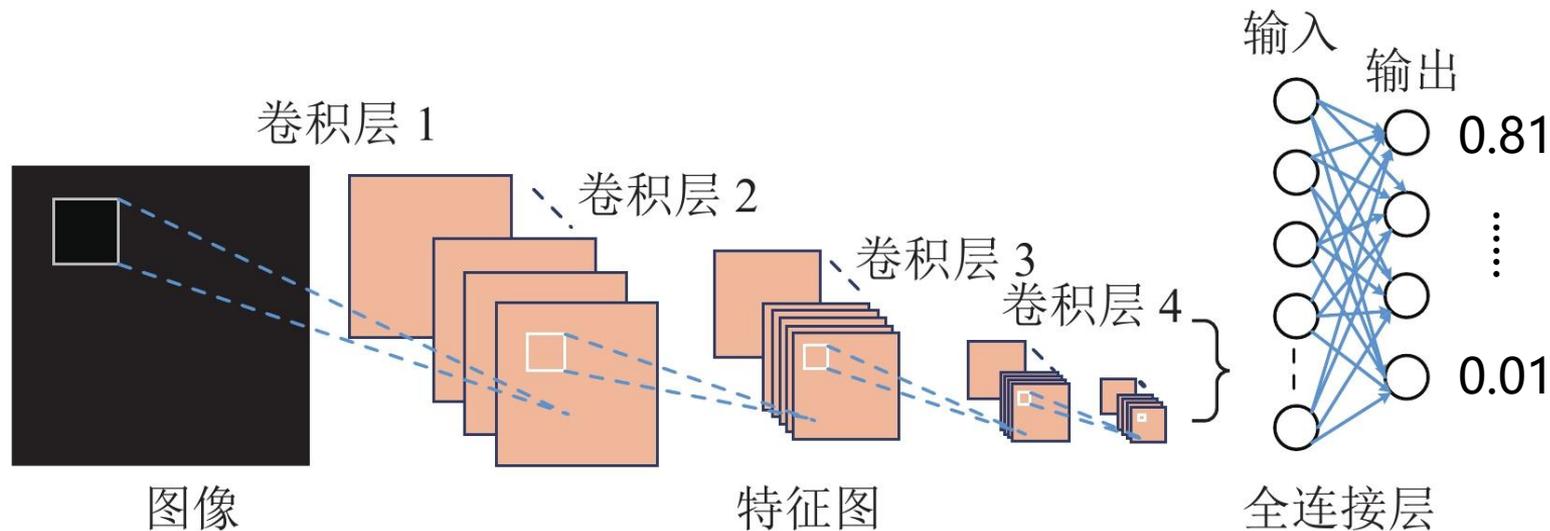
- 输入：一维向量
- 输出：数字（只能表征某类别）



一、输入输出的变化

深度学习——CNN分类

- 输入：二维矩阵
- 输出：向量（能表征属于各个类别的概率）



一、输入输出的变化

深度学习——CNN语义分割

- 输入：二维矩阵
- 输出：0/1矩阵（能细致的表示感兴趣区域mask，但无法区分个体）

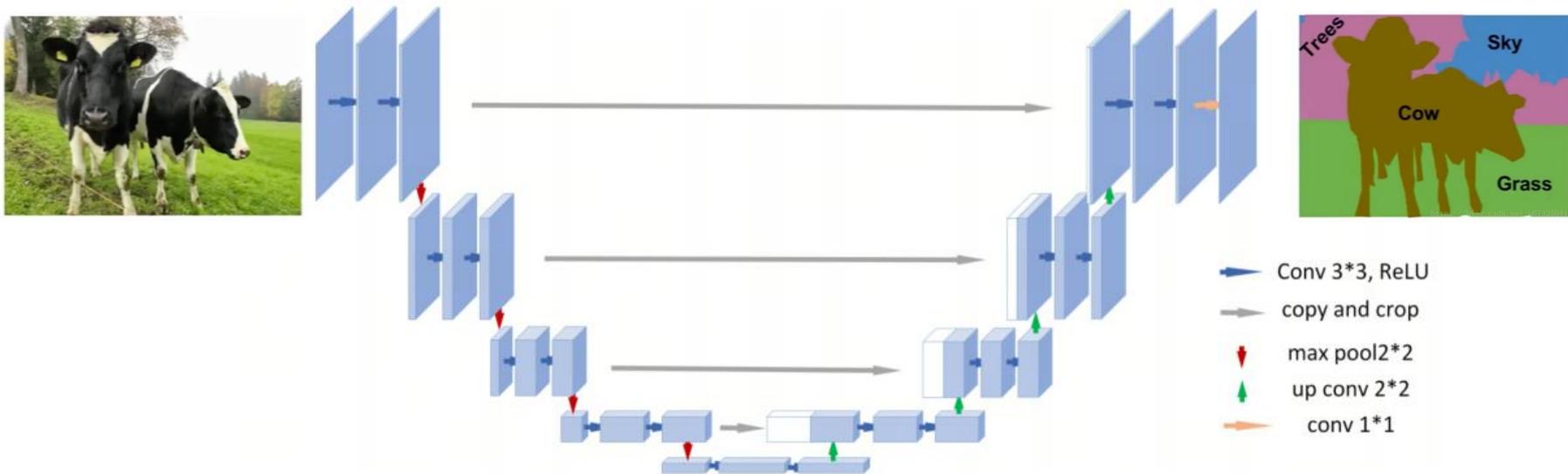
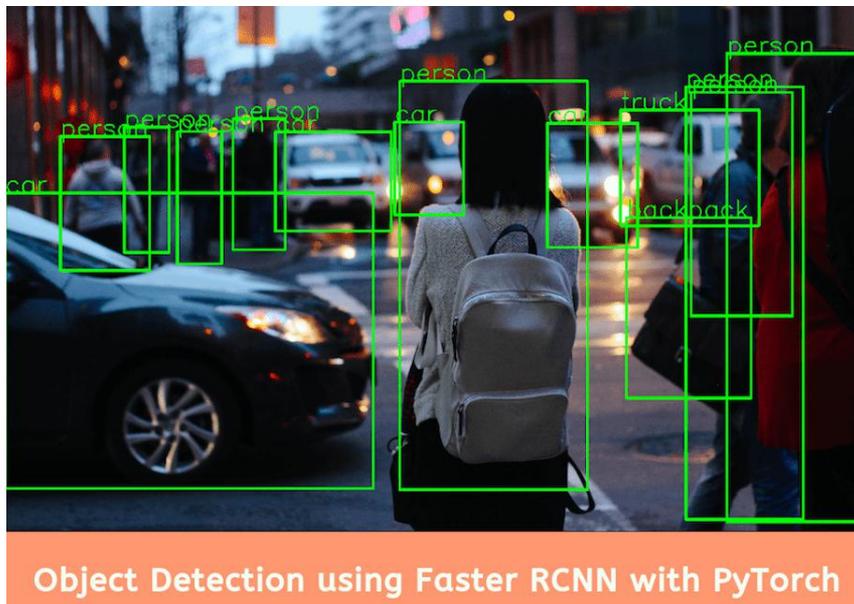


Fig. 1 Structure of U-Net

一、输入输出的变化

深度学习——CNN目标检测

- 输入：二维矩阵
