

人工智能與計算機視覺

王玉璽

中科院香港創新研究院人工智能與機器人創新中心

2022-10-19

1. 人工智能的發展歷程
2. 計算機視覺技術在現實場景中的應用
3. 人工智能技術未來發展方向



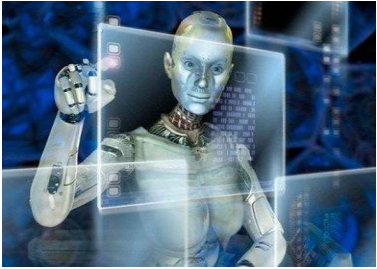
- 那 你在研究什么 - 我展示给你看
- So, what are you been working on? - I'll show you.

人工智能是什麼：模擬、延伸和擴展人類智慧

研究目的： 探尋智慧本質，研製出具有類人智慧的智能機器。

研究內容： 能夠模擬、延伸和擴展人類智慧的理論、方法、技術及應用系統

呈現形式：

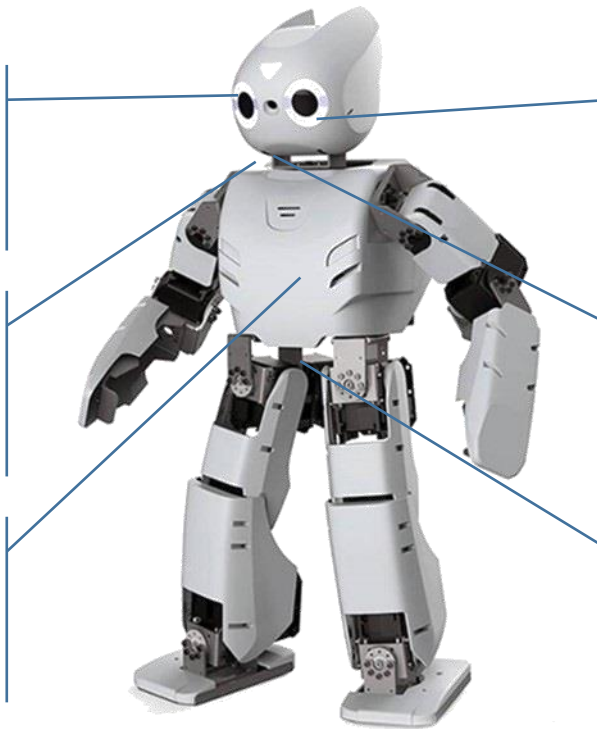


會聽： 語音識別、
說話人識別、機器
翻譯



會思考： 人機對弈、
定理證明、醫療診
斷

會學習： 機器學習、
知識表示



會看： 圖像識別、
文字識別、車牌識
別

會說： 語音合成、
人機對話

會行動： 機器人、
自動駕駛汽車、無
人機

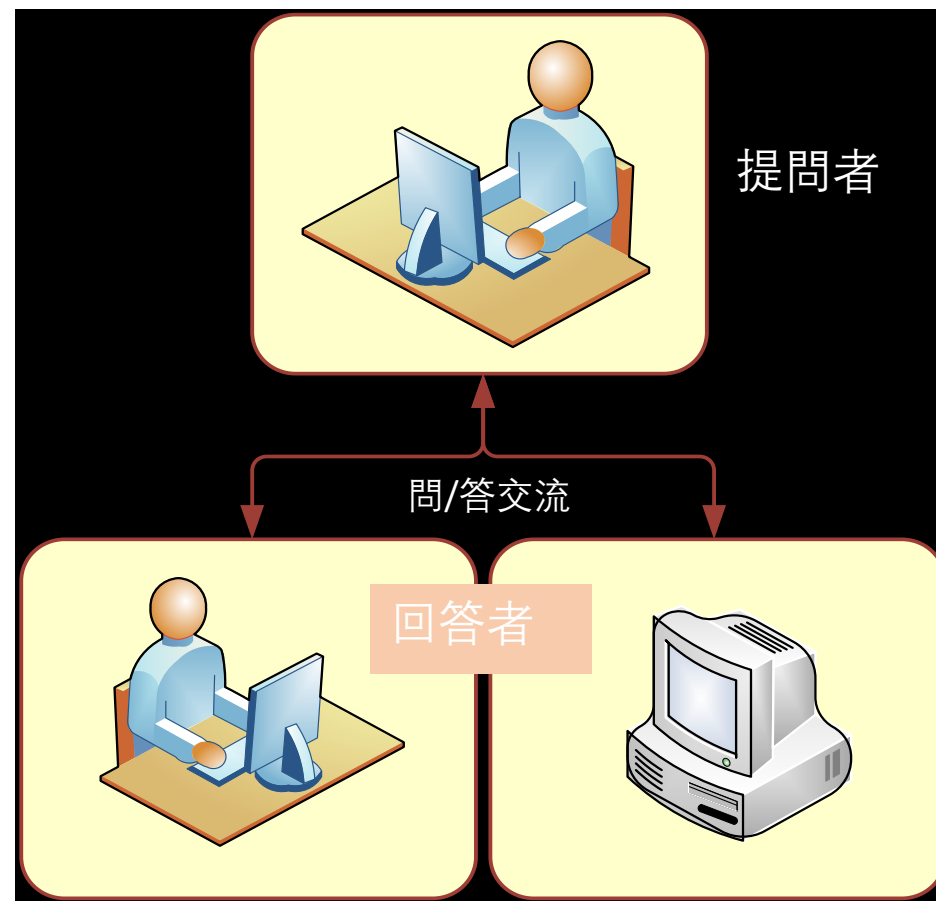


人工智能 (Artificial Intelligence: AI)

人工智能的起源



阿蘭·圖靈：英國數學家、密碼學家、電腦科學家，1936年提出了圖靈機模型，1950年提出了“機器能夠思維”的論述



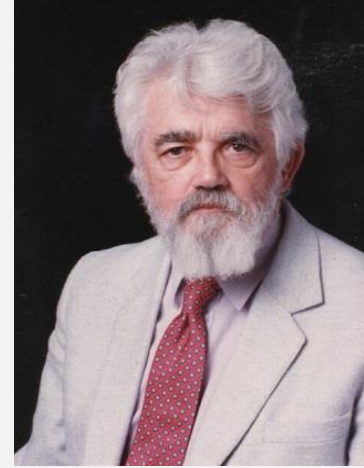
圖靈測試：如果電腦在5分鐘的文字對話中能回答由人類測試者提出的一系列問題，且其超過30%的回答讓測試者誤認為是人類所答，則電腦通過圖靈測試，被認為具有智能。

