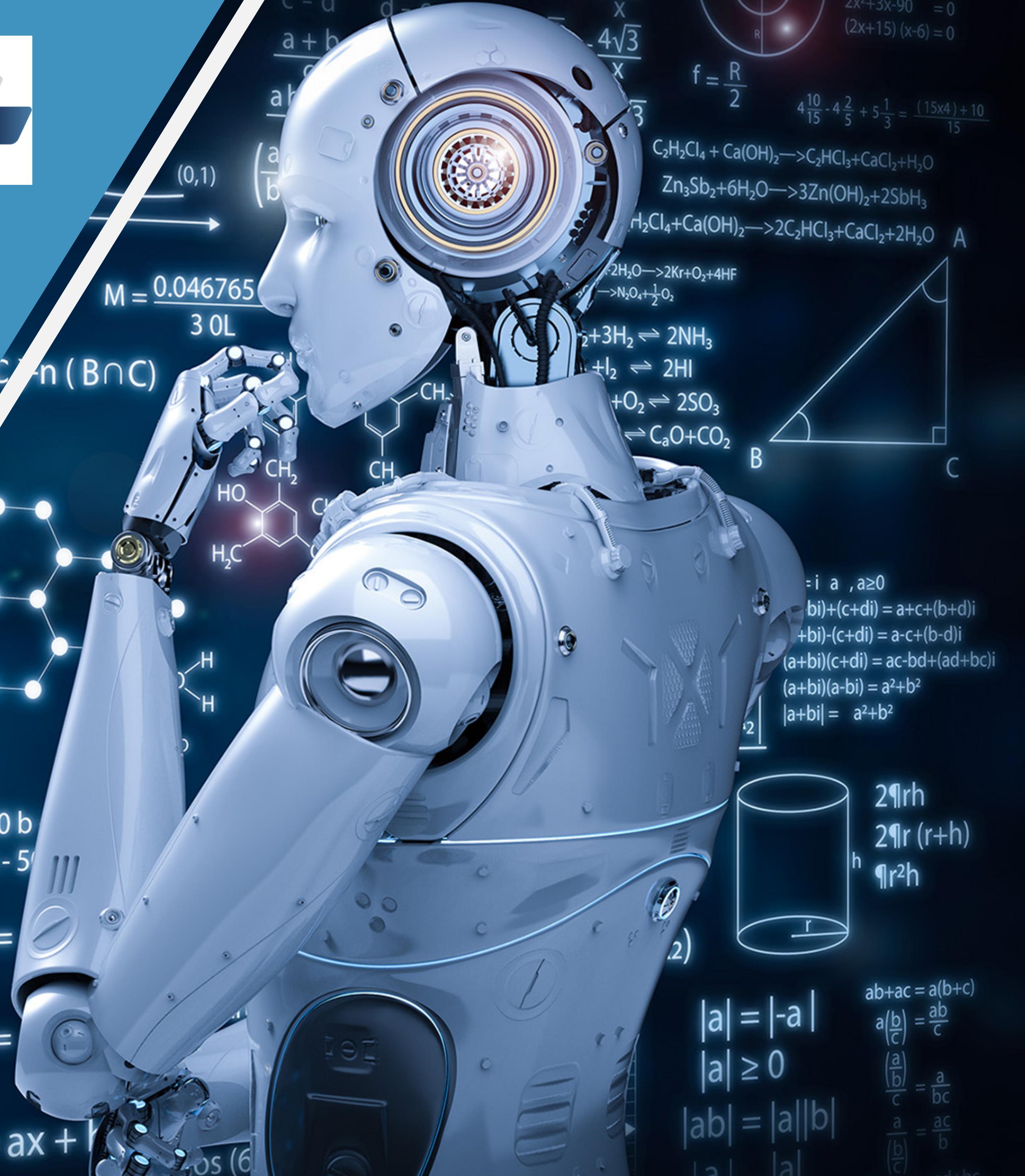




Day 11

深度學習與電腦視覺 學習馬拉松

Cupay 陪跑專家：楊哲寧





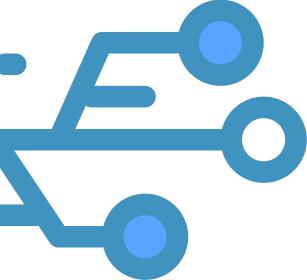
深度學習理論與實作

深度學習簡介- CNN 概念與原理



