正修科技大學
GAI-2D到影音

金石教育科技

四日課程安排

- 第一周：AI數位工具應用與發展
- 第二周：GAI生成文字劇本
- 第三周：文字轉換2D影像
- 第四周：2D影像轉換至影片生成

先來省思

台灣只能做代工
不重研究做不了技術創新，
不重人文，做不了產品創新，
也沒氣質可言所以做不了品牌，
永遠當代工島.

三力

創新力

技術力

執行力

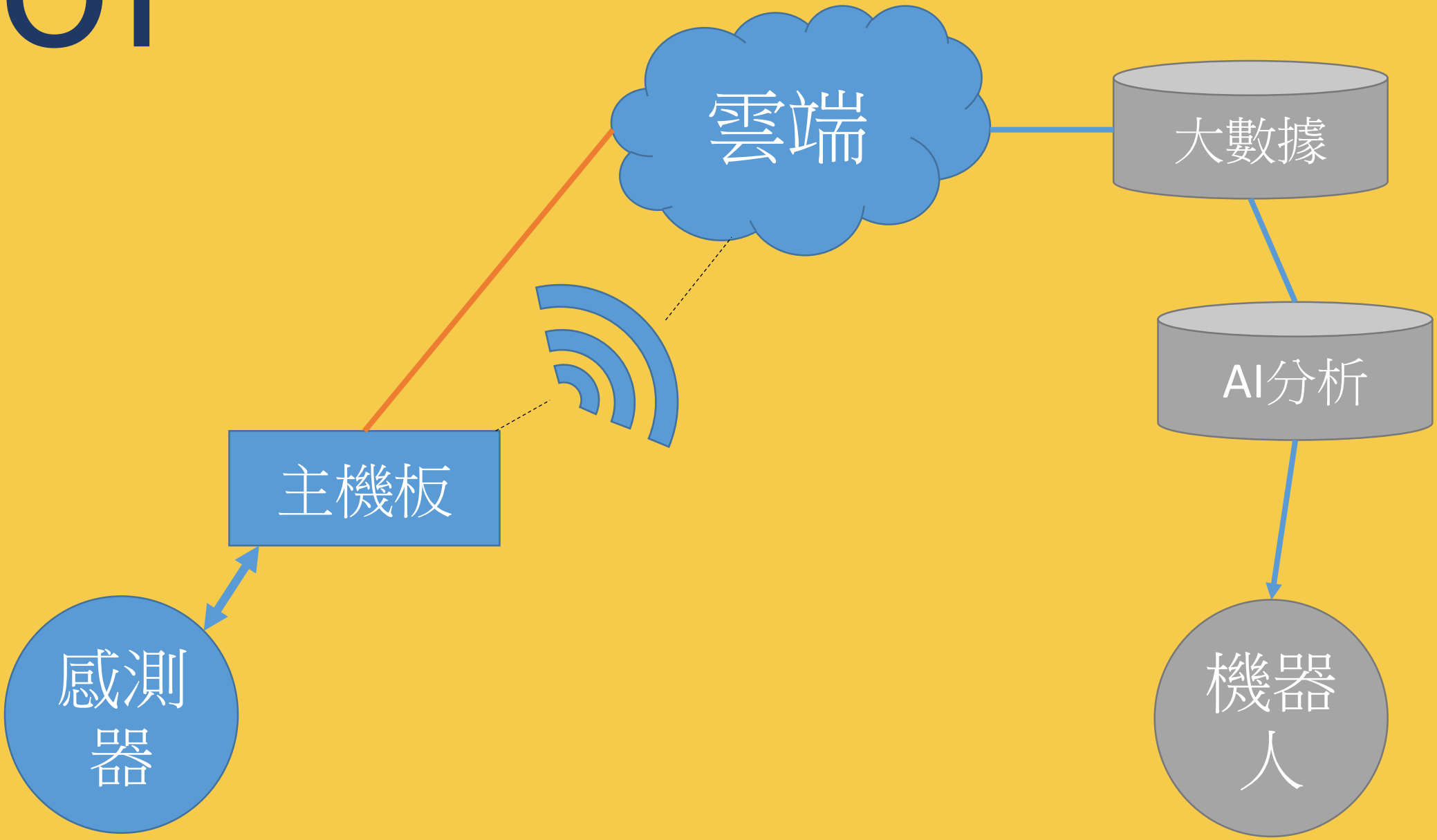
先來看部影片

- https://youtu.be/nJNdgqizrtM

工業4.0

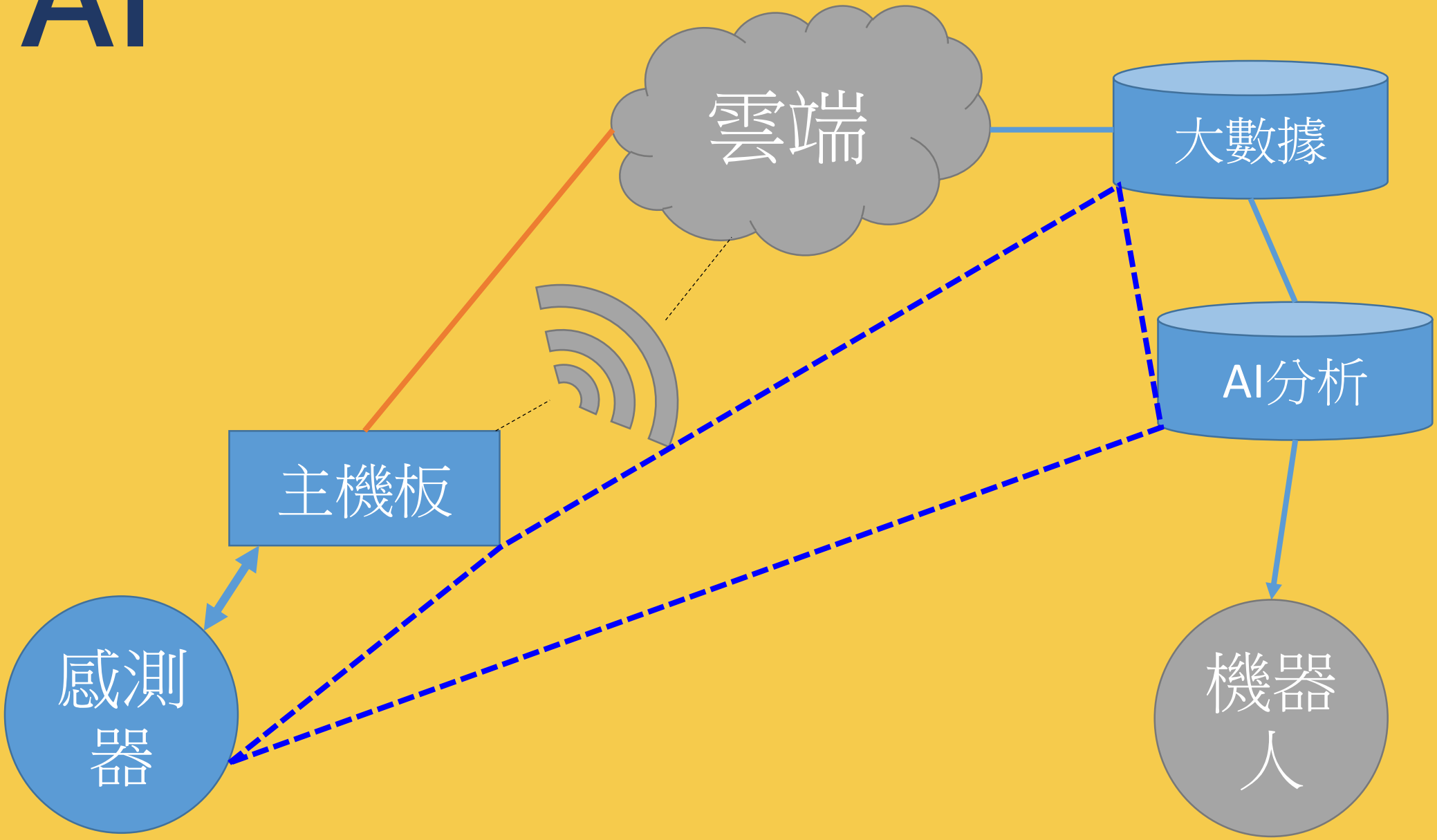
工業4.0,鄭朝元,2019/09

IOT



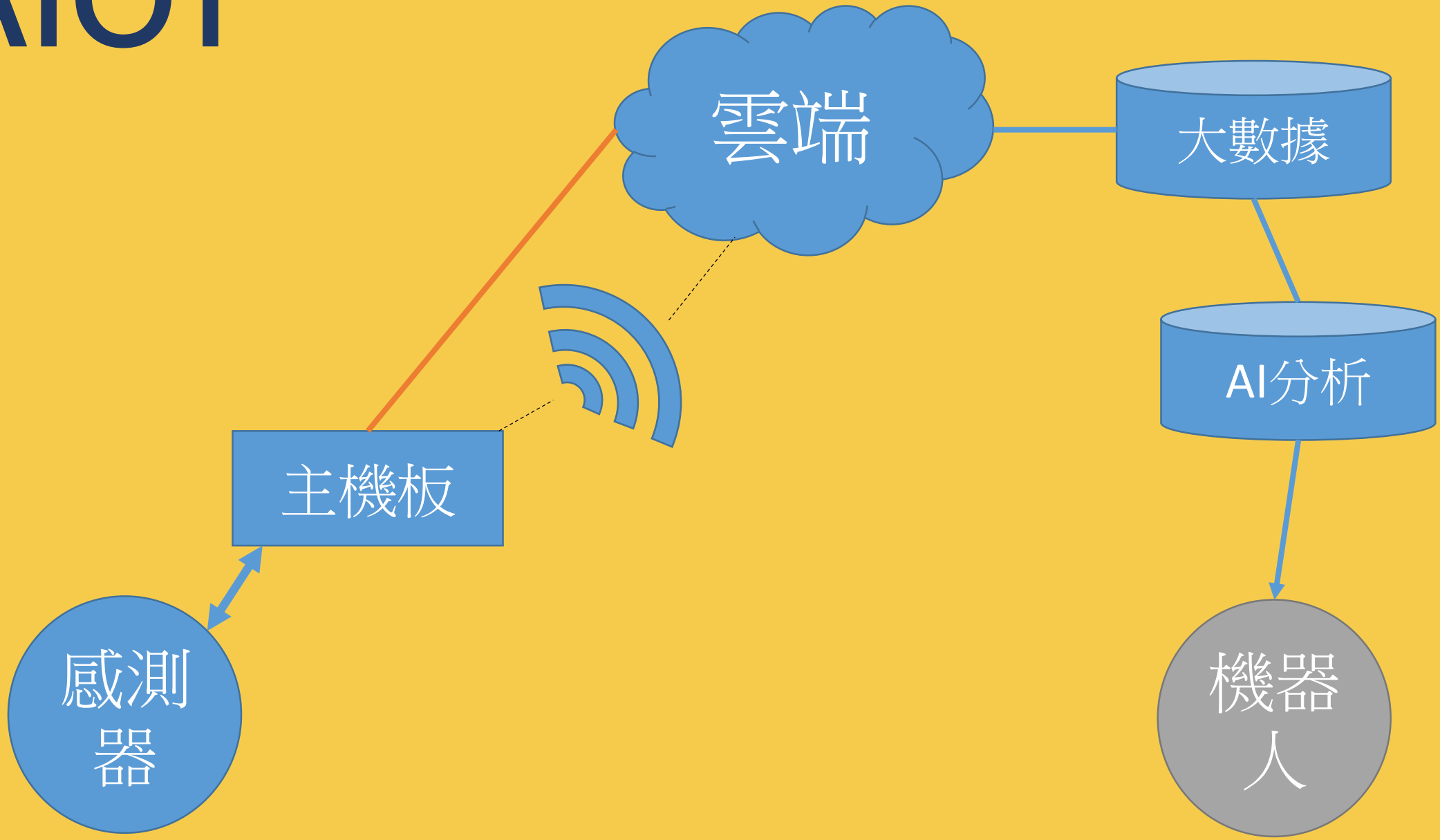
工業4.0,鄭朝元,2019/09

AI



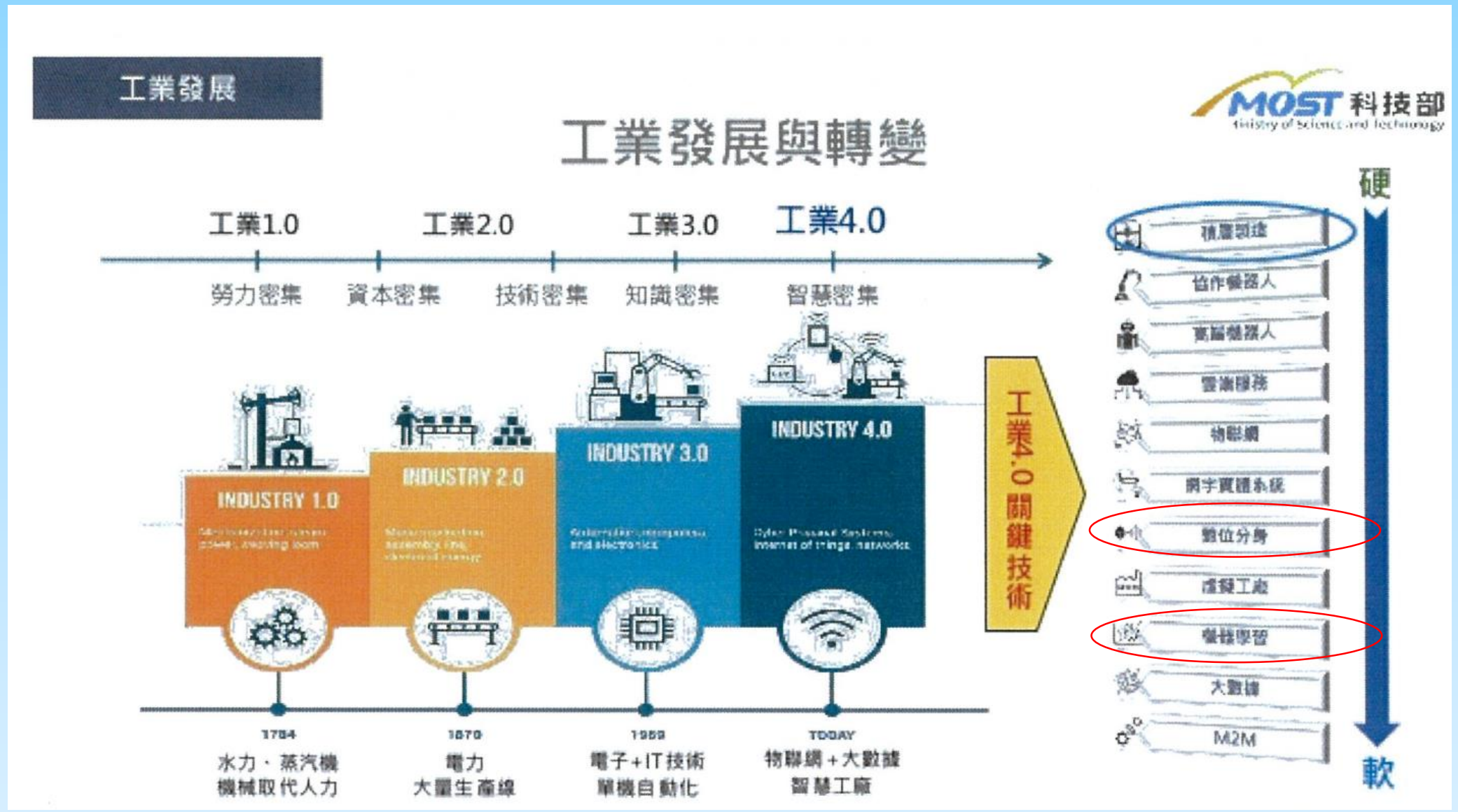
工業4.0,鄭朝元,2019/09

AIOT



工業4.0,鄭朝元,2019/09

3D-GAI重要性


工業發展
工業發展與轉變
MOST 科技部
工業1.0
工業2.0
工業3.0
工業4.0
勞力密集
資本密集
技術密集
知識密集
智慧密集
INDUSTRY 1.0
INDUSTRY 2.0
INDUSTRY 3.0
INDUSTRY 4.0
工業4.0關鍵技術
硬
軟
積層製造
協作機器人
高階機器人
雲端服務
物聯網
網宇實體系統
數位分身
虛擬工廠
機器學習
大數據
M2M
1784
水力、蒸汽機
機械取代人力
1870
電力
大量生產線
1969
電子+IT技術
單機自動化
TODAY
物聯網+大數據
智慧工廠

1958年

POWER COMPANIES BUILD FOR YOUR NEW ELECTRIC LIVING

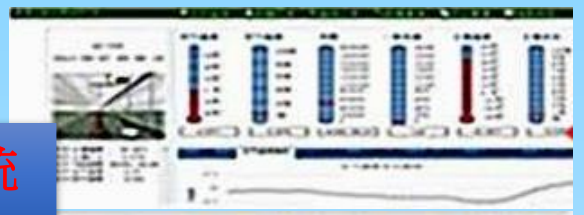
智慧溫室系統

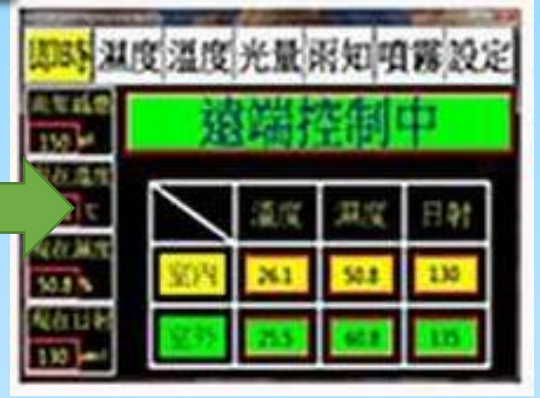
2 遠端APP監控系統

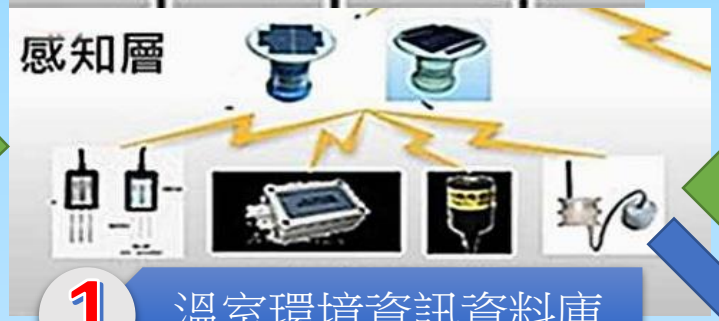
管理者

7 自適應系統

3 人機設備管理界面







1 溫室環境資訊資料庫

網路層 AIoT自適應演算平台

GOOGLE GCP

wss / https
加密傳輸

無線感測網絡

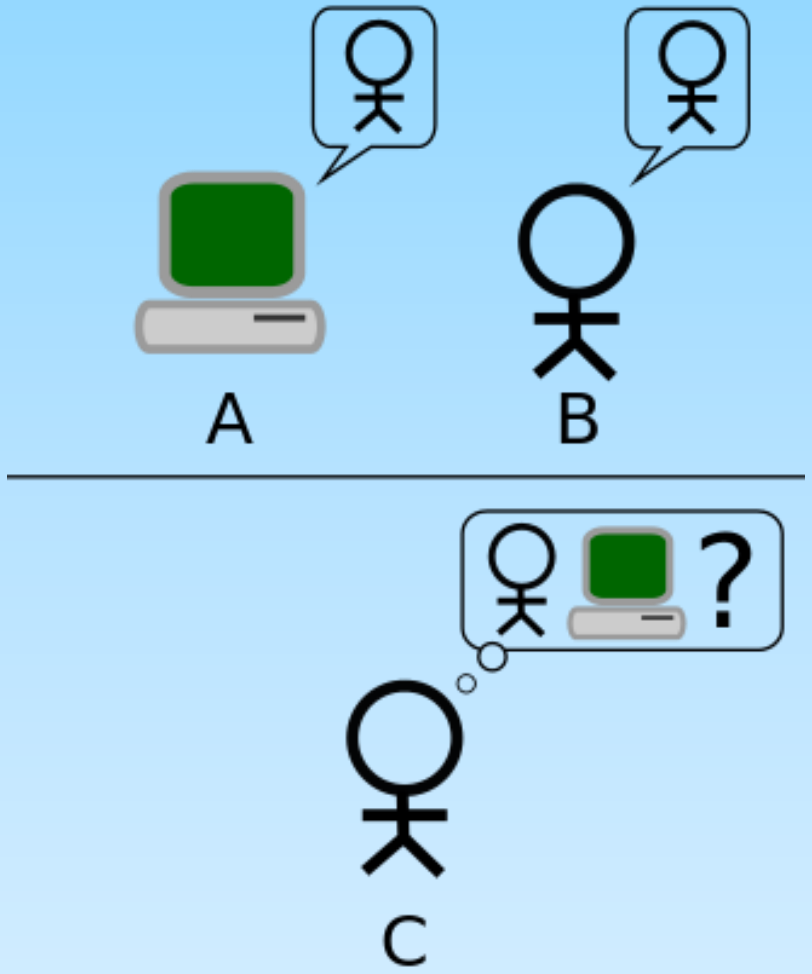
5 專家變項參數M2M演算

4 APP 影像監控與警示

6 複製的溫室增量數據庫

你覺得什麼是AI？ 1950(Turing test）

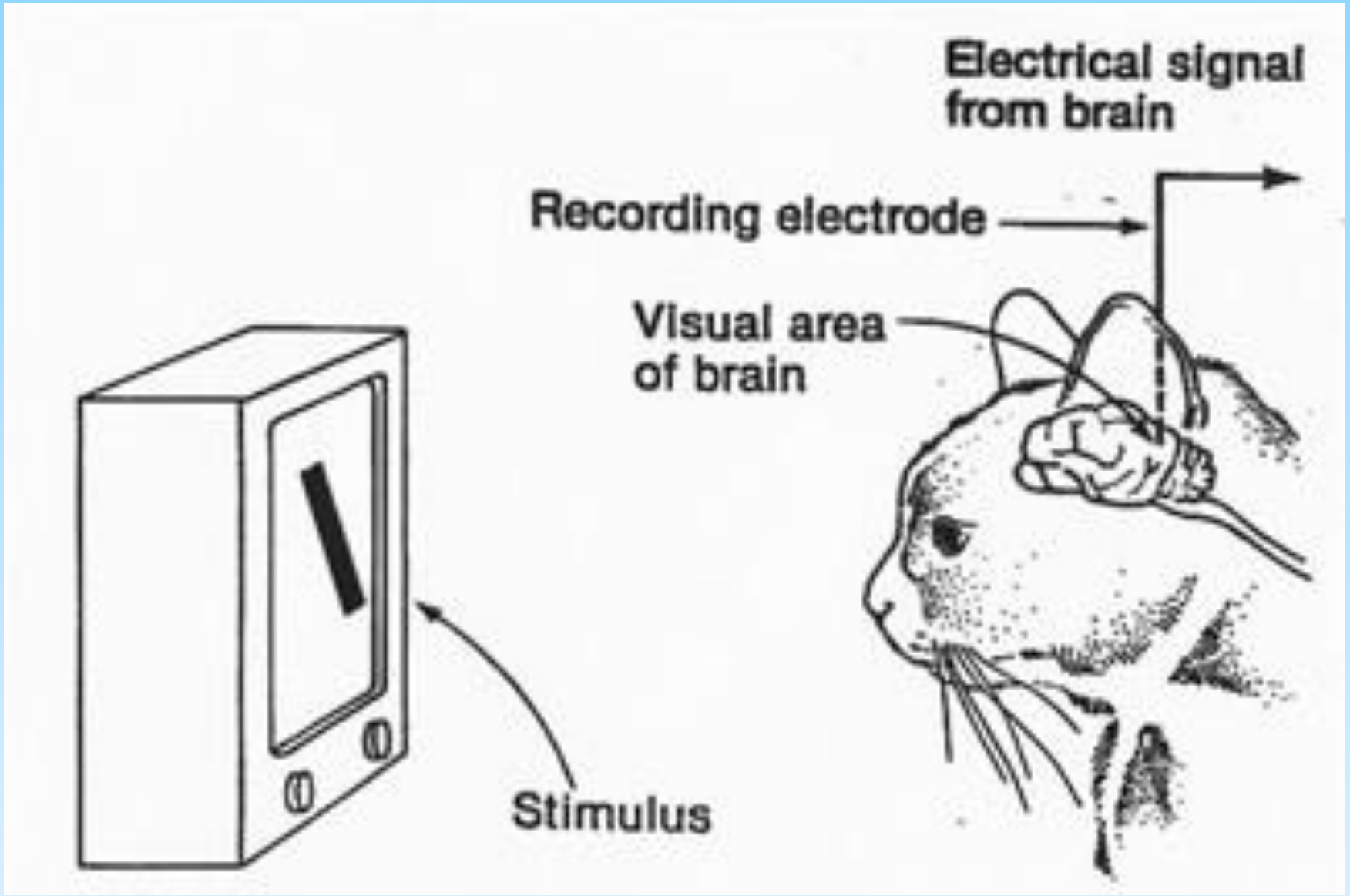
- 如果一個人（代號C）使用測試對象皆理解的語言去詢問兩個他不能看見的對象任意一串問題。對象為：一個是正常思維的人（代號B）、一個是機器（代號A）。如果經過若干詢問以後，C不能得出實質的區別來分辨A與B的不同，則此機器A通過圖靈測試。



David Hubel和Torsten Wiesel測試了貓的視皮質細胞反應

- 在**螢幕上打出一些光影或者其他圖形時**，貓就用帶子繫好，藉已固定好貓的頭部，研究者就可以知道是網膜上的哪一部分是圖像顯現之處，然後把這個被**刺進的皮質區**進行連接，透過放大器和喇叭，他們可以聽到細胞啟動的聲音。其結果**顯示細胞對一個橫向的線或者邊緣有強烈反應**，但對點、斜線或直線只有非常微弱的反應，或者根本就沒有反應，之後的研究繼續顯示：**有些細胞對某些處在一個角度上的線條、垂直線條、直角或者明顯的邊緣線，都有特別的反應。**

1981諾貝爾獎

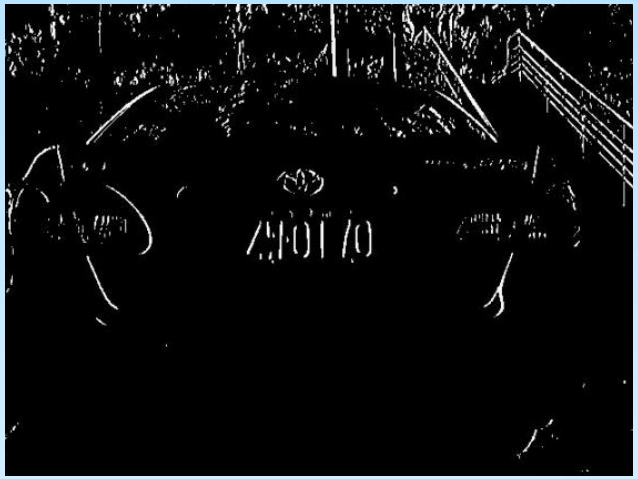

Electrical signal
from brain
Recording electrode
Visual area
of brain
Stimulus

可生活應用如車牌辨識

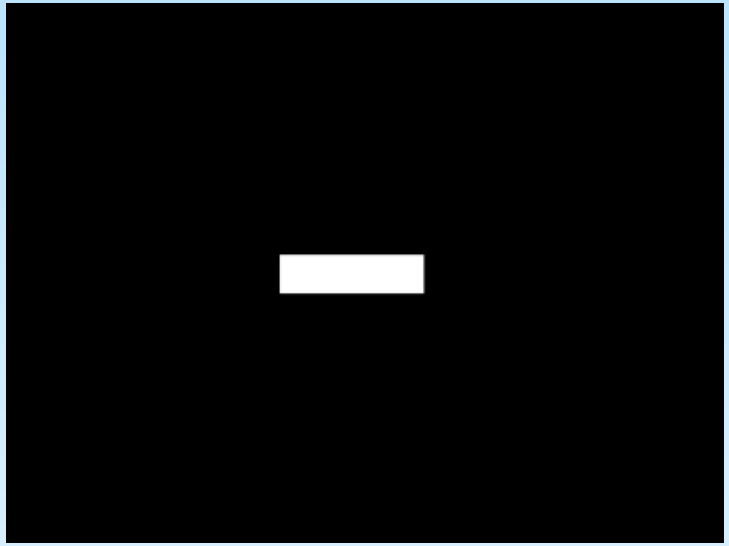

讀取車牌影像
灰階處理
邊緣偵測
二值化
中值濾波
膨脹及侵蝕處理
標籤化及條件篩選
擷取車牌
邊界切除
切割字元
字元正規化
樣本比對
輸出結果


ZW-0170


ZW-0170



可生活應用如車牌辨識



可生活應用如車牌辨識

0	1	2	3	4	5	6	7
8	9	A	B	C	D	E	F
G	H	I	J	K	L	M	N
O	P	Q	R	S	T	U	V
W	X	Y	Z				

為了後續與樣本字元比對，必須將切割出來的字元正規化，本專題採用線性正規化，將字元正規化成 15*45 的大小，正規化的方法是以 15*45 的座標推斷欲正規化字元在此大小裡的座標是黑點或白點。

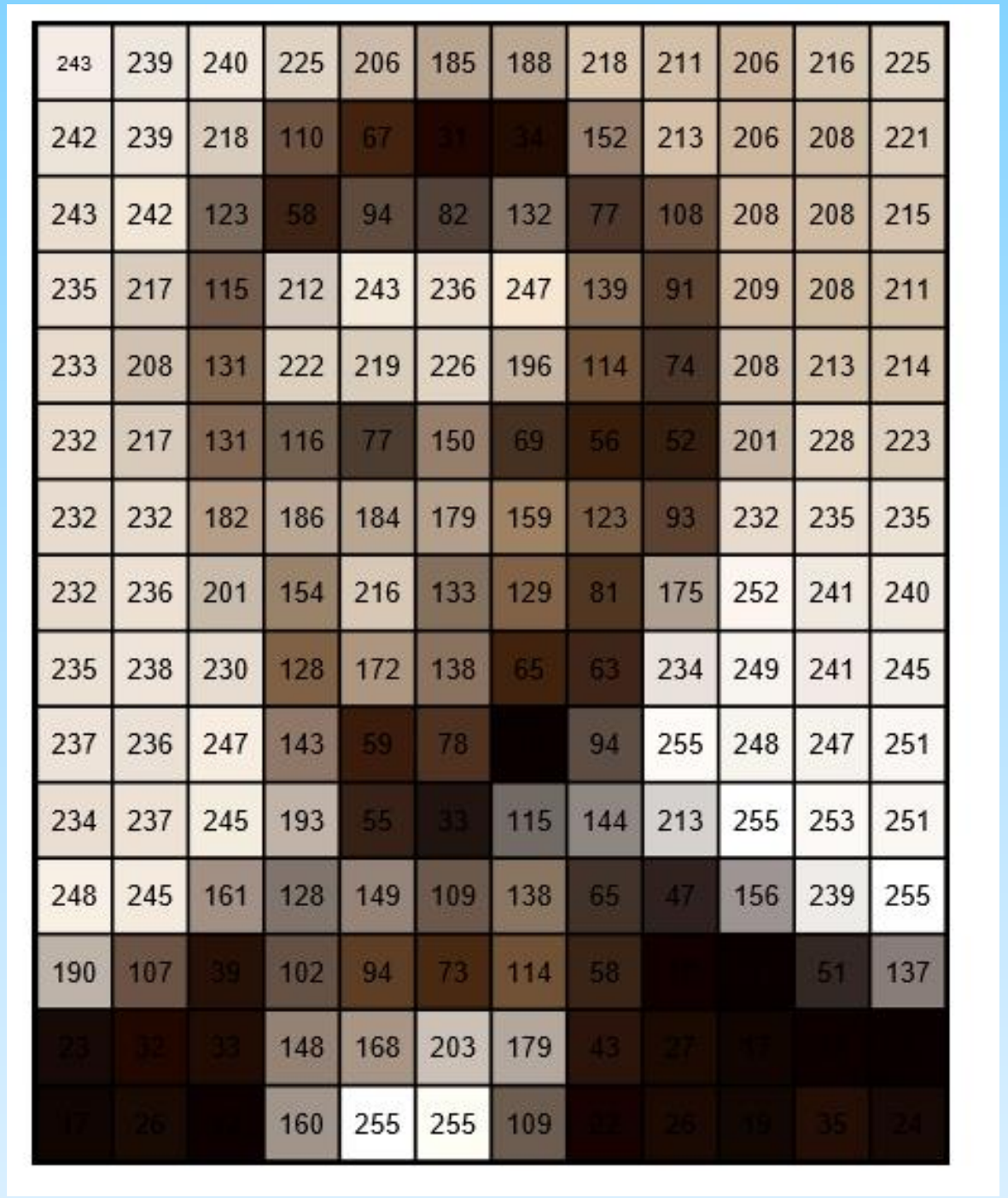
可生活應用如車牌辨識

什麼是影像？

數位相片？

243	239	240	225	206	185	188	218	211	206	216	225
242	239	218	110	67	31	34	152	213	206	208	221
243	242	123	58	94	82	132	77	108	208	208	215
235	217	115	212	243	236	247	139	91	209	208	211
233	208	131	222	219	226	196	114	74	208	213	214
232	217	131	116	77	150	69	56	52	201	228	223
232	232	182	186	184	179	159	123	93	232	235	235
232	236	201	154	216	133	129	81	175	252	241	240
235	238	230	128	172	138	65	63	234	249	241	245
237	236	247	143	59	78	10	94	255	248	247	251
234	237	245	193	55	33	115	144	213	255	253	251
248	245	161	128	149	109	138	65	47	156	239	255
190	107	39	102	94	73	114	58	17	7	51	137
23	32	33	148	168	203	179	43	27	17	12	8
17	26	12	160	255	255	109	22	26	19	35	24