人工智能的起源

1956年夏天約翰.麥卡錫等人在美國達特茅斯學院開會研討“如何用機器模擬人的智慧”，會上提出“人工智慧”這一概念，標誌著人工智慧學科的誕生。



達特茅斯學院



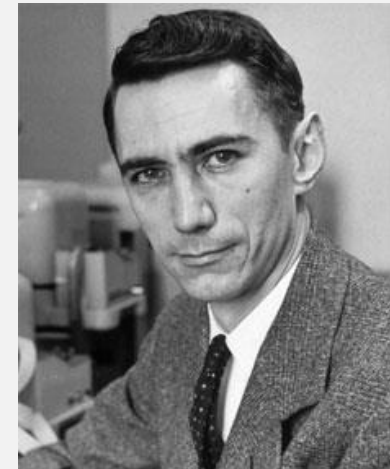
麥卡錫



明斯基



羅徹斯特



香農

人工智能的發展歷程：推理-知識-深度神經網路

■ 第一次人工智慧浪潮 (1956-1976)：推理



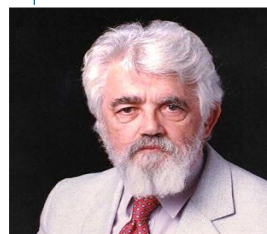
1954 IBM-701首次完成
英俄機器翻譯



1952 圖靈提出讓電腦來冒
充人



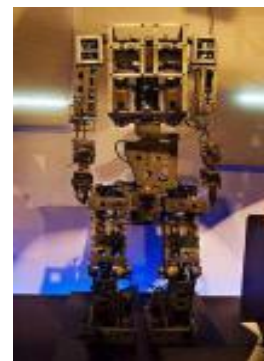
1946 第一台電子電腦
ENIAC誕生



1958, 約翰·麥卡錫開
發Lisp語言



1957
Rosenblatt提
出感知器模型



1970 日本早稻田大學
造出第一個人形機器人



1969 馬文·明斯基和西摩
爾·帕普特提出了簡單神經網路
的局限性



1973 英國劍橋大學數學家詹
姆斯·萊特希爾稱人工智慧沒
帶來任何影響



1956 達特茅斯會議

1946

1954

1956

1970

1976

人工智能的發展歷程：推理-知識-深度神經網路

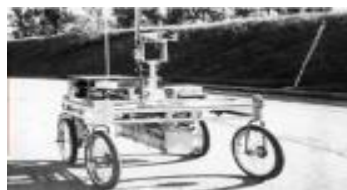
■ 第二次人工智慧浪潮 (1976-2006)：知識



1977, 吳文俊首次發表定理的機械化證明的論文, 開闢全新的方向



1981 日本國際貿易和工業部提供8.5億美元用於第五代電腦專案研究



1979 斯坦福大學的自動駕駛汽車成功駛過一個充滿障礙的房間



1980 卡耐基梅隆大學為DEC製造出了專家系統



1990 863計畫支持研發農業專家系統



IBM、蘋果等PC性能超越Symbolics等通用電腦



1993, 中國研製成功首台採用微處理器的並行處理機系統“曙光一號”, 定點速度每秒6.4億次



1984 在年度AAAI會議上, 羅傑·單克和馬文·明斯基警告“AI之冬”即將到來

1976

1982

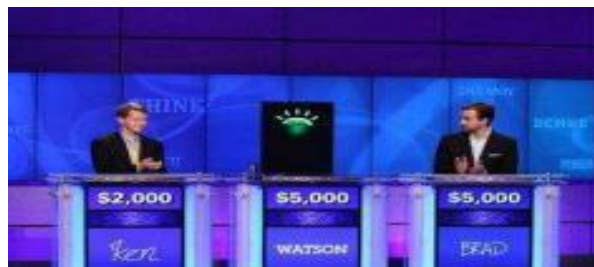
2006

人工智能的發展歷程：推理-知識-深度神經網路

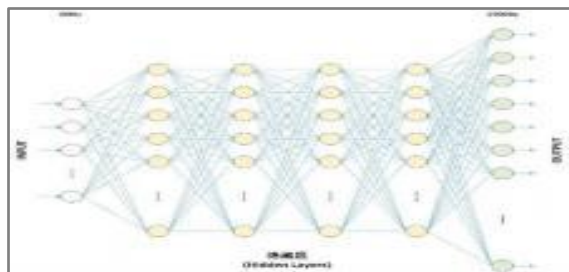
■ 第三次人工智慧浪潮（2006-至今）：深度神經網路



2007 李飛飛和普林斯頓大學的同事開始建立ImageNet



2011 IBM超級電腦沃森在美國老牌益智節目擊敗人類



2006 傑夫·辛頓提出深度學習的理論框架，提出學習生成模型



2009 穀歌汽車在內華達州通過自動駕駛汽車測試



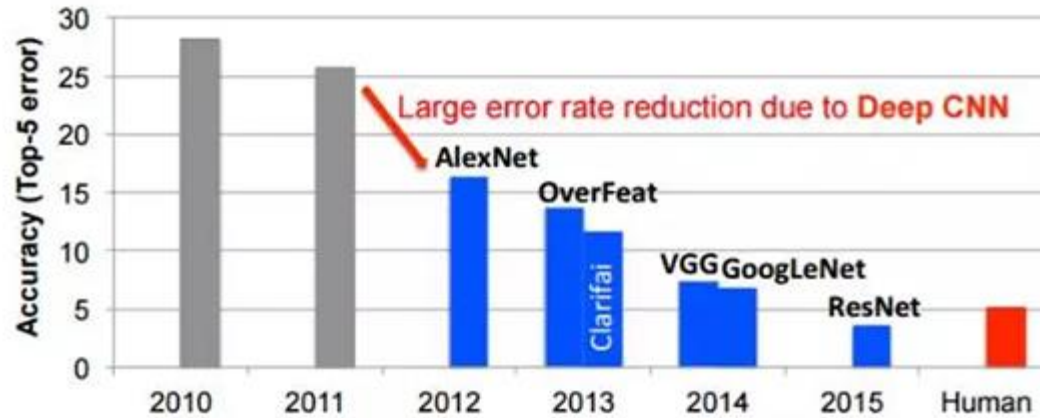
2016-2017 穀歌研發的AlphaGo連續擊敗頂尖人類棋手

2006

至今

人工智慧的發展歷程

- 深度神經網路（深度學習）



深度學習在ImageNet上的識別準確率已經超過了人類水準



穀歌研發的AlphaGo、AlphaZero連續擊敗人類頂尖棋手

人工智慧的發展歷程

■ 深度學習演算法簡介

用資料優化
電腦程式的
模型參數



历史数据

训练

模型

预测

未知
属性

通過經驗自
動改進的電
腦演算法



经验

归纳

规律

预测

未来

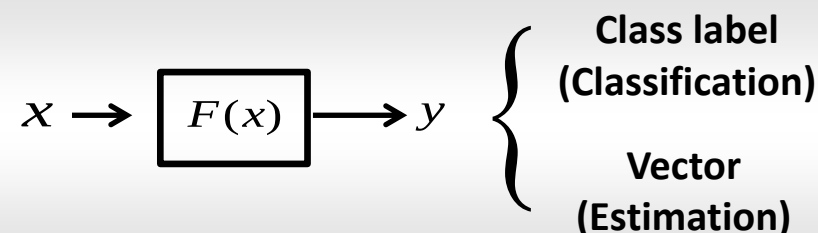
• 機器學習的基本定理

模型的出錯率 $\propto \frac{\text{模型的複雜程度}}{\text{樣本的大小}}$

推論：

模型複雜-》大樣本

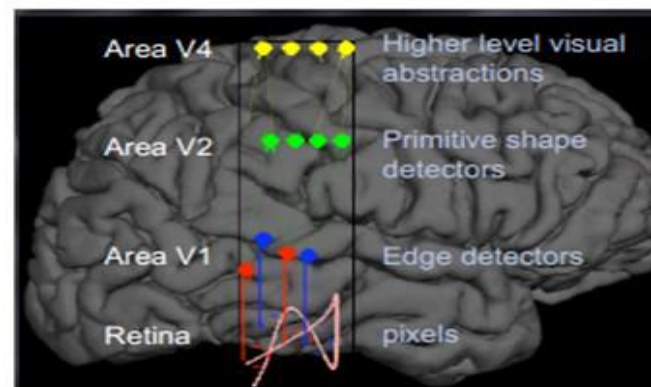
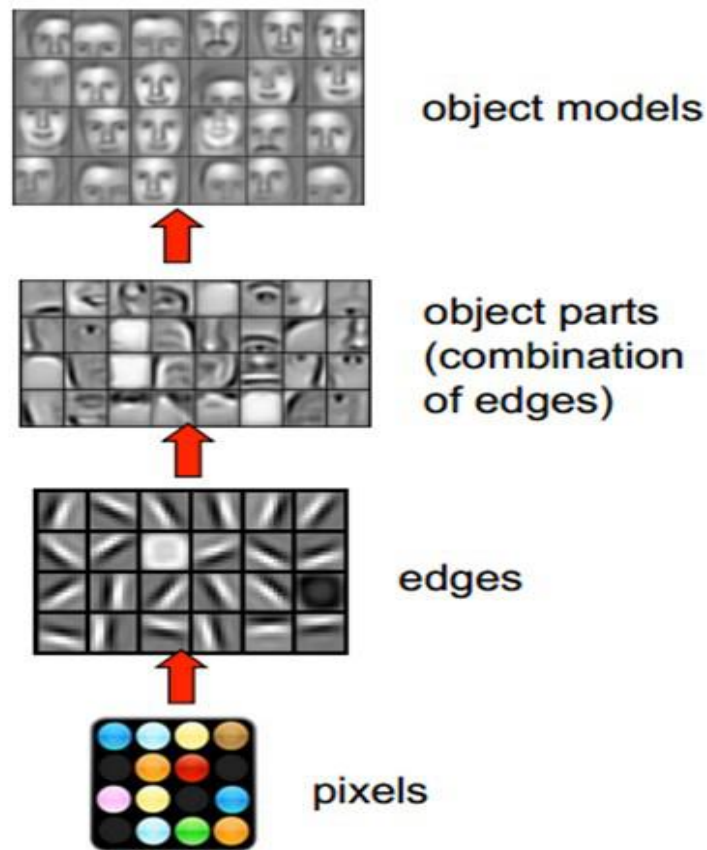
樣本小-》簡化模型



- 深度學習是使用包含複雜結構或由多重非線性變換構成的多個處理層對資料進行高層抽象的演算法
- 深度學習是一種基於對資料進行表徵學習的方法。觀測值（例如一幅圖像）可以使用多種方式來表示，如每個圖元強度值的向量，或者更抽象地表示成一系列邊、特定形狀的區域等。而使用某些特定的表示方法更容易從實例中學習任務（例如，人臉識別或面部表情識別）。深度學習的好處是用非監督式或半監督式的特徵學習和分層特徵提取高效演算法來替代手工獲取特徵

人工智慧的發展歷程

■ 深度神經網路 (深度學習)



- 人腦是通過分級的、多層網路模型來識別
- 減少資料量，保留物體的有用資訊

- 對效果影響極大
- 手動化特徵工程
- 非常耗時

低層信息

預處理

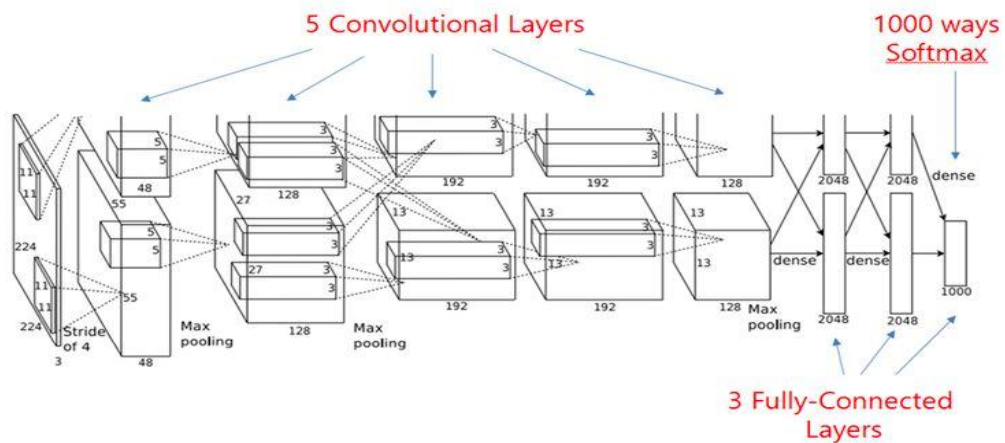
特徵提取

識別分類

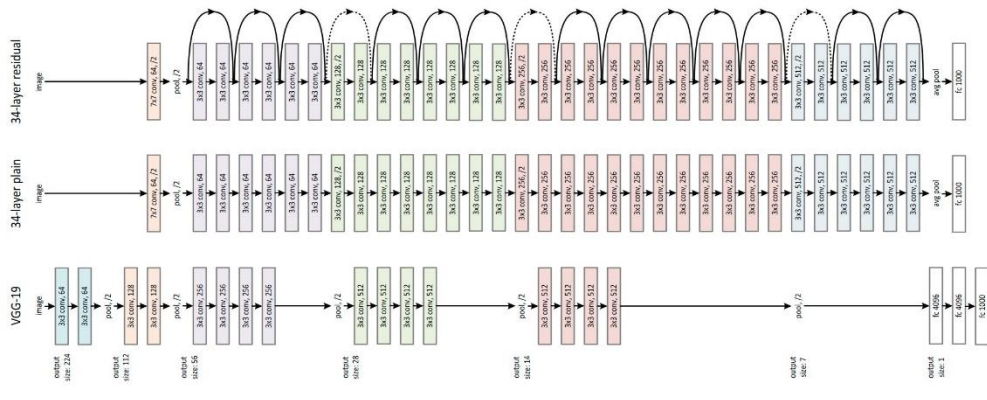
圖像識別的一般流程

人工智慧的發展歷程

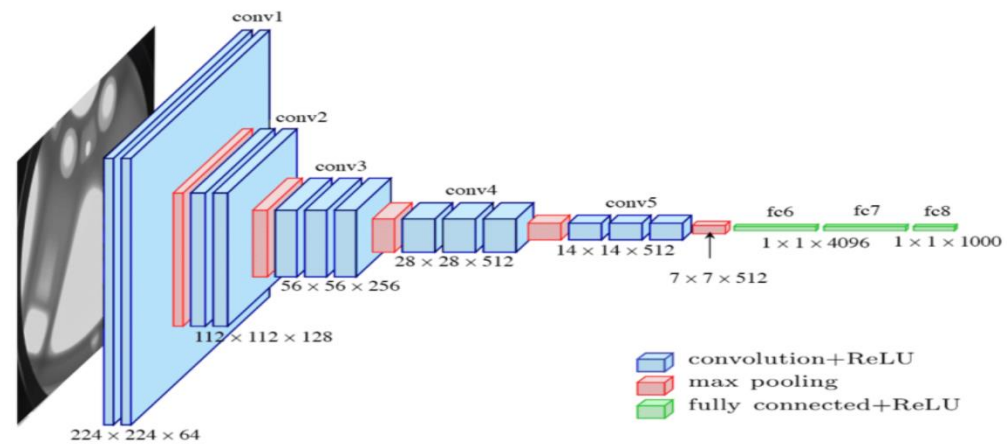
■ 深度神經網路 (深度學習)