重要知識點

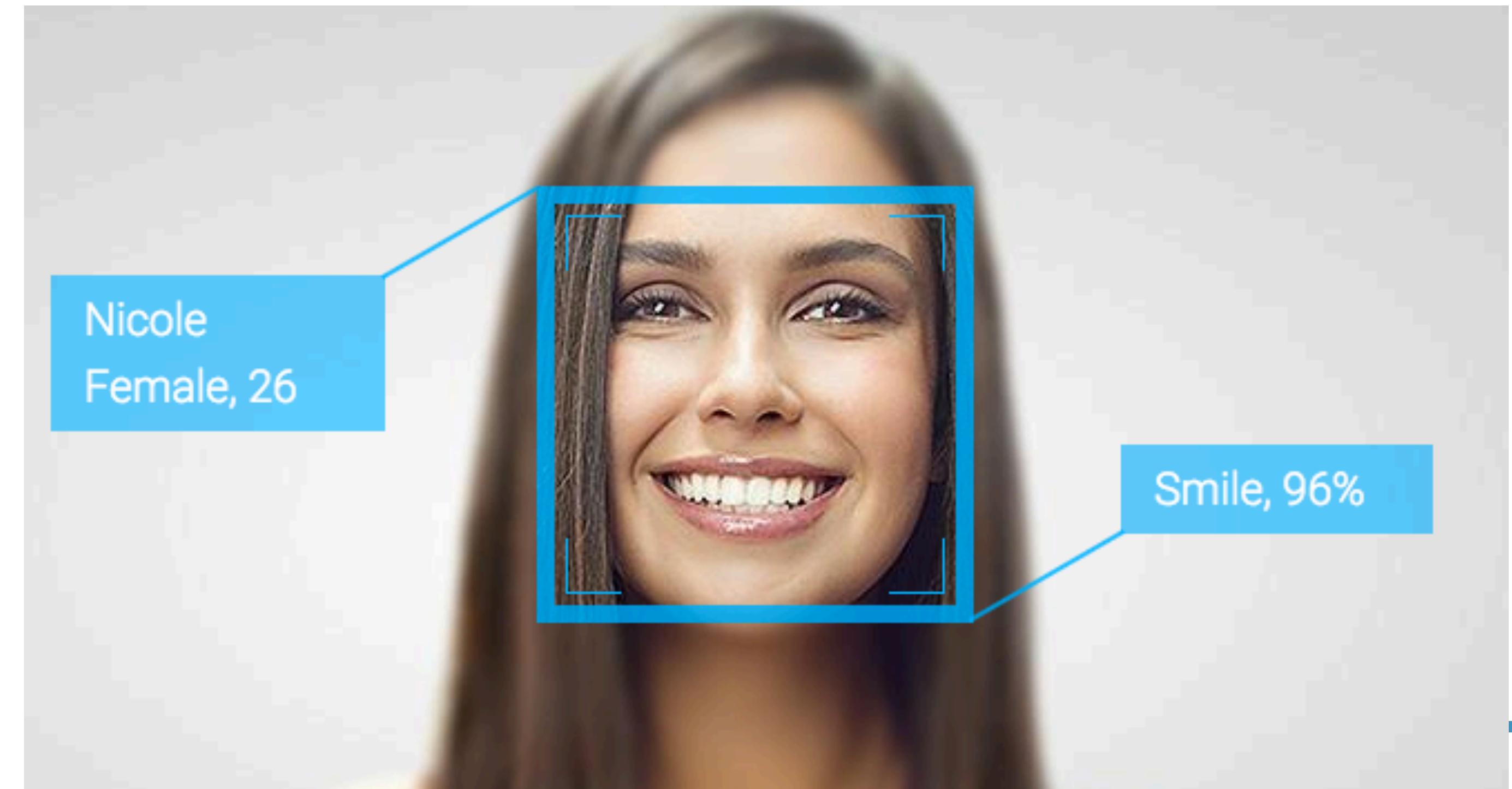
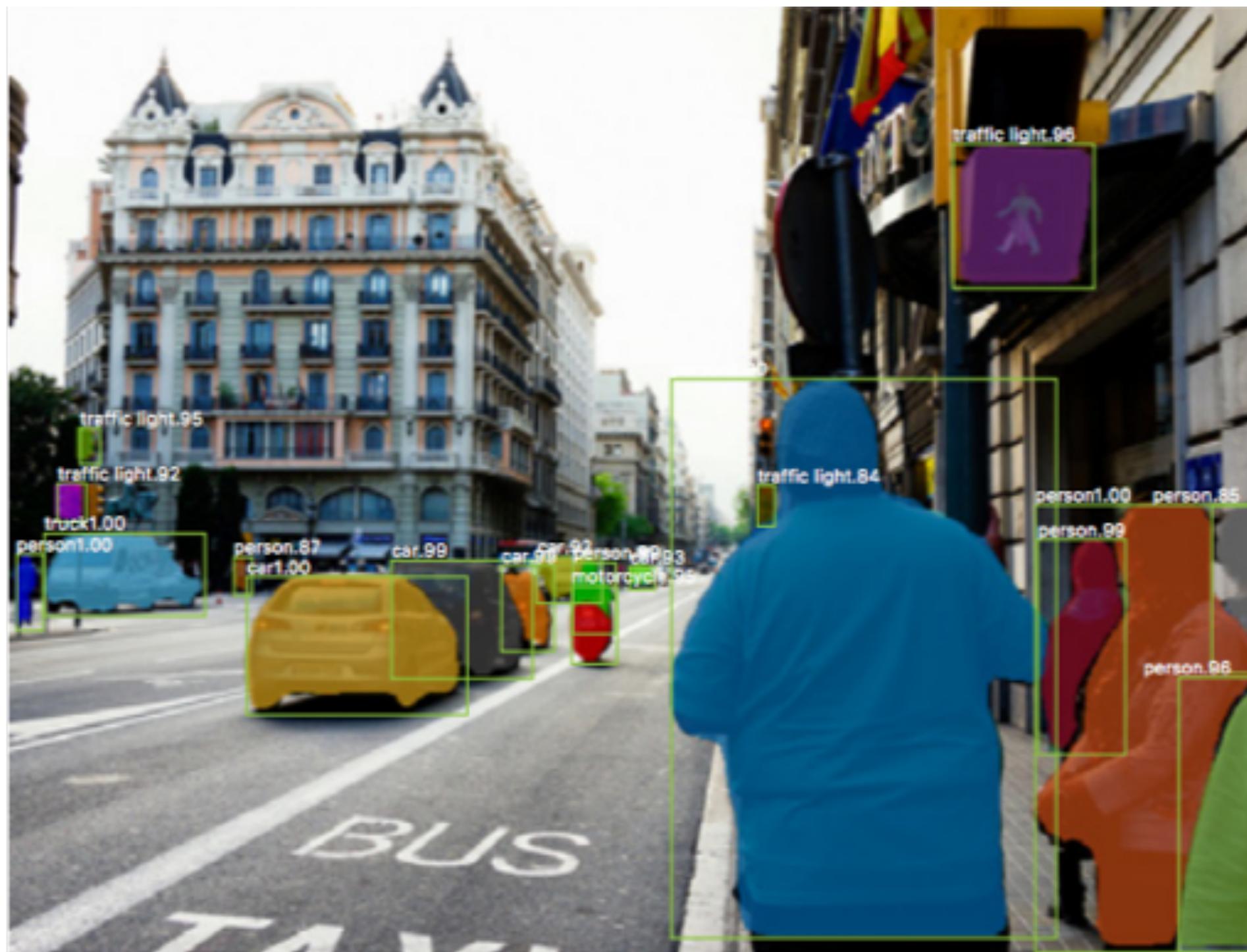
- 卷積神經網路 (CNN) 能用來解決怎樣的問題？
- 全連接神經網路與卷積神經網路 (CNN) 的差異？
- 卷積神經網路 (CNN) 的原理與優勢