數位相片？


243 239 240 225 206 185 188 218 211 206 216 225
242 239 218 110 67 31 34 152 213 206 208 221
243 242 123 58 94 82 132 77 108 208 208 215
235 217 115 212 243 236 247 139 91 209 208 211
233 208 131 222 219 226 196 114 74 208 213 214
232 217 131 116 77 150 69 56 52 201 228 223
232 232 182 186 184 179 159 123 93 232 235 235
232 236 201 154 216 133 129 81 175 252 241 240
235 238 230 128 172 138 65 63 234 249 241 245
237 236 247 143 59 78 94 255 248 247 251
234 237 245 193 55 33 115 144 213 255 253 251
248 245 161 128 149 109 138 65 47 156 239 255
190 107 39 102 94 73 114 58 51 137
148 168 203 179 43
160 255 255 109

彩色照片？

Convolution

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

input image

$*$

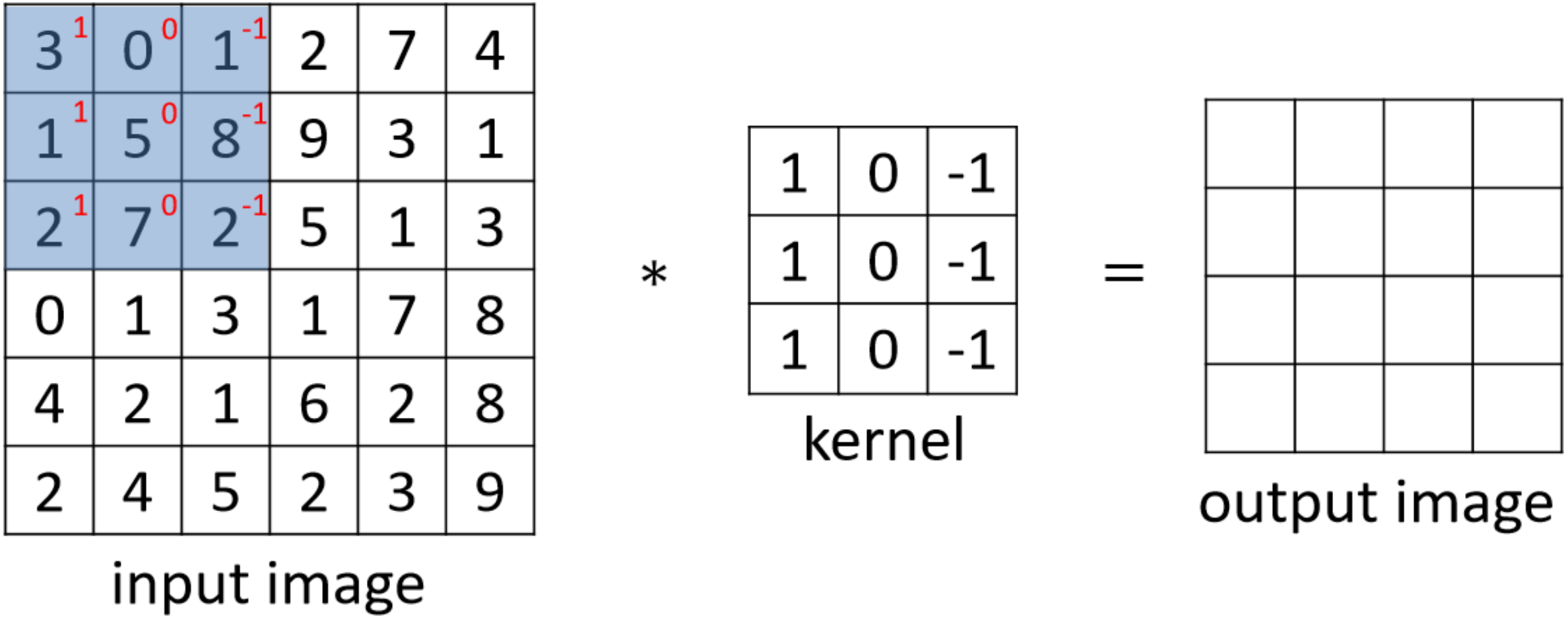
1	0	-1
1	0	-1
1	0	-1

kernel
(filter)

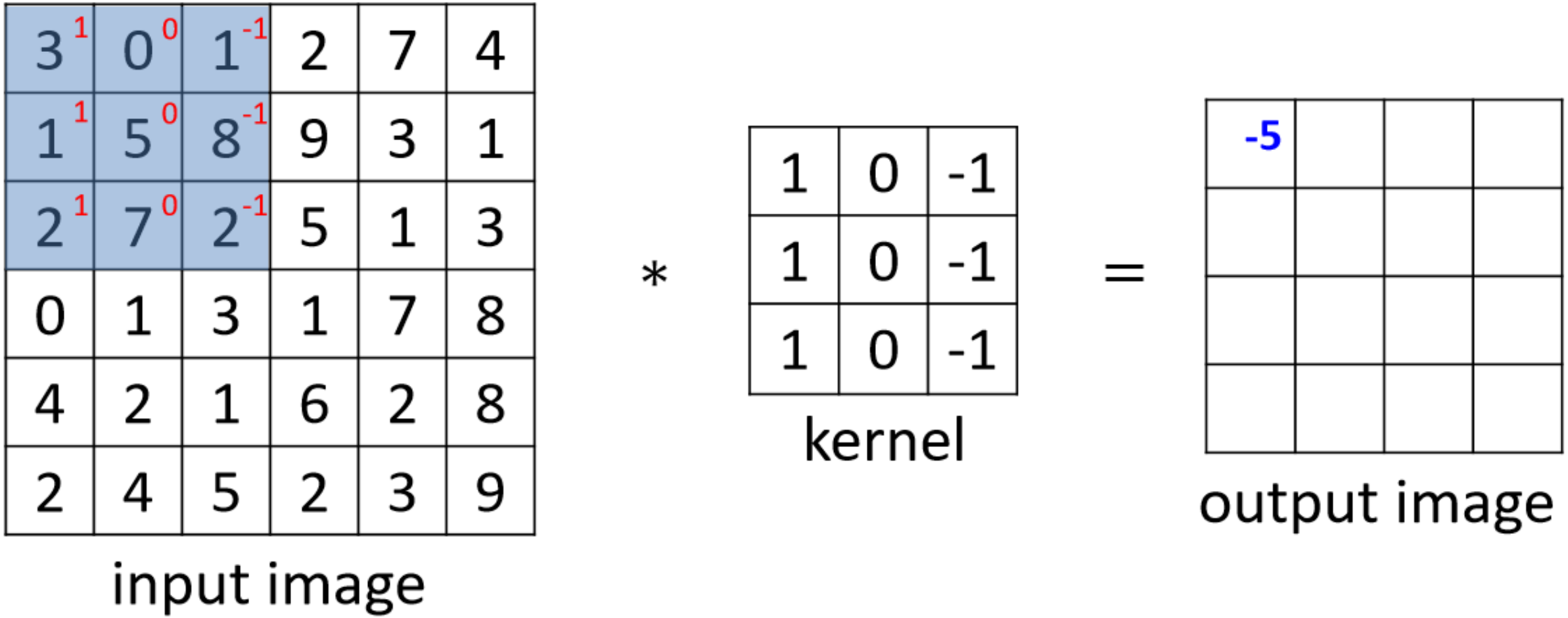
$=$

output image

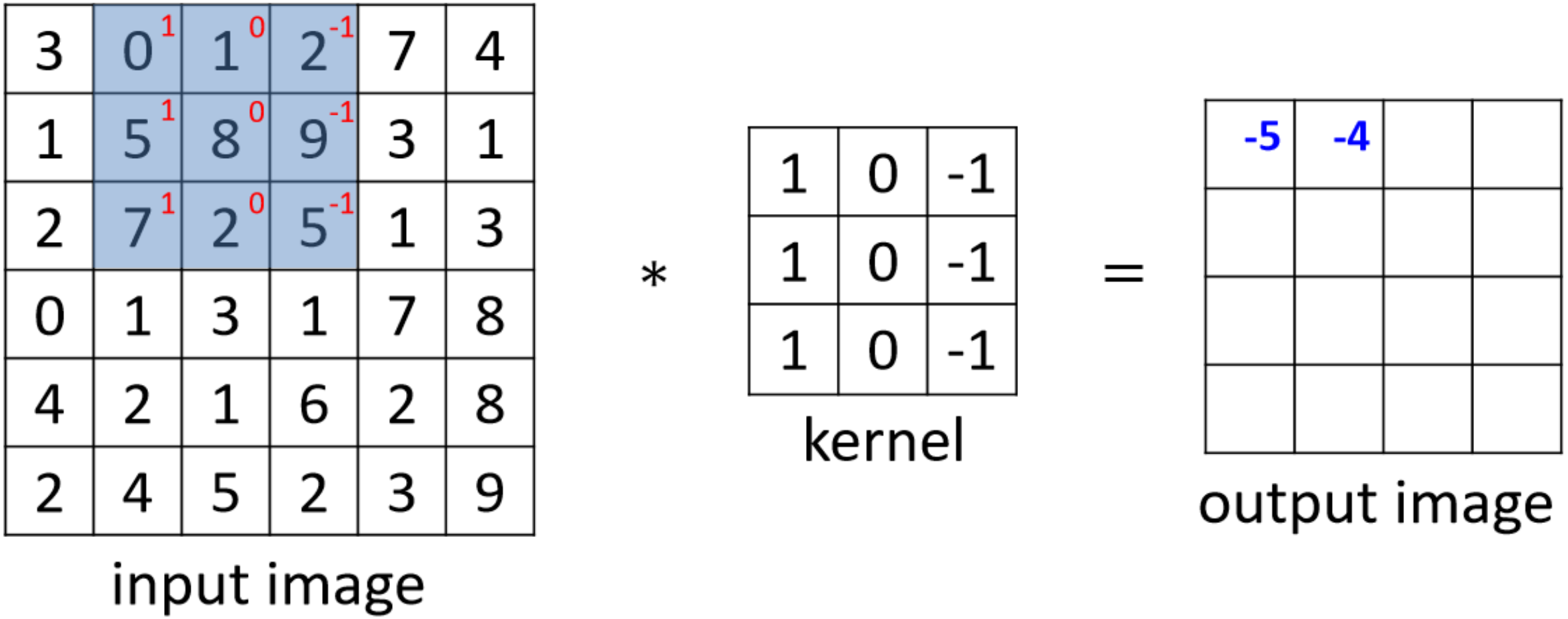
Convolution



Convolution



Convolution



Convolution

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

input image

$*$

1	0	-1
1	0	-1
1	0	-1

kernel

$=$

-5	-4	0	8
-10	-2	2	3
0	-2	-4	-7
-3	-2	-3	-16

output image

Convolution

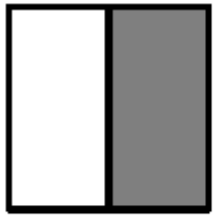
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

*

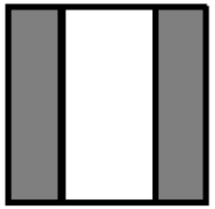
1	0	-1
1	0	-1
1	0	-1

=

0	30	30	0
0	30	30	0
0	30	30	0
0	30	30	0



*

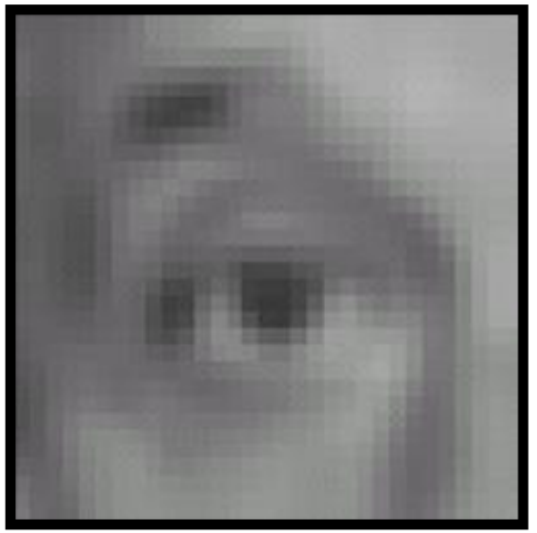


Convolution

original

$$\frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}$$

Box filter



result

Convolution

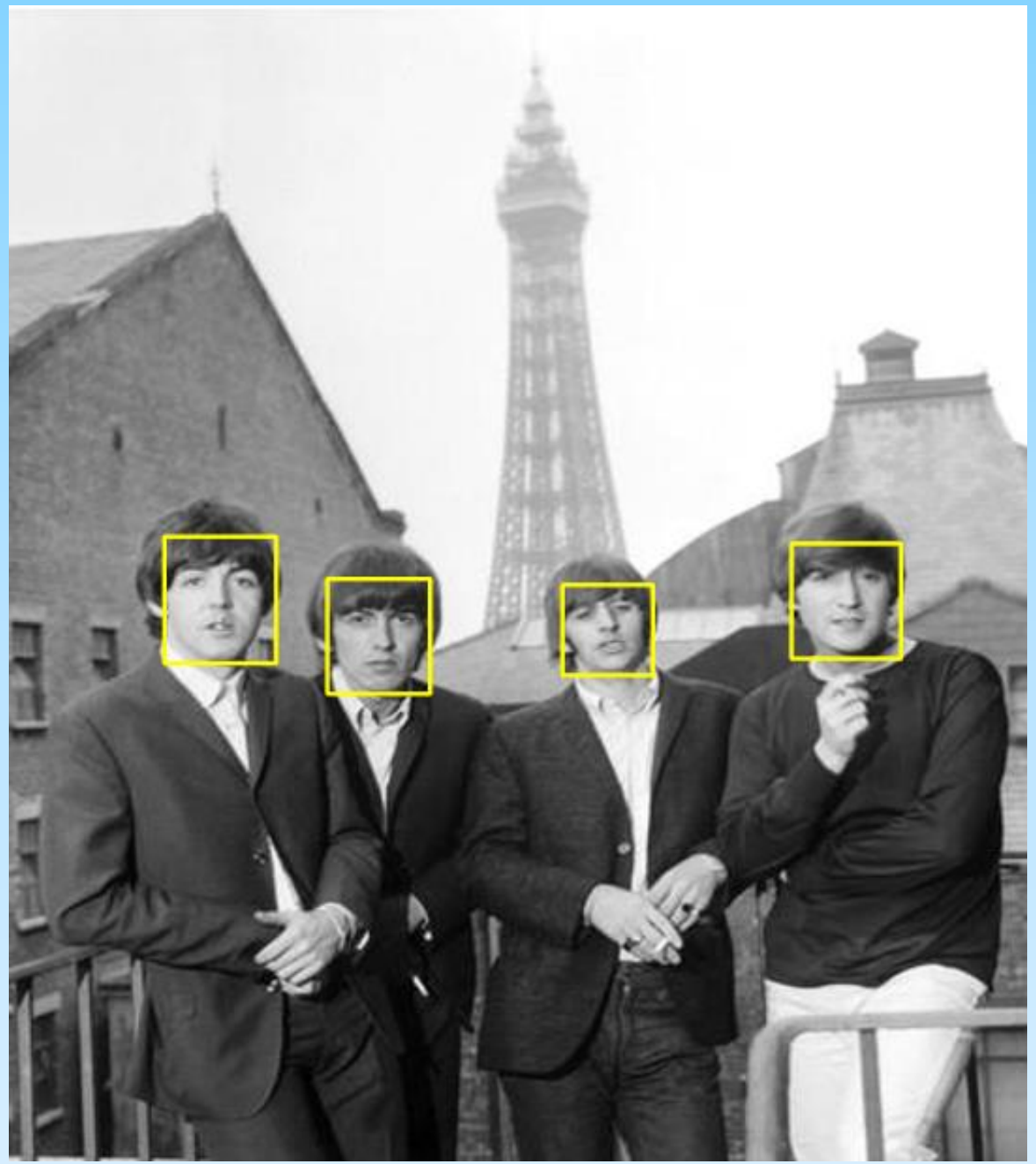
original

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} - \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

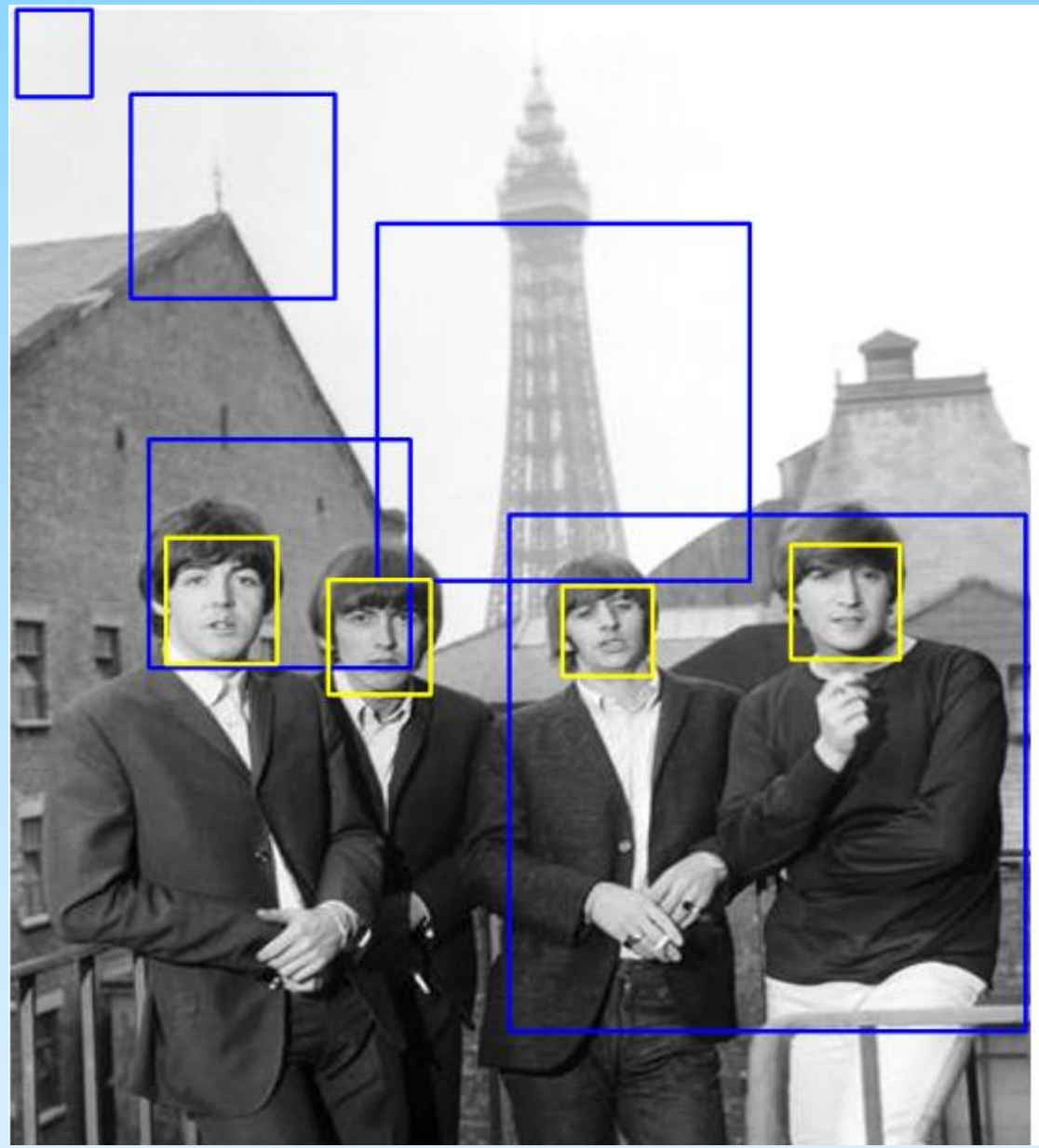
Sharpening filter

result

Face detection



Sliding windows

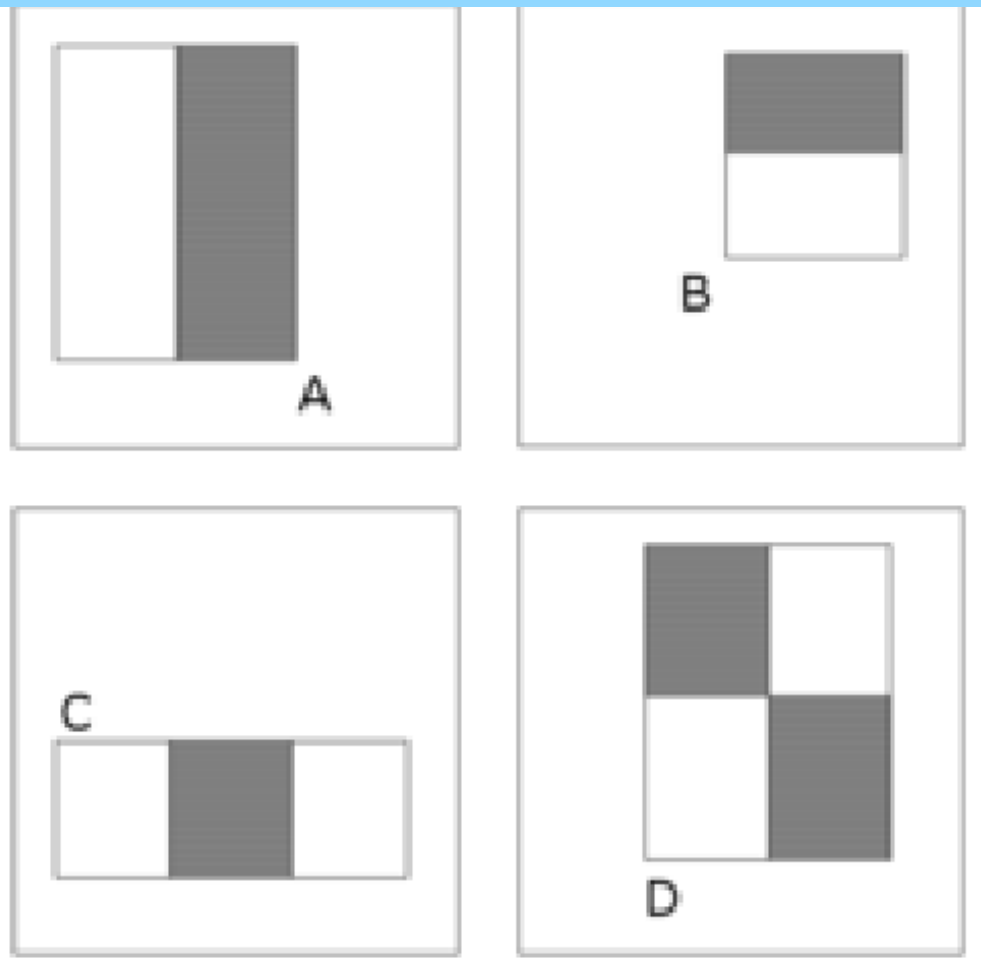


1. hypothesize:
try all possible rectangle locations, sizes

2. test:
classify if rectangle contains a face (and only the face)

Note: 1000's more false windows than true ones.

Features

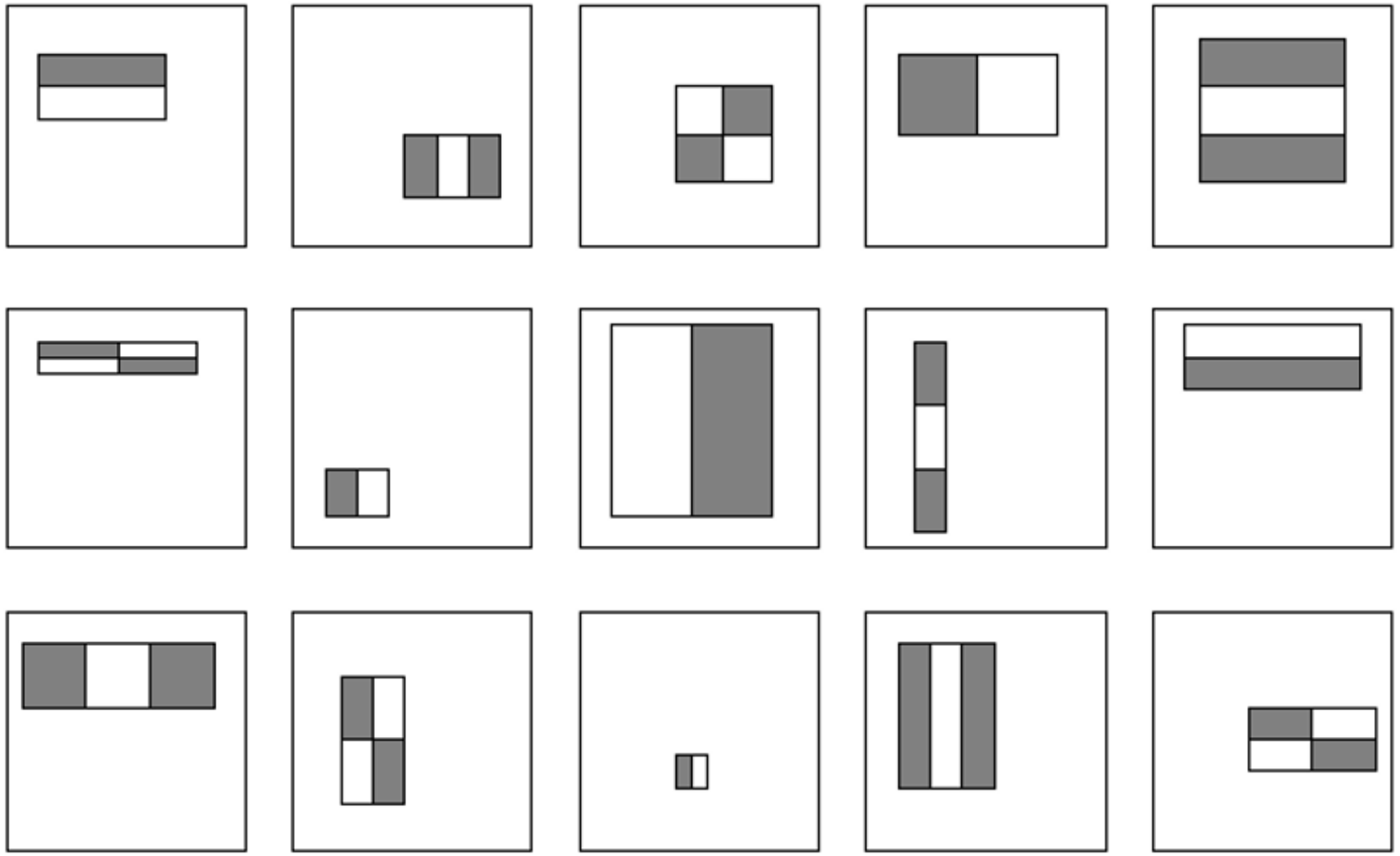


4 Types of “Rectangle filters” (Similar to Haar wavelets)

$$g(x) = \text{sum(WhiteArea)} - \text{sum(BlackArea)}$$

Features

For a 24x24 grid: 160,000+ features to choose from



The learned features

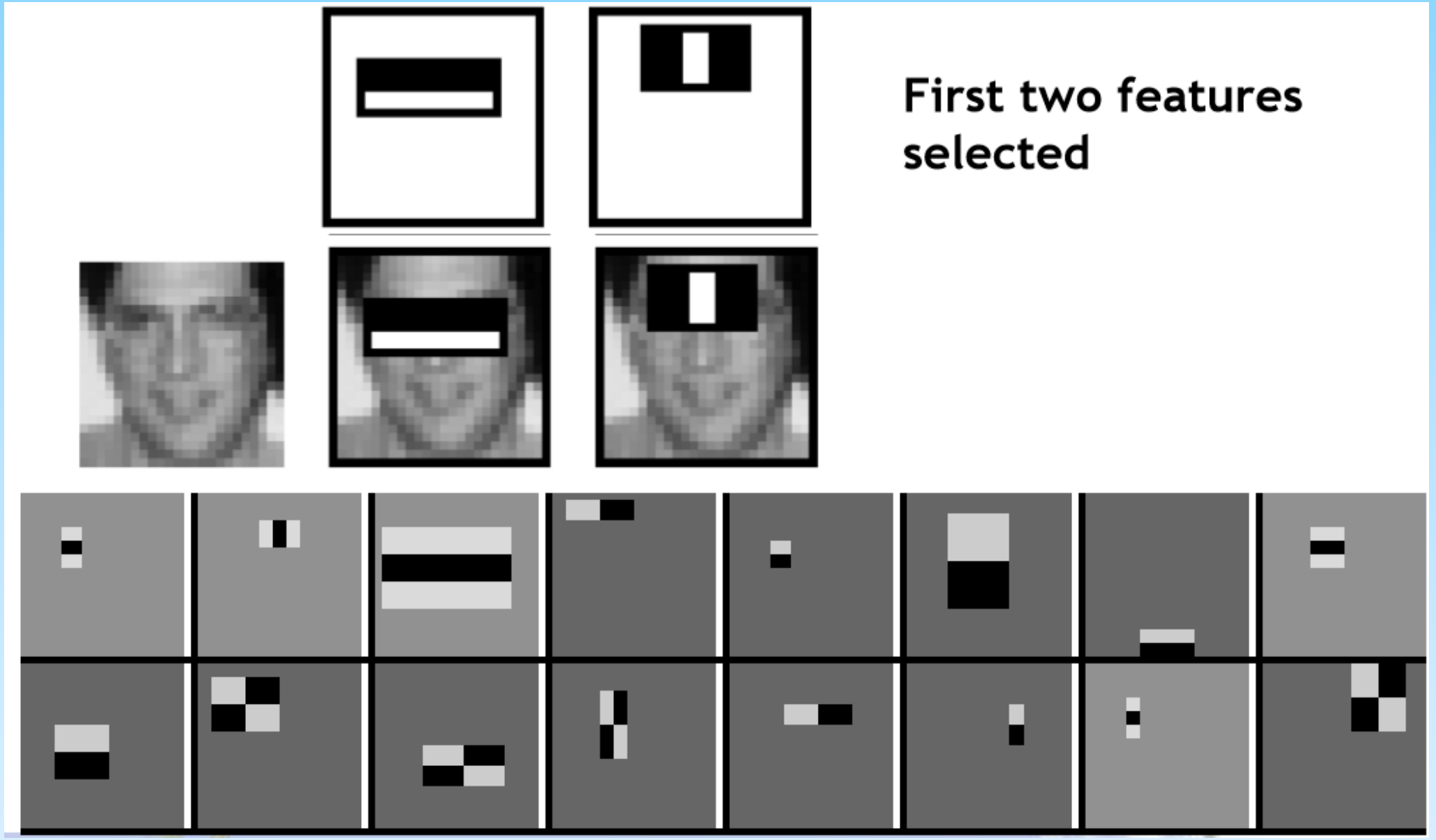


Image classification

cat?