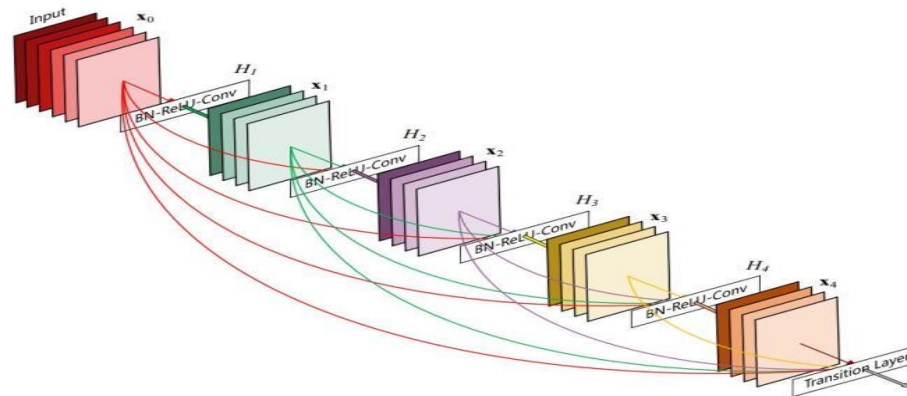
2012年, AlexNet



2016年, ResNet

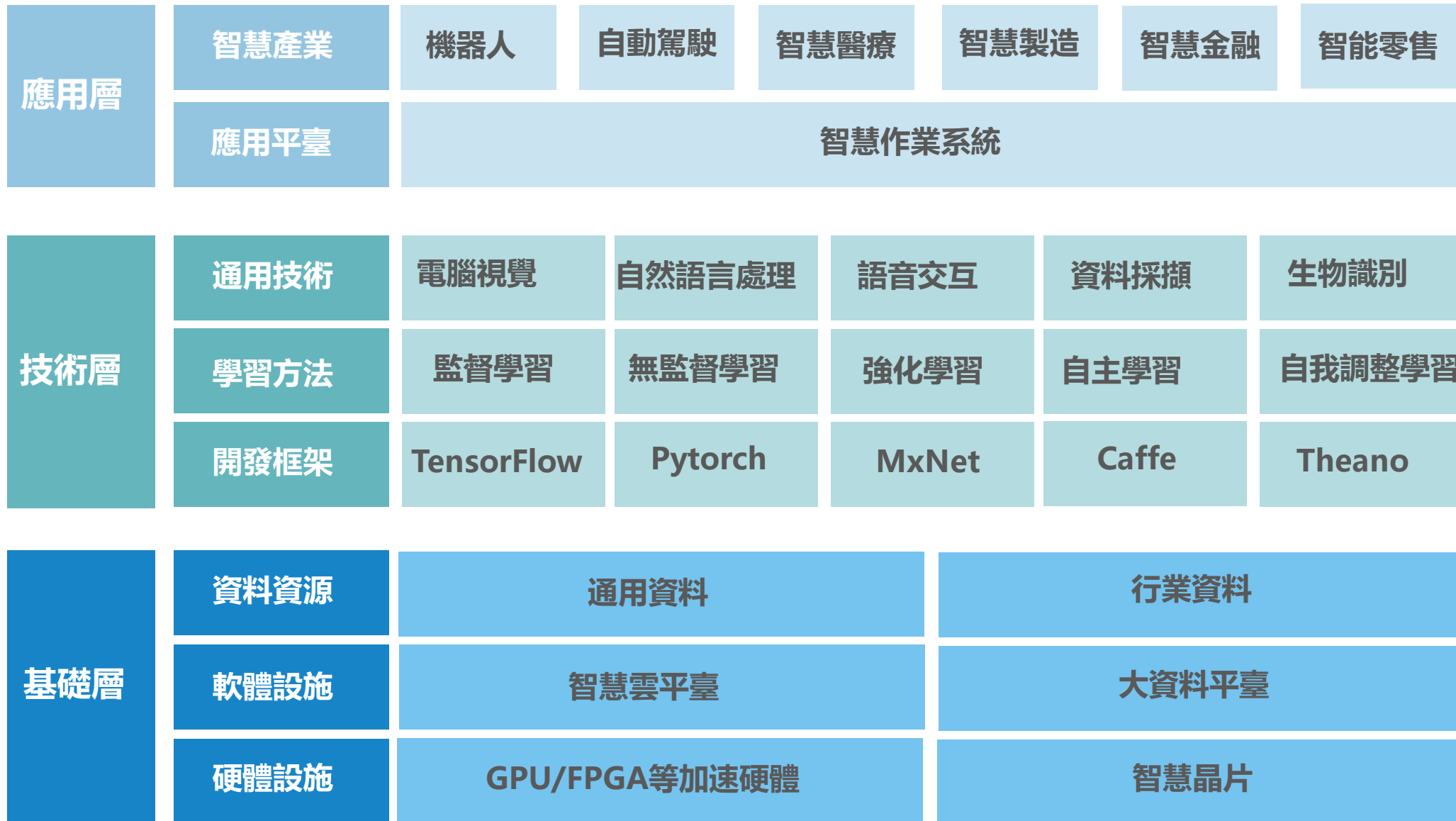


2014年, VGGNet



2017年, DenseNet

人工智慧的研究領域



人工智能的研究領域

■ 手機中的AI

Source: <http://www.163.com>



识别:
语音、人脸、视频内容
应用: 安防、客服。。。

BI、商业流程自动化

互联网数据
购买转化、商品推荐、定价、精准营销、
社交媒体营销

券商股票:
智能+量化交易、智能投顾、机器人理财

银行、保险应用:
针对性电话营销、
贷款审批、信用卡欺诈

医学:
自动读片、辅助诊断、个性化诊断、基因排序

教育:
学习外语、智能选题

人工智慧的研究領域

- 自動駕駛—Google Waymo



人工智慧的研究領域

- 快速藝術圖像生成



人工智慧的研究領域

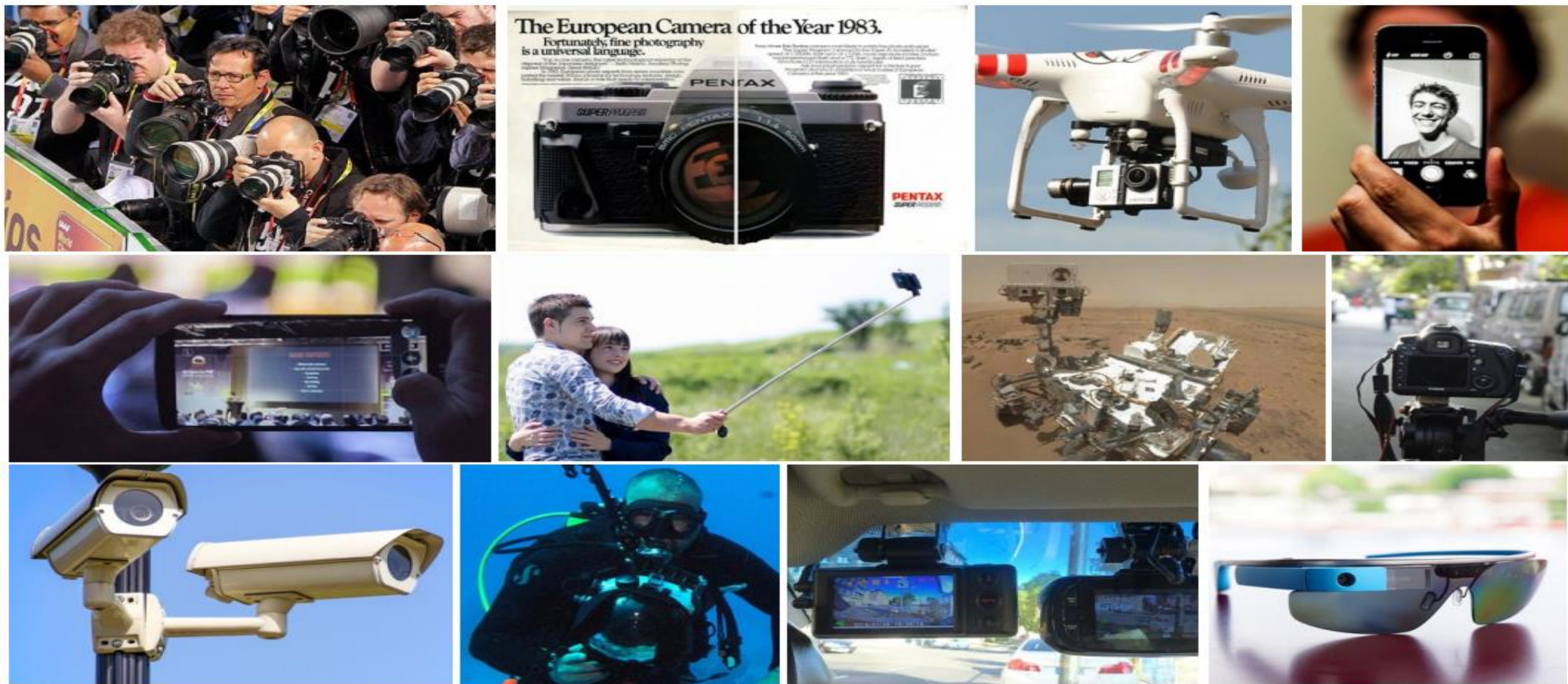
專用人工智慧在多領域取得成功應用：人機對話領域



快剪辑

計算機視覺技術

又叫機器視覺，是一門讓機器學會像人一樣看



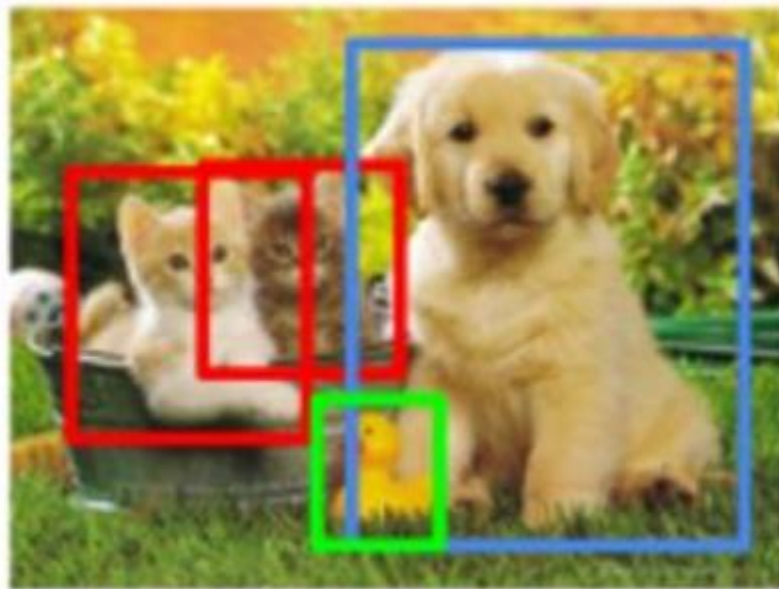
基於深度學習技術，電腦視覺近年來取得了突飛猛進的進展

圖像分類



貓

目標檢測



貓、鴨子、狗

語義分割



貓、鴨子、狗

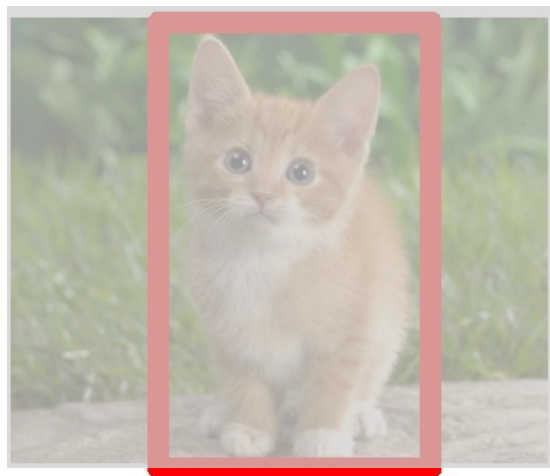
圖像分類：給定一張圖片，識別類別

分類
(Classification)