- 输出：每个个体bbox坐标和类别（向量、能表征不同目标的不同“粗糙”区域）



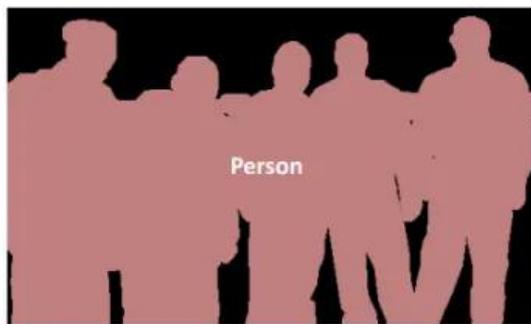
一、输入输出的变化

深度学习——CNN实例分割

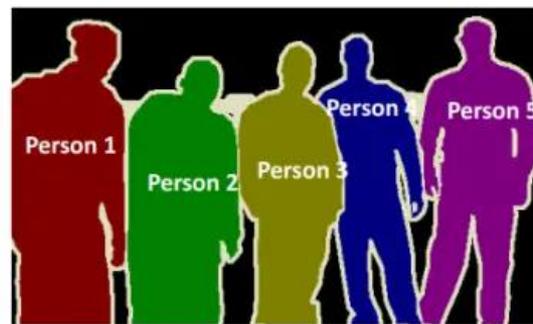
- 输入：二维矩阵
- 输出：每个个体mask和类别（0/1矩阵和向量，能表征不同目标的不同“细致”区域）



Object Detection



Semantic Segmentation



Instance Segmentation

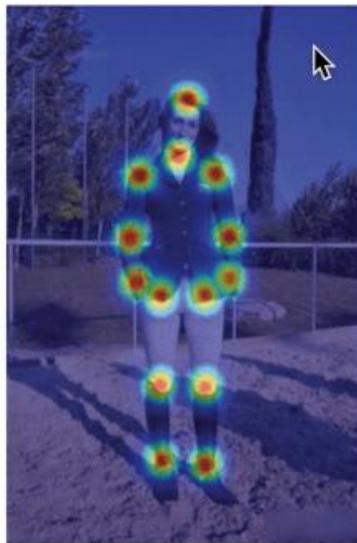
一、输入输出的变化

深度学习——CNN关键点检测

- 输入：二维矩阵
- 输出：0/1矩阵



(a)