Image classification

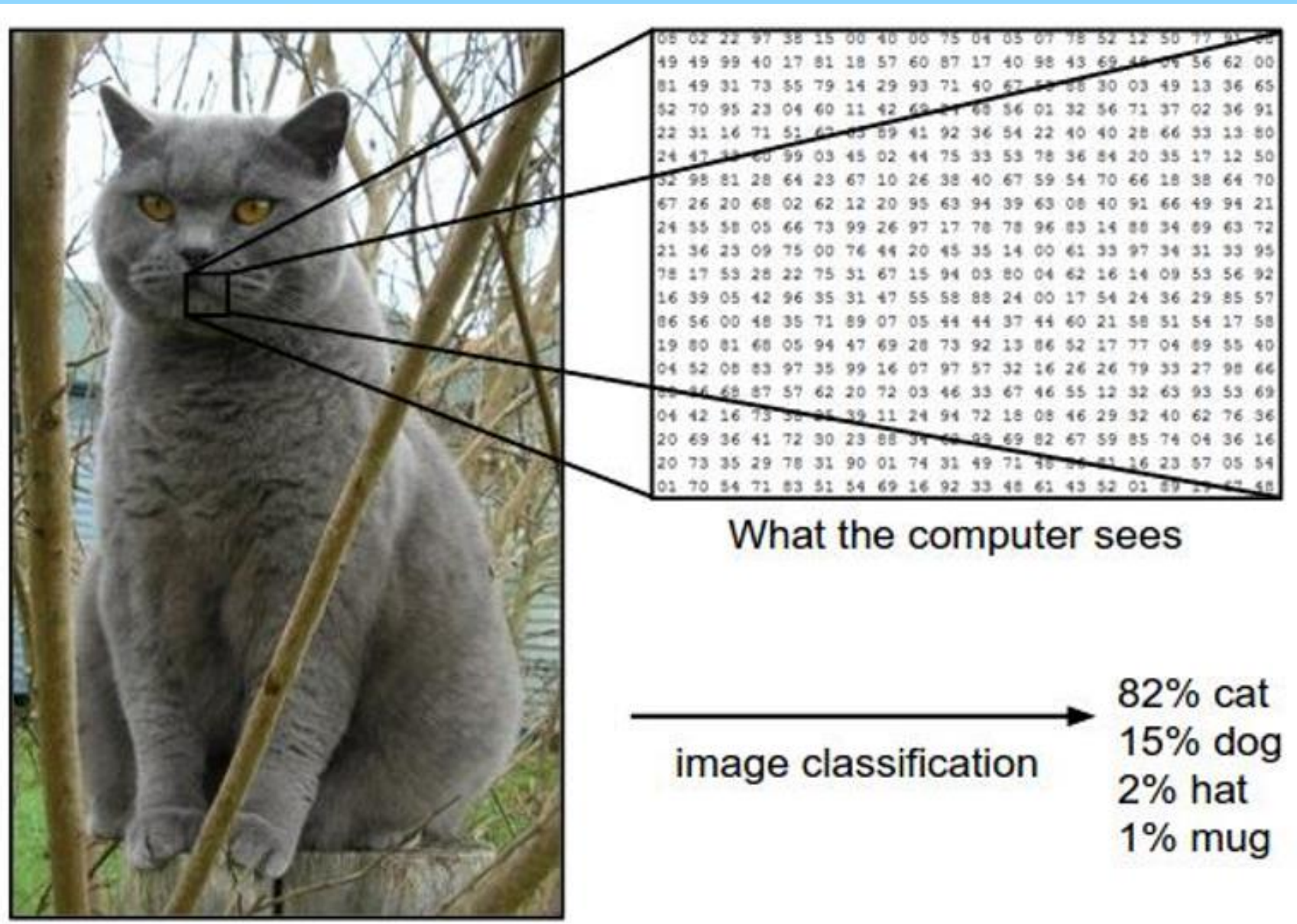
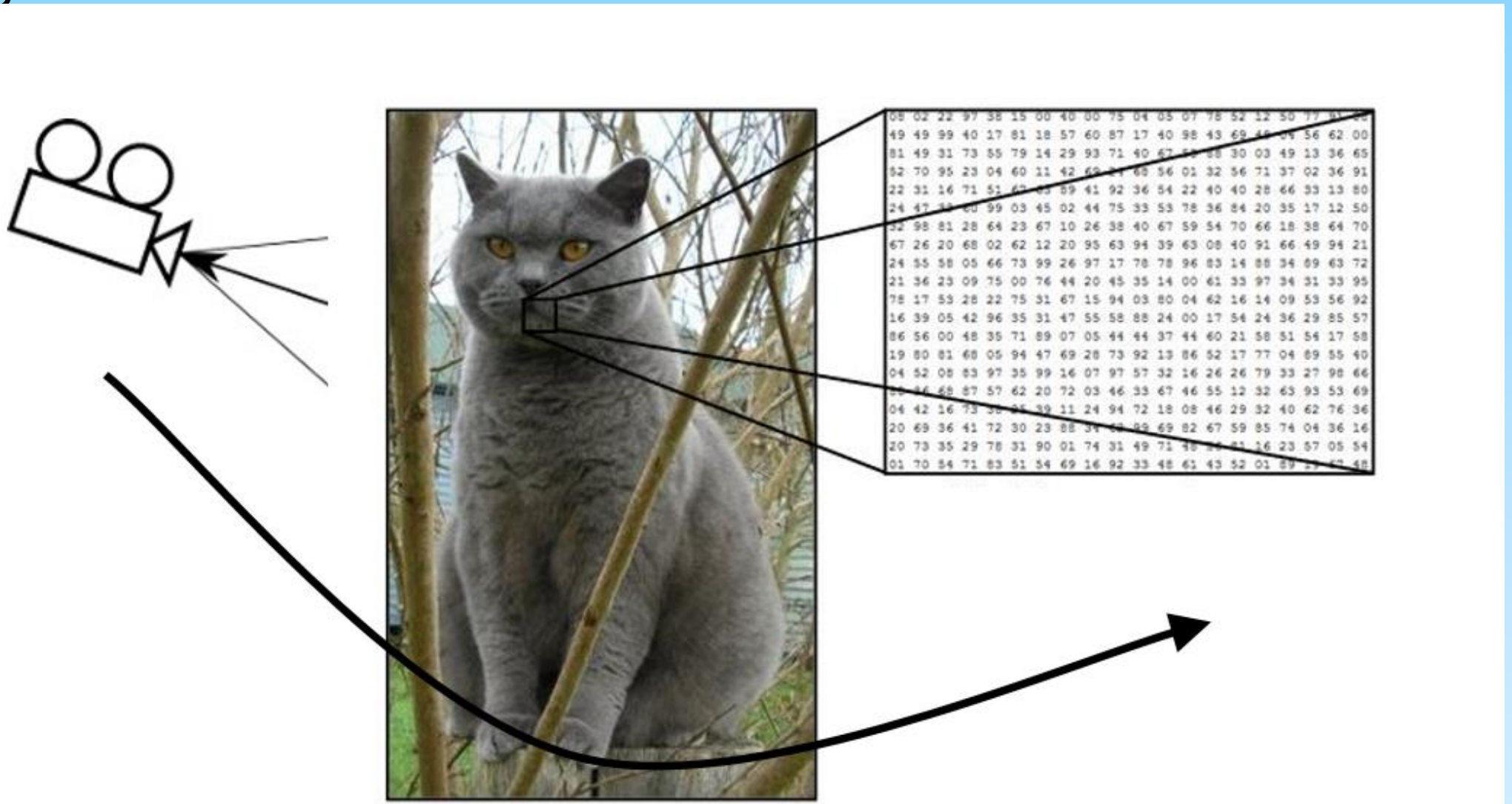
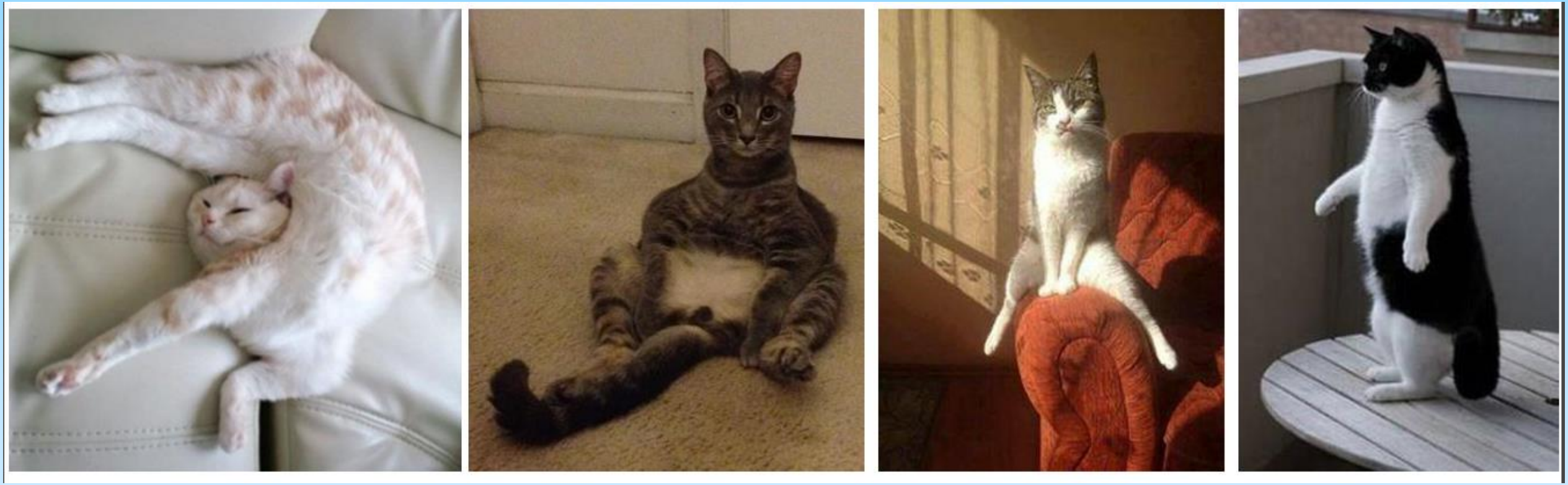


Image classification



Challenge: deformation



Challenge: occlusion

Challenge: background clutter

Challenge: intraclass variation

Classical pipeline for classification

- Classical CV

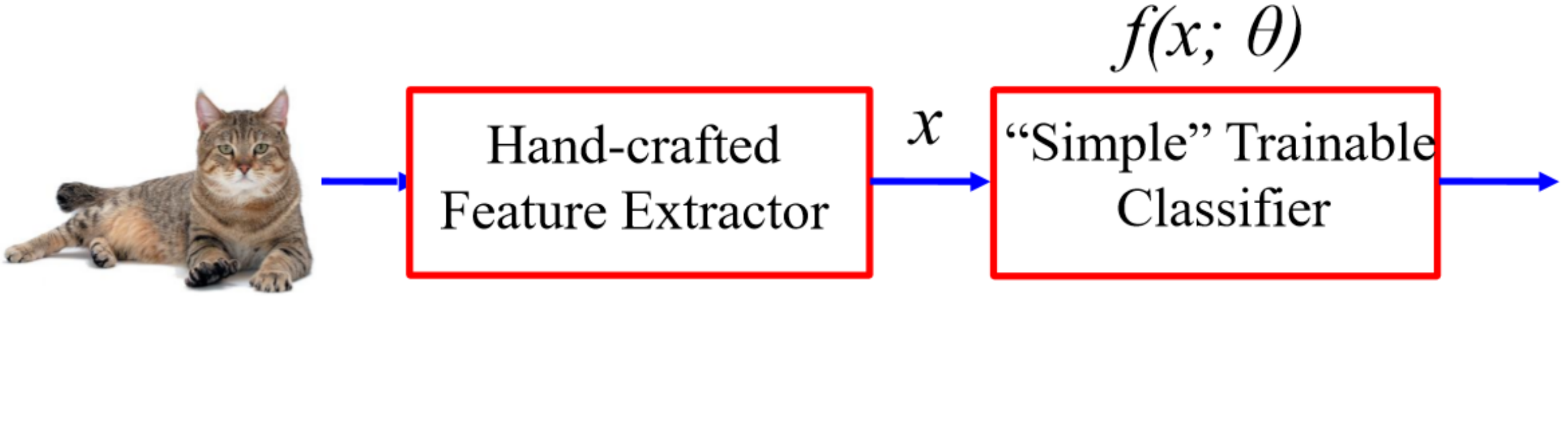
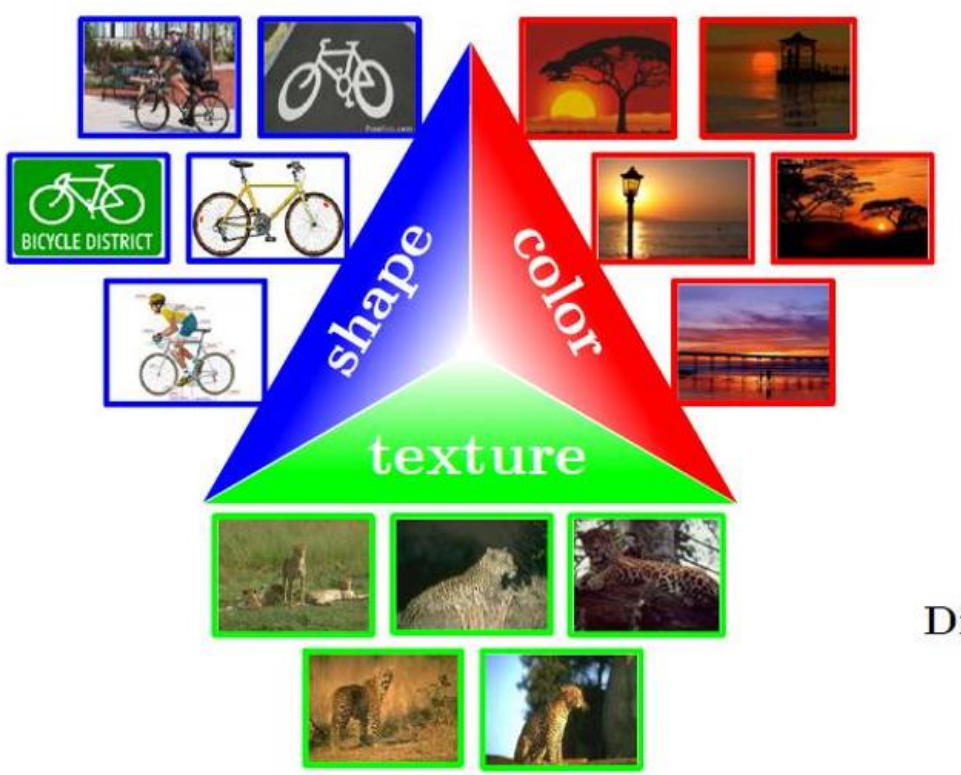
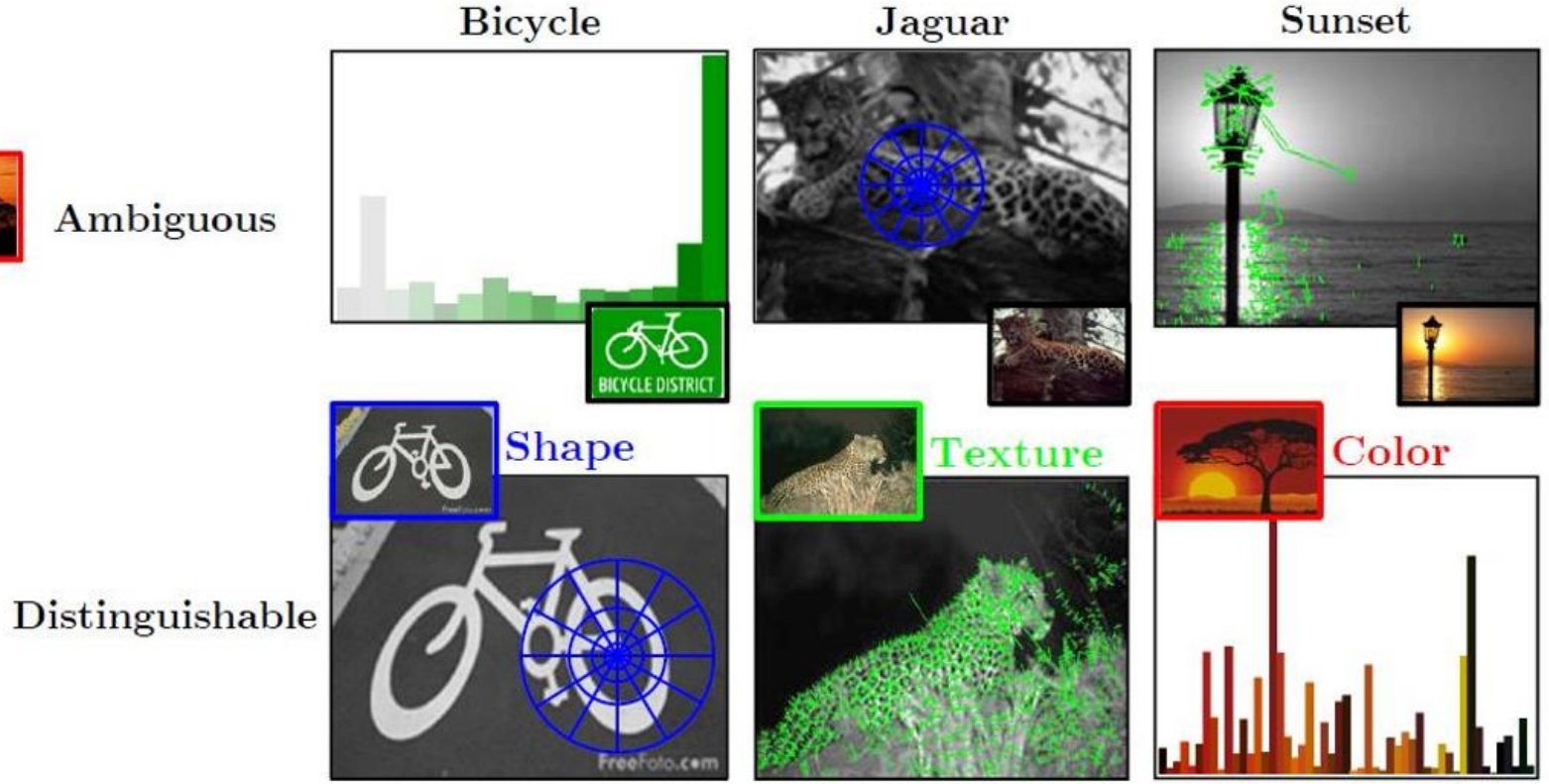


Image classification

- 1/特徵選定是分類器發展的關鍵。
- 2/手工製作的標準答案功能是否最佳?
- 3/分類器的最佳解，通常因目標選定而異，甚至因類別而異。





Classification pipelines

- Classical CV

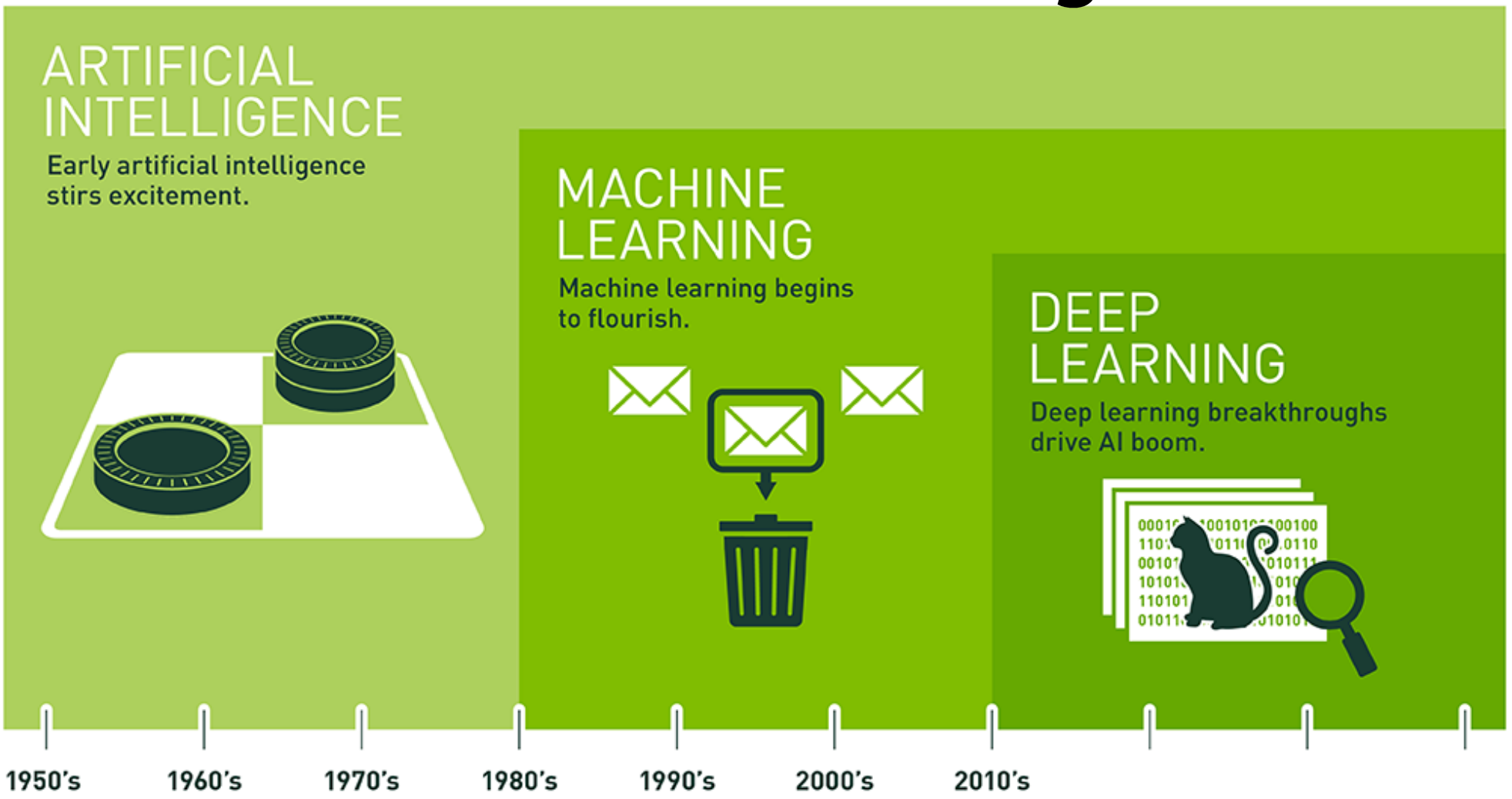
Hand-crafted Feature Extractor → "Simple" Trainable Classifier →

- Deep learning

Trainable Feature Extractor → **Trainable** Classifier →

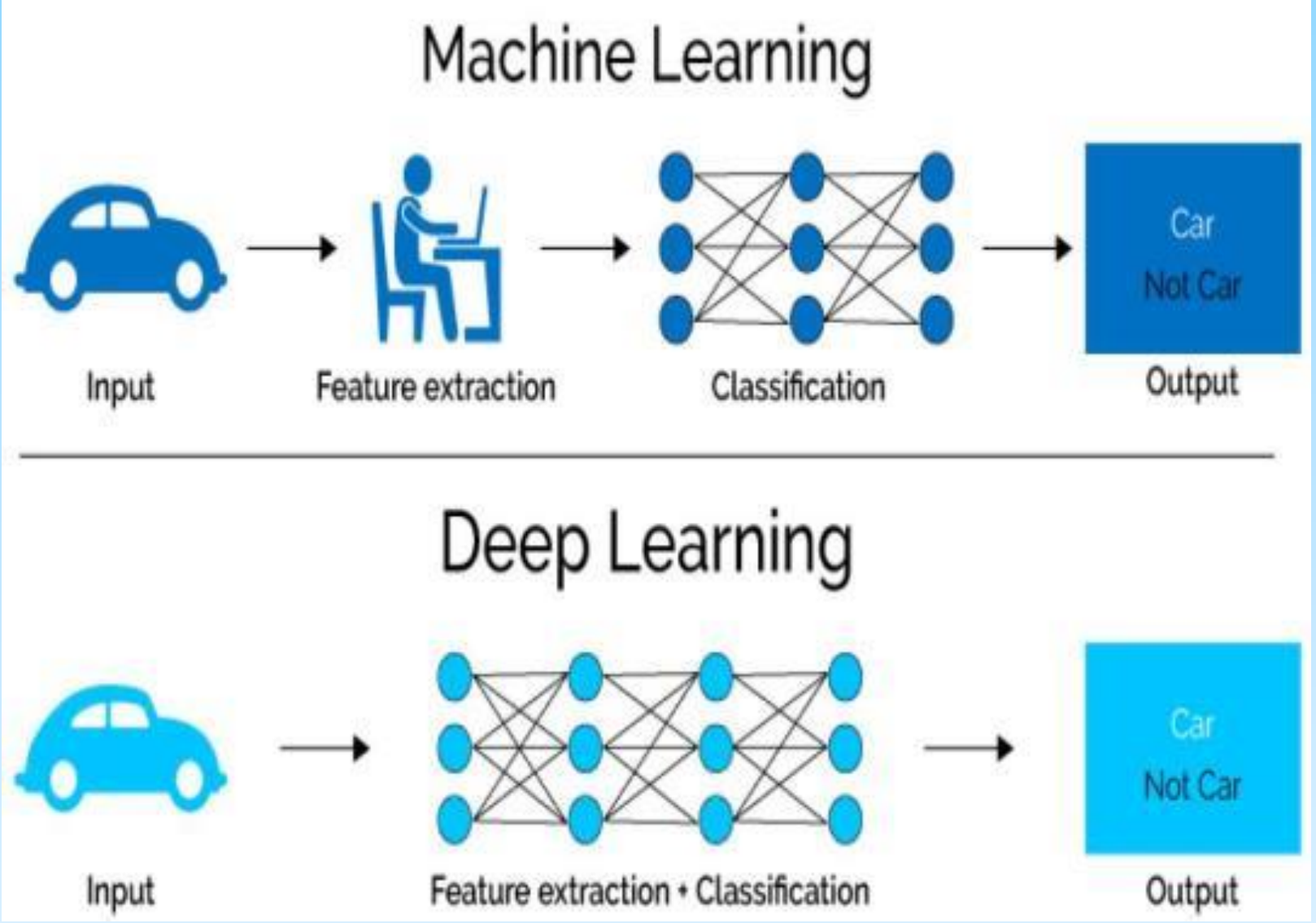
End-to-End Learning

AI人工智慧（Artificial Intelligence）



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

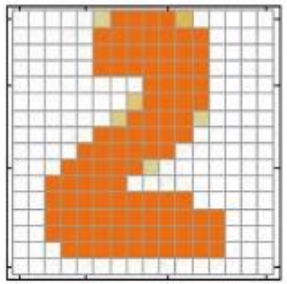
ML v.s DL


Machine Learning
Input
Feature extraction
Classification
Car
Not Car
Output
Deep Learning
Input
Feature extraction • Classification
Car
Not Car
Output

深度神經網路

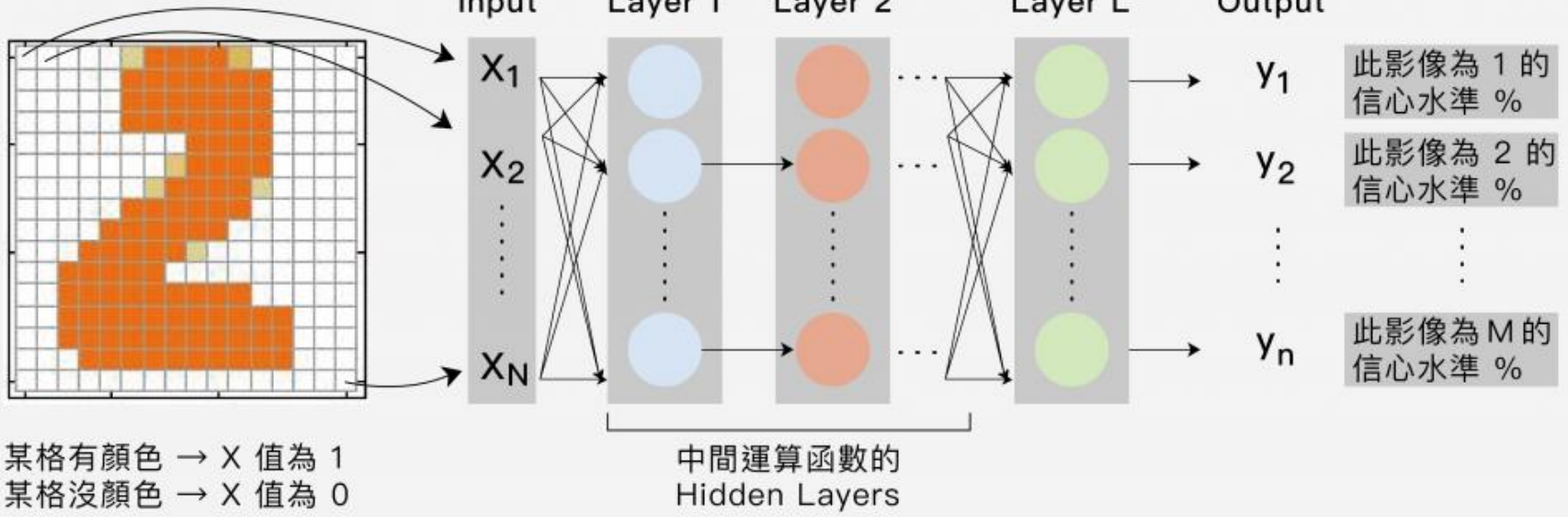
目的

教電腦學會
辨識數字 2 的影像



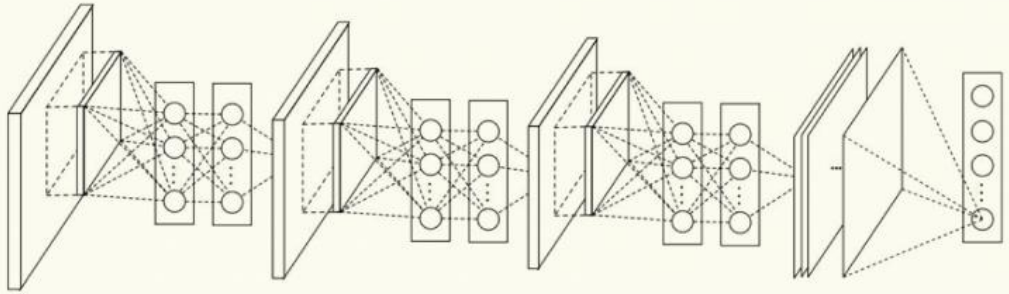
Machine → “2”