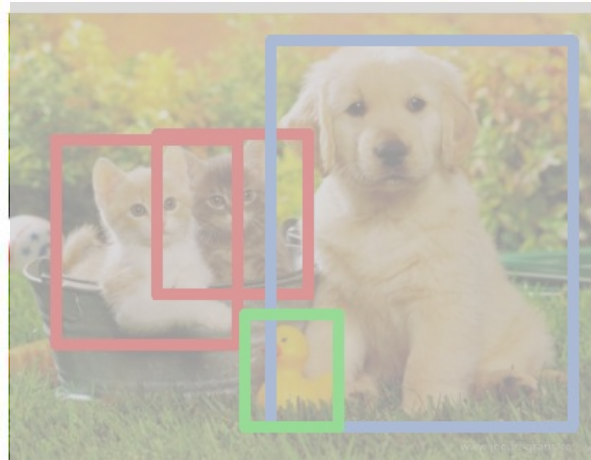
CAT

分類 + 定位
(Classification+Localization)



CAT

物體檢測
(Object Detection)



CAT, DOG, DUCK

實例分割
(Instance Segmentation)



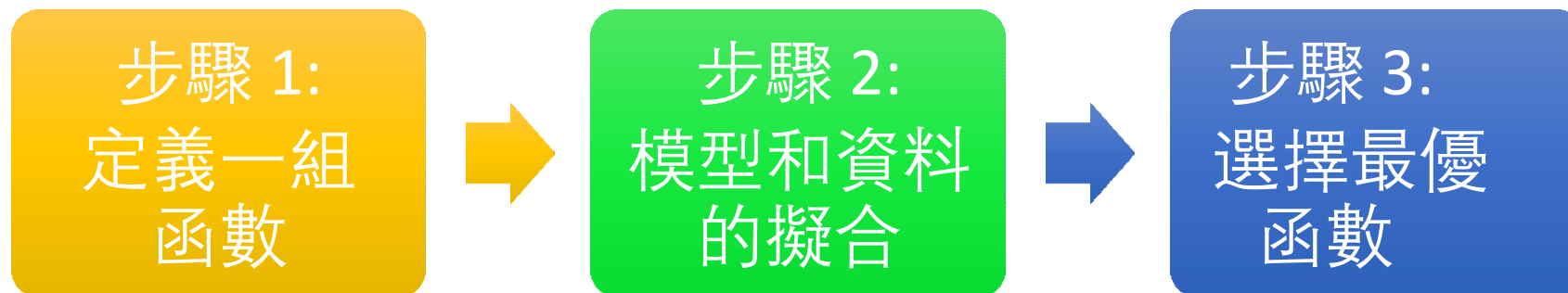
CAT, DOG, DUCK

單個物體

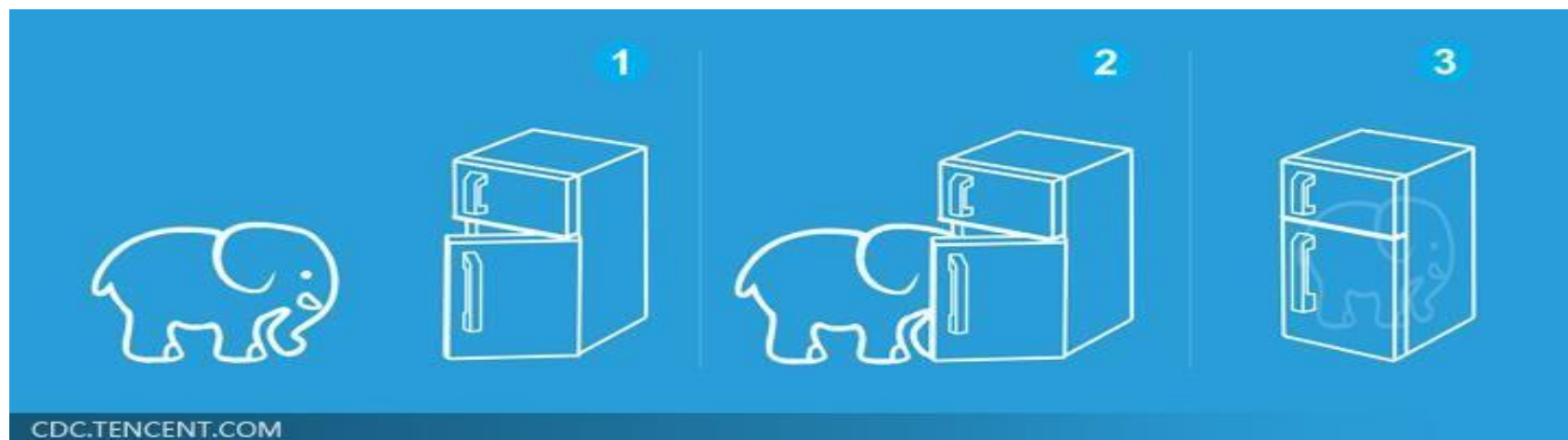
多個物體

圖像分類：給定一張圖片，識別類別

深度學習三個步驟.....



深度學習是如此簡單.....

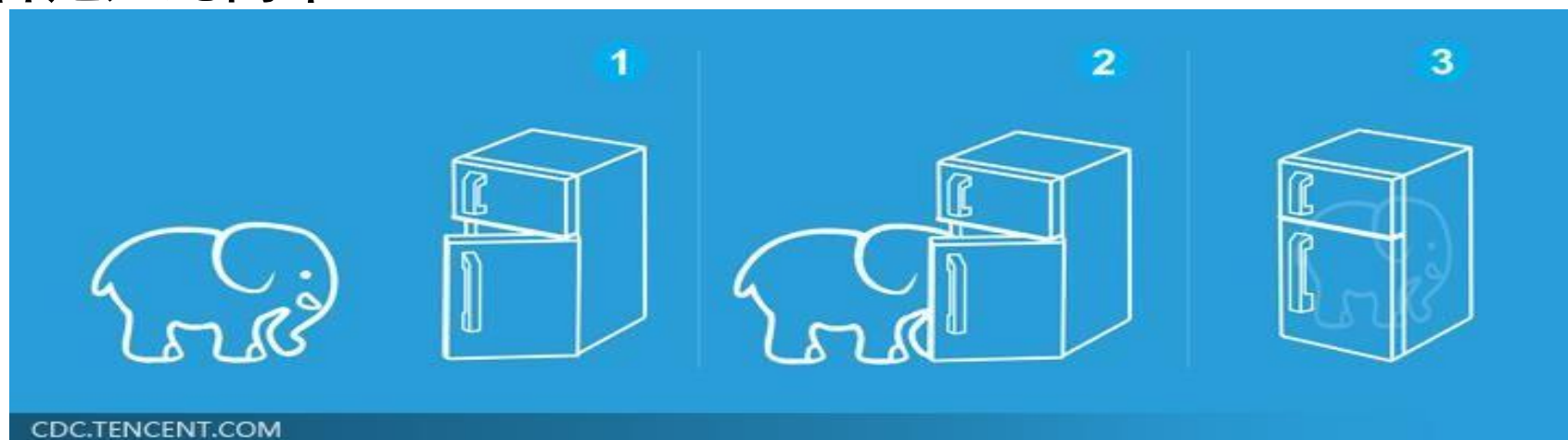


圖像分類：給定一張圖片，識別類別

深度學習三個步驟.....



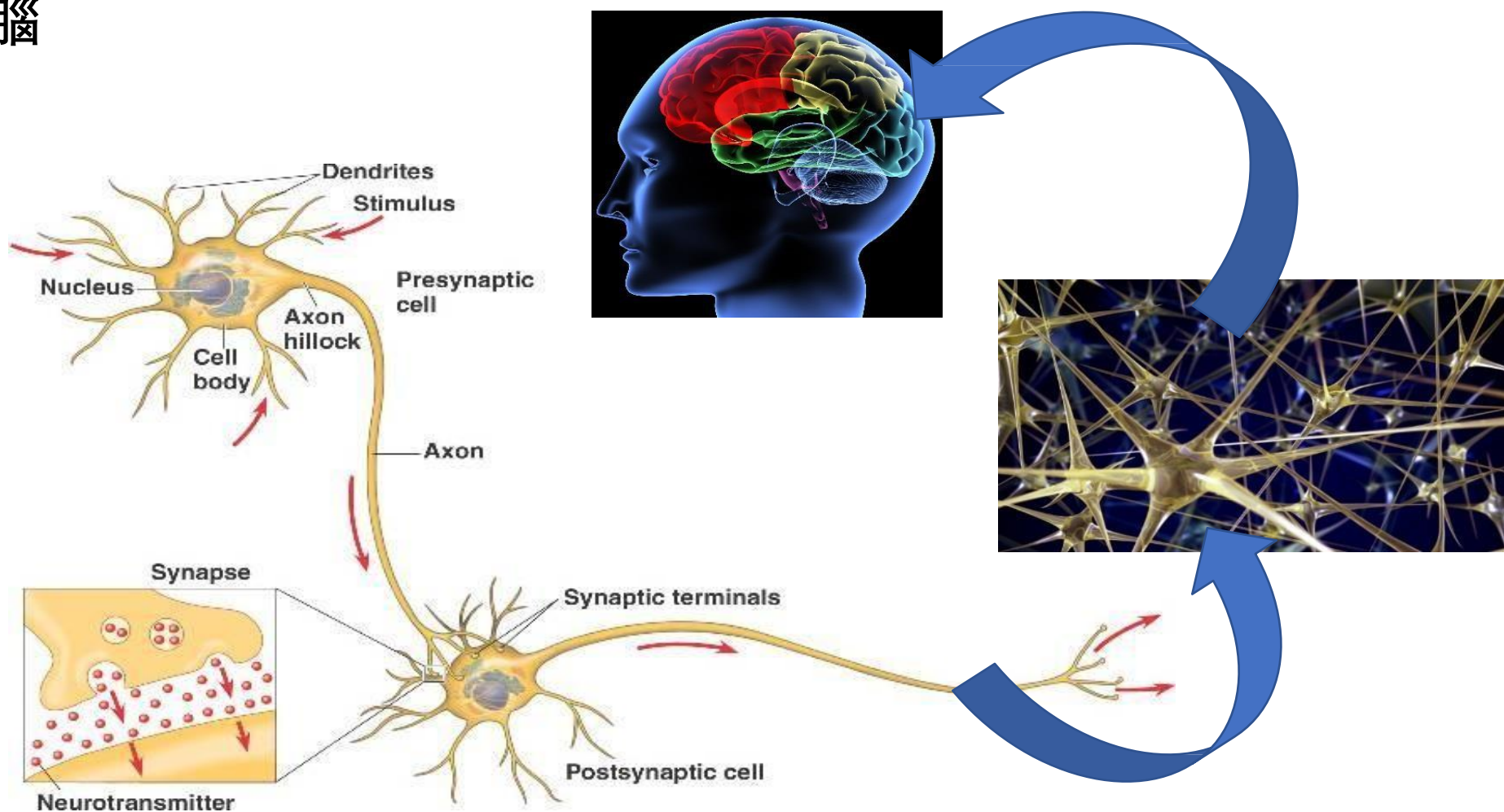
深度學習是如此簡單.....