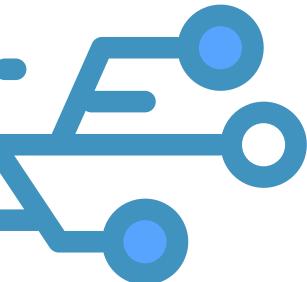


卷積神經網路 (CNN)



CNN 為深度學習 Computer Vision(CV) 領域的入門必備知識，諸多應用都是由CNN的基礎衍伸，因此搞懂 CNN 的原理相信對之後的課程與實際上的應用都能有更深入的了解。



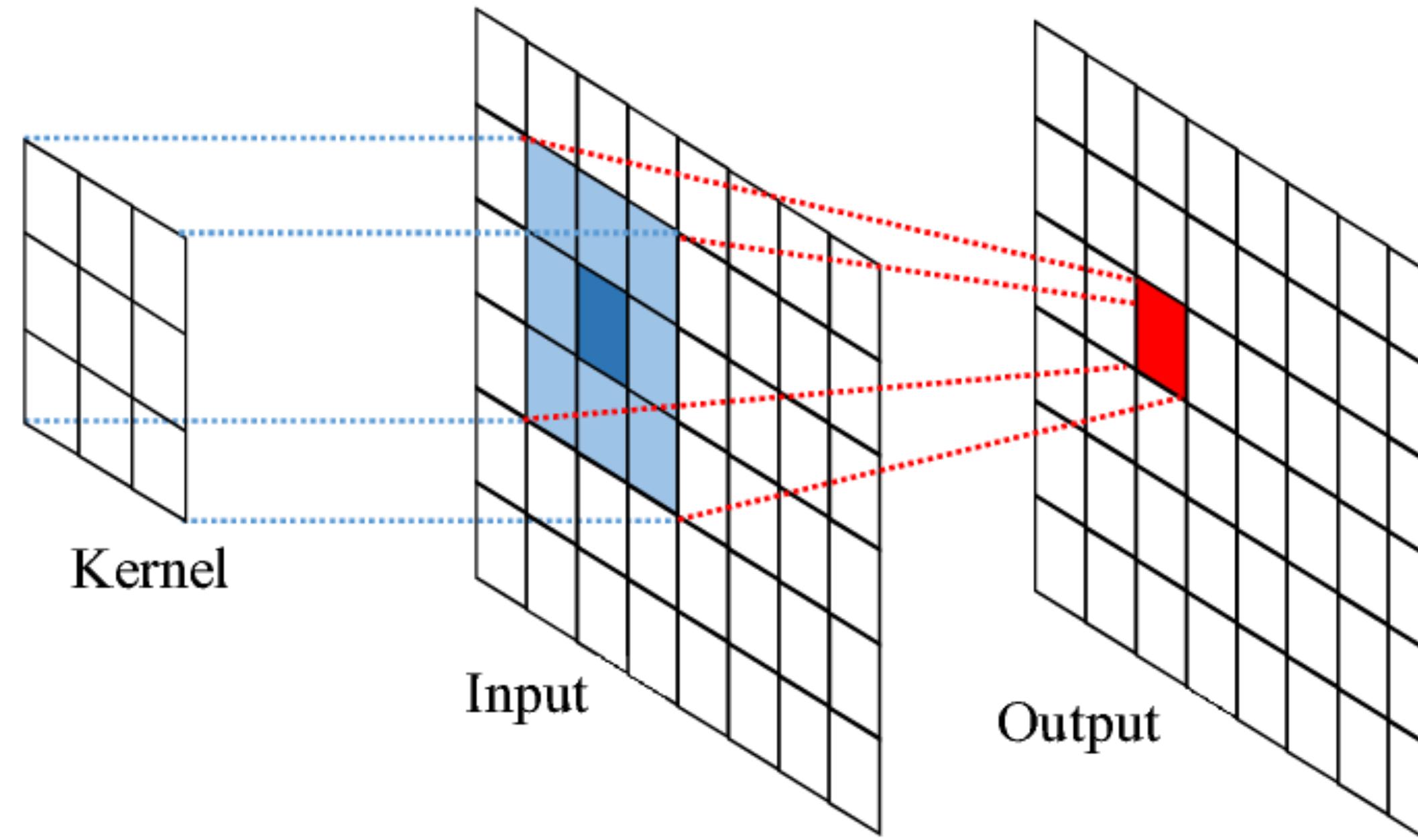


卷積神經網路 (CNN)

