

The background features a dark blue gradient with faint, light blue concentric circles and degree markings (40, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260) on the left side, suggesting a technical or scientific theme.

多媒體系統

MULTIMEDIA SYSTEM

張家瑋 博士

助理教授

國立臺中科技大學資訊工程系

FOUNDATIONS OF COMPUTER VISION

電腦視覺的原理與應用



HOW WE TEACH COMPUTERS TO UNDERSTAND PICTURES



電腦視覺

- 利用攝影機和電腦代替人眼對目標進行識別、跟蹤和測量等機器視覺，並做圖像處理，用電腦處理為更適合人眼觀察或傳送給儀器檢測的圖像。
- 人工智慧的主要研究問題是：如何讓系統具備「計劃」和「決策能力」，使之完成特定的動作，如移動機器人通過特定環境。
 - 此問題中，電腦視覺可作為感知器，為決策提供資訊。其中研究方向包括模式識別和機器學習，因此電腦視覺被看作人工智慧的分支。

電腦視覺應用

- 作為一個工程學科，電腦視覺基於相關理論來建立電腦視覺系統。這類系統的組成部分包括：
 1. 過程控制(Process Control) (如工業機器人和無人駕駛車)
 2. 事件監測(Event Monitoring) (如圖像監測)
 3. 資訊組織(Information Organization) (如圖像資料庫和圖像序列的索引建立)
 4. 物體與環境建模 (如工業檢查，醫學圖像分析和拓撲建模)
 5. 交感互動 (如人機互動的輸入裝置)

電腦視覺的相關領域 (1/2)

- **電腦視覺**的研究物件主要是對映到單幅或多幅圖像上的**三維場景**，如三維場景的重建。電腦視覺的研究很大程度上針對圖像的內容。
- **圖像處理**與**圖像分析**的研究物件主要是**二維圖像**，實現圖像的轉化，尤其針對像素級的操作，例如提高圖像對比度，邊緣提取，去雜訊和幾何變換如圖像旋轉。這一特徵表明無論是圖像處理還是圖像分析其研究內容都和圖像的具體內容無關。

電腦視覺的相關領域 (2/2)

- **機器視覺**主要是指**工業領域**的視覺研究，如自主機器人的視覺，用於**檢測**和**測量**的視覺。這表明在這一領域通過**軟體硬體**、**圖像感知與控制理論**與**圖像處理**緊密結合來實現高效的**機器人控制**或各種**實時操作**。
- **模式識別**使用各種方法從訊號中提取資訊，主要運用**統計學**的理論。此領域的一個主要方向便是從圖像資料中提取資訊。

電腦視覺的步驟 (1/3)

1. 圖像取得：數位圖像是由一或多個圖像傳感器產生，傳感器可以是各種攝錄影機，包括X-Ray斷層掃描，雷達，超聲波等，圖片可以是二維、三維圖組或者一個圖像序列。圖片像素值對應在一個或多個光譜上（如灰階圖）。
2. 預處理：對圖像提取某種特定的資訊，使圖像滿足後繼方法的要求。如：
 - 二次取樣保證圖像坐標的正確
 - 平滑去噪來濾除傳感器引入的裝置雜訊
 - 提高對比度來保證實現相關資訊可以被檢測到
 - 調整尺度空間使圖像結構適合局部應用

電腦視覺的步驟 (2/3)

3. 特徵提取：從圖像中提取各種複雜度的特徵。

- 線、邊的提取
- 局部的特徵點檢測，如邊角、斑點檢測
- 更複雜的特徵可能與紋理或形狀有關。

4. 檢測/分割：對圖像分割並提取用於後繼處理的部分。

- 篩選特徵點
- 分割一或多幅圖片中含有特定目標的部分

電腦視覺的步驟 (3/3)

5. 進階處理：資料往往已經精煉到很小的數量，如含有目標物體的部分。

- 驗證得到的資料是否符合前提要求
- 估測特定係數，比如目標的姿態，體積
- 對目標進行分類

COLOR HISTOGRAM



- 顏色都是由紅綠藍三原色（RGB）構成的，所以左圖共有4張直方圖（三原色直方圖 + 最後合成的直方圖）。
- 每種原色都可以取256個值，那麼整個顏色空間共有1600萬種顏色（256的三次方）。
- 針對1600萬種顏色比較直方圖，計算量太大，因此需要簡化。
 - 可以將0~255分成四個區：0~63為第0區，64~127為第1區，128~191為第2區，192~255為第3區。
 - 紅綠藍分別有4個區，總共可以構成64種組合（4的3次方）。