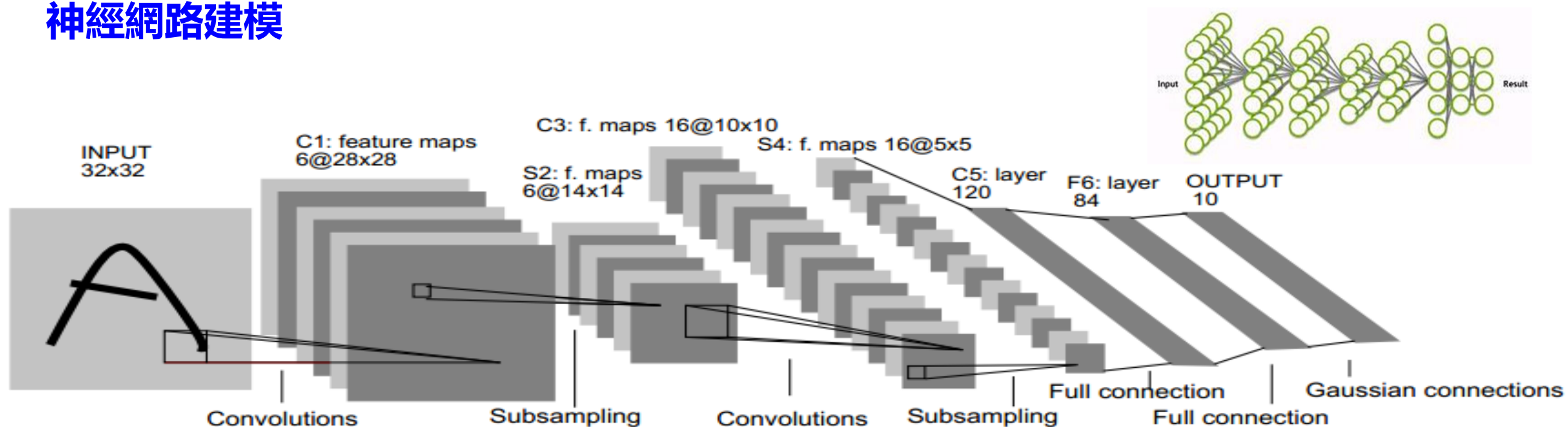
神經網路建模

人類的大腦

120到140億個



神經網路建模



C1層：

輸入圖片大小： 32*32 卷積窗大小： 5*5
卷積窗種類： 6 輸出特徵圖數量： 6
輸出特徵圖大小： 28*28 神經元數量： 4707
連接數： 12304 $[(5*5+1)*6]*(28*28)$
可訓練參數： 156 $[(5*5+1)*6]$

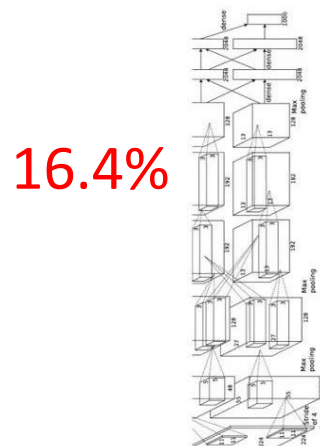
➤卷積過程包括：用一個可訓練的濾波器 f_x 去卷積一個輸入的圖像，然後加一個偏置 b_x ，得到卷積層 C_x 。

➤子採樣過程包括：每鄰域四個圖元求和變為一個圖元，加權再增加偏置，通過一個啟動函數，產生一個縮小四倍的特徵映射圖

計算機視覺技術

神經網路建模

超深網路



16.4%

AlexNet (2012)

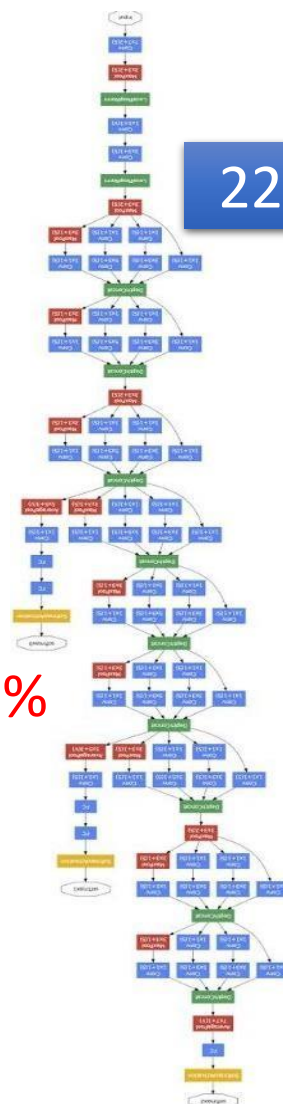
7.3%



VGG (2014)

19 layers

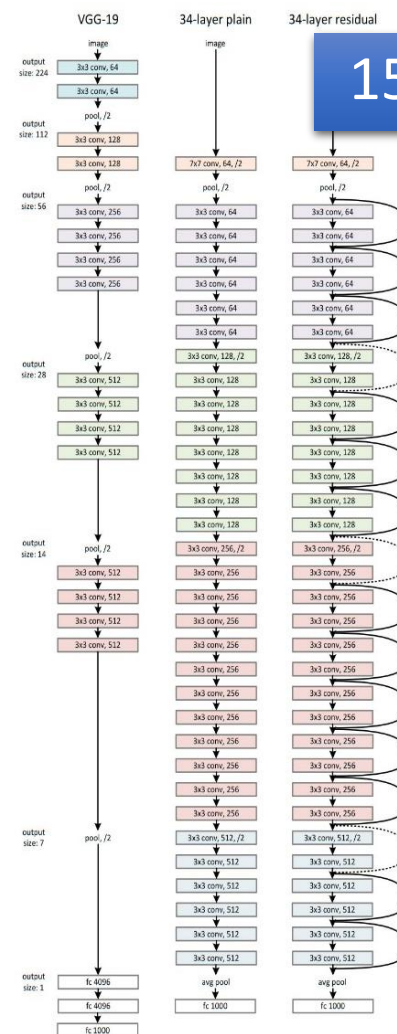
6.7%



22 layers

GoogleNet (2014)

3.5%



152 layers

ResNet (2016)