深度神經網路

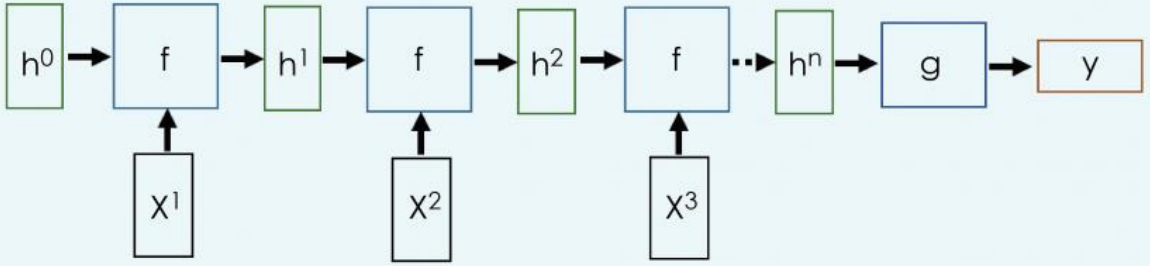


Deep Learning 之下

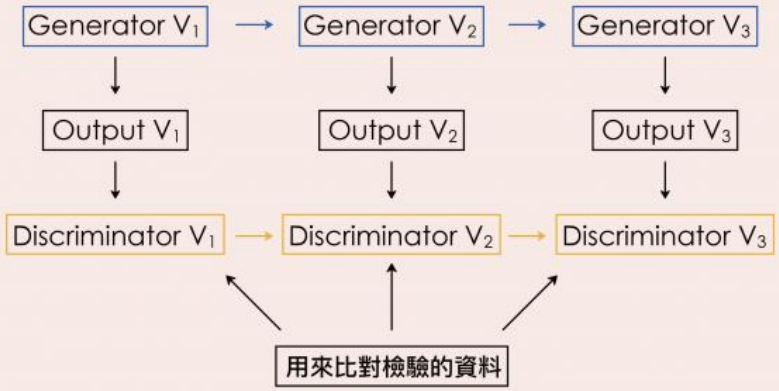
CNN
Convolutional Neural Network
卷積神經網路



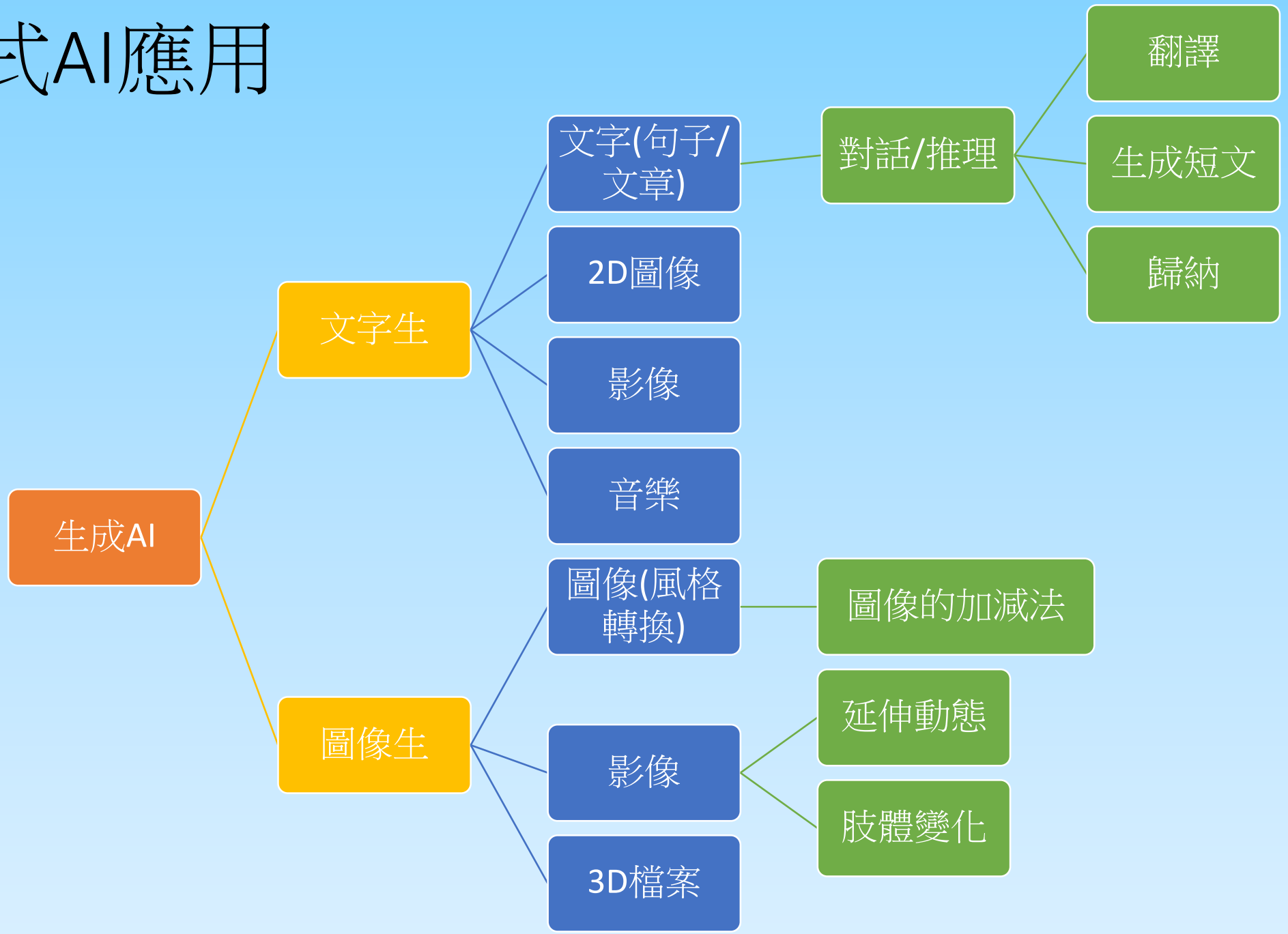
RNN
Recurrent Neural Network
循環神經網路



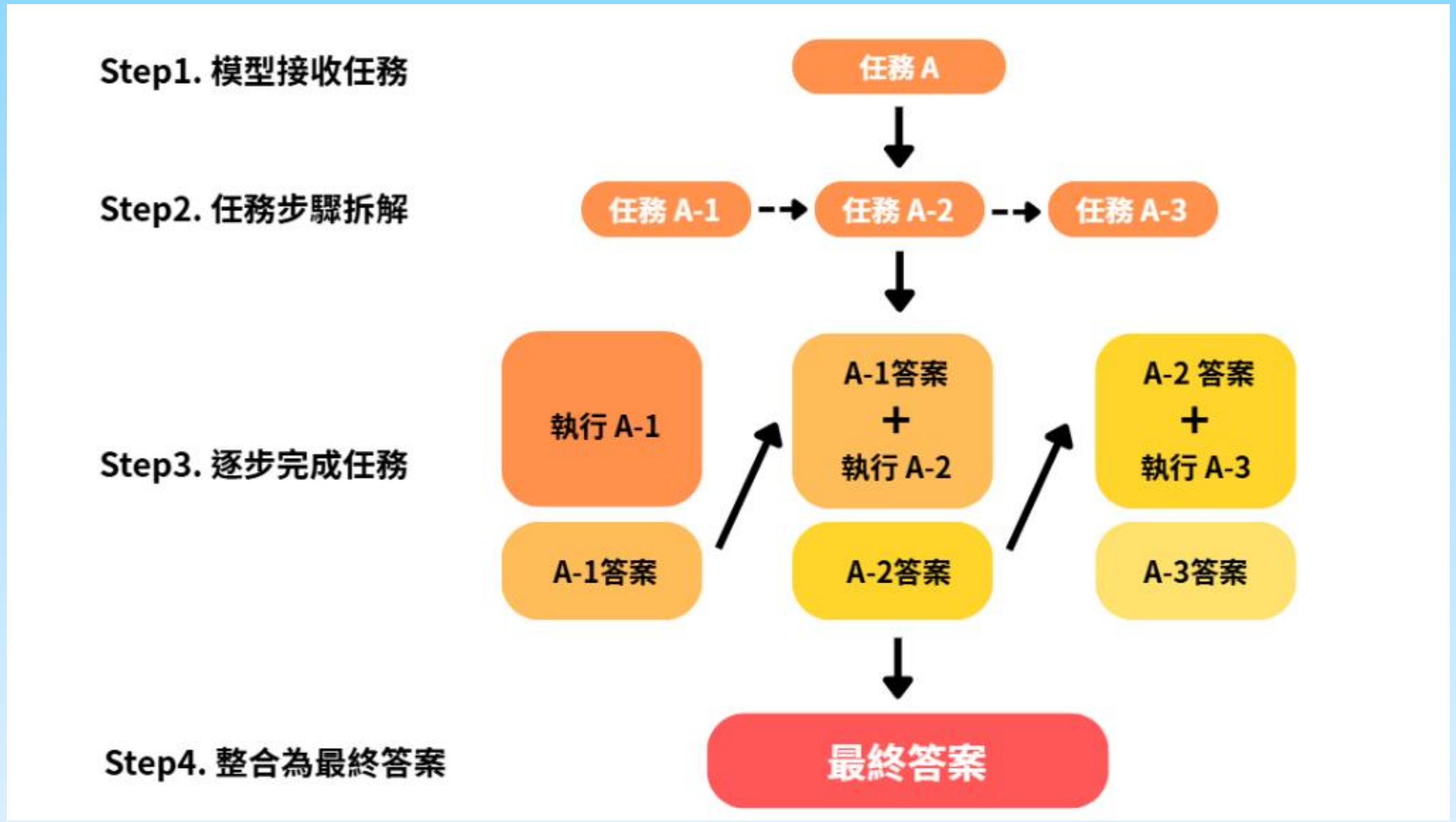
GAN
Generative Adversarial Nets
生成式對抗網路



生成式AI應用


生成AI
文字生
文字(句子/文章)
對話/推理
翻譯
生成短文
歸納
2D圖像
影像
音樂
圖像生
圖像(風格轉換)
圖像的加減法
影像
延伸動態
肢體變化
3D檔案

大公司模型-推理


Step1. 模型接收任務
任務 A
Step2. 任務步驟拆解
任務 A-1
任務 A-2
任務 A-3
Step3. 逐步完成任務
執行 A-1
A-1答案
A-1答案
+
執行 A-2
A-2答案
A-2 答案
+
執行 A-3
A-3答案
Step4. 整合為最終答案
最終答案

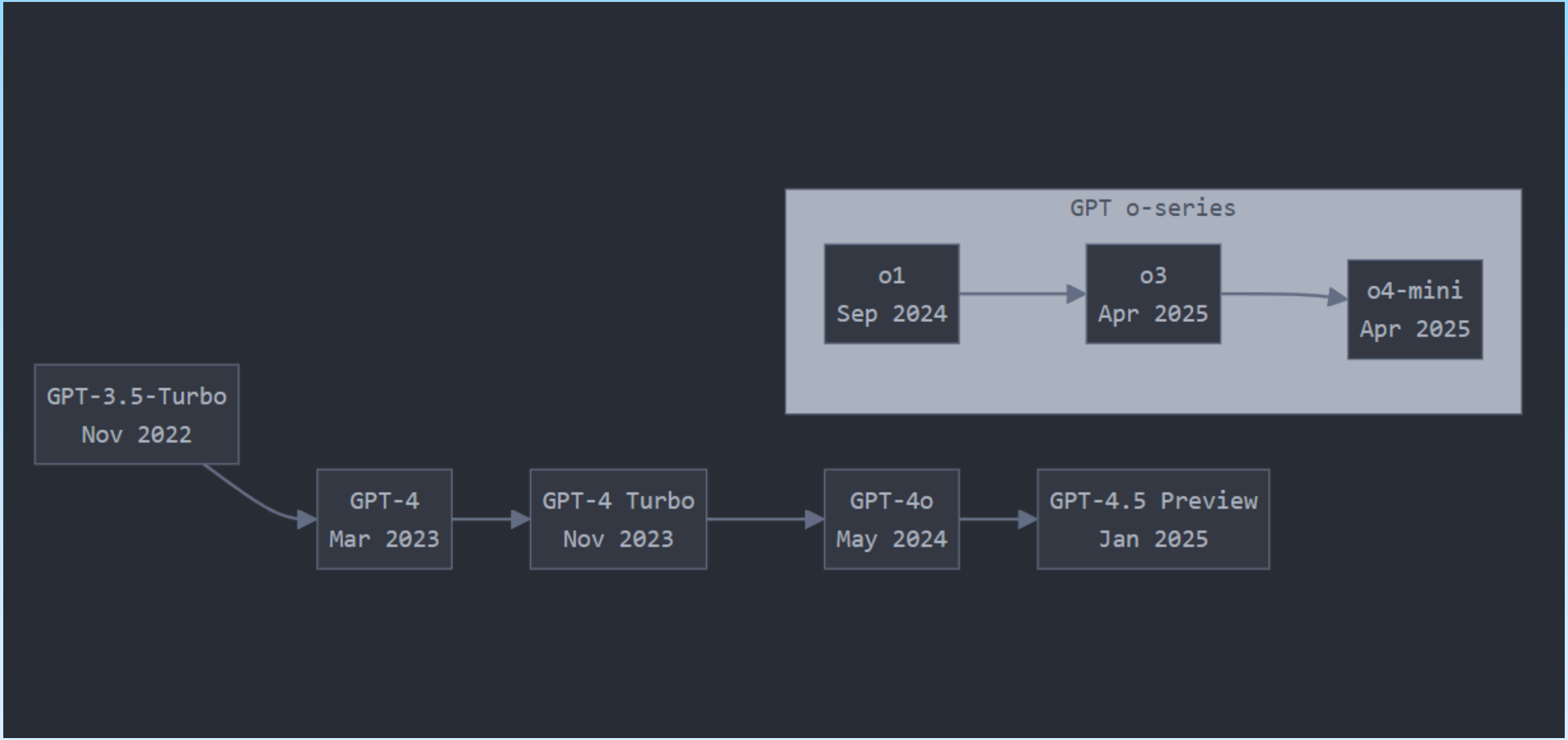
大公司模型-ChatGPT

比較項目	推理模型（o3）	一般模型（GPT-4o）
核心特色	推理**透明且完整**	生成**快速且即時**
運作過程	自動判斷需求， **規劃步驟、主動除錯**	**僅完成部分任務**， **需多次互動**以推進任務執行
結果呈現	內容詳細， 並以表格等視覺化方式清楚說明	條列簡述， 多為分類結果概述
限制	面對複雜問題時， **推理耗時**、處理時間較長	不利處理複雜任務， 需多次釐清需求
適用情境	• 需完整規劃的任務 • 一次性完成的高複雜任務	• 單一步驟的簡單任務 • 需人為逐步監督確認的任務

大公司模型-ChatGPT

比較項目		模型亮點	適用場景	限制
一般模型	**GPT-4o**	• 通用性最高	日常任務 （如 會議記錄、起草郵件）	智力較 GPT-4.5 低
	GPT-4o mini	• 成本效益最高	快速回應簡單任務 （如 大量文件摘要）	智力最低
	GPT-4.5	• 模型規模最大 • 情感互動強	長篇寫作及創意發想	回應速度較慢
推理模型	**o3**	• 通用推理能力最高 • 整合多項應用工具	複雜與多步驟的任務 （如 策略規劃、分析）	幻覺風險較高
	o4 mini	• 成本效益最高 • 整合多項應用工具	快速回應複雜任務 （如 Python 程式碼除錯）	推理能力最低
	o4 mini high	• 最強推理模型	複雜的高階任務 （如 高階數學、程式編碼）	回應速度較慢

ChatGPT模型進展



ChatGPT模型

家族	核心追求	最擅長的事情
GPT 系列	**廣度**：龐大語料預訓練，文字與多語言流暢度	一般對話、內容創作、圖片／聲音理解（4o 起）
o 系列	**深度**：刻意強化「多步推理」與工具調用	進階程式、數學解題、資料分析、自動化流程

兩系列 **並非** 誰淘汰誰，而是依照「成本、反應時間、推理深度」不同場景自選最合適的組合。

C/P值

¦ 價格 × 效能

（每千 token，美元，2025 年 4 月）

模型	In	Out	速度★
GPT-3.5-Turbo	**$0.002**	$0.006	★★★★☆
GPT-4	0.03	0.06	★★☆☆☆
GPT-4 Turbo	0.01	0.03	★★★☆☆
GPT-4o	**0.005**	0.015	★★★★★
GPT-4.5 prev.	0.008	0.024	★★★★☆
o3	0.01	0.04	★★★☆☆
o4-mini	0.003	0.012	★★★★☆

大公司模型


ogle AI Studio
使用 Gemini 實現從提示到生產的最快方式
在遊樂場與模特兒聊天
在 Build 中啟用 Vibe 代碼 GenAI 的應用程式
在儀表板中監控使用情況
什麼是新的
嘗試奈米香蕉
Gemini 2.5 Flash Image，最先進的圖像生成和編輯
Veo 3.1
我們最好的影片生成模型，現在具有音效。
使用 URL 上下文獲取信息
從網路連結取得即時資訊
使用 Gemini 產生母語語音
使用 Gemini 產生高品質的文字轉語音

各平台模型

	OpenAI ChatGPT	Google Gemini	Perplexity AI	xAI Grok	Microsoft Copilot
推理模型	o3 o4 mini o4 mini high	Gemini 2.5 Flash (experimental) Gemini 2.5 Pro (experimental)	OpenAI o4 mini Claude 3.7 Sonnet DeepSeek R1 1776	Grok-3 Beta Grok-3 mini Beta	OpenAI o1
使用方式	**免費用戶：**點擊「推理」啟動推理模型 **付費用戶：**可以選用指定推理模型	選用推理模型即可啟動	所有用戶在對話時即會自動進行推理 付費用戶可以選用指定推理模型	所有用戶透過點按「Think」啟動推理功能	所以用戶皆可點按「Think Deeper」使用推理模型
方案及額度	所有用戶皆可使用，使用額度限制依方案而定	所有用戶皆可使用，使用額度限制依方案而定	所有用戶皆可使用，使用額度限制依方案而定	所有用戶皆可使用，使用額度限制依方案而定	所有用戶皆可使用，使用額度限制依方案而定


時に、西暦2021
主題代號
場景配件
補完計畫
1/64ミニカー
64模型事務所
襲來
第2新屏東市
廢除禁行機車
廢除車種分流
第壹話
記念
文創收錄內容
起動準備完了
文化記憶的再構築を開始せよ。
交通革命
路權平等
行人安全
合理速限
廢除禁行
前行測速照相
命中無效化
起動中
未來を守る
記憶、構築、展示
覺醒
台灣都市交通構造要素改革
交通再構築戰術司令部
機密機關 T.R.U.S.T.


AI
生成搭配示意圖


AI
生成搭配示意圖


AI
生成搭配示意圖

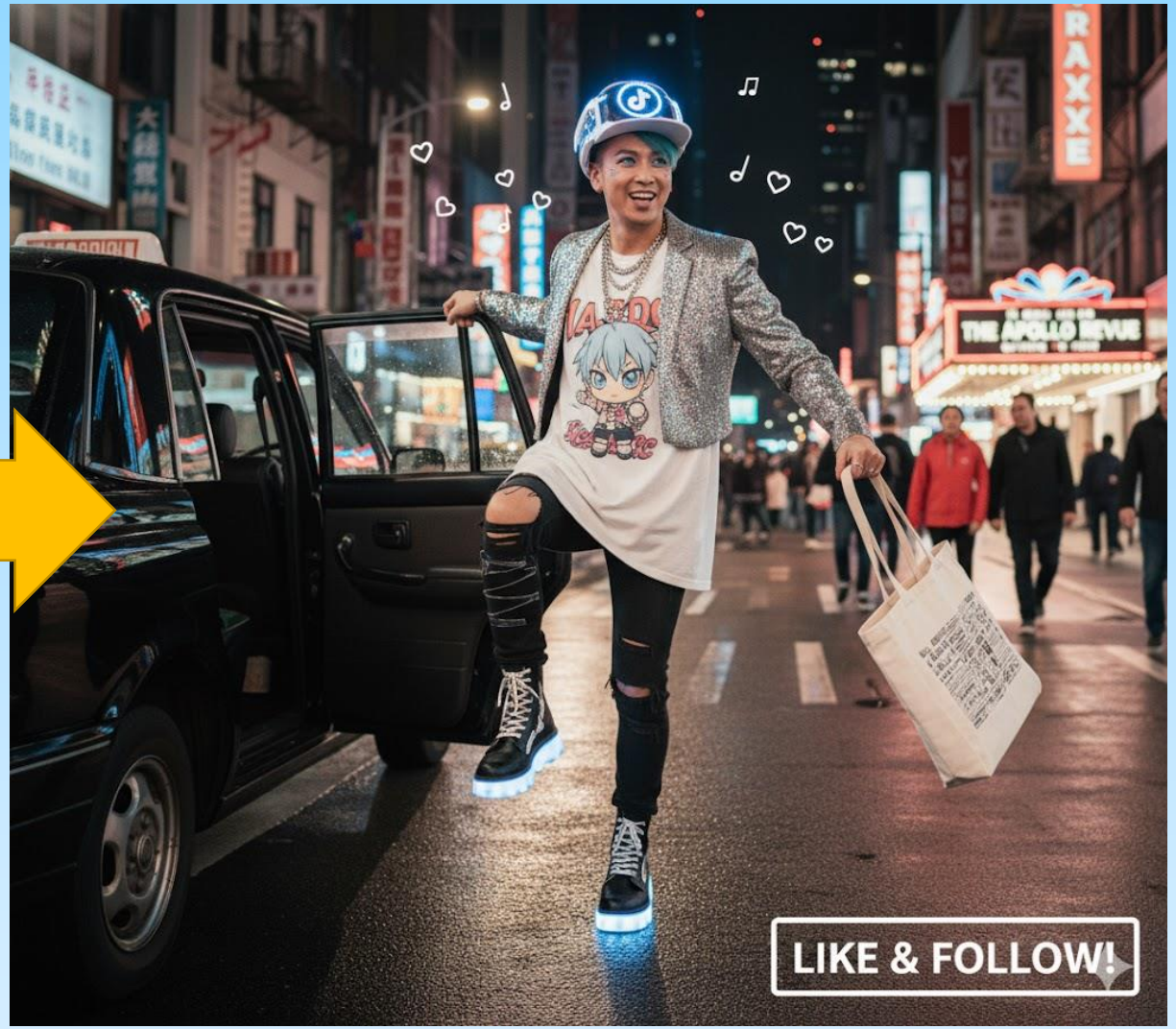
AI生成搭配示意圖

同學開啟Gemini試試吧! [課程小練習]

Gemini 2.5 Flash Image，暱稱「奈米香蕉」（nano-banana）

同學開啟Gemini試試吧! [課程小練習]