COLOR HISTOGRAM

红	绿	蓝	像素数量	红	绿	蓝	像素数量	红	绿	蓝	像素数量	红	绿	蓝	像素数量
0	0	0	7414	1	0	0	891	2	0	0	1146	3	0	0	11
0	0	1	230	1	0	1	13	2	0	1	0	3	0	1	0
0	0	2	0	1	0	2	0	2	0	2	0	3	0	2	0
0	0	3	0	1	0	3	0	2	0	3	0	3	0	3	0
0	1	0	8	1	1	0	592	2	1	0	2552	3	1	0	856
0	1	1	372	1	1	1	3462	2	1	1	9040	3	1	1	1376
0	1	2	88	1	1	2	355	2	1	2	47	3	1	2	0
0	1	3	0	1	1	3	0	2	1	3	0	3	1	3	0
0	2	0	0	1	2	0	0	2	2	0	0	3	2	0	0
0	2	1	0	1	2	1	101	2	2	1	8808	3	2	1	3650
0	2	2	10	1	2	2	882	2	2	2	53110	3	2	2	6260
0	2	3	1	1	2	3	16	2	2	3	11053	3	2	3	109
0	3	0	0	1	3	0	0	2	3	0	0	3	3	0	0
0	3	1	0	1	3	1	0	2	3	1	0	3	3	1	0
0	3	2	0	1	3	2	0	2	3	2	170	3	3	2	3415
0	3	3	0	1	3	3	0	2	3	3	17533	3	3	3	53929

- 左圖是某張圖片的顏色分佈表，將表中最後一欄提取出來，組成一個64維向量(7414, 230, 0, 0, 8, ..., 109, 0, 0, 3415, 53929)。這個向量就是這張圖片的**特徵值**或者叫**"指紋"**。
- 尋找相似圖片就變成**找出與其最相似的向量**，可利用Pearson相關係數或Cosine相似度算出。

PERCEPTUAL HASH ALGORITHM

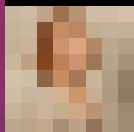
- 利用以下步驟對每張圖片生成“指紋”（fingerprint）字符串來比較不同圖片的指紋。
 1. 縮小尺寸：只保留圖片的結構，明暗等基本資訊，避免不同尺寸的差異。如縮成 8×8 。
 2. 簡化色彩：將圖片轉為64級灰度，所有像素點總共只有64種顏色。
 3. 計算平均值：計算所有64個像素的灰度平均值。
 4. 比較像素的灰度：將每個像素的灰度與平均值比較， \geq 平均值，記為1、 $<$ 平均值，記為0。
 5. 計算 Hash 值：這64位的整數組合，就是圖片的指紋。組合的次序並不重要，只要保證所有圖片都採用同樣次序就行了，在此以 Hash 加密方法來組成。
 - 雜湊演算法(Hash Algorithm)是一種從資料中建立「數位指紋(Digital fingerprint)」的方法，可以將任何長度的資料轉換成一個長度較短的「雜湊值(Hash value)」，又稱為「訊息摘要(MD：Message Digest)」。

PERCEPTUAL HASH ALGORITHM

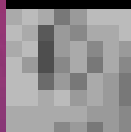
原圖



1. 縮小 (8*8)



2. 簡化色彩 (64級灰度)



8級灰度



32級灰度



3. 計算64個像素的灰度平均值

4. 每個像素與灰度平均值比較



5. 得出 Hash 值

8f373714acfcf4d0

- 得到指紋以後，就可以對比不同的圖片，看看64位中有多少位是不一樣的。如果不相同的數據位不超過5，就說明兩張圖片很相似；如果大於10，就說明這是兩張不同的圖片。