(b)



(c)

一、输入输出的变化

缺点

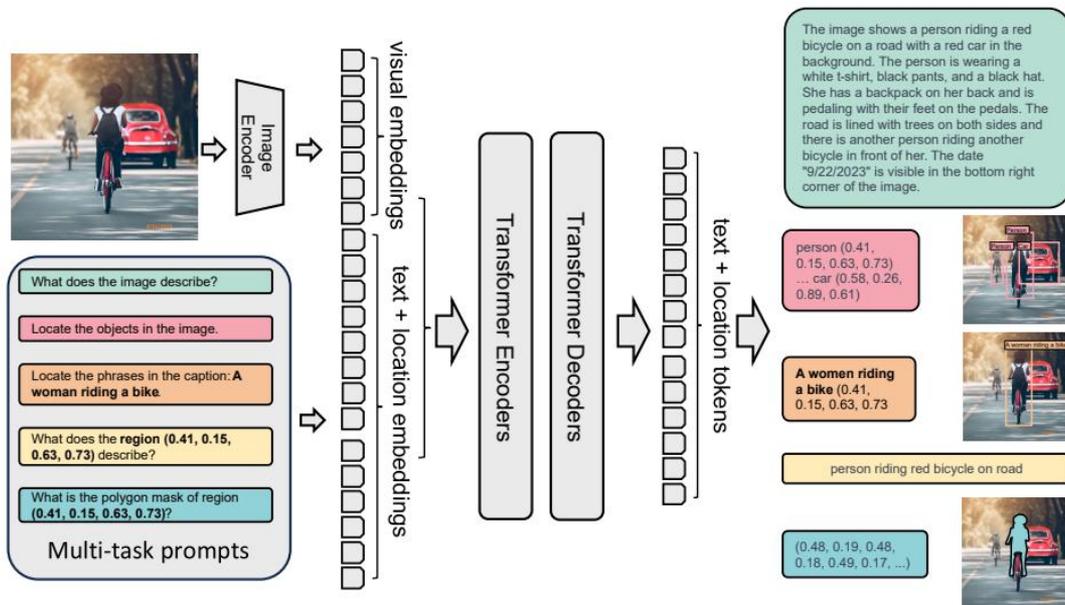
- 输入固定、输出固定
- 一种模型及其输出对应一个任务

任务	机器学习		深度学习CNN			
	输入	输出	输入	输出		
图像分类	向量	数字	矩阵	向量		
语义分割			矩阵	矩阵		
目标检测			矩阵	向量		
实例分割			矩阵	向量+矩阵		
关键点检测			矩阵	矩阵		

二、大模型的突破

深度学习——视觉大模型VLM

- 输入：图像矩阵+文本（非固定）
- 输出：文本（非固定）



二、大模型的突破

大模型优点

- 输入灵活、输出灵活
- 文本可以涵盖所有输出方式

任务	机器学习		深度学习CNN		深度学习VLM	
	输入	输出	输入	输出	输入	输出
图像分类	向量	数字	图像	向量	图像+文本	文本
语义分割			图像	矩阵	图像+文本	
目标检测			图像	向量	图像+文本	
实例分割			图像	向量+矩阵	图像+文本	
关键点检测			图像	矩阵	图像+文本	

二、大模型的突破

大模型优点

- 输入灵活、输出灵活
- 文本可以涵盖所有输出方式

文本才是我们理解世界的根本方式？

任务	机器学习		深度学习CNN		深度学习VLM	
	输入	输出	输入	输出	输入	输出
图像分类	向量	数字	图像	向量	图像+文本	文本
语义分割			图像	矩阵	图像+文本	
目标检测			图像	向量	图像+文本	
实例分割			图像	向量+矩阵	图像+文本	
关键点检测			图像	矩阵	图像+文本	