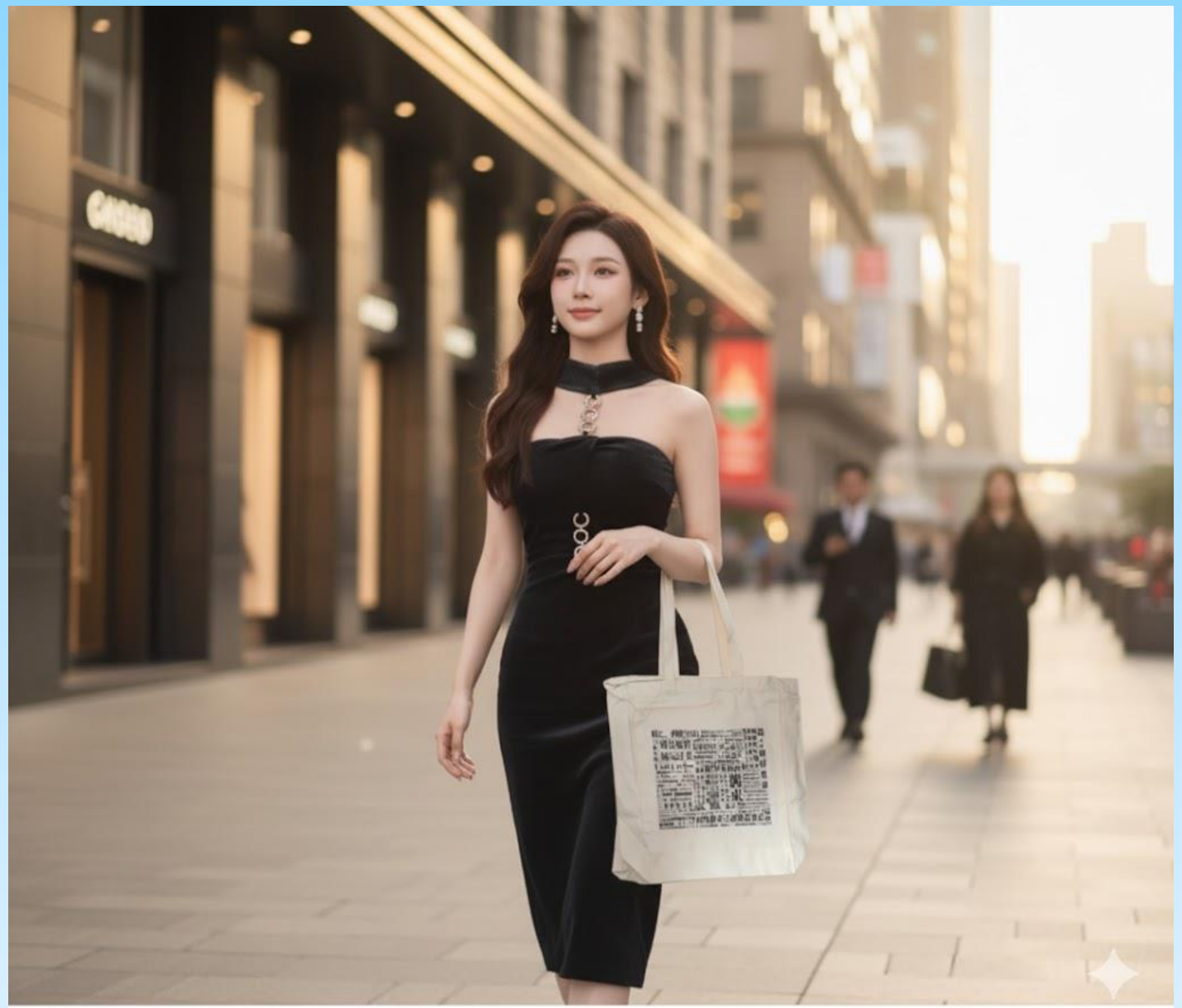
我已經調整了表演者的風格，以符合抖音上流行的表演者形象。



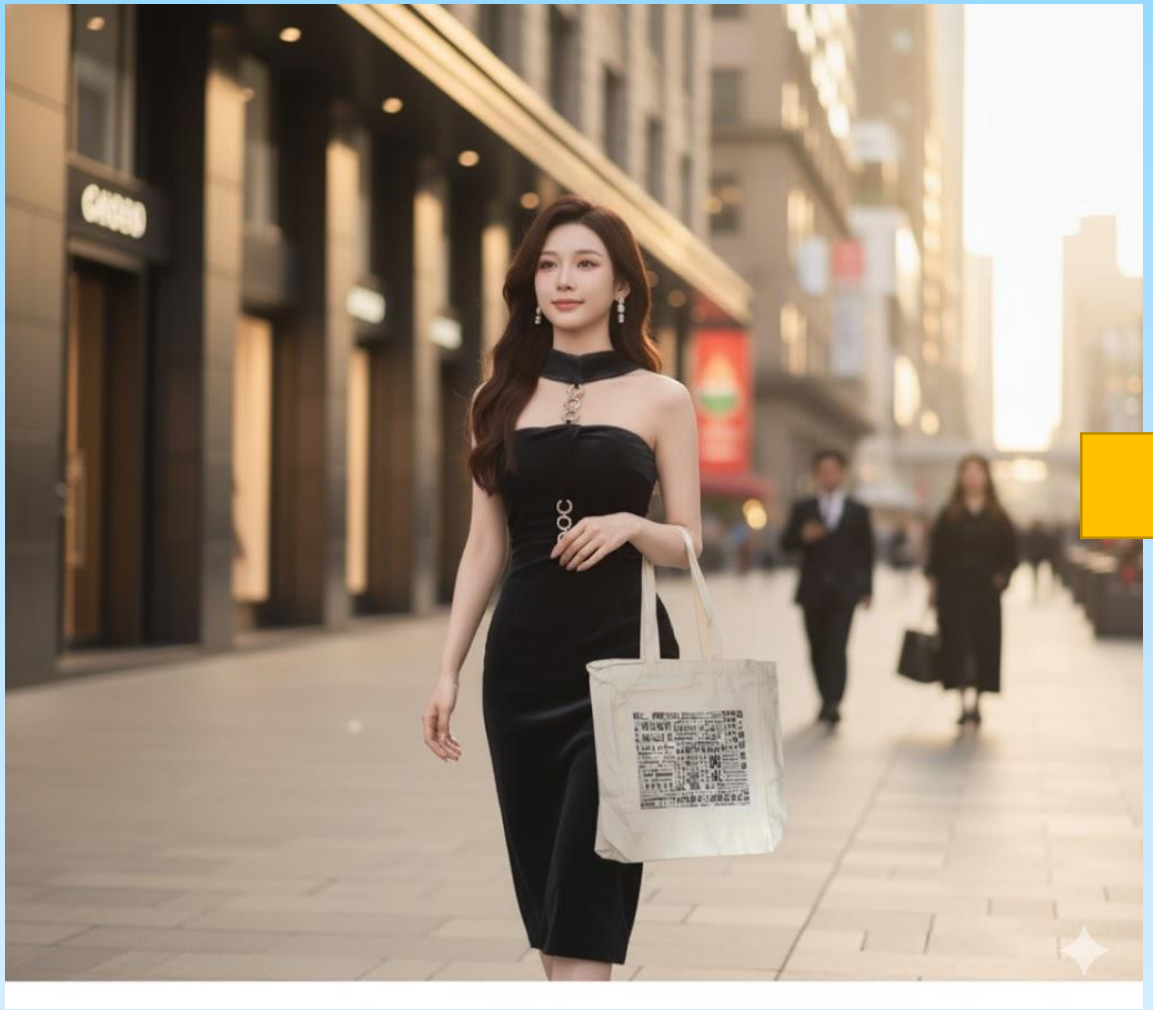
加法練習

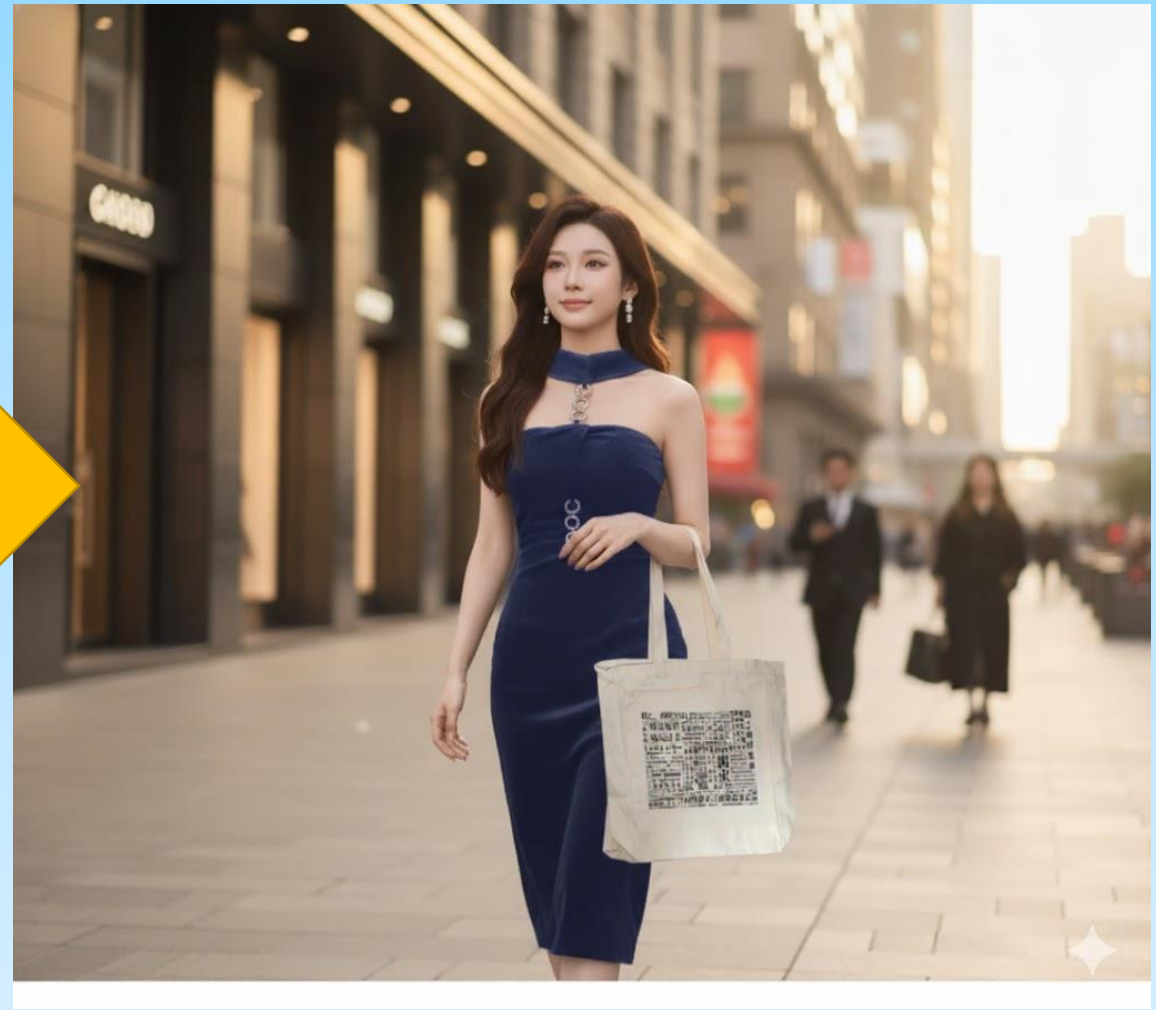
我需要一張此照片女生提這個袋子優雅走在路上

加法練習

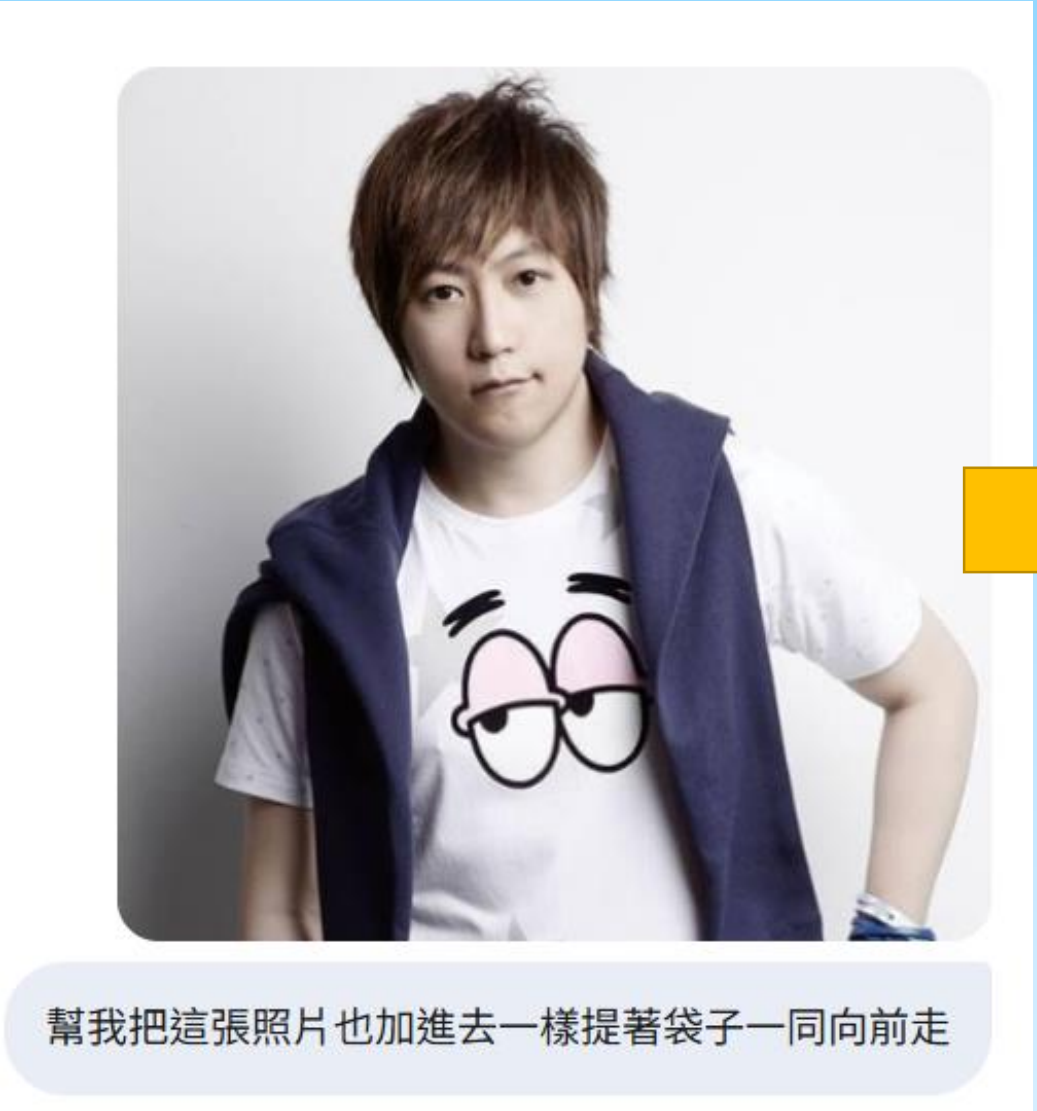


減法練習





人物加法



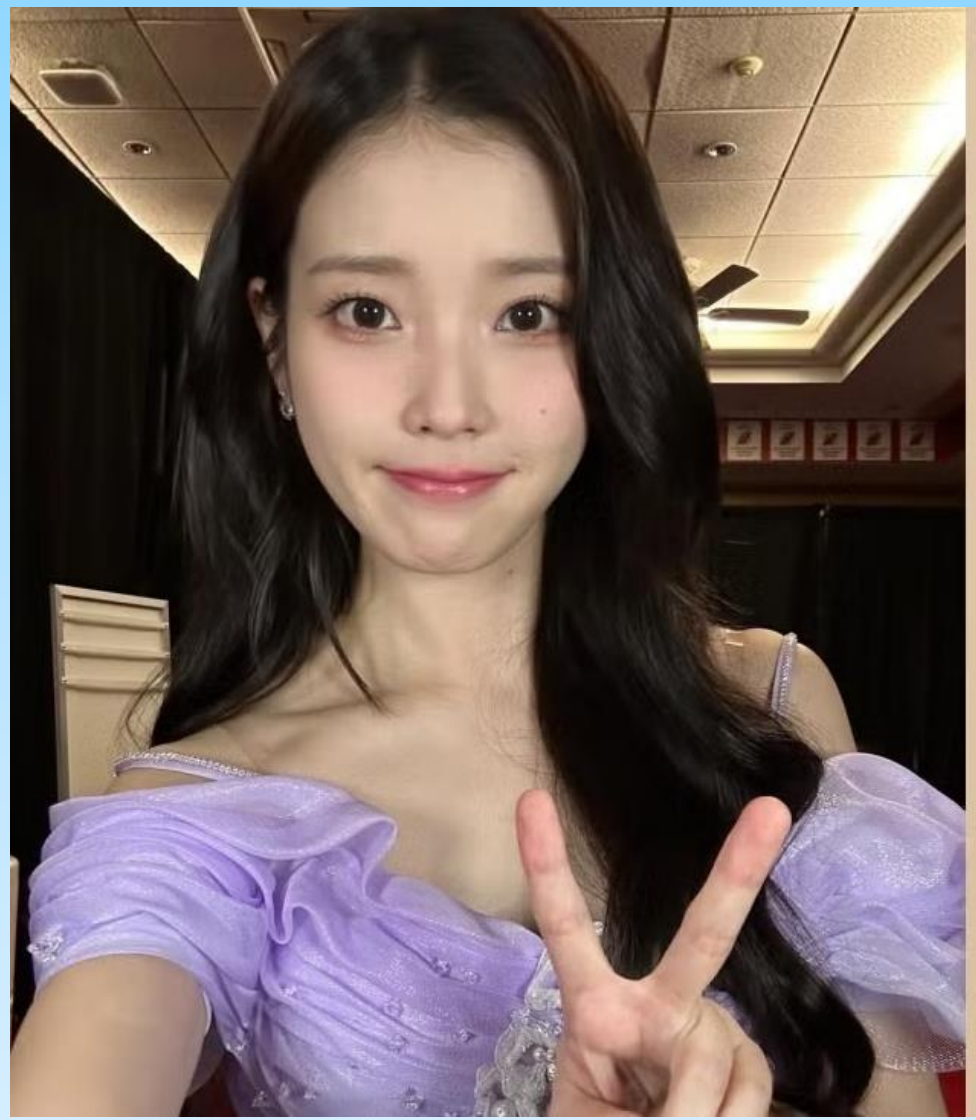
?

人物加法

時尚圈

- 在南韓就有一位名為Sae Na的虛擬網紅，與韓國最強女團BLACKPINK來自同一個經紀公司YG娛樂，虛擬網紅Sae Na不僅受邀出席Jimmy Choo快閃店活動、走訪Maison Margiela咖啡廳，還與YG旗下明星李洙赫、Treasure同框拍照。

3D公仔盒



3D公仔盒

- 下指令範本：

- 請參考此照片 生成一張公仔的圖片（記得要上傳照片）

- 奶茶色系的公仔包裝設計
- 上方要標有IU字樣
- 包裝內應包含專屬區塊表示「FOODIE」
- 娃娃感十足的圓圓大眼睛，但是也要有照片 中人物的神韻
- 紫色髮帶
- 質感韓系陳列，配件整齊擺放

3D公仔盒

- 配件如下：

- 一支麥克風
- 一束花
- 一盒義美小泡芙
- 一台粉色富士相機
- 一隻棕色泰迪熊絨毛娃娃
- 包裝要有掛勾設計，符合零售陳列需求

Convolutional Neural Network (CNN) **Binary Classification**

鄭朝元

V09

Binary Classification

- 二元分類(Binary Classification)是根據分類規則將一組數據（data points）分為兩組（classes）的任務。以下為常見二元分類演算法。

  Decision trees (決策樹)

  Random forests (隨機森林)

  Bayesian networks (貝氏網路)

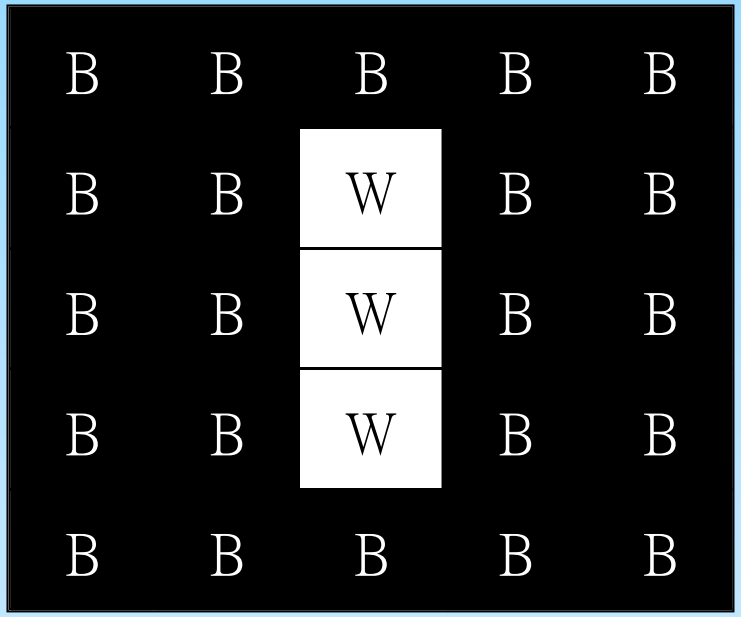
  Support vector machines (SVM,支持向量機)

  Neural networks (NN,類神經網路)

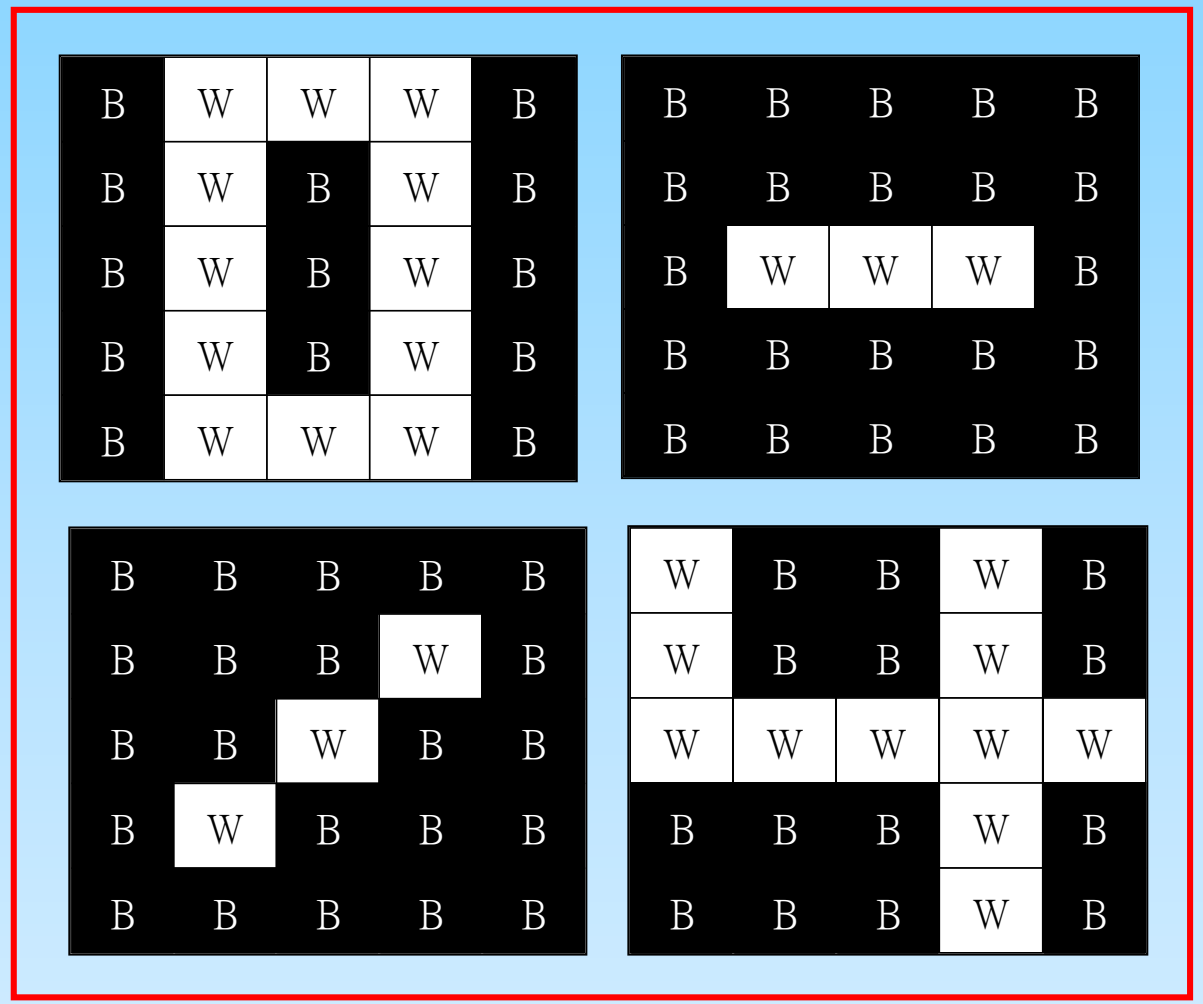
  Logistic regression (邏輯回歸)

  Probit model (機率模型)

二元分類器以分類圖像上所描述數字1為例



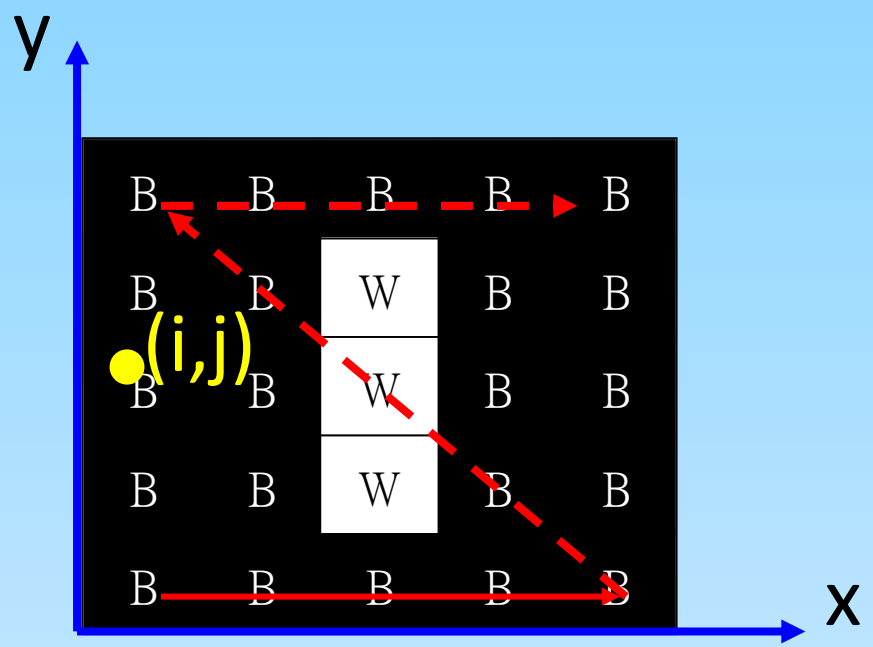
GTT.bmp此為5x5 pixels之圖像，其所代表為數值1。



GTF.bmp此群為5x5 pixels之圖像，其所代表為數值非1。

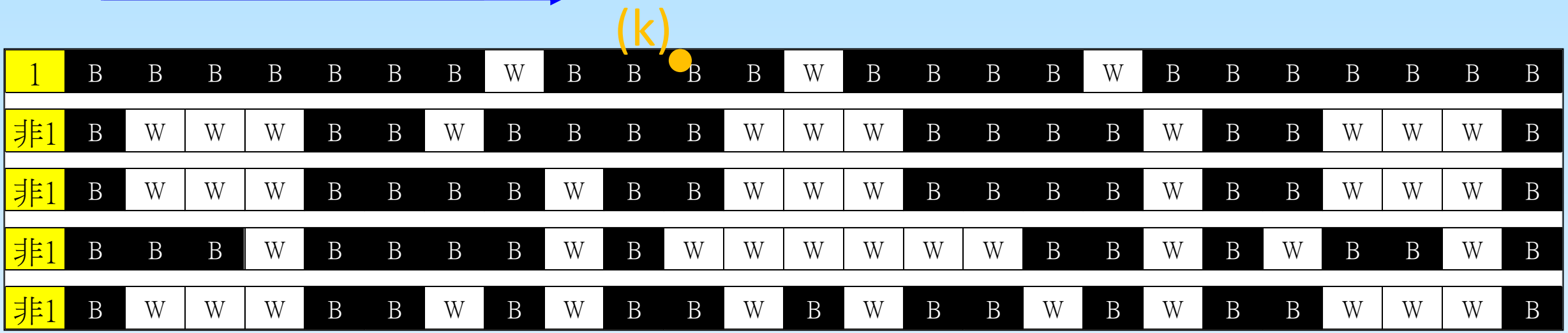
附註：B(Black)，W(White)所代表為圖像之intensity值。

二維編碼轉換為一維編碼

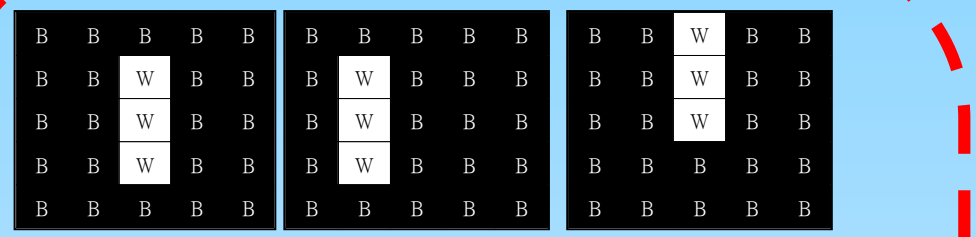


GTT.bmp與GTF.bmp為5x5 pixels，為使運算神經網路中做為單一節點個體單元運算，其需將原二維陣列圖像座標傳換成為一維陣列圖像座標，本圖像採用BMP圖檔座標系(i, j)，轉換為一維座標系(k)，而原座標所具有的intensity不會有所改變。

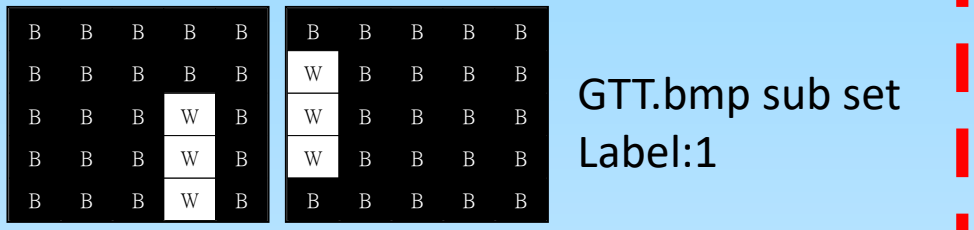
$$k = i + j \times imgW$$



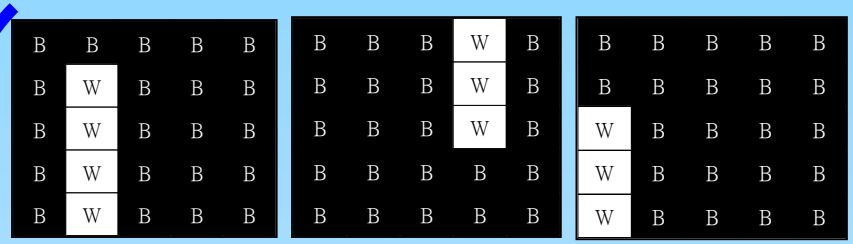
二元分類Data Set



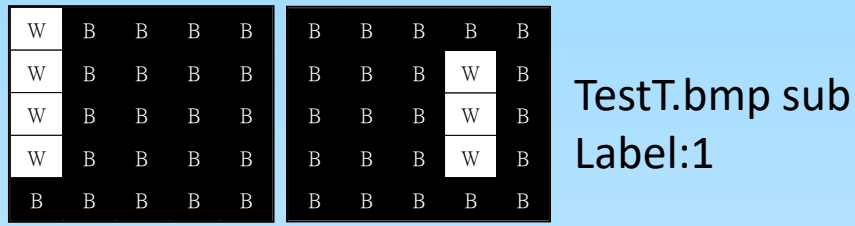
GTT.bmp sub set
Label:1



GTF.bmp sub set
Label:0



TestT.bmp sub set
Label:1



TestF.bmp sub set
Label:0

Training Data

Data Set

Testing Data

圖像Intensity Mapping對應關係轉換

Data set中之各圖像intensity為8位元之灰階圖像，其對應值關係如下：

人眼視覺感受	B	往左漸暗，往右漸亮	W
8位元灰階圖數值	0		255
神經網路節點數值 映射至0- 1區間	0		1

B	B	B	B	B
B	B	W	B	B
B	B	W	B	B
B	B	W	B	B
B	B	B	B	B

人眼視覺感受

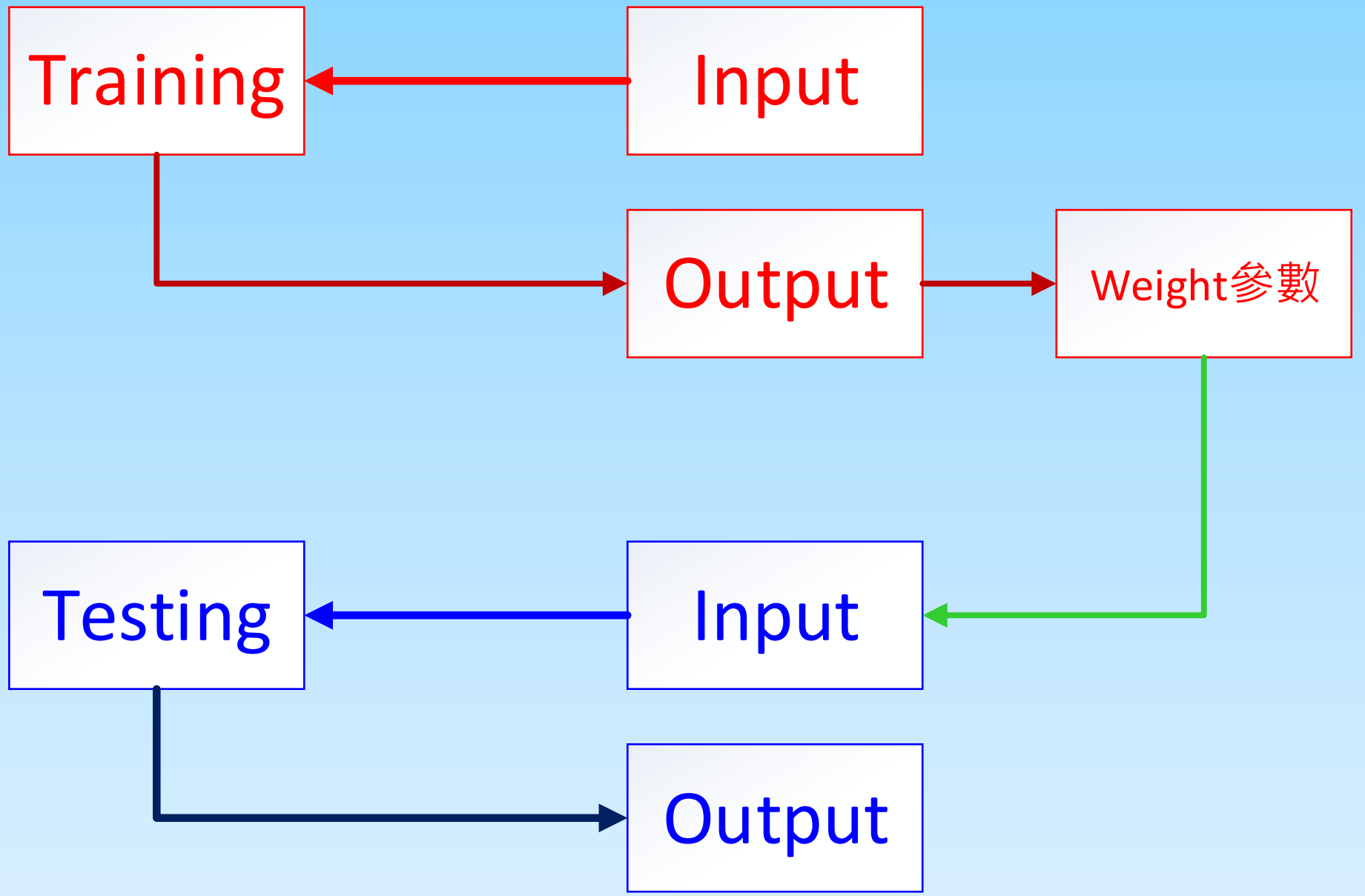
26	0	0	0	0
8	0	252	0	23
0	0	250	0	0
0	36	255	0	0
26	0	28	0	0

8位元灰階圖數值

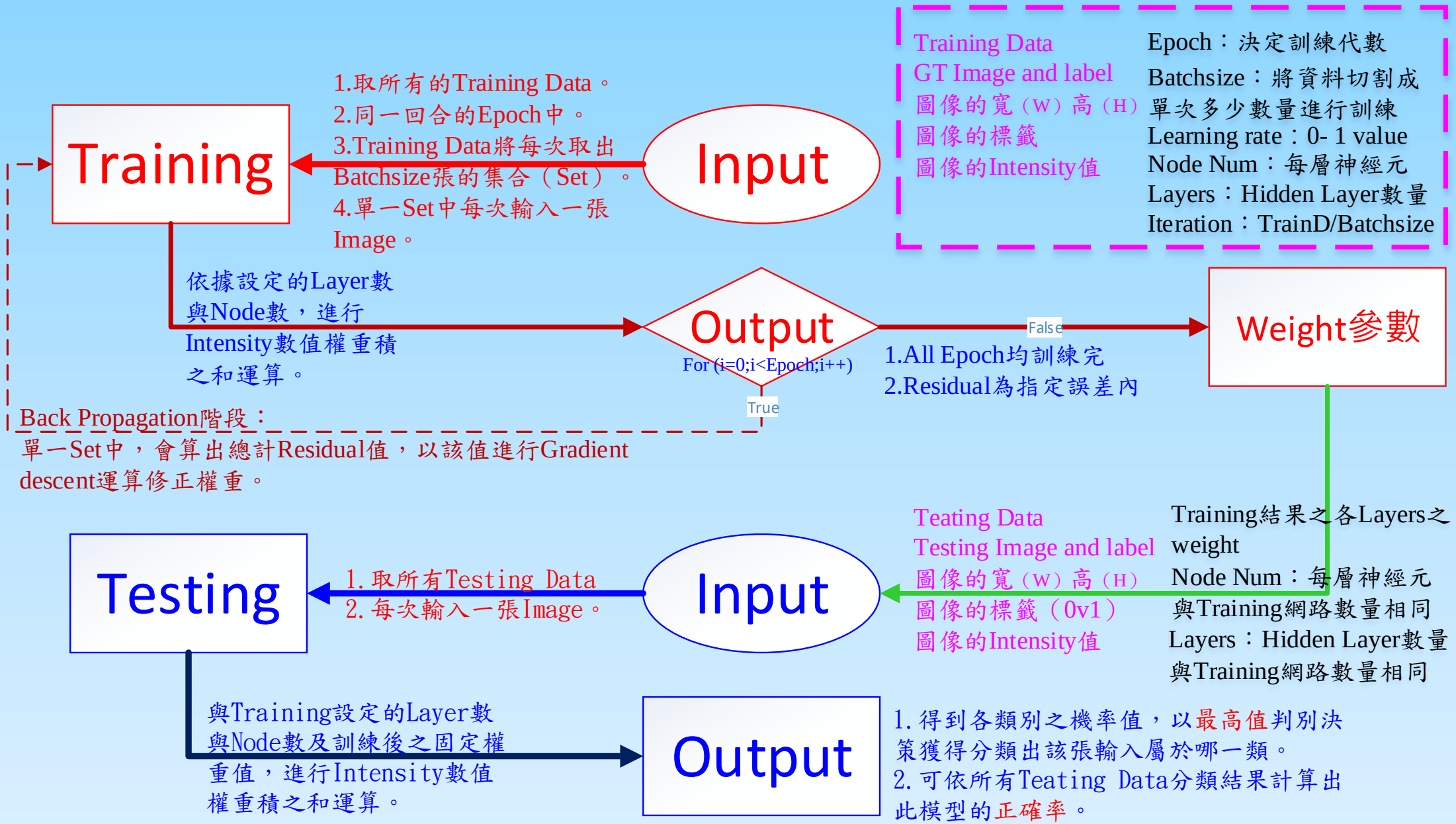
0.1	0	0	0	0
0.03	0	0.99	0	0.09
0	0	0.98	0	0
0	0.14	1	0	0
0.1	0	0.11	0	0

神經網路節點數值
映射至0- 1區間

Training與Testing相互網路關係簡圖



Training與Testing相互網路關係簡圖-Zheng註解



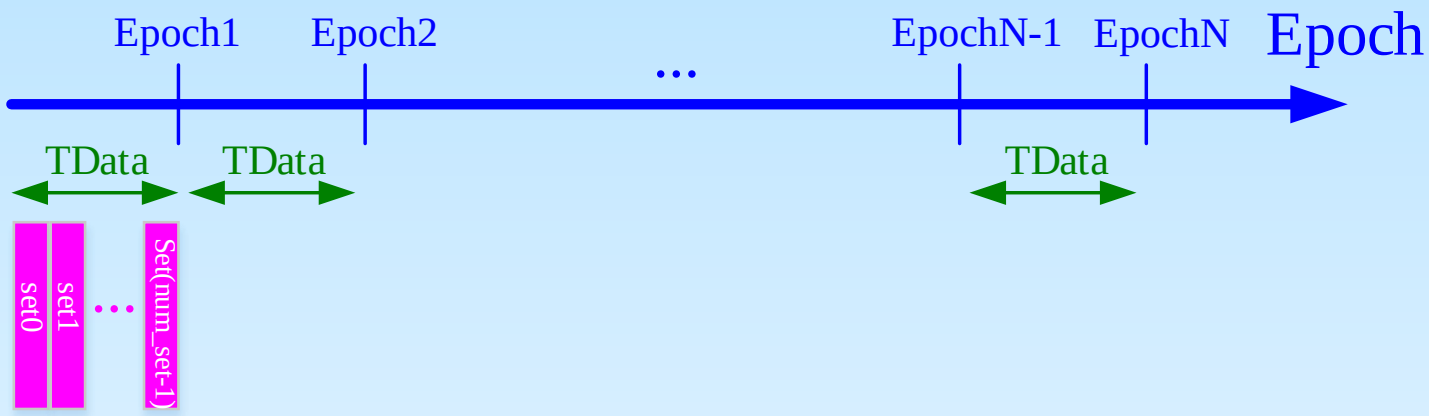
重要名詞定義

Training階段，吾人需準備Training Data，而同一筆Data中，皆須經過多代訓練，其所訓練之代數稱為Epoch。

時常，Training Data過多，需將其分割成多個集合，此定義為Set，每個Set中，均分Training Data中的Image，其具有張數稱為Batchsize。