卷積神經網路

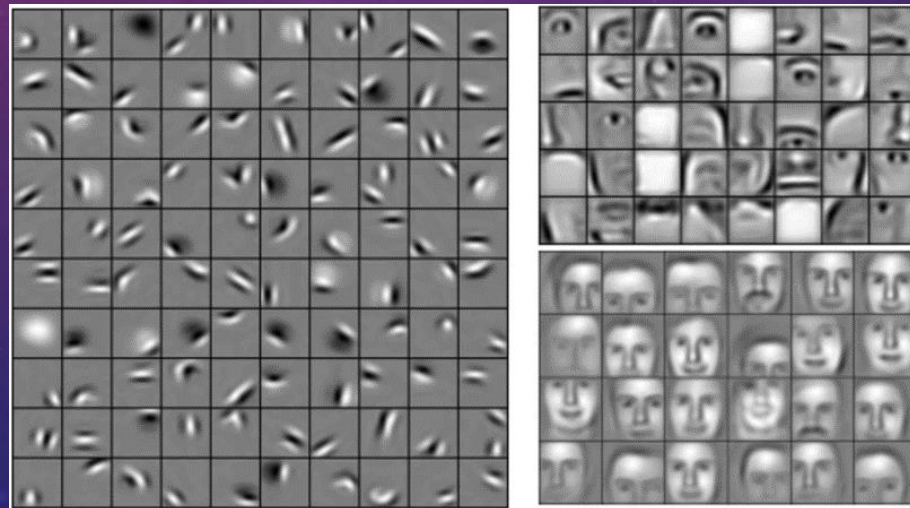
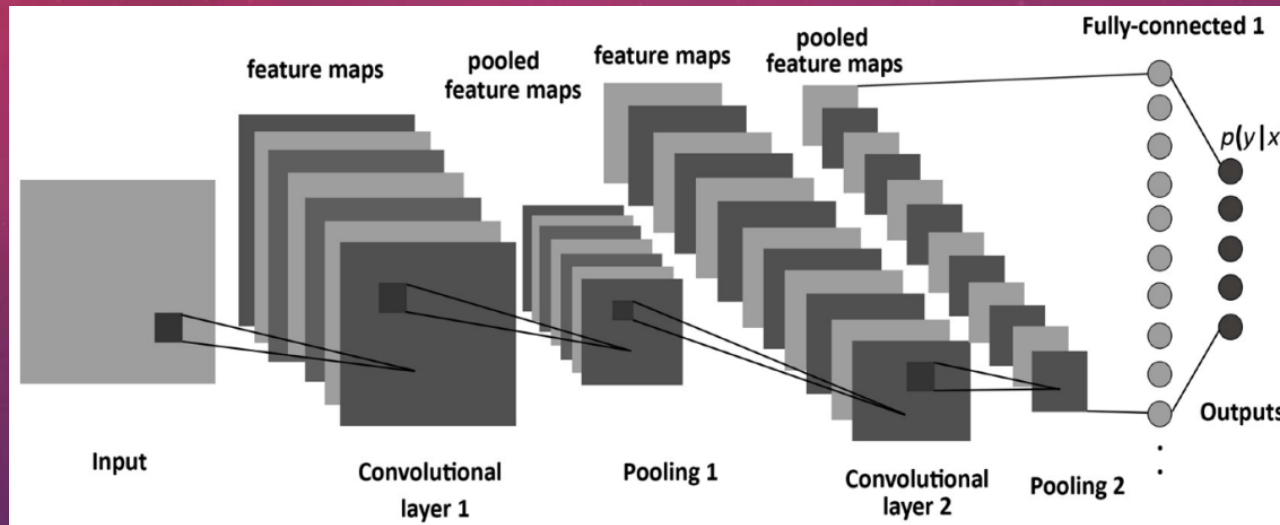
CONVOLUTIONAL NEURAL NETWORK

參考文獻

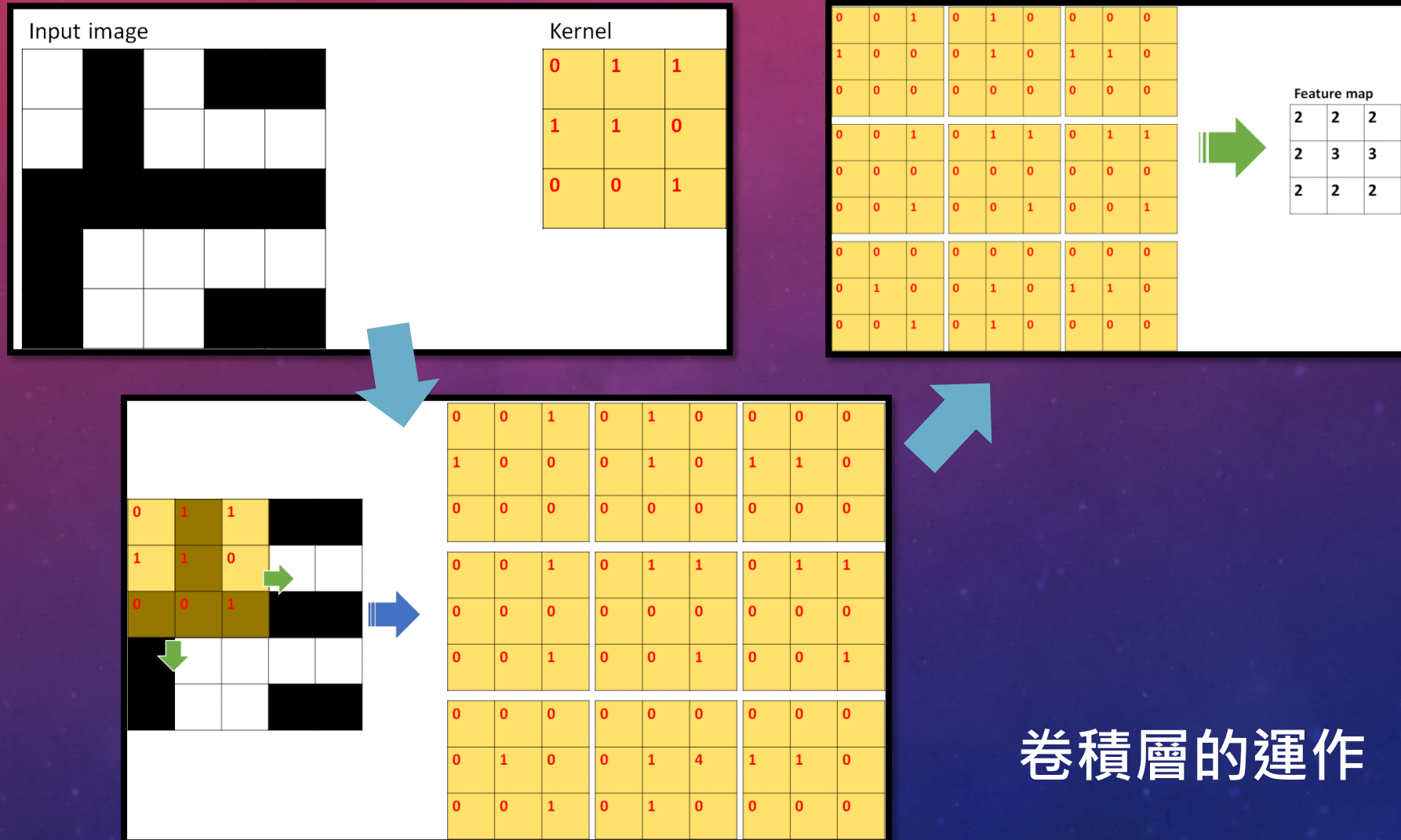
[HTTPS://GOO.GL/FJVJG1](https://goo.gl/FJVJG1)