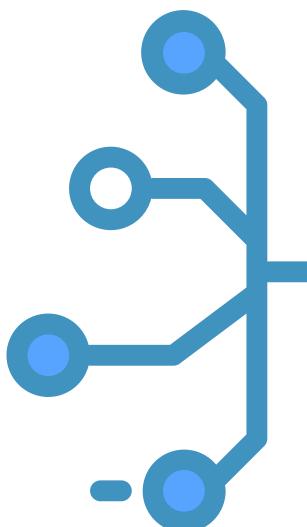


我們先用簡短兩句話描述卷積的概念：

「透過卷積核 (Kernels) 滑動對圖像做訊息提取，
並藉由步長 (Strides) 與填充 (Padding) 控制圖像的長寬。」



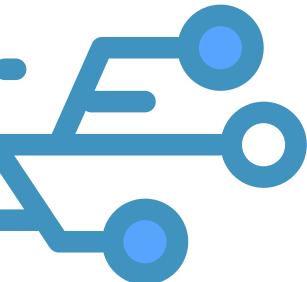
參考來源：[卷積的原理\[Source\]](#)





卷積核

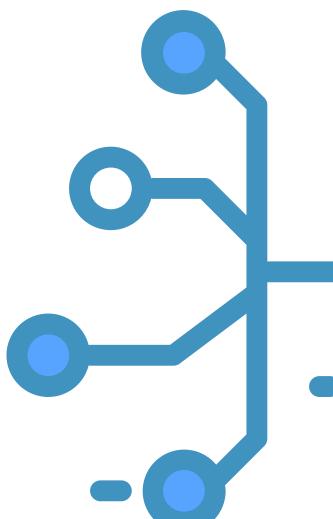
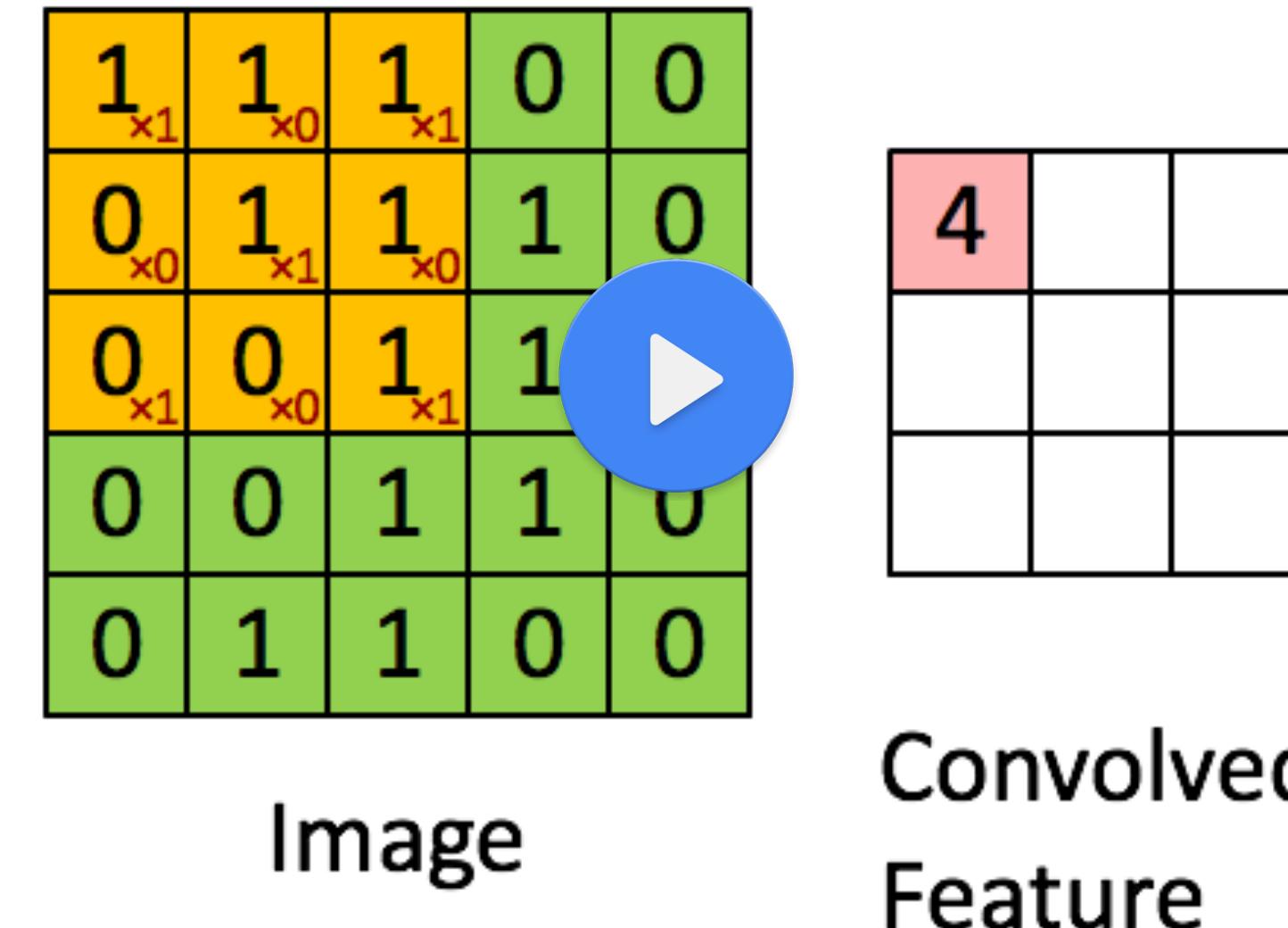