每個Set均會經過Forward propagation，獲得一定統計之殘差值，其需以Back propagation修正各權重，此時每次修正稱為Iteration。

每個Epoch有多少set則就要相對Iteration多少次，故最低每個Epoch必有一次Iteration。

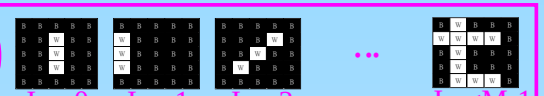


Epoch：訓練代數

Training Data

GTNum=num_set x Batchsize張

Set0



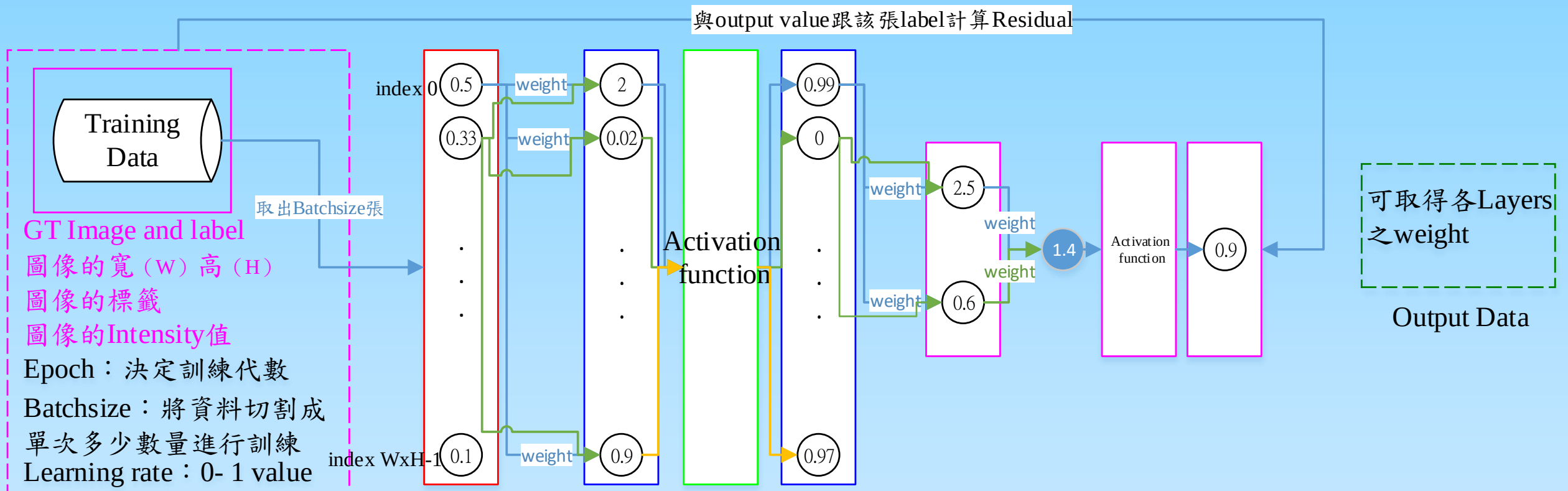
Batchsize張＝M張

Set1

set的數量：
num_set

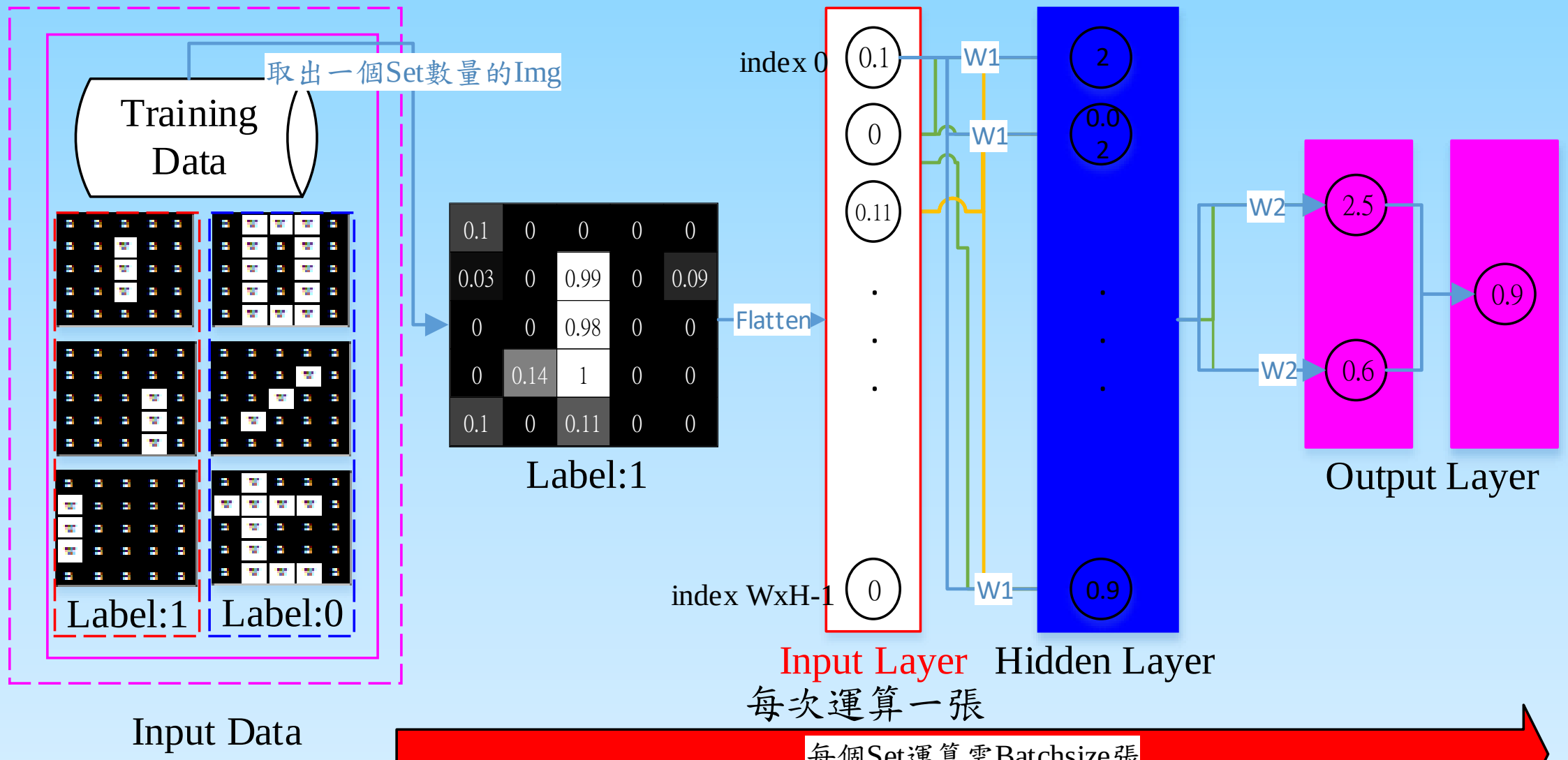
Set num_set-1

Training Stage



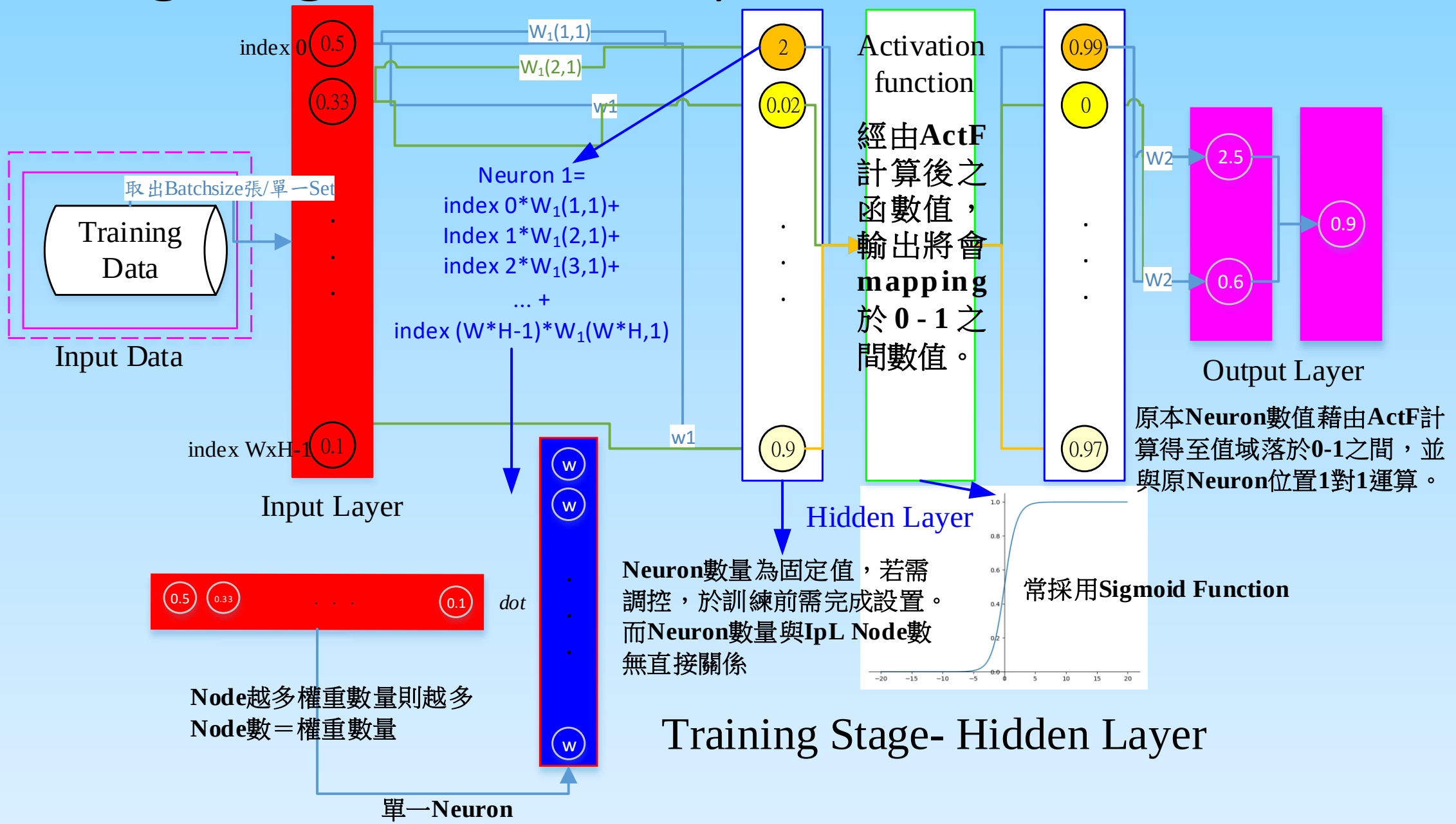
Training Stage

Training Stage- Input Layer



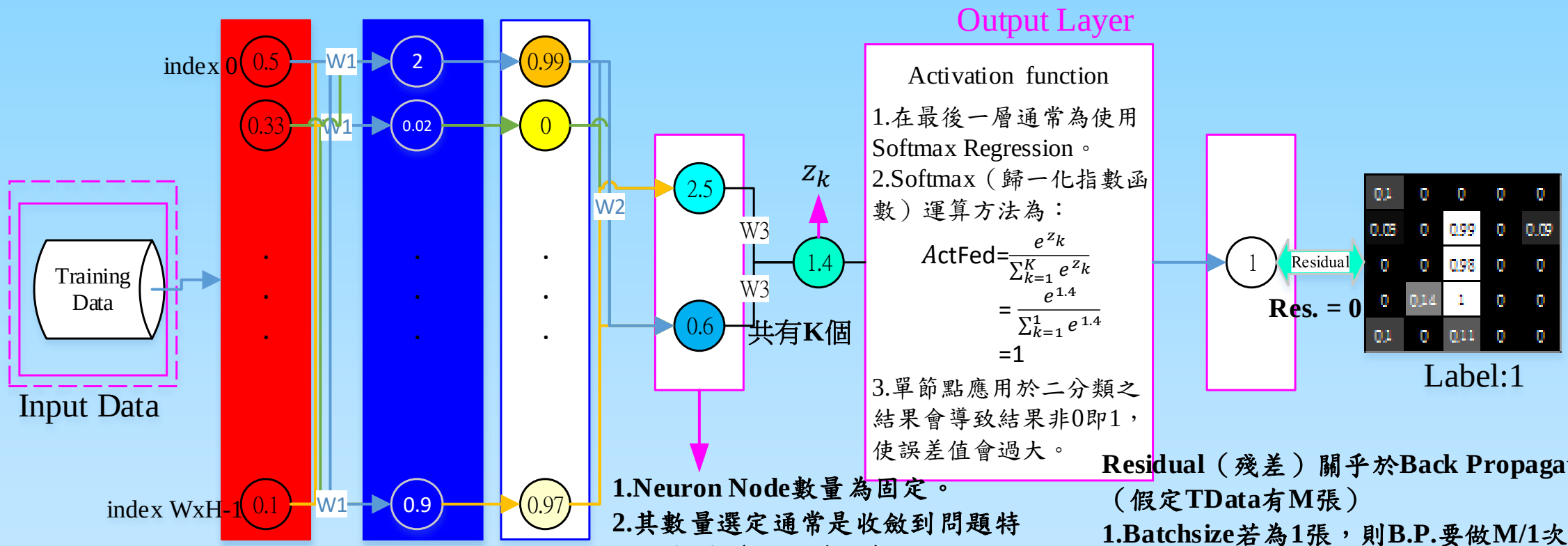
Training Stage- Input Layer

Training Stage- Hidden Layer



Training Stage- Hidden Layer

Training Stage- Output Layer



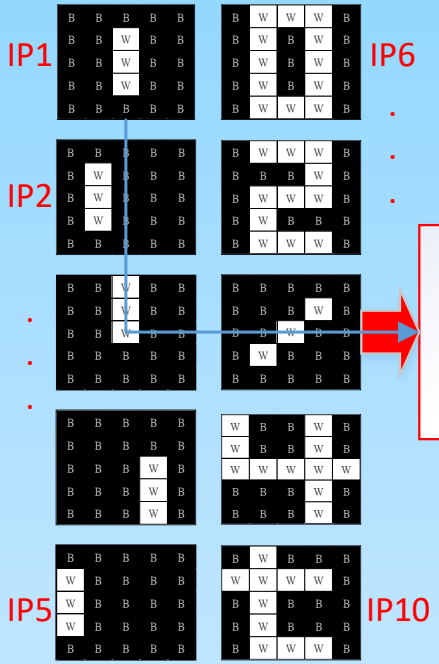
1.Neuron Node數量為固定。
2.其數量選定通常是收斂到問題特性之數量（二元/多元）分類。
3.假若H.L.是有10個Node，即代表此層有10個權重值，計算權重總和之值所代表該張輸入圖對於此O.L.的Node神經激發程度。

Residual（殘差）關乎於Back Propagation運算。（假定TData有M張）
1.Batchsize若為1張，則B.P.要做M/1次，會因為圖形Res.過大使得權重更新跳動過大，難以收斂。
2.Batchsize若為100張，則B.P.要做M/100次，可以採用區域性集合與有限記憶體進行運算，而權重會跟圈選子集合特性有關。
3.Batchsize若為M張，則B.P.要做M/M次，可以一次做完，更新一次權重，其可以考量所有圖像，但缺點為運算記憶體需較大，另運算時間過長。

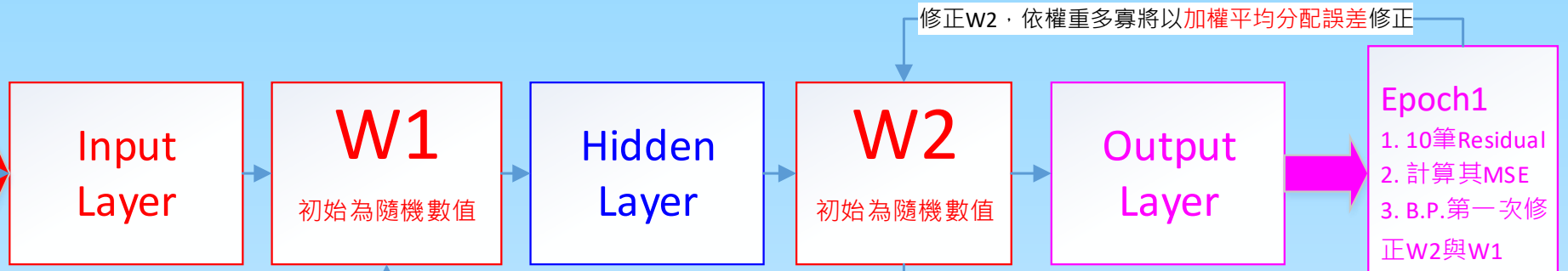
Training Stage- Output Layer

Training動態變化-以GTNum取全部Training Data一次全訓練

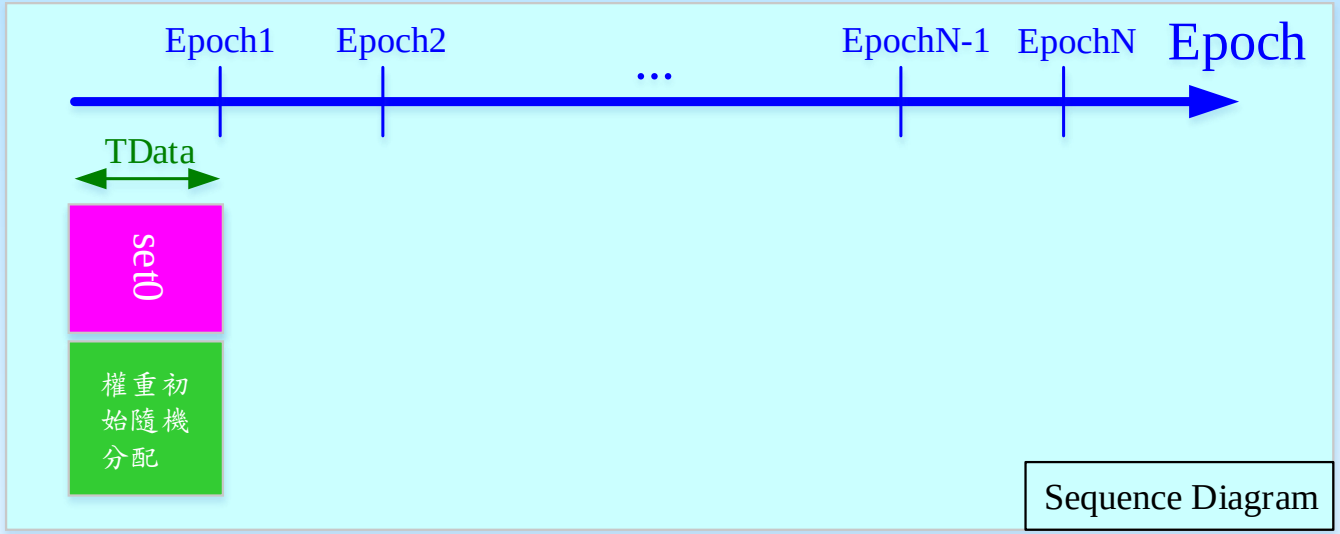
Training Stage- Epoch1



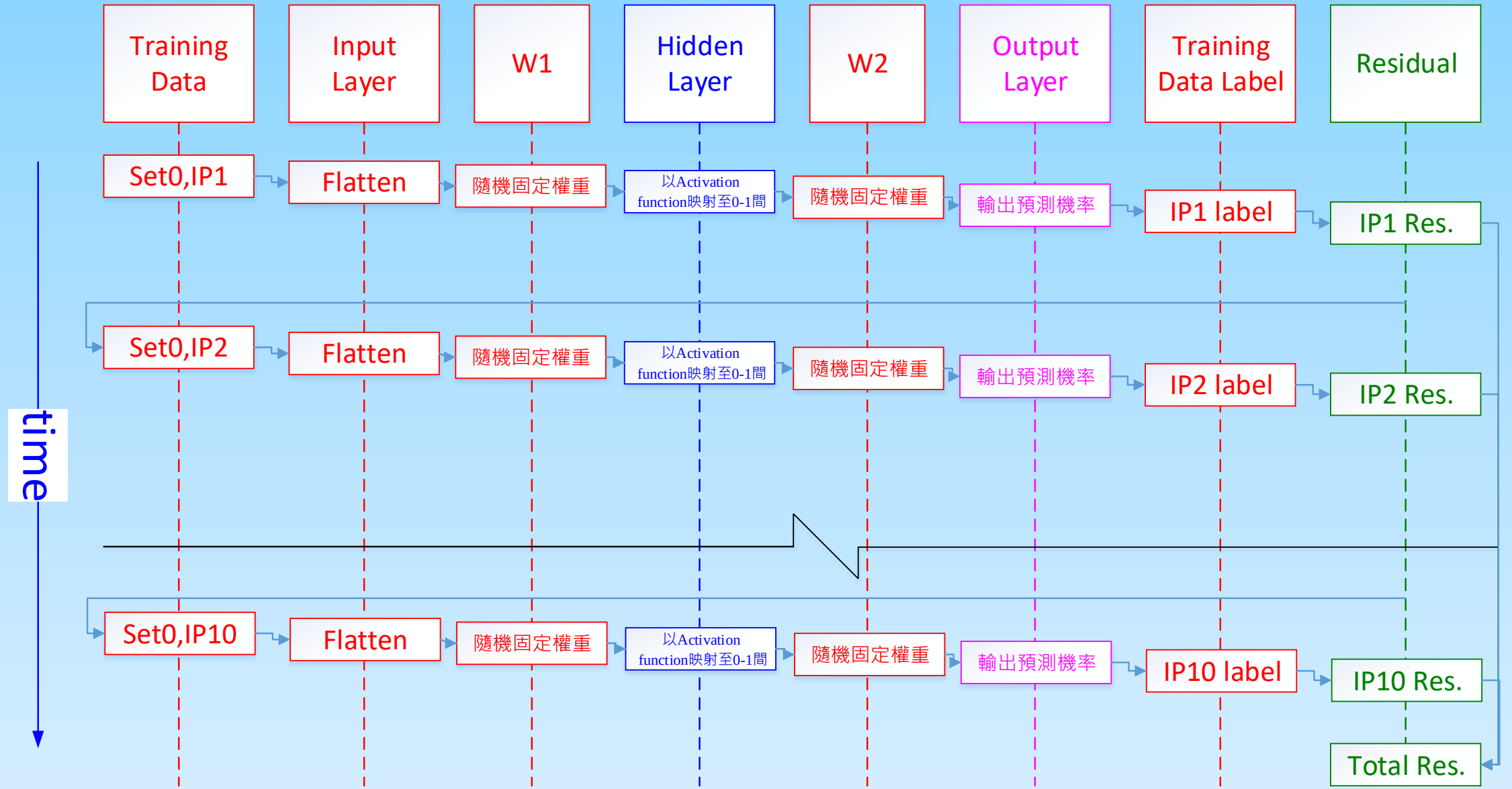
GTNum：10
Batchsize：10
num_set：10/10＝1
故set只有1組。
Iteration：1次



IP1 Res.=0.9
⋮
IP10 Res.=0.82

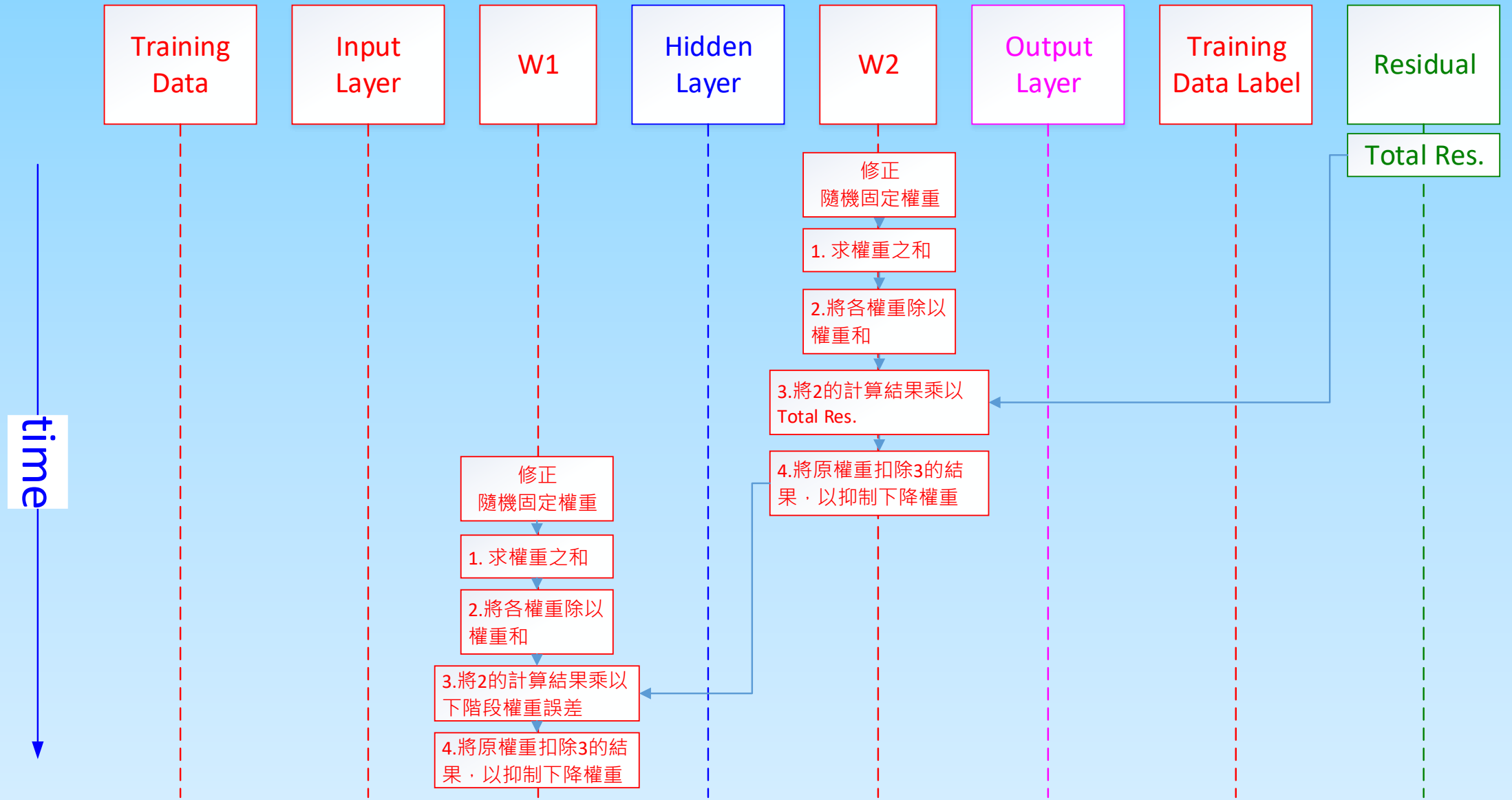


Epoch1 timing diagram



Training Stage- Epoch1 timing diagram

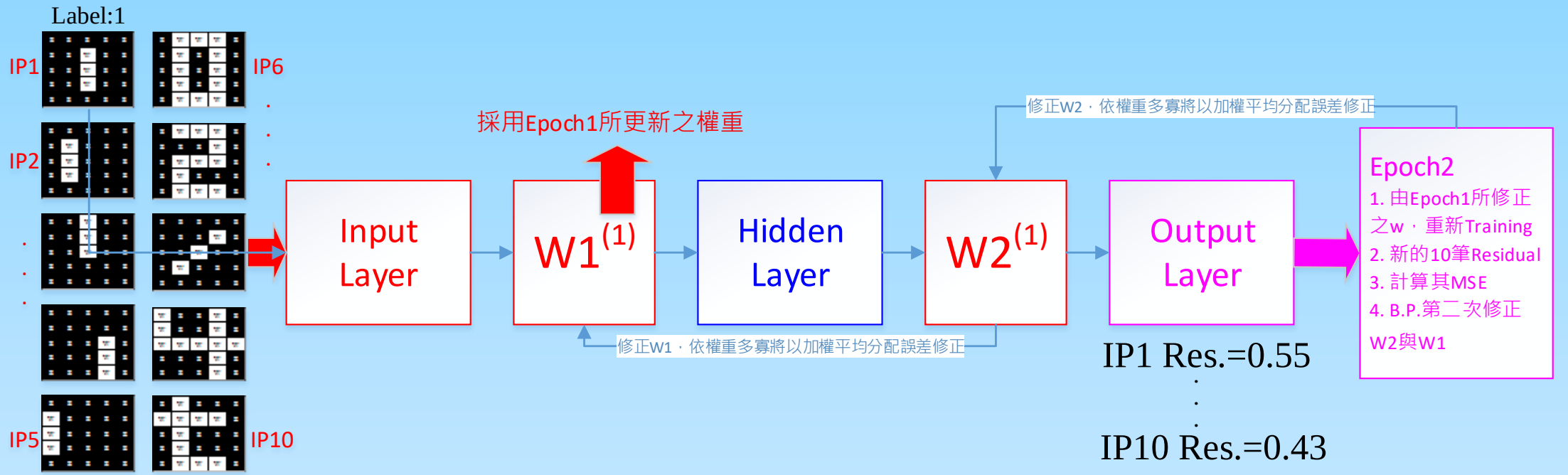
Epoch1 Back propagation timing diagram



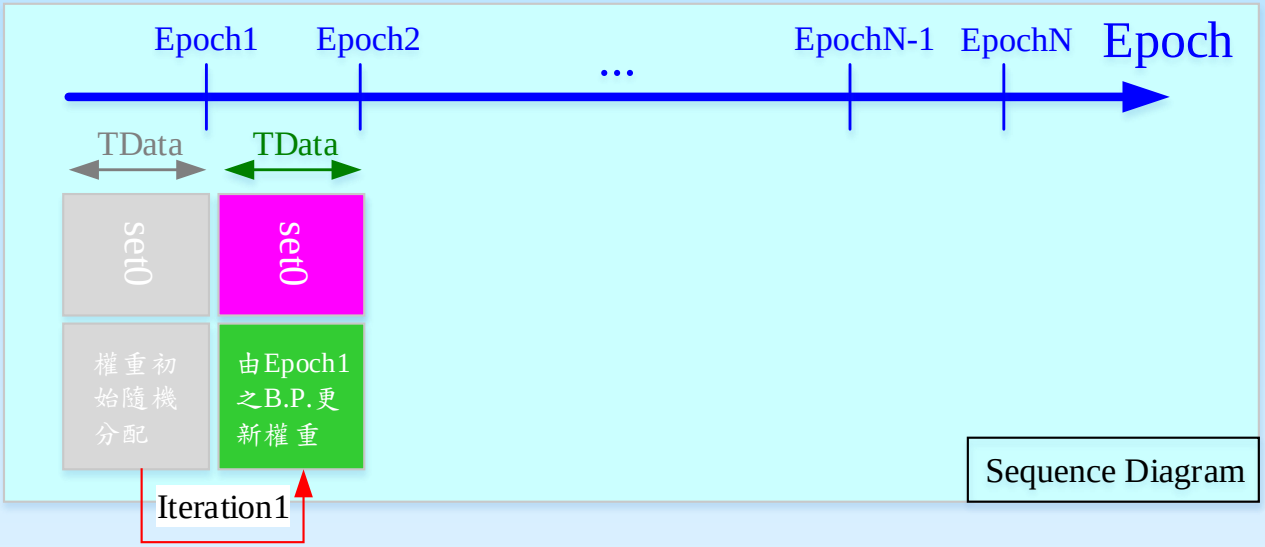
Training Stage- Epoch1 Back Propagation timing diagram