[HTTPS://GOO.GL/Q5YKPK](https://goo.gl/Q5YKPK)

CONVOLUTIONAL NEURAL NETWORK



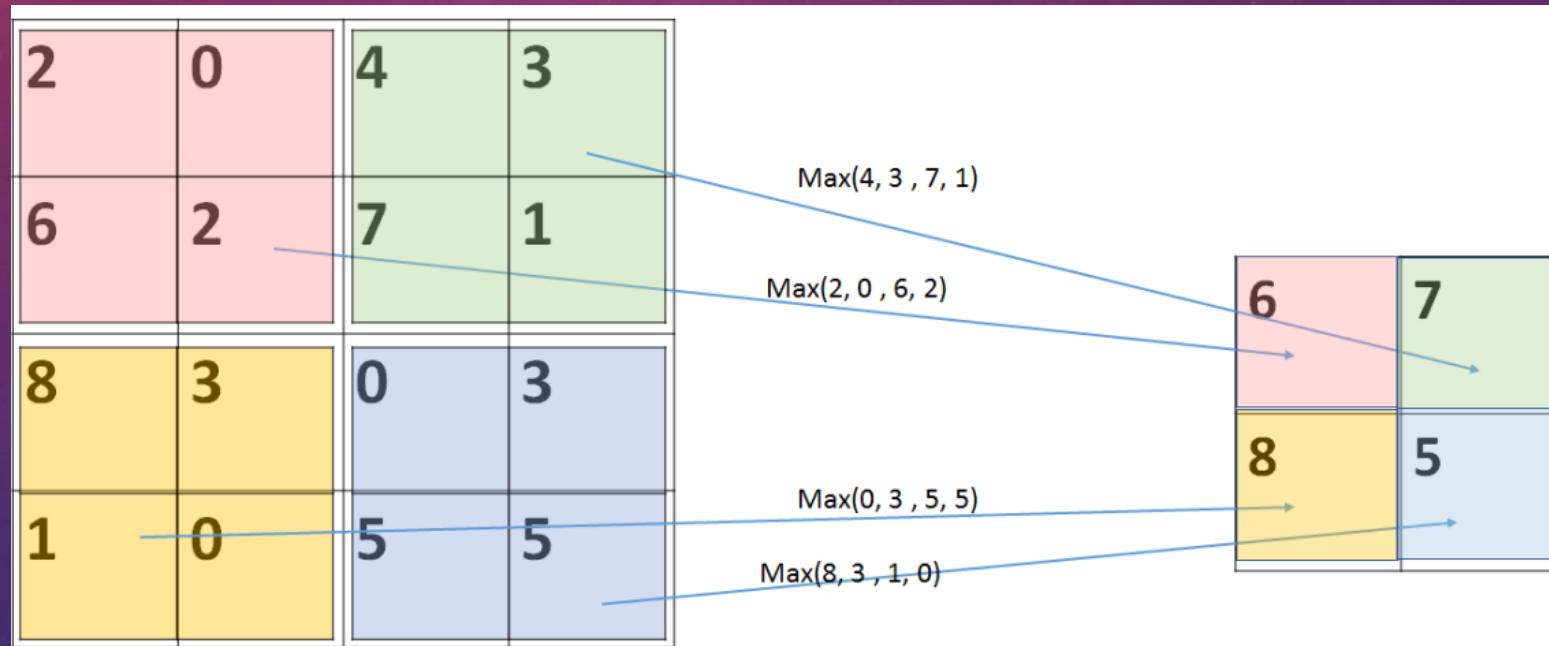
CONVOLUTIONAL NEURAL NETWORK



卷積層的運作

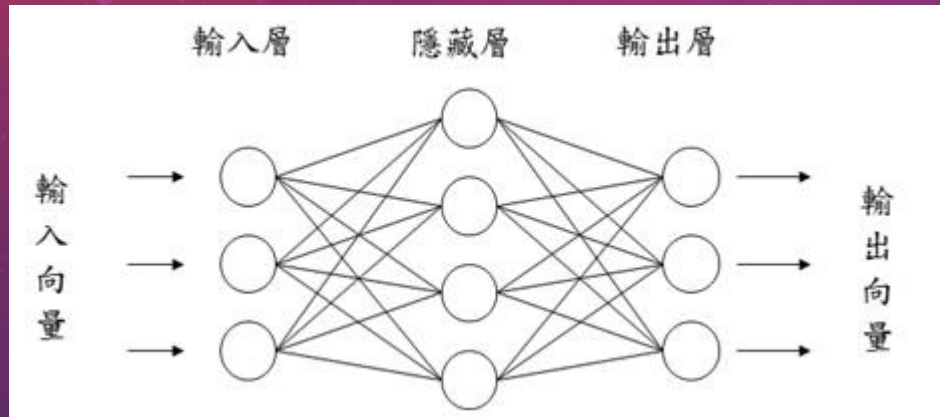
CONVOLUTIONAL NEURAL NETWORK