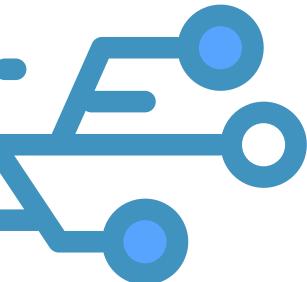


卷積核

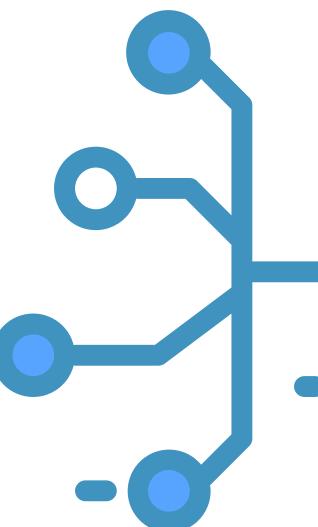


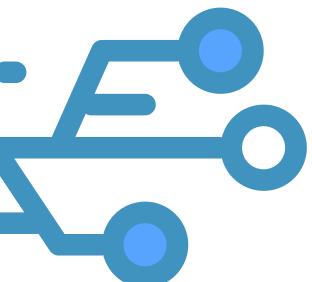
- 又稱為 **Filter**、**Kernel**、**feature detector** (下圖黃色區塊)
- 卷積核大小(**kernel_size**)：自定義，常見大小如 **3*3**、**5*5**。
- 用途：**Kernel** 用來學習圖像的特徵，**Kernel** 的張數 (**filters**) 決定學習的參數量，**Kernel** 大小決定 **Kernels** 的特徵接受域 (Receptive field)，也就是看到特徵的大小。





- **起始值**：上圖中黃色區塊就是一個 $3*3$ 的 Kernel，其中的值就是我們要訓練的權重，通常用常態分佈隨機產生，再經由訓練更新。
- **張數**：控制張數主要就是控制學習的參數量，常見是16、32或64，如果使用16張 Kernels 就能達到很好的效果，也就不需要浪費額外的參數去增加學習與計算量。

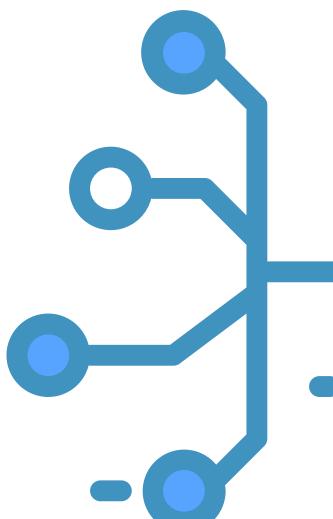
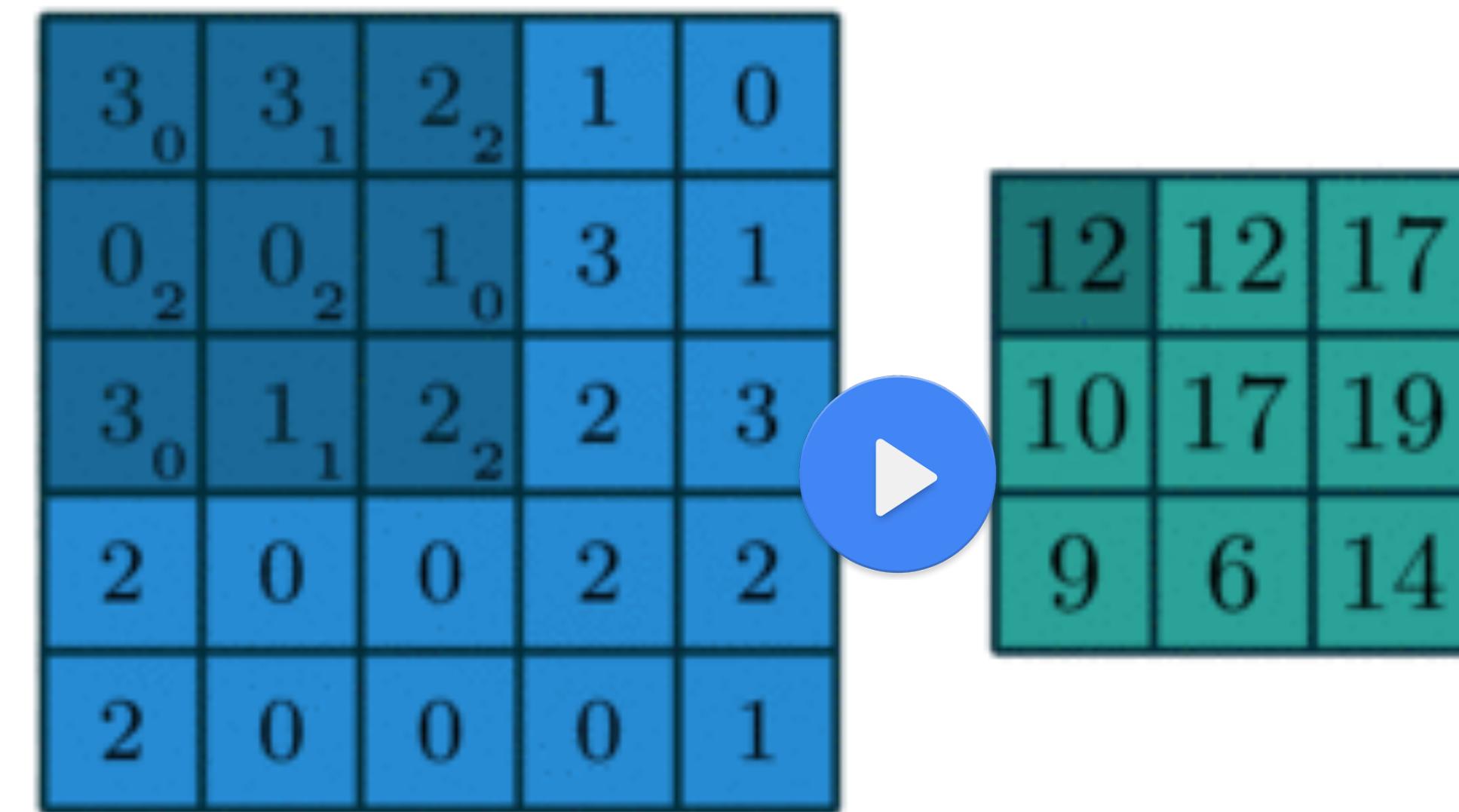