由Epoch1修正w，Train第2次

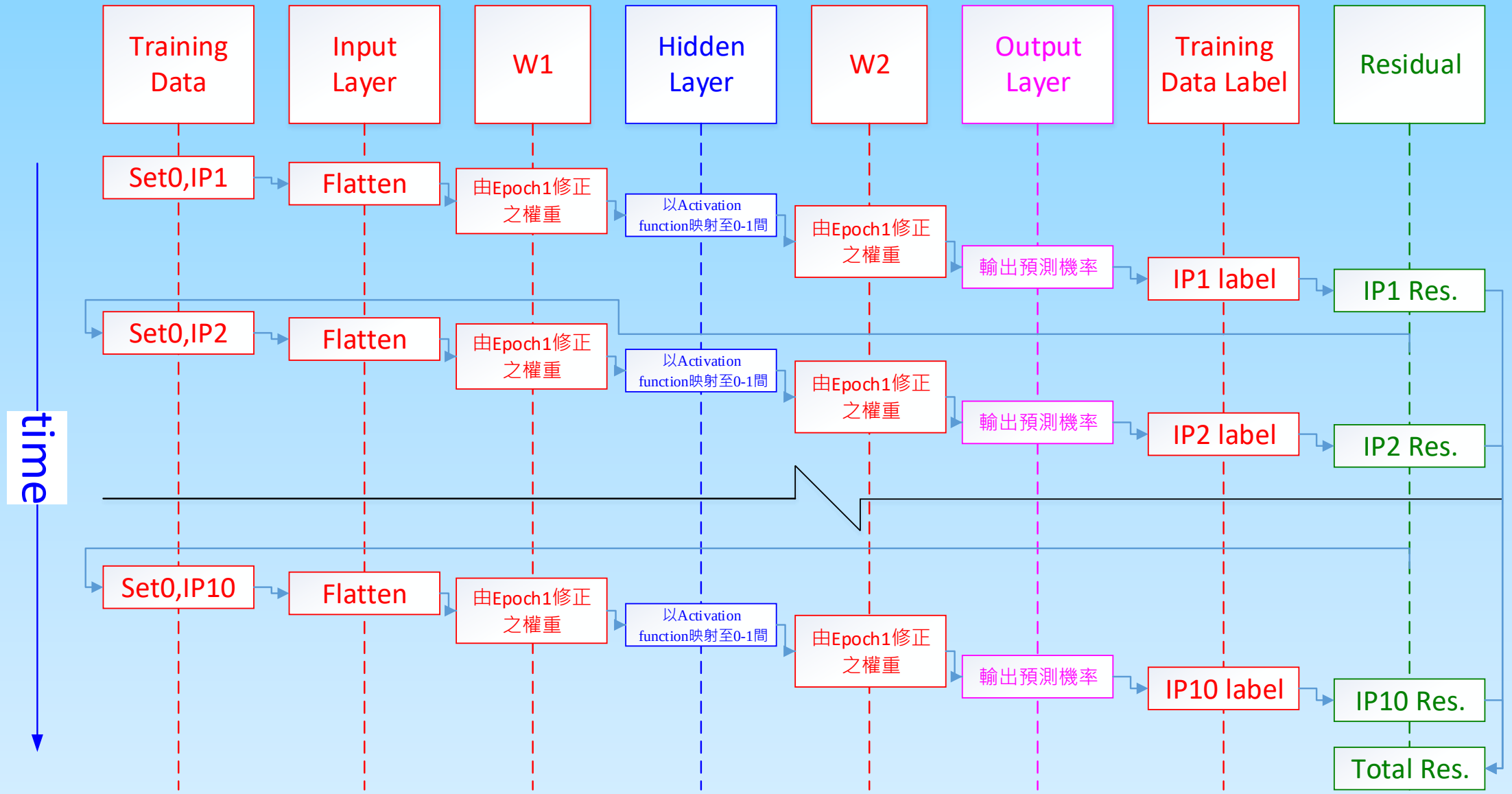
Training Stage- Epoch2



註：$Wp^{(q)}$，p為目前屬於第幾層（會伴隨Layers），q為更新權重次數。



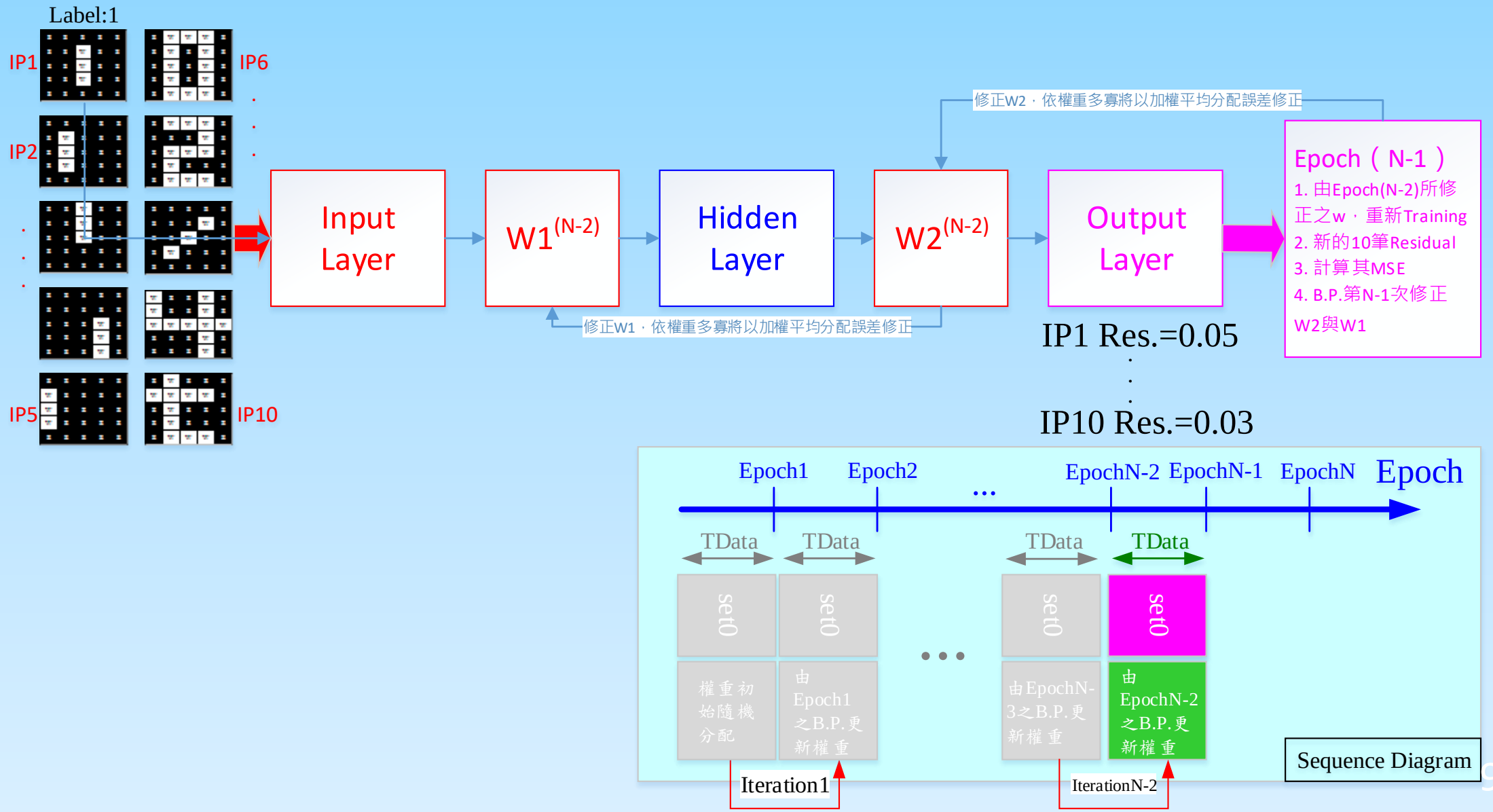
Epoch2 timing diagram



Training Stage- Epoch2 timing diagram

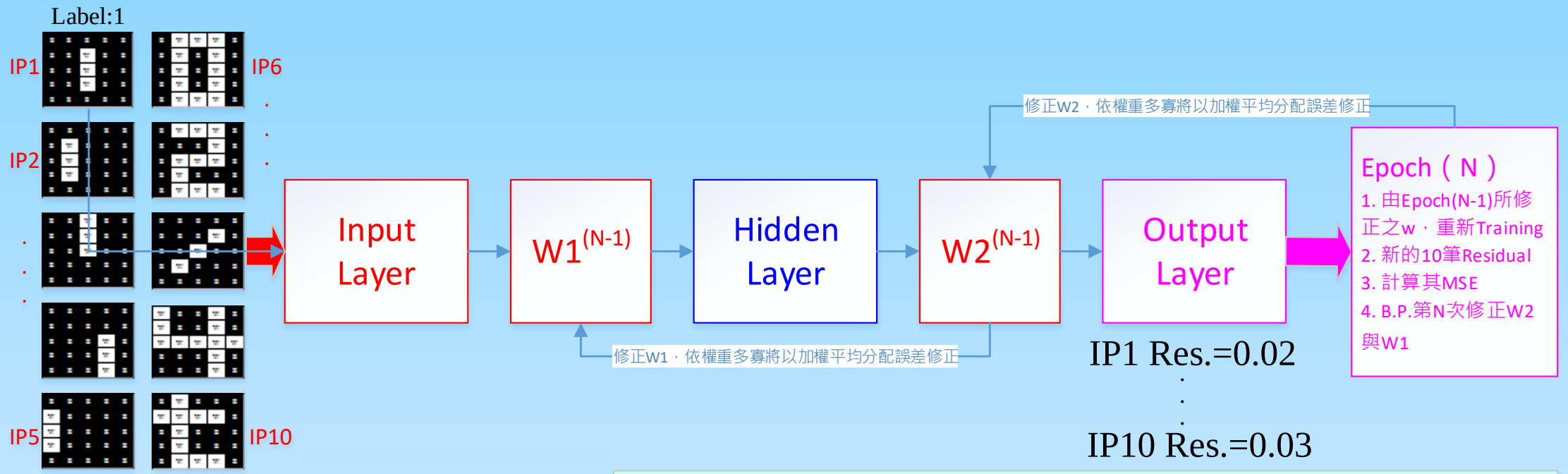
由Epoch（N-2）修正w，Train第N-1次

Training Stage- Epoch(N-1)

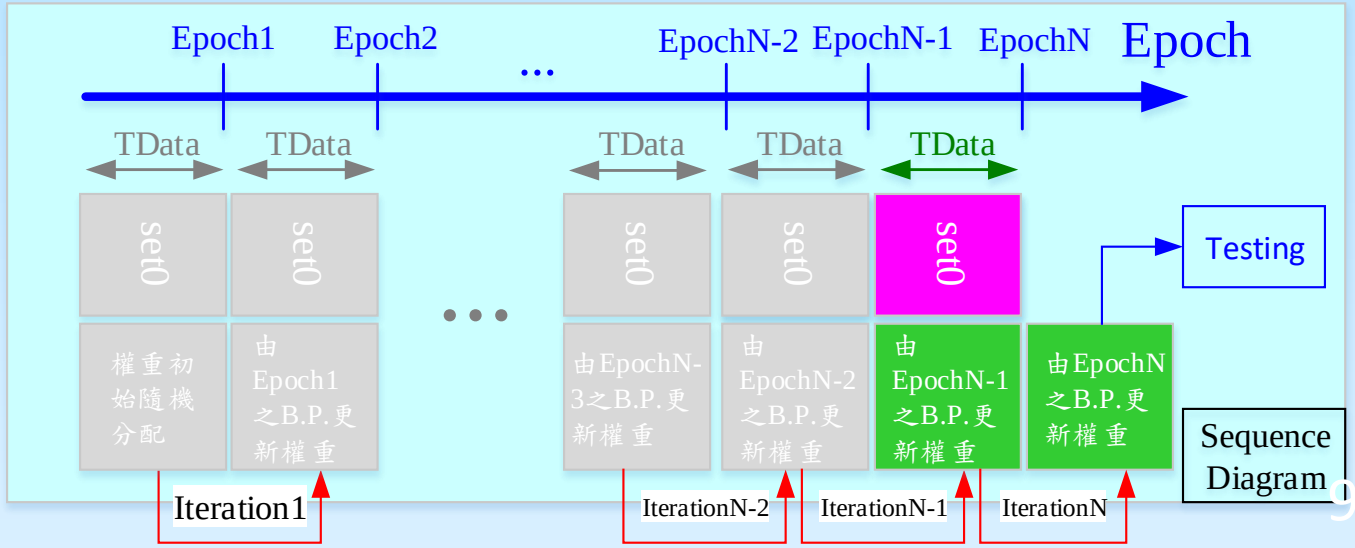


由Epoch（N-1）修正w，Train第N次，同時最後一次修正w

Training Stage- EpochN

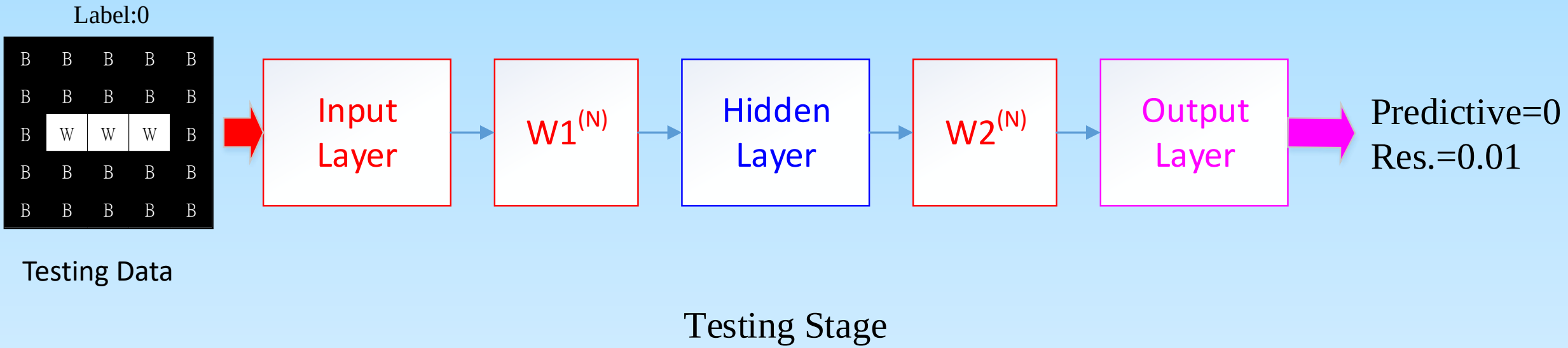


最後一次修正w，即為固定，修正完的權重值，也不會以IP1- IP10再次訓練，故沒有Res.。

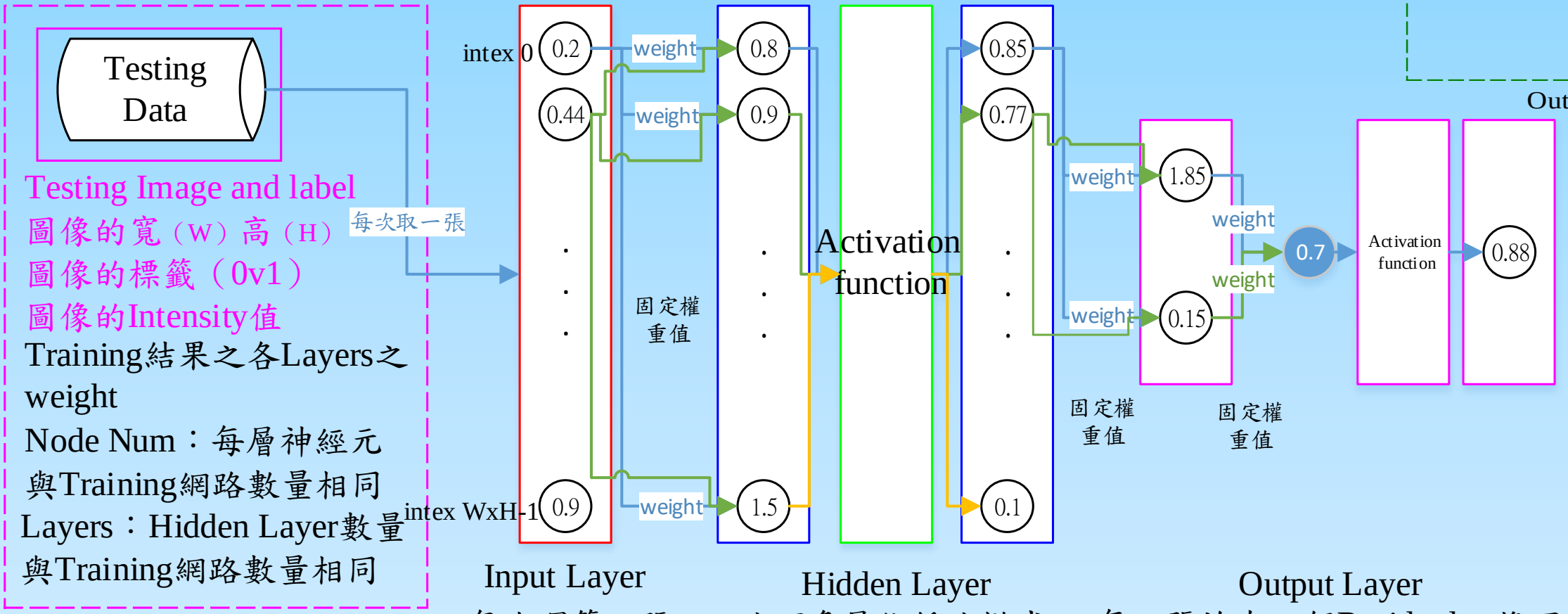


由訓練最後一次已經固定之權重值，進行測試

1. 目前W1（N）與W2（N）皆為定值。
2. Testing Data基本定義為不可是與Training Data重覆的。
3. 經過Testing Stage將能得Predictive value。



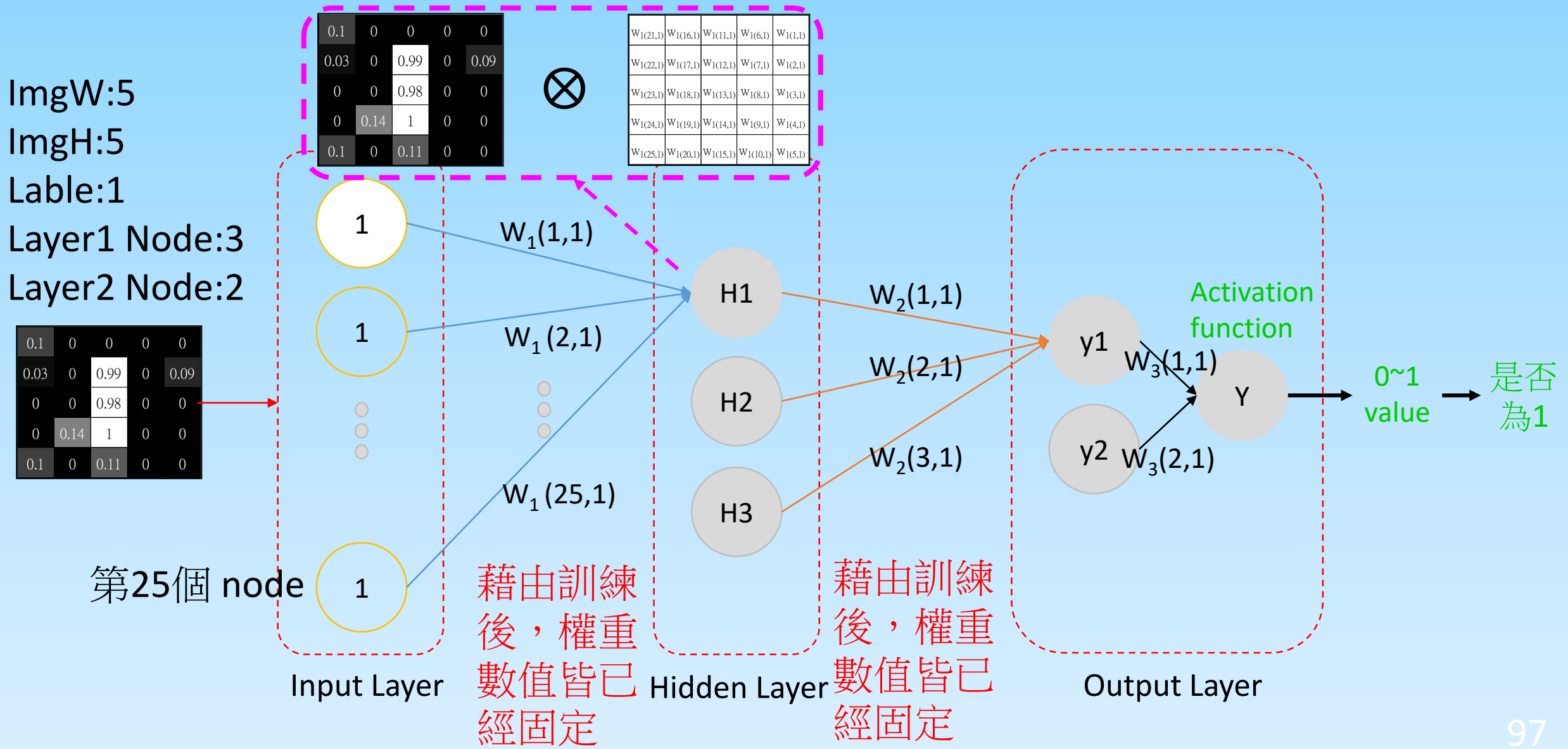
Testing Stage



獲得單一圖像判別分類結果
藉由Residual獲得測試準確率
藉由真實答對樣本數獲得測試準確率

Output Data

ANN(Artificial Neural Network) 測試階段

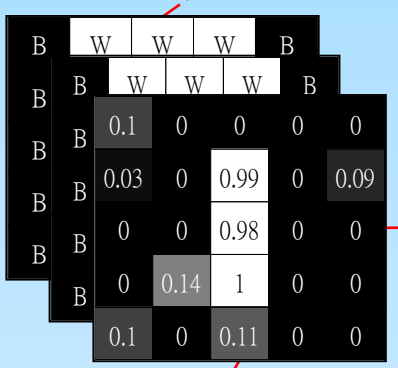


ANN(Artificial Neural Network) 訓練階段

ImgW:5,ImgH:5
Layer1 Node:3,Layer2 Node:2
LearnR=0.1,TrainD=1000
Batchsize=100

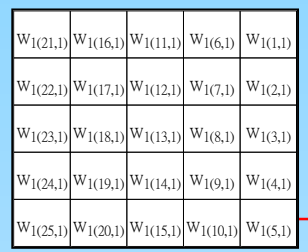
一個set輸入固定張數。如:100張

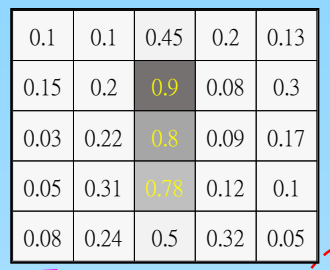
負樣本



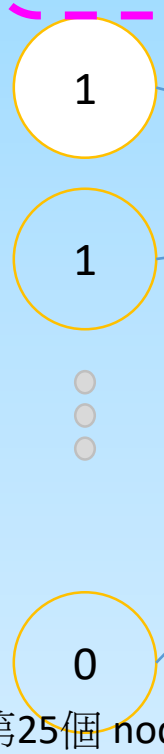
正樣本

Input Data





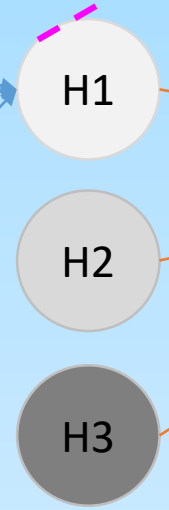
期待將神經元(neuron)訓練出具有顯著協助偵測特徵的權重



$W_1(1,1)$

$W_1(2,1)$

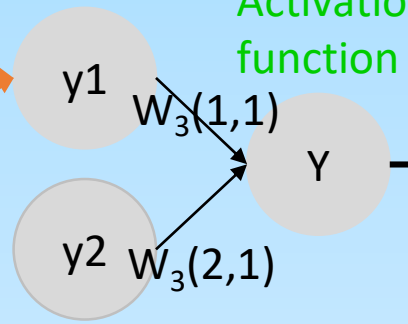
$W_1(25,1)$



$W_2(1,1)$

$W_2(2,1)$

$W_2(3,1)$



Activation function

GT label

正樣本為1
負樣本為0

0~1 value

每一張均可計算出與GT相比之誤差，而累積100張誤差，藉由該誤差值進行BP演算，修正權重值。

以原權重之佔比以總誤差將新權重做更新

BP(每一set修正一次)

Input Layer

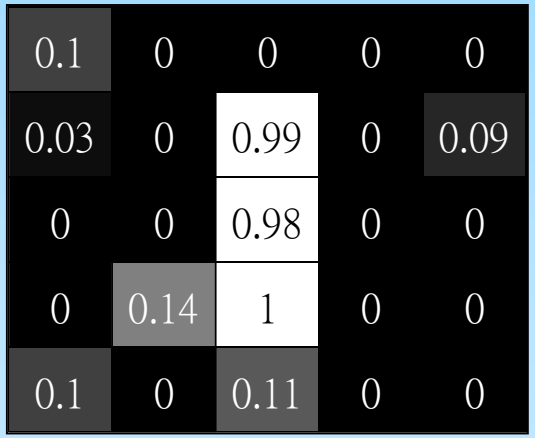
第一Epoch之第一個set將以亂數呈現權重值,依序將100張輸入

Hidden Layer

第一Epoch之第一個set將以亂數呈現權重值,依序將100張輸入

Output Layer

Hidden Layer GTT.bmp



0.1	0.1	0.45	0.2	0.13
0.15	0.2	0.9	0.08	0.3
0.03	0.22	0.8	0.09	0.17
0.05	0.31	0.78	0.12	0.1
0.08	0.24	0.5	0.32	0.05

→

0.01	0	0	0	0
0	0	0.9	0	0.03
0	0	0.8	0	0
0	0.04	0.78	0	0
0.01	0	0.06	0	0

→ Sum=2.6279

0.1	0.1	0.45	0.2	0.13
0.15	0.2	0.9	0.08	0.3
0.03	0.9	0.8	0.78	0.17
0.05	0.31	0.78	0.12	0.1
0.08	0.24	0.89	0.32	0.05

→

0.01	0	0	0	0
0	0	0.9	0	0.03
0	0	0.8	0	0
0	0.04	0.78	0	0
0.01	0	0.1	0	0

→ Sum=2.67

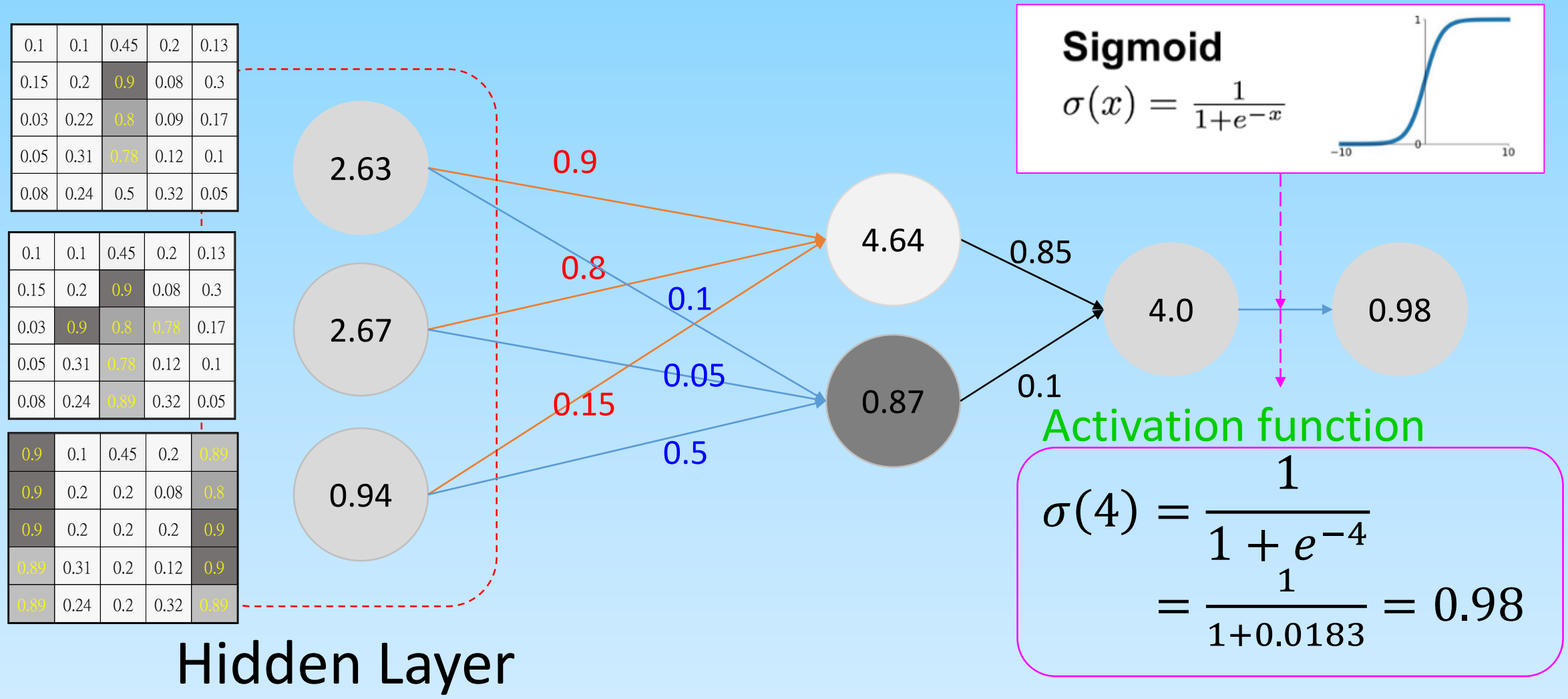
0.9	0.1	0.45	0.2	0.89
0.9	0.2	0.2	0.08	0.8
0.9	0.2	0.2	0.2	0.9
0.89	0.31	0.2	0.12	0.9
0.89	0.24	0.2	0.32	0.89

→

0.09	0	0	0	0
0.03	0	0.2	0	0.07
0	0	0.2	0	0
0	0.04	0.2	0	0
0.09	0	0.02	0	0

→ Sum=0.94

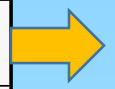
Output Layer- GTT.bmp



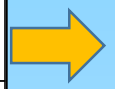
Hidden Layer GTF.bmp

0.1	0.99	0.91	0.99	0.1
0.03	0.96	0.09	0.96	0.03
0	0.99	0	0.99	0
0	0.91	0	0.88	0
0.1	0.99	0.91	0.99	0.1

0.1	0.1	0.45	0.2	0.13
0.15	0.2	0.9	0.08	0.3
0.03	0.22	0.8	0.09	0.17
0.05	0.31	0.78	0.12	0.1
0.08	0.24	0.5	0.32	0.05

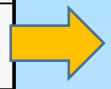


0	0.1	0.45	0.2	0.01
0.03	0.18	0.09	0.08	0.02
0	0.19	0.02	0.09	0
0.01	0.31	0.04	0.11	0.01
0	0.22	0.5	0.32	0



Sum=2.98

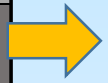
0.1	0.1	0.45	0.2	0.13
0.15	0.2	0.9	0.08	0.3
0.03	0.9	0.8	0.78	0.17
0.05	0.31	0.78	0.12	0.1
0.08	0.24	0.89	0.32	0.05