ImageNet 挑戰賽

IMAGENET Large Scale Visual Recognition Challenge

Steel drum

The Image Classification Challenge:
1,000 object classes
1,431,167 images



Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

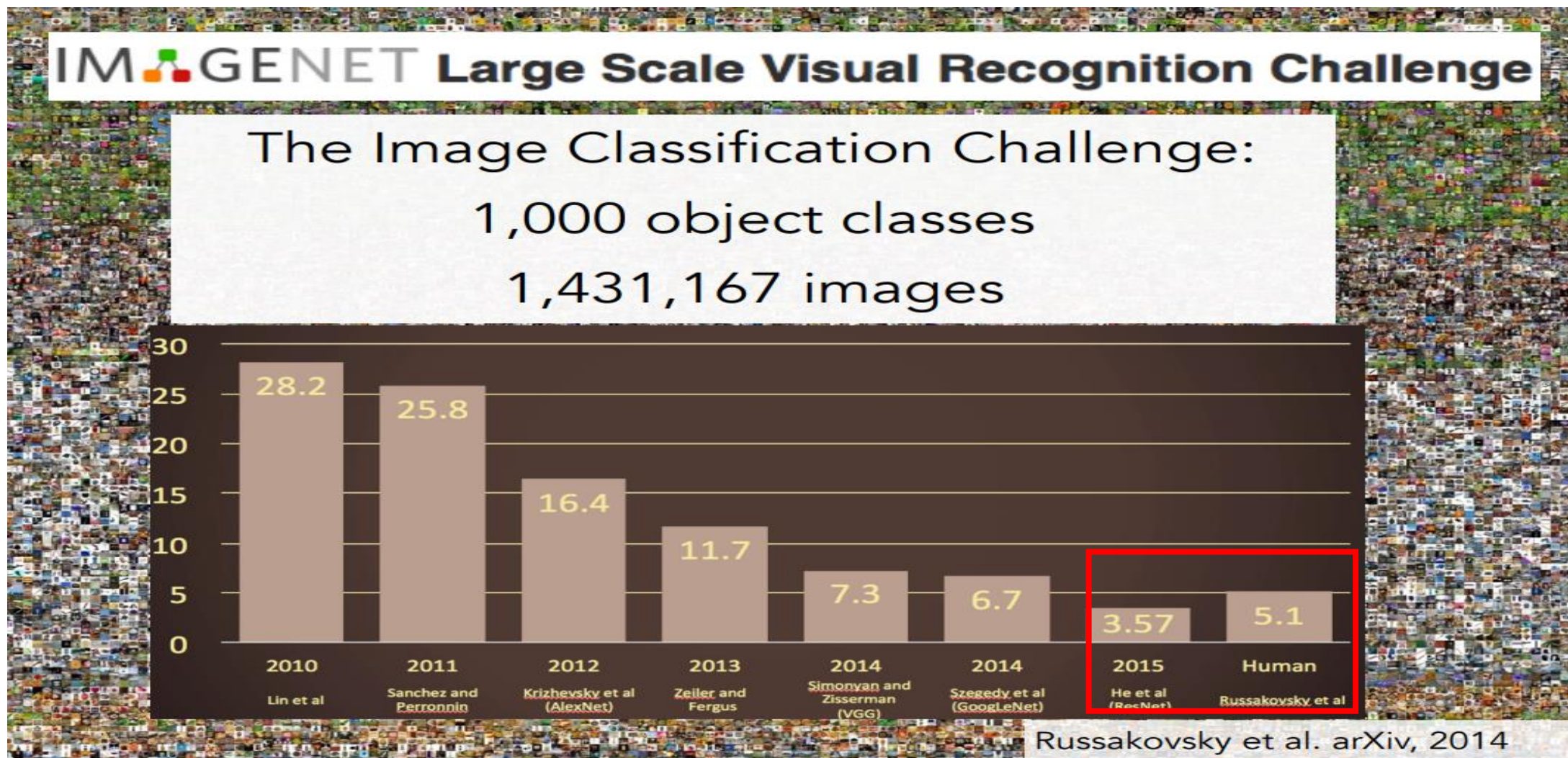
✓

Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

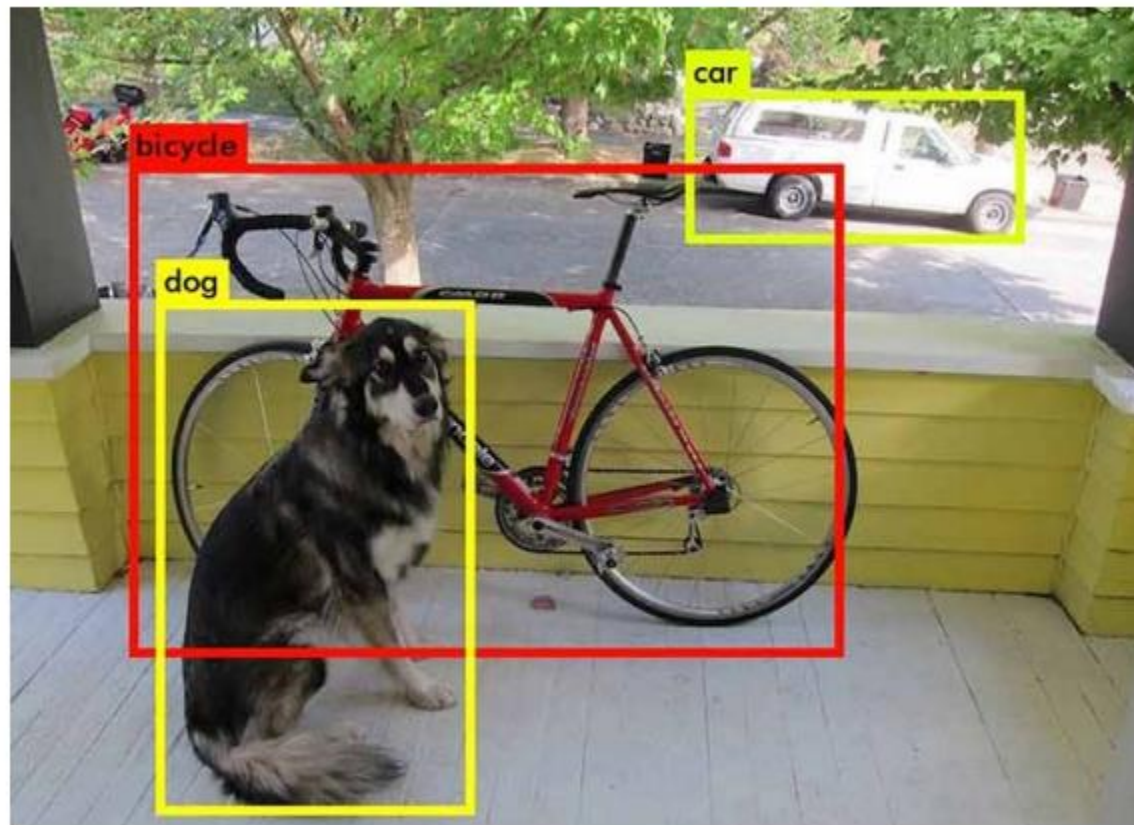
✗

Russakovsky et al. arXiv, 2014

ImageNet 挑戰賽

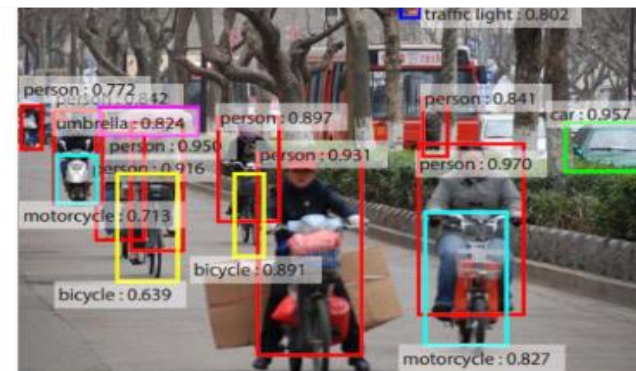
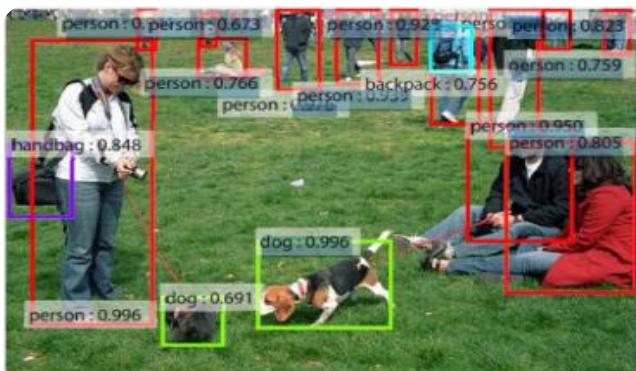


物體檢測



物體檢測的定義

- 物體檢測：判斷一副圖像上是否存在感興趣的物體，如果存在，就給出所有感興趣物體的類別和位置（**What** and **Where**）



物體檢測相關任務對比

分類
(Classification)



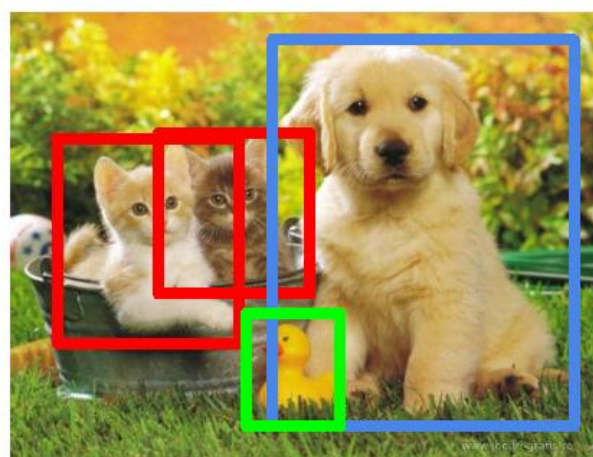
CAT

分類 + 定位
(Classification+Localization)



CAT

物體檢測
(Object Detection)



CAT, DOG, DUCK

實例分割
(Instance Segmentation)



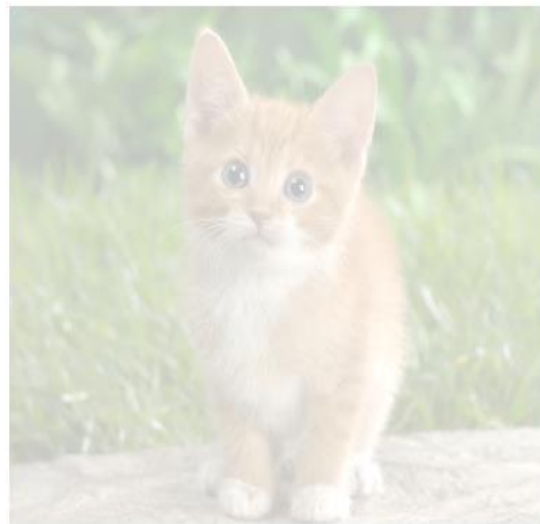
CAT, DOG, DUCK

單個物體

多個物體

物體檢測相關任務對比

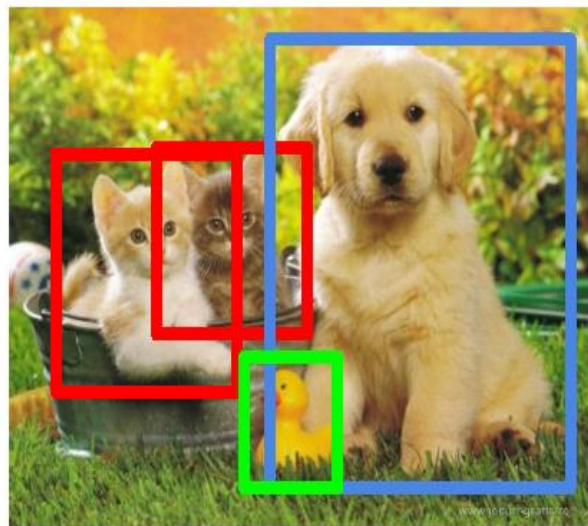
分類
(Classification)



分類 + 定位
(Classification+Localization)



物體檢測
(Object Detection)

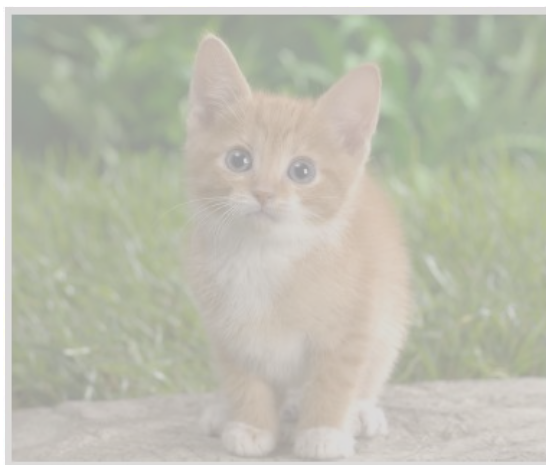


實例分割
(Instance Segmentation)



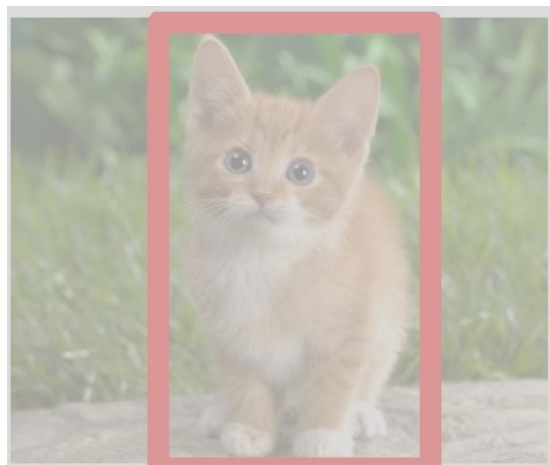
語義分割：在圖元級別上的分類，屬於同一類的圖元都要被歸為一類。

分類
(Classification)