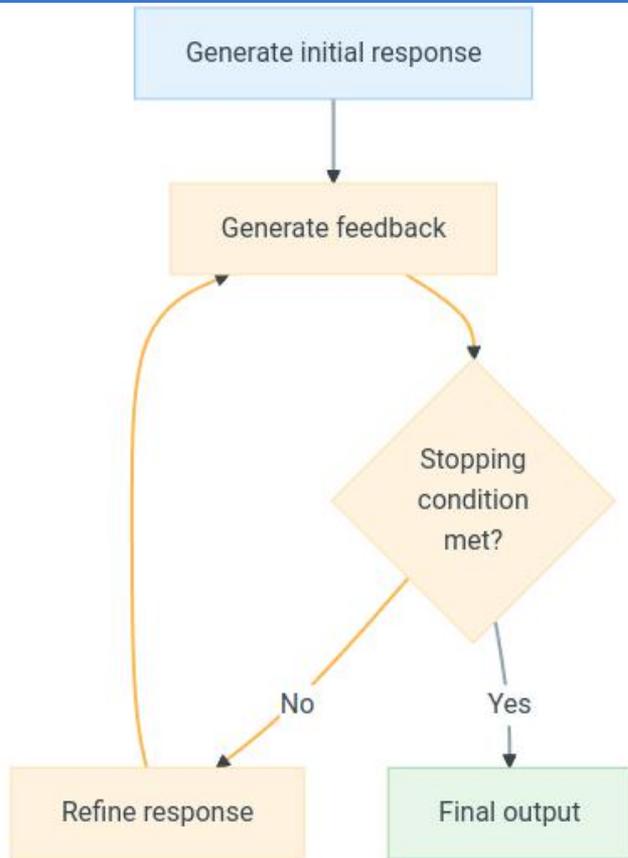
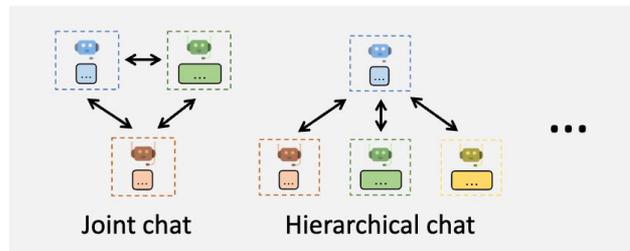
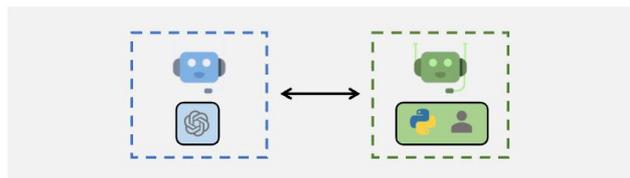
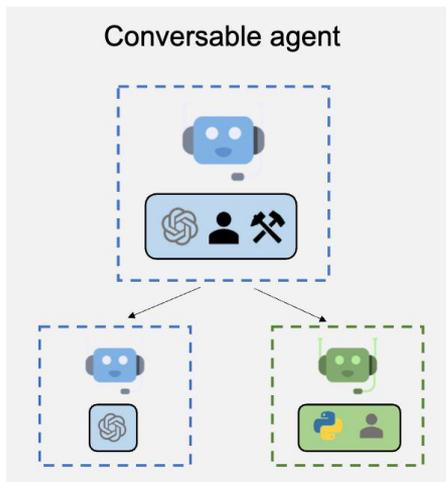
03

AI未来重要方向

三、AI未来重要方向

自我监督

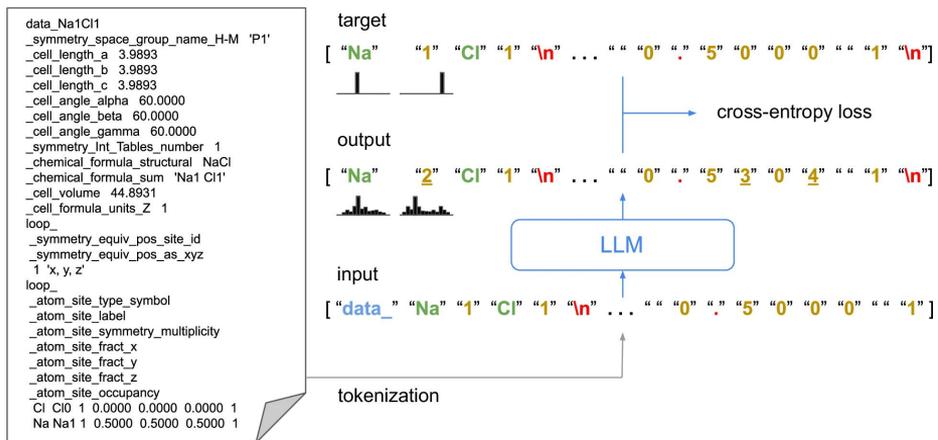
- AutoGen
- Instructor



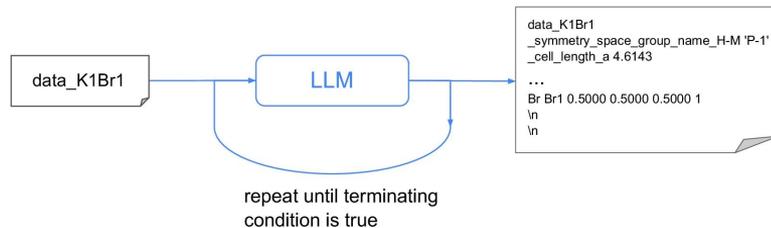
三、AI未来重要方向

交叉学科

a



b





THE END
THANKS

黄恩待

智能医学与生物医学工程研究院

huangendai@nbu.edu.cn