最大池化層的運作



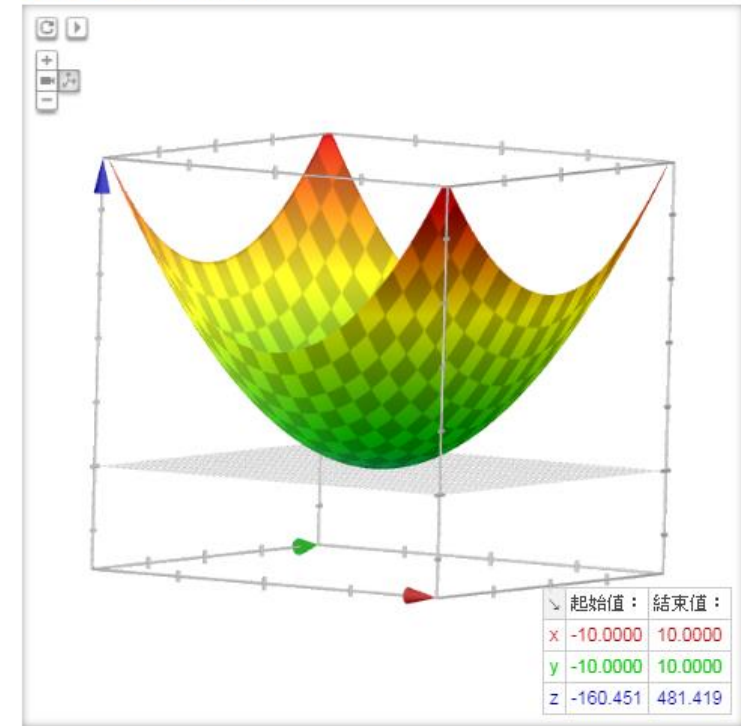
CONVOLUTIONAL NEURAL NETWORK

全連接層的運作



$$3x^2 + 2y^2$$

$3x^2 + 2y^2$ 的圖表

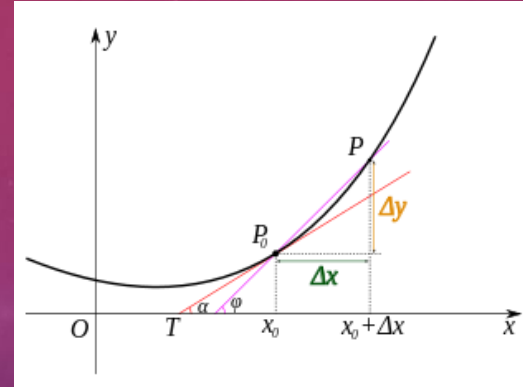
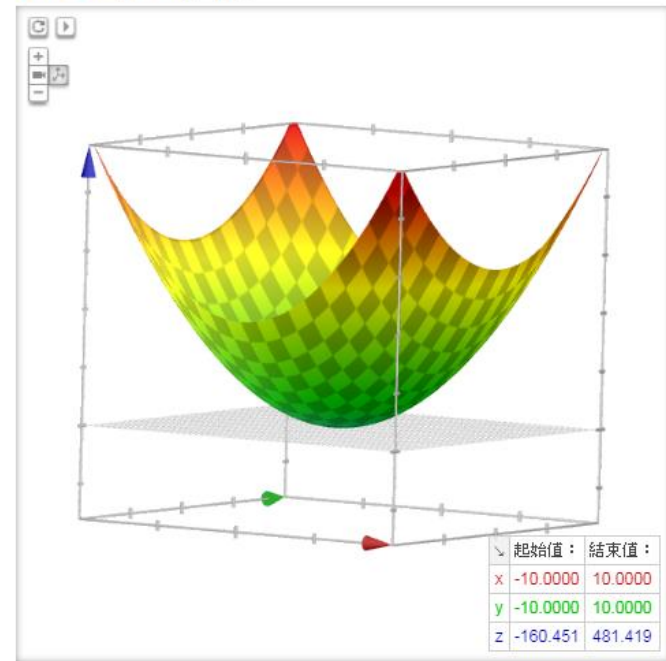