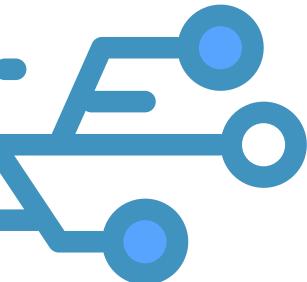


卷積核



運算：我們觀察圖中可以發現 3*3的Kernel 對到我們圖像中的像素也是 3*3，藉由各點相乘後，相加算出一個值，再往下一步走(先往右走，再往下走)，直到走完整張的圖像。

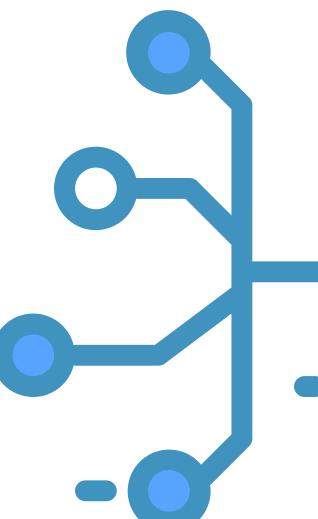


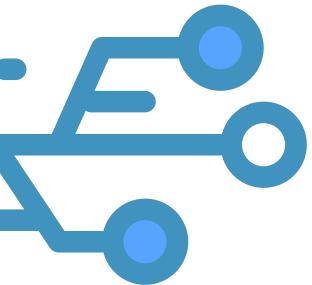


卷積核



- **Kernel大小影響**：Kernel 大小與其 **Receptive field** 有關，Receptive field 直觀來說就是 Kernel 提取資訊的尺度，現在普遍流行的方式是選用不同大小的 Kernel 對圖像做卷積後，再把輸出的 **Feature maps** 合併或平均。
- 常見的Kernel大小有 **1*1, 3*3, 5*5, 7*7**。然而也有人提出，兩層 3*3 的 Kernel 卷積與一層的 **5*5 Kernel** 卷積擁有相近的Receptive field，並且使用較少的參數量，因此大家不妨去嘗試看看不同組合的效果。