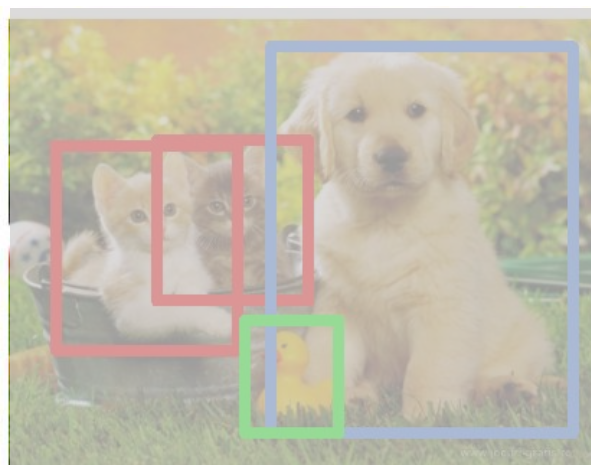
CAT

分類 + 定位
(Classification+Localization)



CAT

物體檢測
(Object Detection)



CAT, DOG, DUCK

實例分割
(Instance Segmentation)

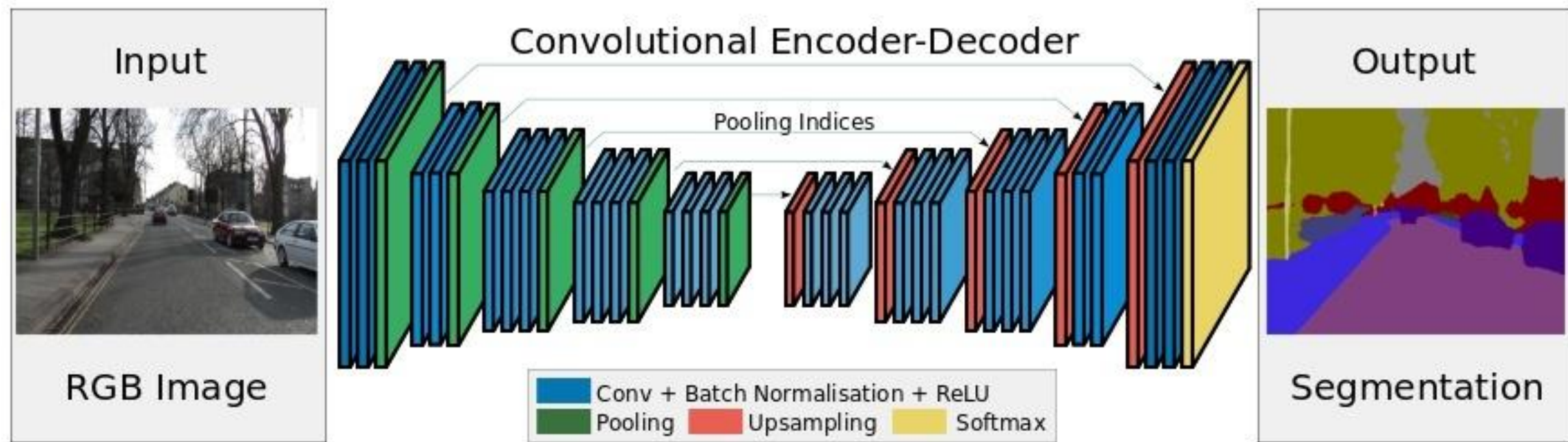


CAT, DOG, DUCK

單個物體

多個物體

語義分割：在圖元級別上的分類，屬於同一類的圖元都要被歸為一類。

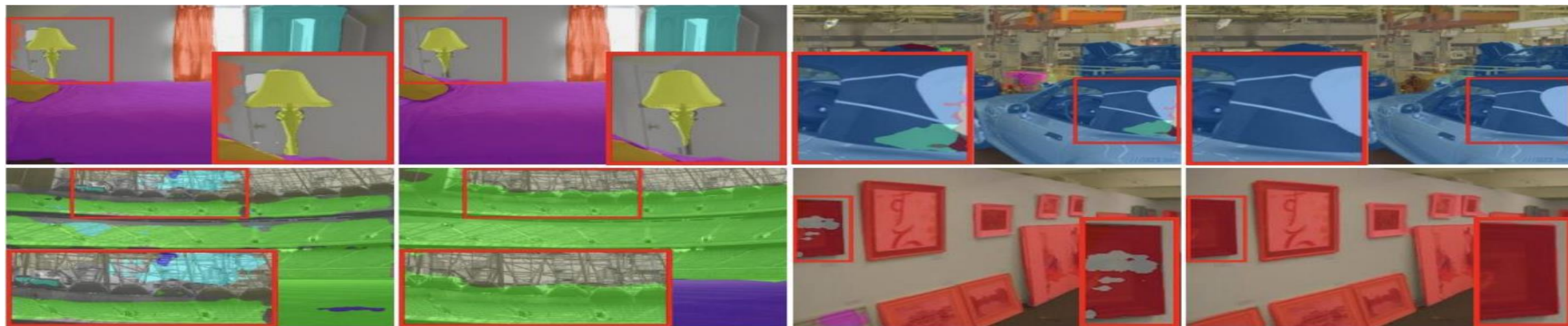


語義分割步驟：

- ① 輸入圖像：輸入一張待分割的圖像
- ② 特徵編碼器：使用卷積神經網路，將待分割圖像編碼到特徵空間
- ③ 圖像解碼器：將編碼特徵進行反卷積操作，映射到原圖像大小，得到輸出語義分割結果

計算機視覺技術

語義分割：在圖元級別上的分類，屬於同一類的圖元都要被歸為一類。





YOLOV5同时做目标检测和语义分割

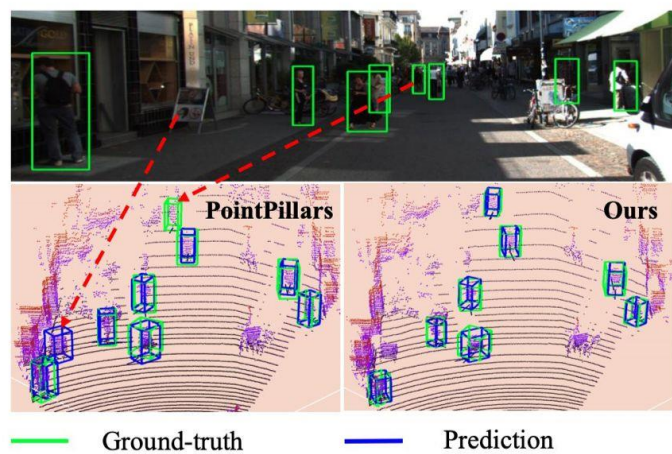
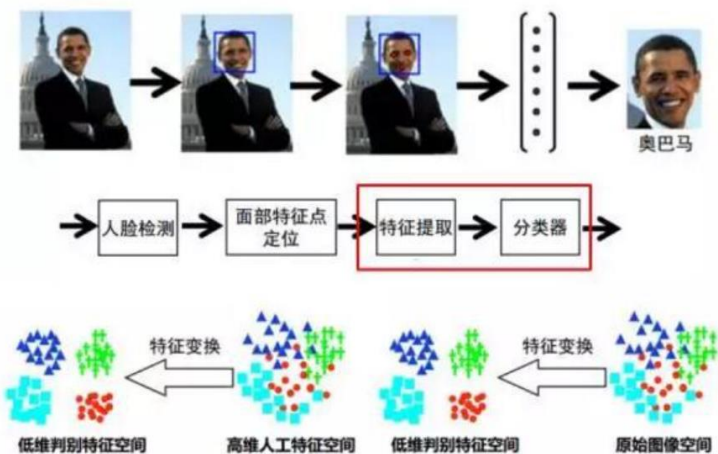