CONVOLUTIONAL NEURAL NETWORK

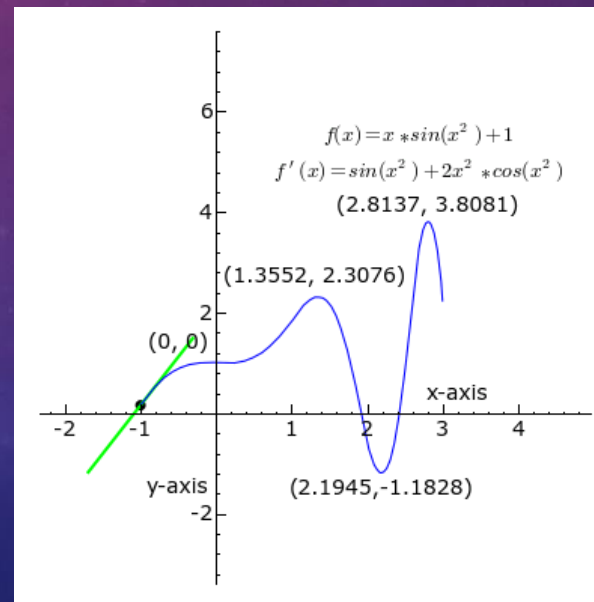
全連接層的運作

$$3x^2 + 2y^2$$

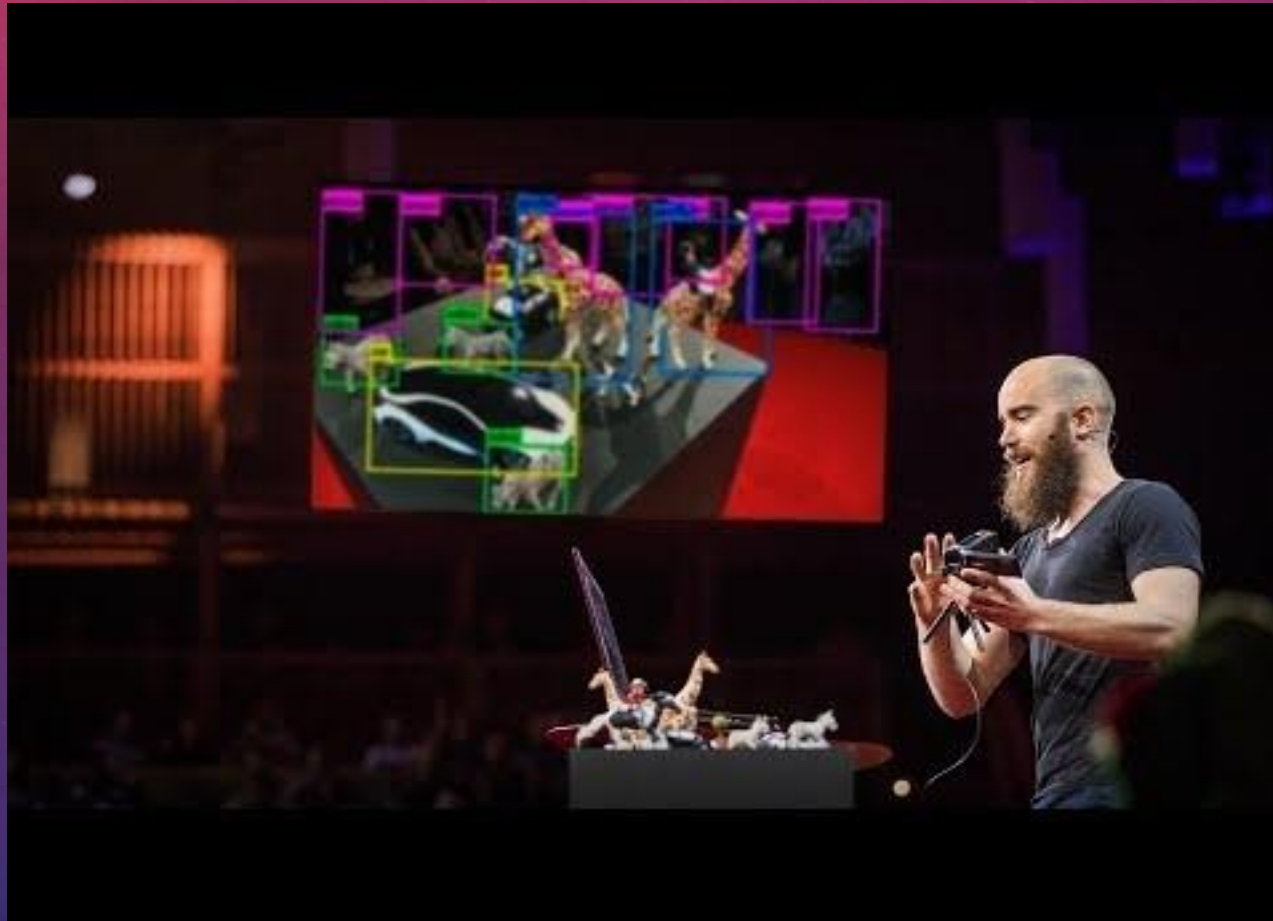
$3x^2 + 2y^2$ 的圖表



$$\tan \alpha = \lim_{\Delta x \rightarrow 0} \tan \varphi = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$