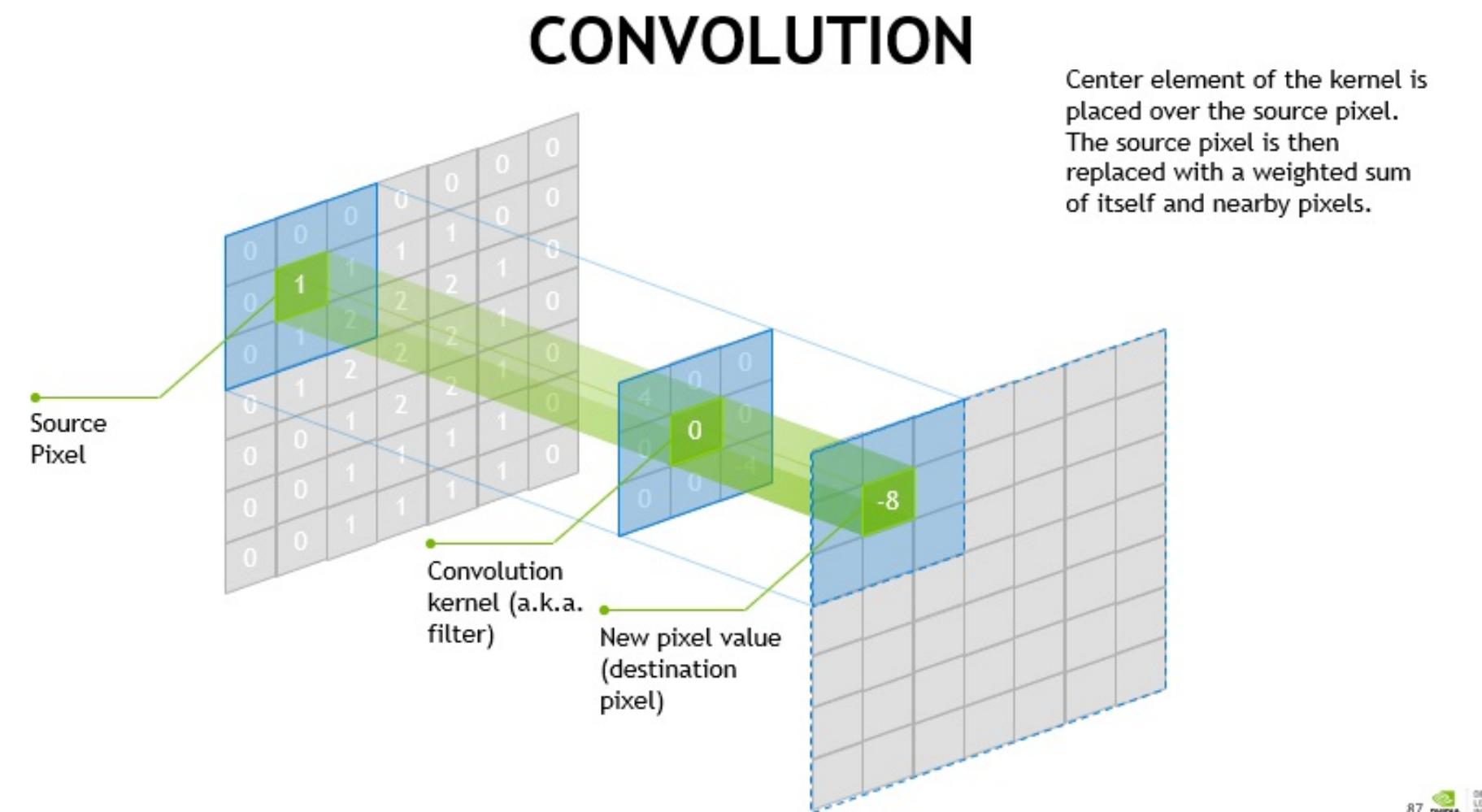


卷積核

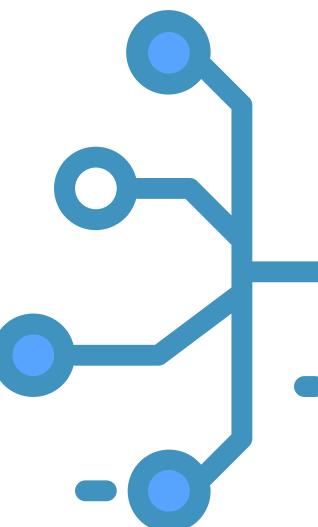


為什麼Kernel都是奇數的呢？

「基數Kernel有幾個先天上的優勢，第一個原因是由於基數的卷積核有中心點，較容易對齊確認位置資訊，再者是因為基數的卷積核能確保Padding的對稱性。」

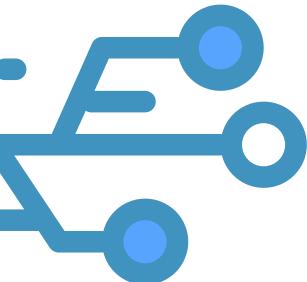


參考來源：[奇數Kernel有中心點\[Source\]](#)





CNN重點與優勢



CNN 的重點與優勢



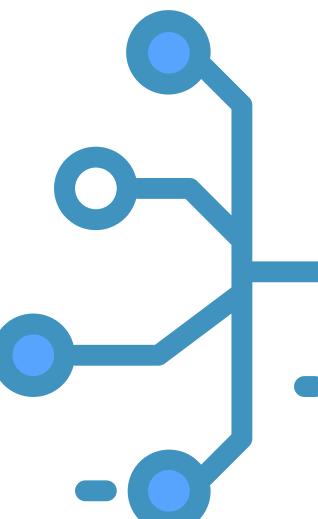
權值共享

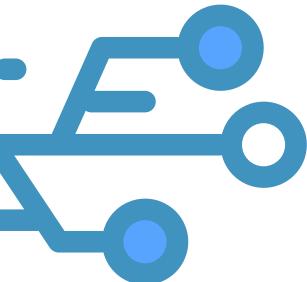
試想今天假如是一般的FC架構，我們輸入一張 $28*28*1$ 的灰階照片(784個特徵)，隱藏層使用100個神經元，那麼我們需要多少個參數？

答案是：

$$100 * 784 \text{ (weights)} + 100 \text{ (bias)} = 78500 \text{ 個參數}$$

單單一層隱藏層參數就瞬間爆量，然而，如果我們使用CNN層的話情況就有所不同了。



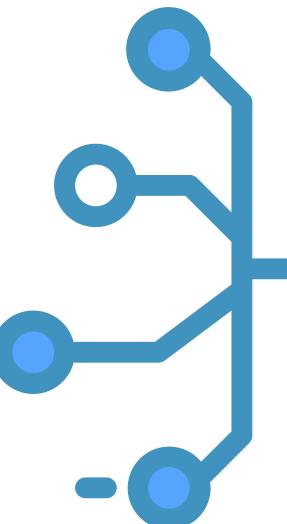
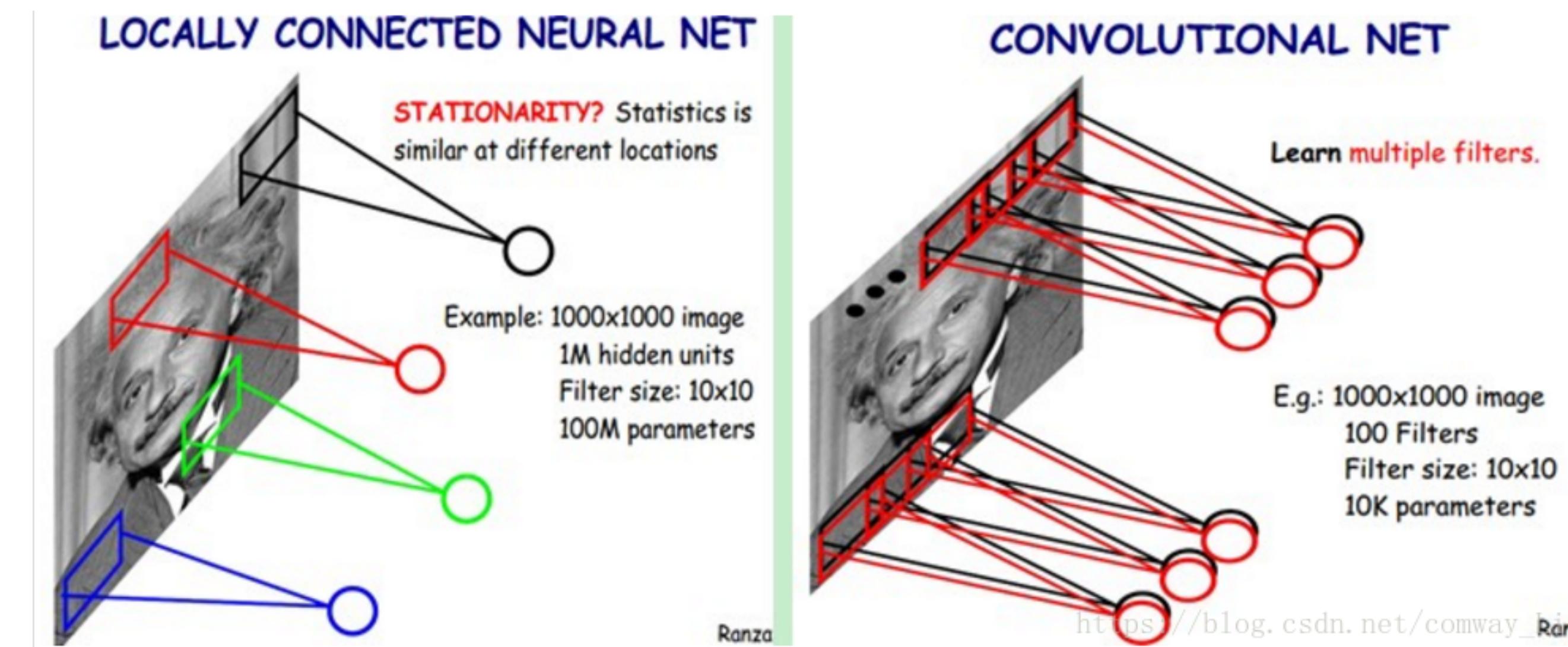