截图打开哔哩哔哩APP
立即观看完整视频

bilibili

計算機視覺技術

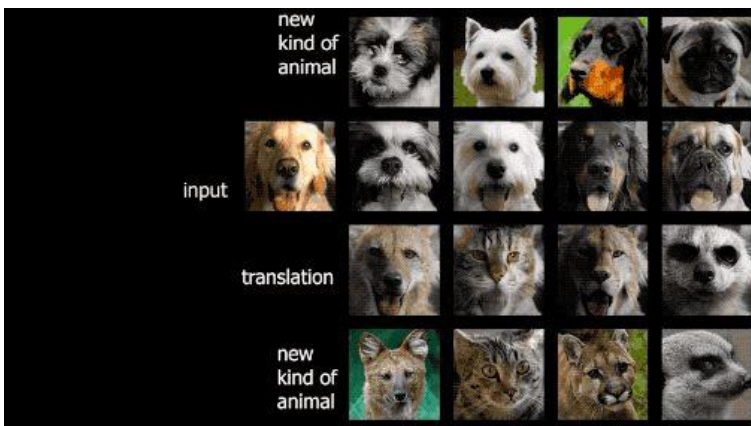
電腦視覺技術的其他應用



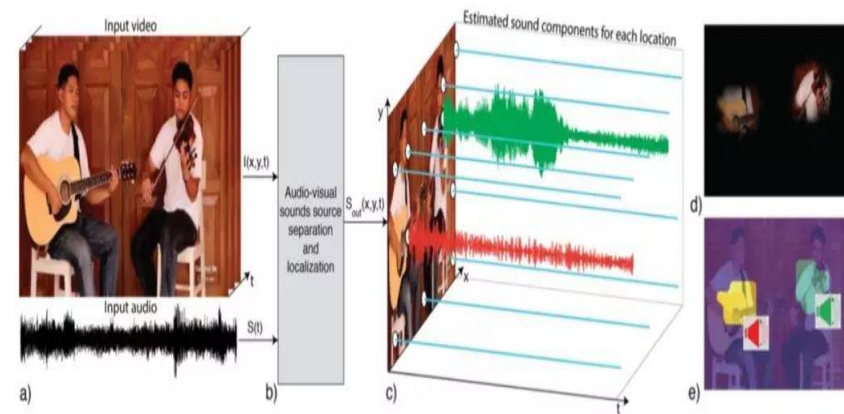
風格遷移



行人重識別



圖像生成



聲源定位

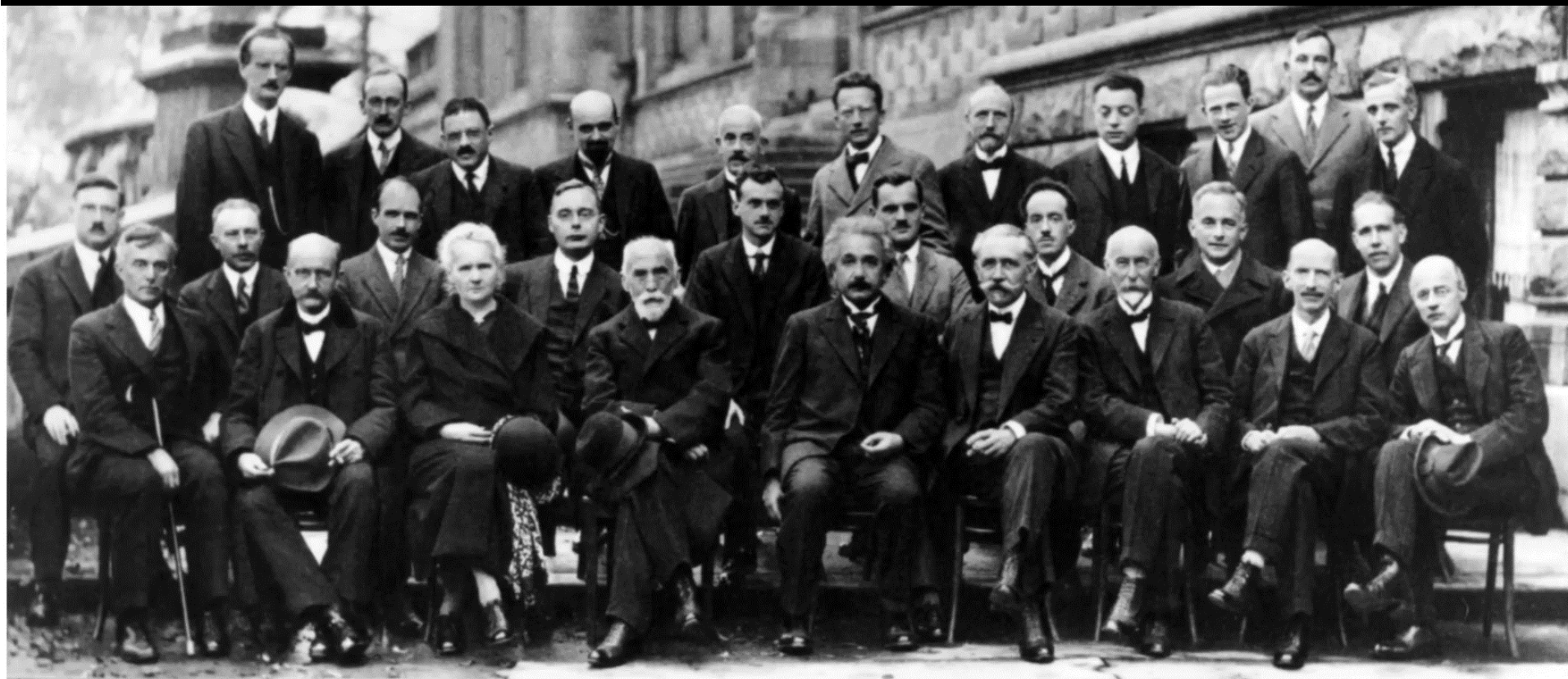


Flute



Violin

Face Restoration



数字孪生系统

計算機視覺技術

小智駕駛助手APP

小智駕駛助手是一款基於手機端的輔助駕駛軟體。

專案簡介

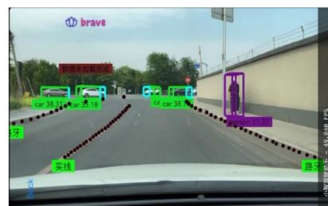
輔助駕駛需要檢測視覺場景中的車輛、行人、車道線等物體，特別是物體準確的3維位置，小智駕駛助手可以利用演算法做出危險駕駛預警，安全距離預警、車道線偏離預警等輔助駕駛功能。

技術特點

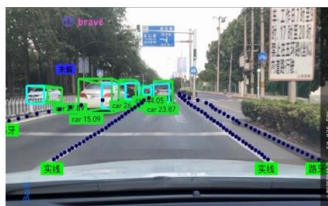
- 即時高效的3D目標檢測
- 多屬性車道線檢測
- ADAS輔助駕駛功能

應用領域

- 智能駕駛
- 輔助駕駛場景
- 智慧交通、智慧城市



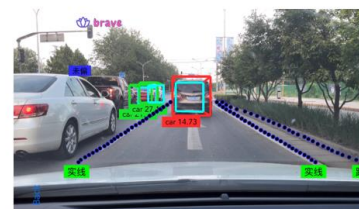
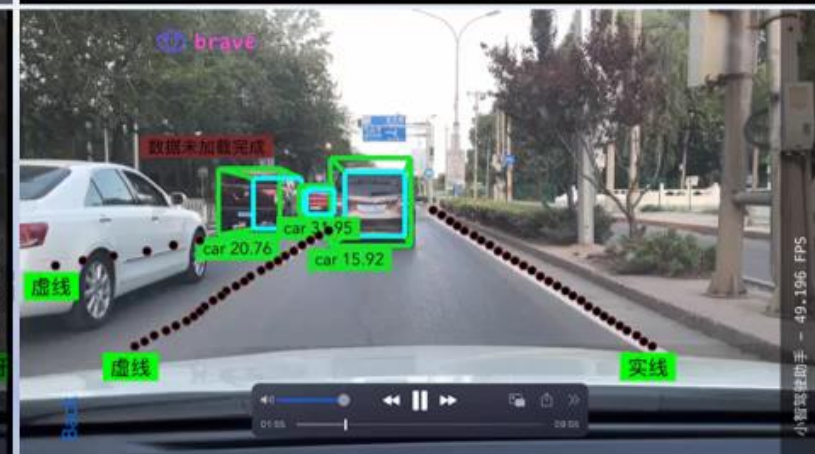
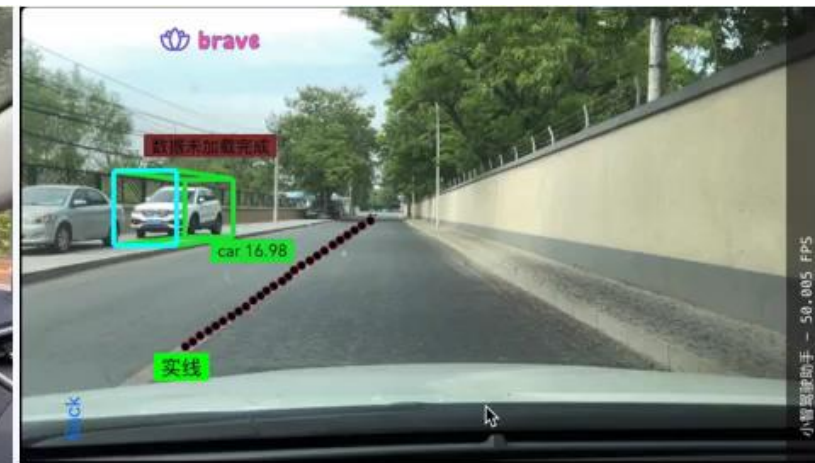
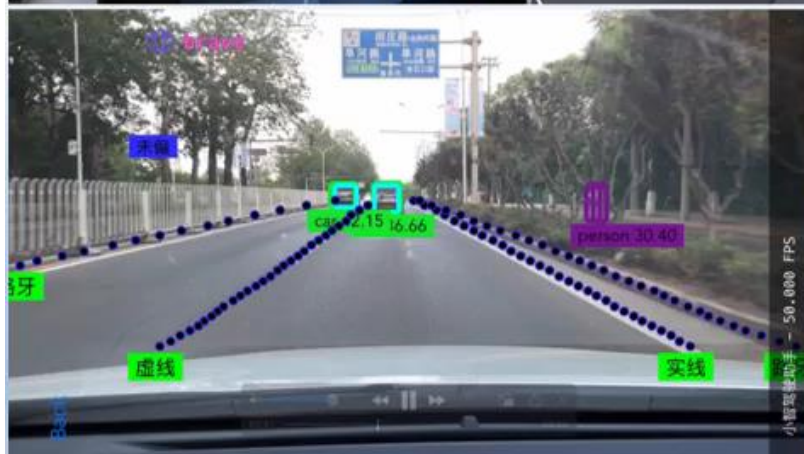
車輛、行人檢測示例



多屬性車道線檢測示例



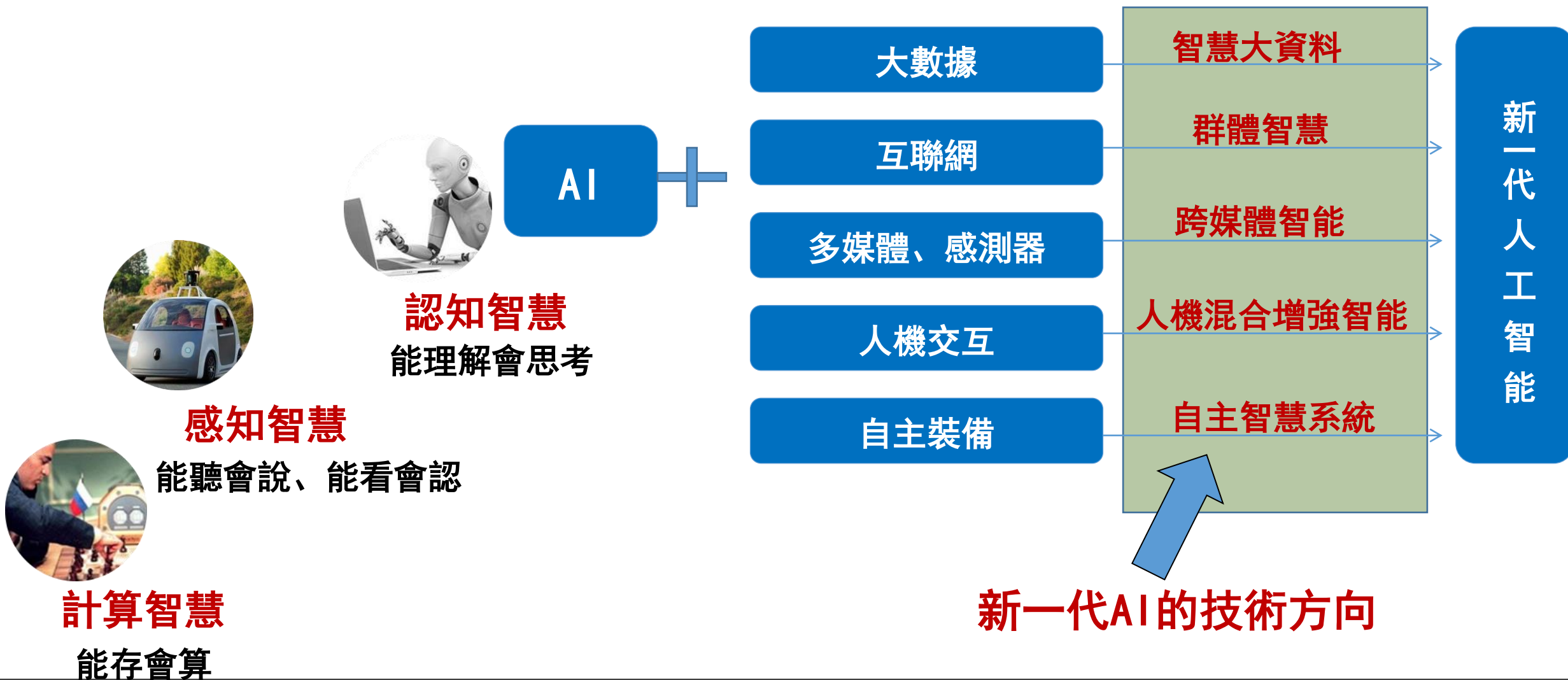
APP Store搜索 - “小智駕駛助手”體驗



車距預警功能

人工智慧未來發展方向

人工智慧從計算智慧到感知智慧，走向認知智慧





謝謝聆聽