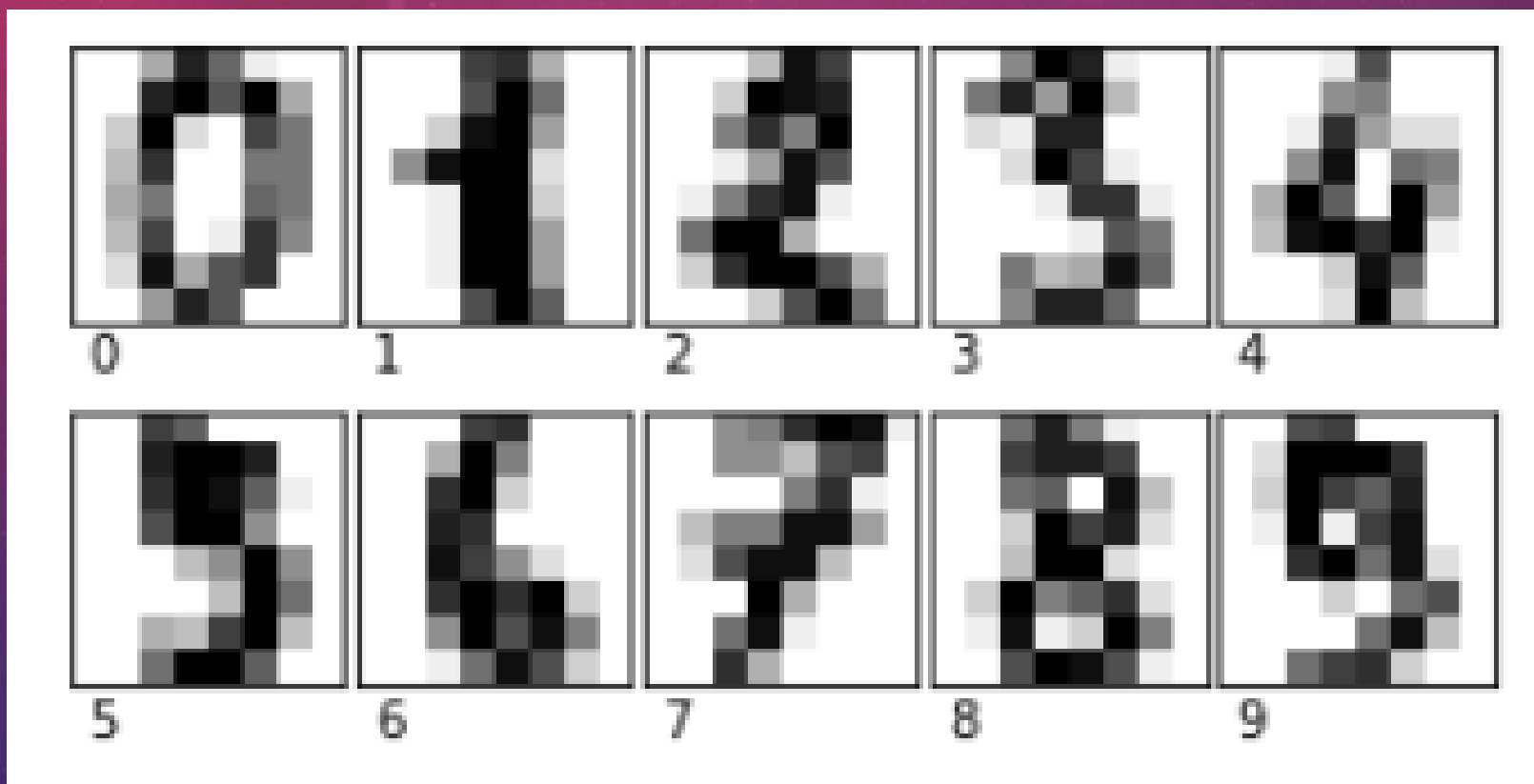
HOW COMPUTERS LEARN TO RECOGNIZE OBJECTS INSTANTLY



The background features a gradient from deep red at the top to dark blue at the bottom, speckled with white star-like dots. Overlaid on this are several faint, white circular and semi-circular lines, some with arrows indicating a clockwise direction. A large circular scale with numerical markings from 140 to 260 is visible on the left side.