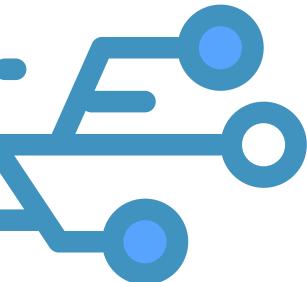
CNN 的重點與優勢



權值共享

在一般的圖像內有許多的特徵是相同的，如特定的輪廓或線條，因此可以用相同幾個神經元組成的卷積核去學這個特徵，透過滑動窗口對整張圖片進行卷積，進而達到節省參數的效果。



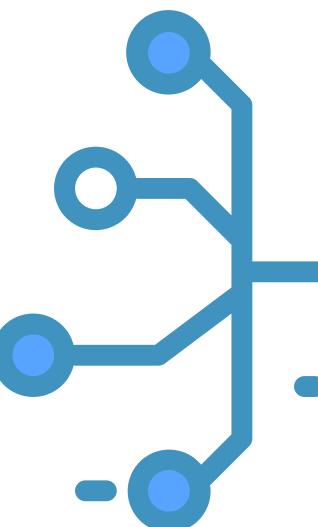


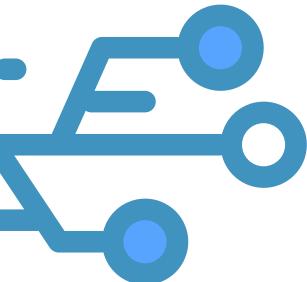
CNN 的重點與優勢



保留空間資訊

CNN 還能保留圖像的位置資訊，圖片中的像素 (Pixels) 與其鄰近的像素會有一定的關聯度，使用 FC 層來訓練圖像資訊的話，要先通過一個展開 (Flatten) 的步驟，把高維的資訊拉成一條一維向量，造成大量空間資訊的流失。

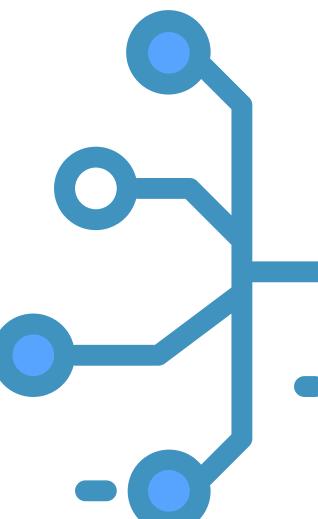




CNN與FC的差異

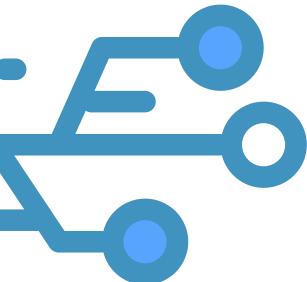


	全連接神經網路 (FC)	卷積神經網路 (CNN)
計算	攤平計算	卷積(保留空間資訊)
參數量	$\text{Input} * \text{output} + \text{output} (\text{bias})$	$(\text{Kernel}_H * \text{W} * \text{channels} + 1) * \text{Kernel numbers}$
優點	保留語意訊息	保留空間訊息



知識點回顧

- CNN 相較全連接網路，不僅運算量較低，更能保留空間資訊，因此較適合使用在影像辨識領域。
- CNN 達到全值共享的效果，相同特徵可藉由特定 Kernel 保留。
- 藉由控制 Kernel 大小，我們可以決定 Kernel 的 Receptive Field，藉此用來偵測度同尺度的物件。
- CNN 還有諸多細節會在下章做跟深入的解釋。



推薦延伸閱讀



DeepLearning.ai- C4W1L01

吳恩達課程

PPT連結

Youtube影片連結



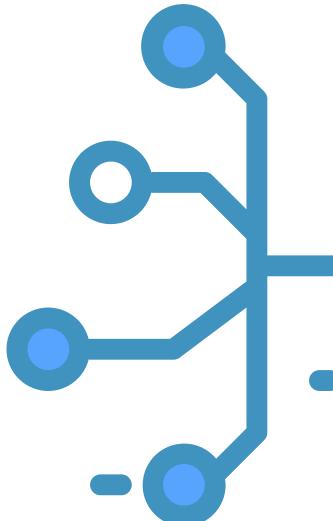
deeplearning.ai

Stanford-CS231n

史丹福大學課程

PDF連結

Youtube影片連結



解題時間 Let's Crack It



請跳出 PDF 至官網 Sample Code & 作業開始解題