實際案例

手寫數字辨識

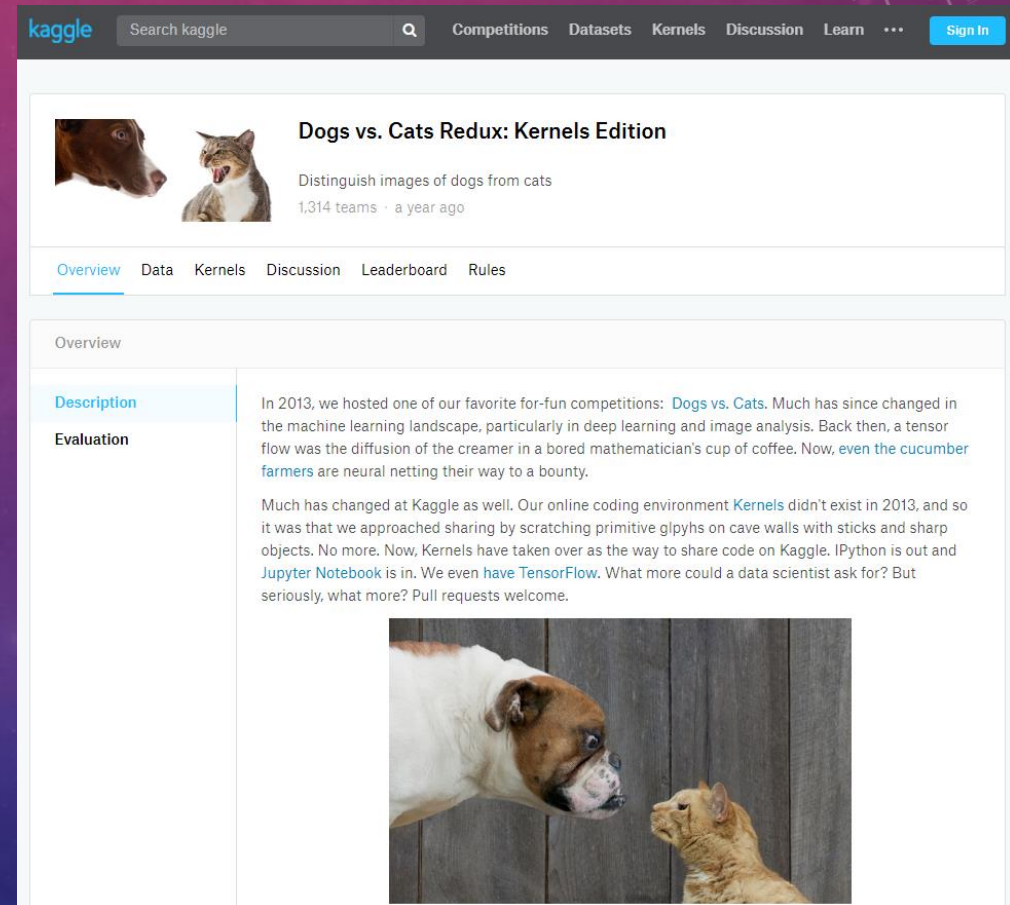


手寫數字辨識

	precision	recall	f1-score	support
0	1.00	0.98	0.99	41
1	0.91	0.95	0.93	42
2	0.94	0.92	0.93	37
3	0.89	0.95	0.92	44
4	0.96	1.00	0.98	49
5	0.96	0.91	0.93	47
6	0.98	1.00	0.99	45
7	0.98	0.98	0.98	60
8	0.87	0.85	0.86	40
9	0.98	0.91	0.94	45
avg / total	0.95	0.95	0.95	450

DOGS VS. CATS

- <https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition>
 - Training data: 25000 images
 - Test data: 12500 images
 - For each image in the test set, you should predict a probability that the image is a dog (1 = dog, 0 = cat).



The screenshot shows the Kaggle website interface for the 'Dogs vs. Cats Redux: Kernels Edition' competition. The header includes the Kaggle logo, a search bar, and navigation links for Competitions, Datasets, Kernels, Discussion, Learn, and a Sign In button. The competition title is 'Dogs vs. Cats Redux: Kernels Edition', with a subtitle 'Distinguish images of dogs from cats' and '1,314 teams · a year ago'. Below the title are tabs for Overview, Data, Kernels, Discussion, Leaderboard, and Rules. The 'Overview' tab is selected, showing a 'Description' section with text about the competition's history and the machine learning landscape in 2013. Below the description is an 'Evaluation' section. At the bottom of the overview, there is a photograph of a white and brown dog and a ginger cat facing each other.

貓狗辨識的結果

[0.14723705	0.98256]
[0.04469623	0.99915826]	
[0.99998796	0.00834211]	
[0.99993527	0.01661933]	
[0.99877125	0.05415697]	
[0.43036035	0.6431105]
[0.20522958	0.95668554]	
[0.8880429	0.28058085]	
[0.08643436	0.9956601]
[0.99908304	0.04826429]	
[0.9999168	0.01841162]	
[0.1397599	0.98483086]	
[0.99558544	0.08876048]	
[0.43750185	0.6269166]
[0.10178897	0.9934009]
[0.10858331	0.9922093]



中文OCR辨識為例



- Lee, M. C., Chiu S. Y., & Chang, J. W. (2017) A Deep Convolutional Neural Network based Chinese Menu Recognition App. Information Processing Letters, 128, 14-20. <https://doi.org/10.1016/j.ipl.2017.07.010> (SCI, COMPUTER SCIENCE, INFORMATION SYSTEMS)

VIDEO TO VIDEO



VIDEO-TO-VIDEO SYNTHESIS

The paper "Video-to-Video Synthesis" and its source code is available here:
<https://tcwang0509.github.io/vid2vid/>
<https://github.com/NVIDIA/vid2vid>

智慧視覺系統機器人



視覺整合信號轉換的體感操控機械手臂





THANK YOU