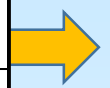
0	0.1	0.45	0.2	0.01
0.03	0.18	0.09	0.08	0.02
0	0.79	0.02	0.78	0
0.01	0.31	0.04	0.11	0.01
0	0.22	0.89	0.32	0

Sum=4.66

0.9	0.1	0.45	0.2	0.89
0.9	0.2	0.2	0.08	0.8
0.9	0.2	0.2	0.2	0.9
0.89	0.31	0.2	0.12	0.9
0.89	0.24	0.2	0.32	0.89

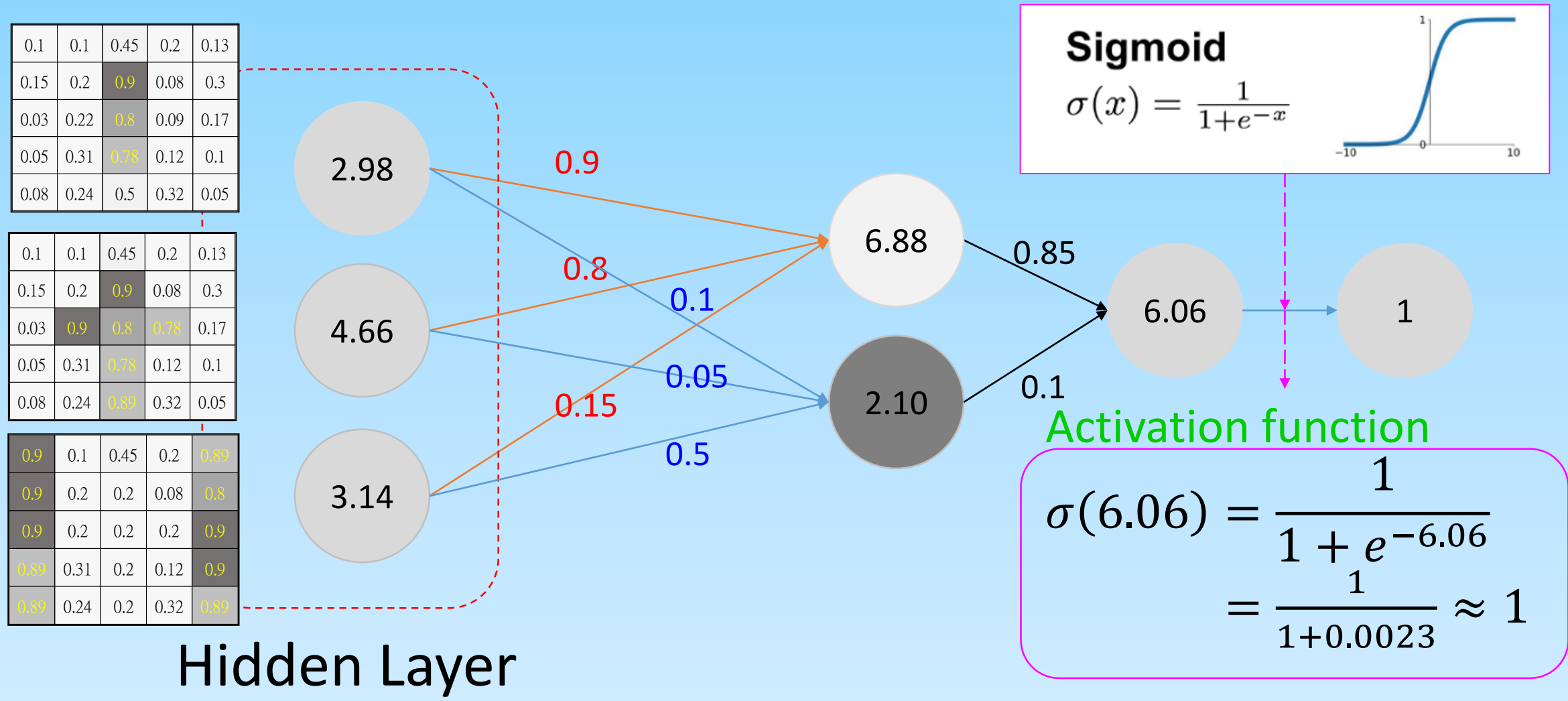


0	0.1	0.45	0.2	0.09
0.18	0.18	0.02	0.08	0.04
0	0.18	0.01	0.2	0
0.13	0.31	0.01	0.11	0.12
0	0.22	0.2	0.32	0

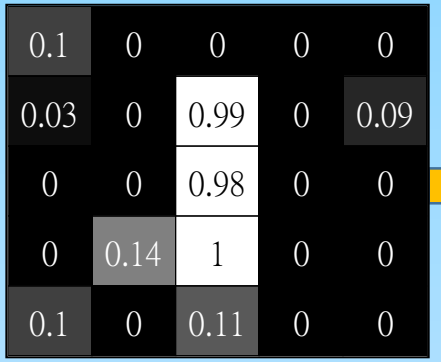


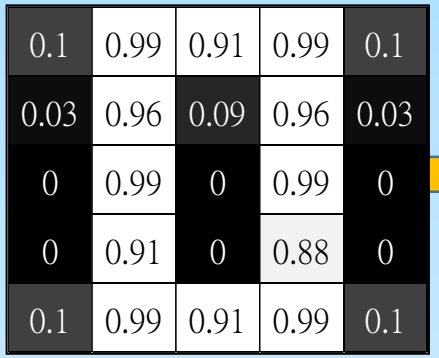
Sum=3.14

Output Layer- GTF.bmp



Loss function





0.98

1

$\hat{y}$

GT label
{正樣本為1
負樣本為0

y

	殘差 Residual	殘差平方 Residual^2
1	$1 - 0.98 = 0.02$	0.0004
0	$0 - 1 = -1.00$	1.0000

均方差 MSE

$$\frac{0.0004 + 1}{2} = 0.5002$$

需越小越好

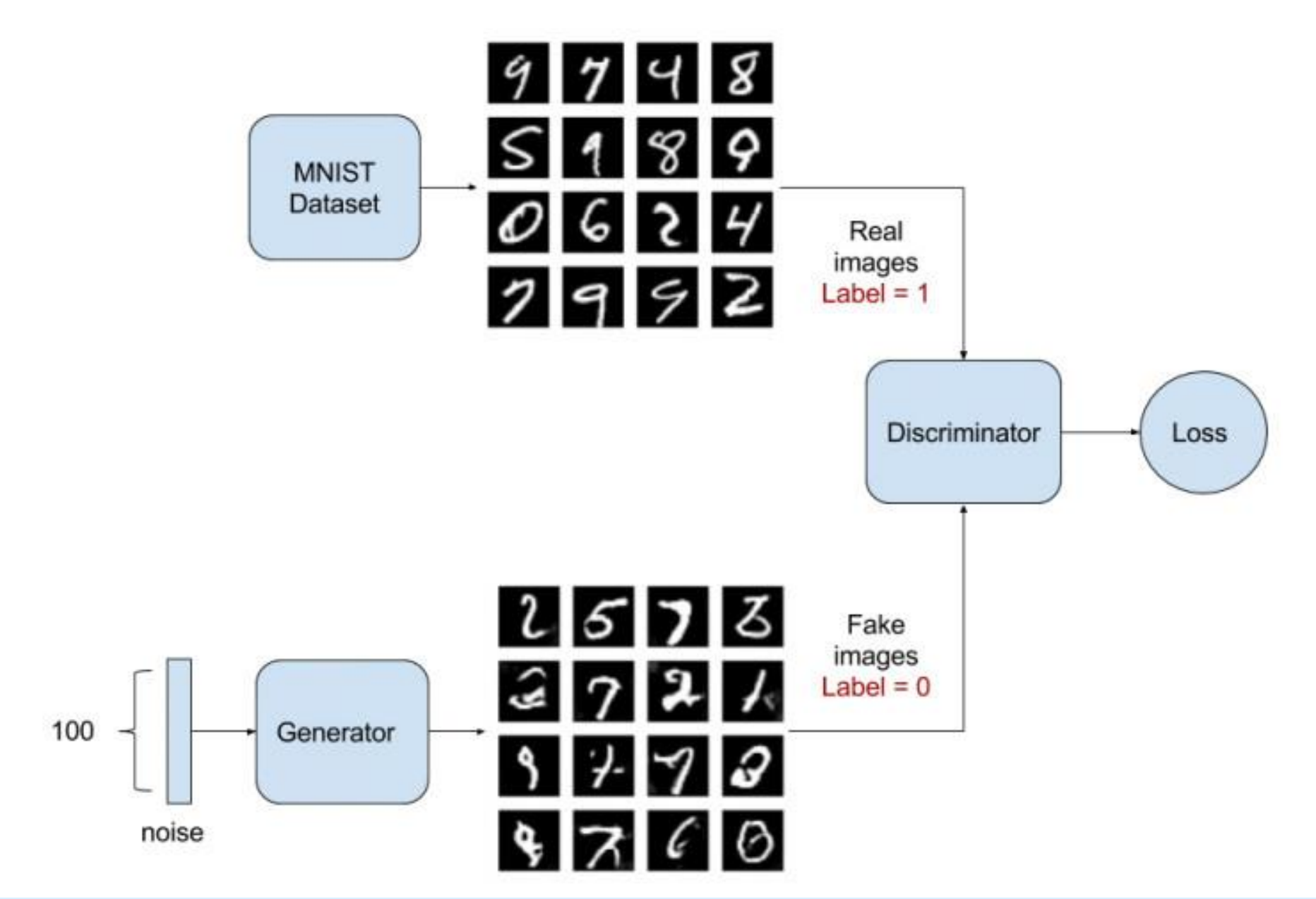
GAN
(Generative Adversarial Networks)

I. J. Goodfellow, J. Pouget-Abadie et al., “Generative adversarial nets,” In Proceedings of NIPS, pp. 2672–2680, 2014.

Generative Adversarial Networks

- 生成對抗網絡（Generative Adversarial Networks，GAN）最早由蒙特婁大學 Ian Goodfellow 在2014 年所提出。主要觀念為：同時訓練兩個相互合作、同時又相互競爭的深度神經網絡（一個稱為生成器Generator，另一個稱為判別器Discriminator）來處理無監督學習的相關問題。

Generative Adversarial Networks

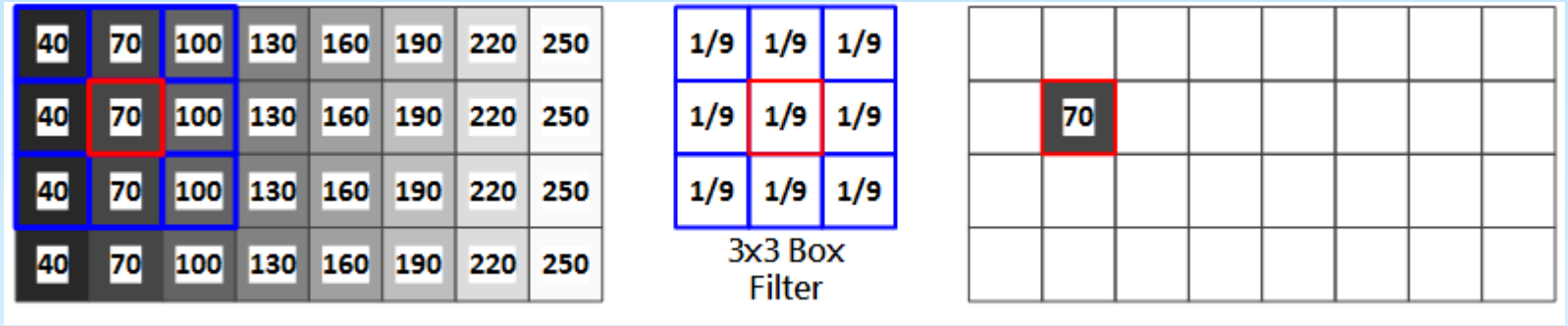


體驗手算的Convolution

II. 解題思考方向

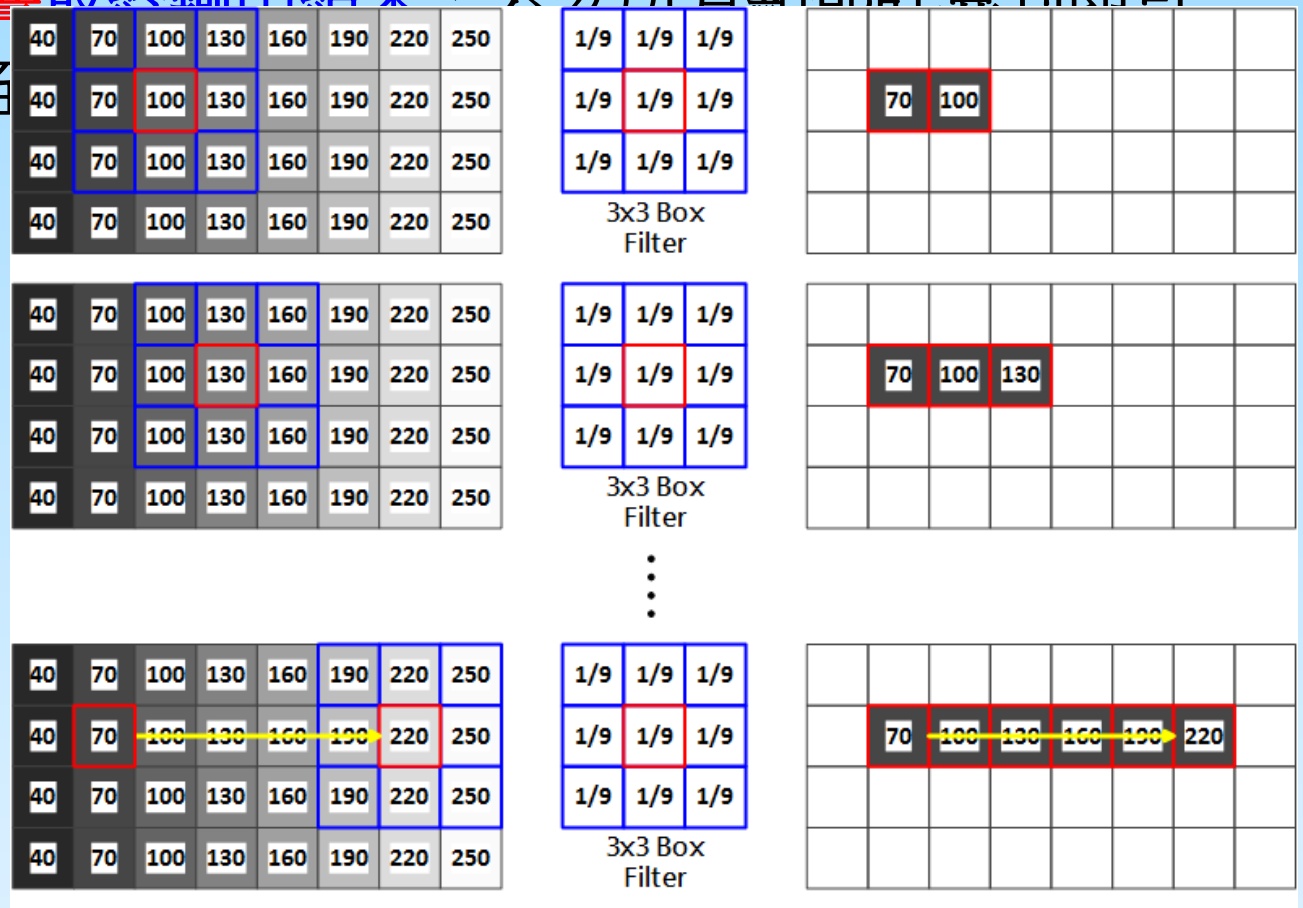
II-3. 數學

- 舉例Convolution計算原理與步驟：
- 以8*4的灰階影像，3*3的Box Filter，當作範例說明，以Mask覆蓋，計算結果。接著移動到下一個計算。如範例：
  - 第一步計算下圖顏色較深的點，其算法如下：
    - $40 \times {}^{1}/_{9} + 70 \times {}^{1}/_{9} + 100 \times {}^{1}/_{9} +$
    - $40 \times {}^{1}/_{9} + 70 \times {}^{1}/_{9} + 100 \times {}^{1}/_{9} +$
    - $40 \times {}^{1}/_{9} + 70 \times {}^{1}/_{9} + 100 \times {}^{1}/_{9} = 70$
- 算法：
  - 將遮罩所對應到影像的位置與遮罩權重相乘，並加總起來，最後將計算結果填進相對的輸出位置。



II. 解題思考方向
II-3. 數學

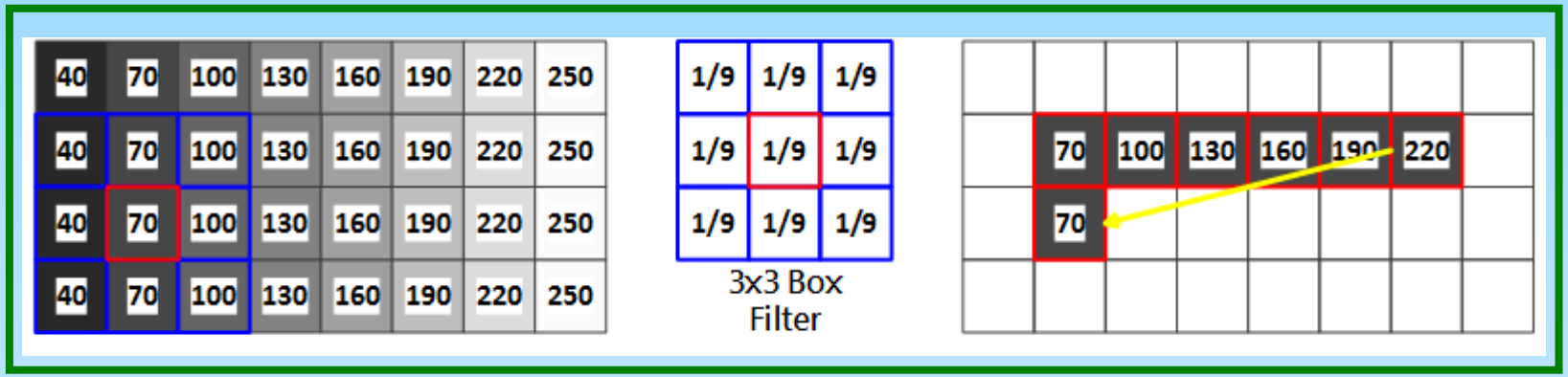
- 依序計算完整張影像就完成了一次Convolution計算。
- 計算順序不影響最終輸出結果，只要所有點都計算到即可。
- 如下圖：由左至

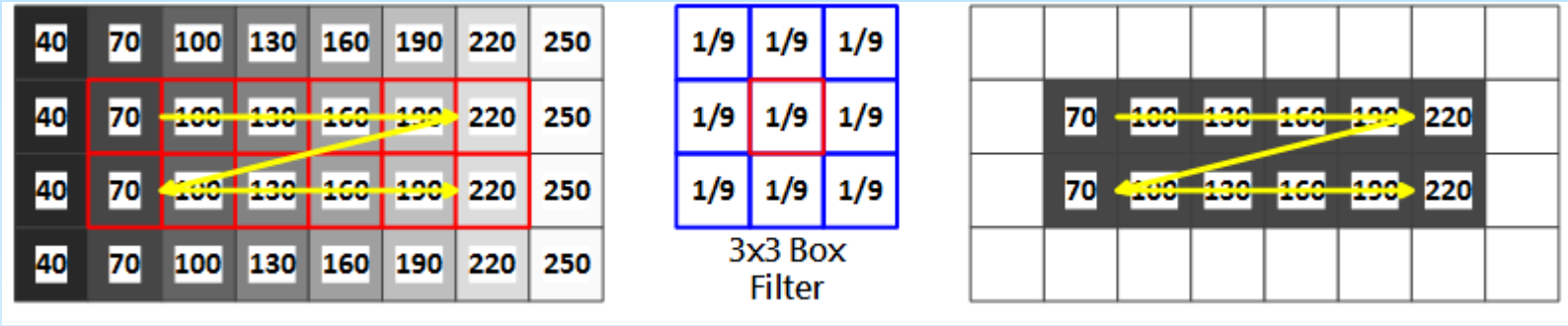


II. 解題思考方向

II-3. 數學

- 如下圖：由左至右計算完之後，換行，再次進行由左至右的計算，直到所有的點都計算完。

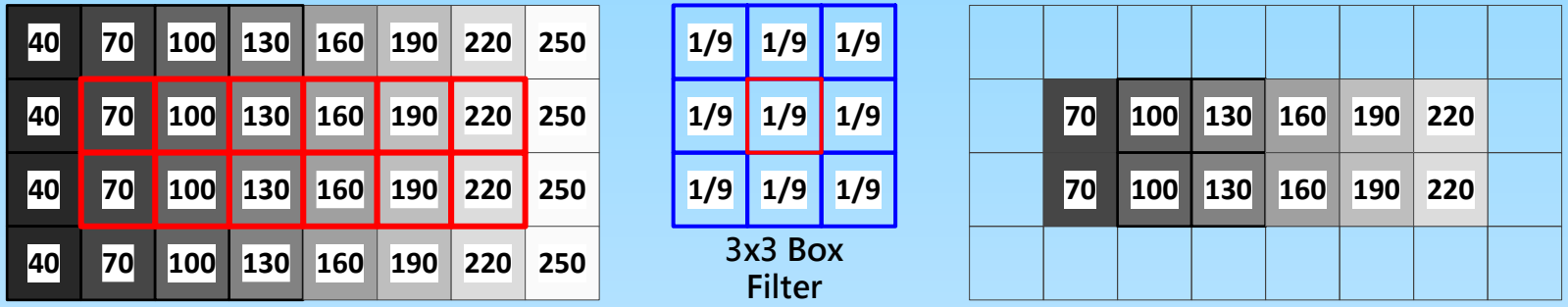




II. 解題思考方向

II-3. 數學

- 整張影像完成Convolution的結果：



- 由圖可看出影像週圍是無法計算的，如3*3大小的遮罩，影像的上下各有一列，左右各有一行無法計算，為了解決這個問題衍生出了三種方法：
  1) **Convolution_Shrink**
  2) **Convolution_ZeroPadding**
  3) **Convolution_AveragePadding**

課堂練習

10	10	10	1	1
2	2	2	2	1
10	10	10	10	1
1	1	1	1	1

Image1

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

Kernel1

Output1

Challenge 1

課堂練習

10	10	10	1	1
2	2	2	2	1
10	10	10	10	1
1	1	1	1	1

Image2

-1	0	1
-2	0	2
-1	0	1

Kernel2

Output2

Challenge 2

課堂練習

10	10	10	1	1
2	2	2	2	1
10	10	10	10	1
1	1	1	1	1

Image3

-1	-2	-1
0	0	0
1	2	1

Kernel3

Output3

Challenge 3

課堂練習

10	10	10	1	1
2	2	2	2	1
10	10	10	10	1
1	1	1	1	1

Image4

0	-1	0
-1	5	-1
0	-1	0

Kernel4

Output4

Challenge 4

體驗程式化的Convolution
要下載程式

先安裝套件

```
import os
os.system('pip install Pillow matplotlib numpy scipy')
```

裝完註解掉

```
1 #import os
2 #os.system('pip install Pillow matplotlib numpy scipy')
```

Convolution Kernel

```
# Step 2: 定義卷積核心（例如銳化）
kernel = np.array([
        [0, -1, 0],
        [-1, 5, -1],
        [0, -1, 0]
])
```

Convolution Kernels

類型	說明	Kernel 數值
模糊（平均）	平均平滑影像	np.ones((3, 3)) / 9
高斯模糊	權重模糊（中央重）	1/16 * np.array([[1, 2, 1], [2, 4, 2], [1, 2, 1]])
銳化	強化邊緣與細節	np.array([[0, -1, 0], [-1, 5, -1], [0, -1, 0]])
邊緣檢測	Laplacian 樣式	np.array([[-1, -1, -1], [-1, 8, -1], [-1, -1, -1]])
Sobel X	偵測垂直邊界	np.array([[-1, 0, 1], [-2, 0, 2], [-1, 0, 1]])
Sobel Y	偵測水平邊界	np.array([[-1, -2, -1], [0, 0, 0], [1, 2, 1]])
emboss 浮雕	做出浮雕效果	np.array([[-2, -1, 0], [-1, 1, 1], [0, 1, 2]])
自訂亮度強化	強調中心亮度，抑制周邊	np.array([[0, -1, 0], [-1, 10, -1], [0, -1, 0]])

智慧這件事-回到過去

Can Machine Think？

- 1950年，圖靈〈**運算機器與智慧**（Computing Machinery and Intelligence）〉「圖靈測試」：如果機器與人類進行非面對面（例如在中間以布幕隔離）對話（例如使用文字訊息），**人類卻無法辨認出對方是機器，那麼這台機器就具有智慧。**
- 人工智慧一詞直到1956年，才在美國新罕布夏州一場為期兩個月的研究工作坊「達特茅斯暑期人工智慧研究計畫（The Dartmouth Summer Research Project on Artificial Intelligence）」上，由負責組織會議的電腦高階語言LISP之父約翰・麥卡錫（John McCarthy）正式定名。這場工作坊所討論的問題：「**計算機、自然語言處理、神經網絡、計算理論、抽象化與隨機創造**」後來都成為人工智慧研究發展的重要領域，而達特茅斯會議也因此成為人工智慧領域的經典起源。

1951

- 1951年，科學家馬文·閔斯基（Marvin Minsky）第一次嘗試建造了世上第一個神經元模擬器：Snarc（Stochastic Neural Analog Reinforcement Calculator），它能夠在其40個「代理人」和一個獎勵系統的幫助下穿越迷宮。六年後，康乃爾航空工程實驗室的法蘭克·羅森布拉特（Frank Rosenblatt）設計、發表神經網絡的感知器（Perceptron）實作後，**人工神經網絡（或稱類神經網絡）**學者曾經一度振奮，認為這個突破終將帶領人工智慧邁向新的發展階段。

1970

- 人工智慧領域的研究在1970年代因為**缺乏大規模數據資料、計算複雜度無法提升**，無法把小範圍的問題成功拓展為大範圍問題，導致計算機領域無法取得更多科學研究預算的投入而沉寂。

1980

- 1980年代，科學家首先透過思考上的突破，設計出新的演算方法來**模擬人類神經元**，迎來神經網絡發展的文藝復興時期。物理學家約翰．霍普費爾德（John Hopfield）在1982率先發表Hopfield神經網絡，開啟了神經網絡可以**遞迴設計**的思考。四年後，加州大學聖地牙哥分校教授大衛．魯梅爾哈特（David Rumelhart）提出了反向傳播法（Back Propagation），透過每次資料輸入（刺激）的變化，**計算出需要修正的權重回饋給原有函數，進一步刷新了機器「學習」的意義。**科學家更進一步把神經元延伸成為神經網，透過多層次的神經元締結而成的**人工神經網絡**，在**函數表現上可以保有更多「被刺激」的「記憶」**。

現代

- 目前**多層次的人工神經網絡模型**，主要包含**輸入層（input layer）**、**隱層（hidden layer）與輸出層（output layer）**，另外根據資料輸入的流動方向，又分為單向流動或可以往回更新前一層權值的反向傳播法。由於神經網絡模型非常仰賴計算規模能力，為了增加高度抽象資料層次的彈性，將其複合為更複雜、多層結構的模型，並佐以多重的**非線性轉換**，將其稱之為深度學習（Deep Learning）。

大數據的機器學習

- 1970年代，人工智慧學者從前一時期的研究發展，開始思辯在機器上顯現出人工智慧時，是否一定要讓機器真正具有思考能力？因此，人工智慧有了另一種劃分法：**弱人工智慧（Weak AI）與強人工智慧（Strong AI）**。**弱人工智慧**意指如果一台機器具有博聞、強記（可以快速掃描、儲存大量資料）與**分辨**的能力，它就具有表現出人工智慧的能力。**強人工智慧**則是希望建構出的系統架構可媲美人類，**可以思考並做出適當反應，**真正具有人工智慧。

監督式學習

- 機器學習（Machine Learning）可以視為**弱人工智慧的代表**，只要定義出問題，**蒐集了適當的資料（資料中通常需要包含原始數據與標準答案，例如人像圖片與該圖片內人像的性別、年齡）**，再將資料分做兩堆：**訓練用與驗證用**，以訓練用資料進行學習，透過特定的**分類演算法抽取特徵值**，建構出資料的**數學模型**，以該數學模型**輸入驗證用資料**，**比對演算的分類結果**是否與真實答案一樣，如果該數學模型能夠**達到一定比例的答對率**，則我們認為這個機器學習模型是有效的。這種具有標準答案，並以計算出的預期結果進行驗證的機器學習，通常被稱為**監督式學習**。

監督式學習

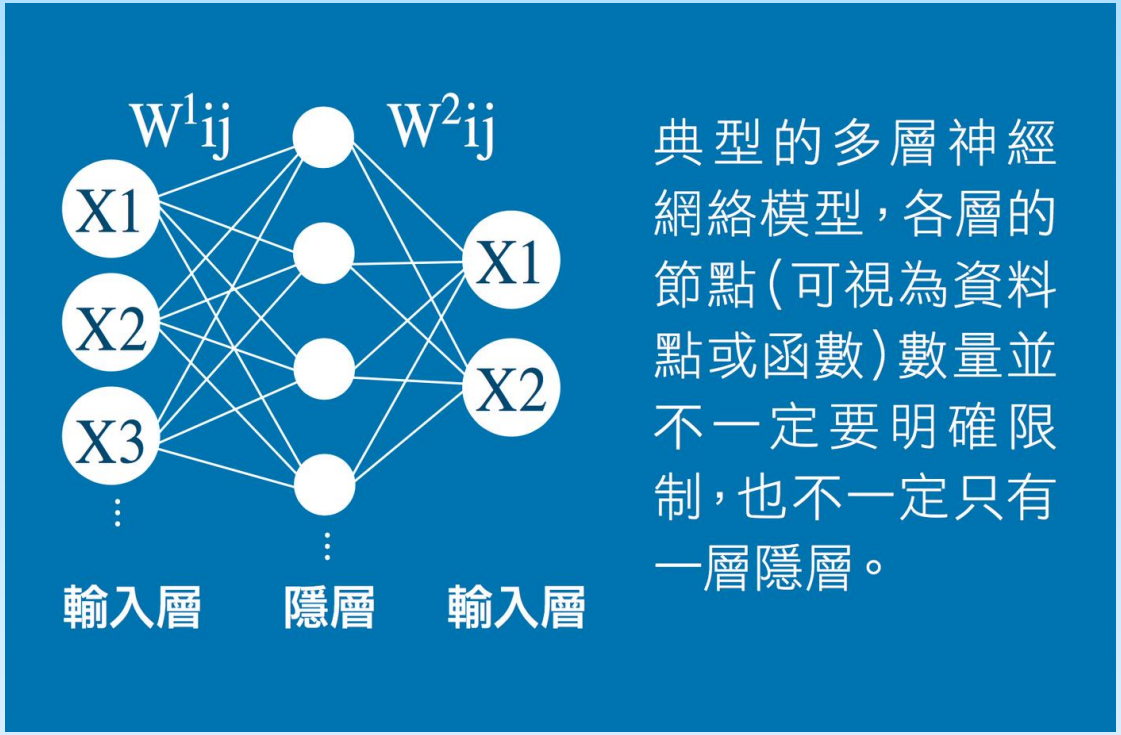
- 分類演算法抽取特徵值
- 建立數學模型
- 輸入驗證用資料
- 比對演算的分類結果
- 達到一定比例的答對率

非監督式學習

- 非監督式學習則**強調不知道資料該如何分類的機器學習**，換句話說，我們提供電腦大量資料，但不告訴它（或許我們也真的不知道）這些資料該用什麼方式進行分類，然後電腦透過演算法將資料分類，**人類只針對最終資料分類進行判別**，在數據尋找規律就是機器學習的基礎。

深度學習Deep Learning

- 深度學習是**機器學習**的一種分支，也是目前機器學習發展方向的主流。其概念主要是複合多層複雜結構的人工神經網絡，並將**其中函數作多重非線性轉換**，使之增加**高度抽象化**資料、**記憶資料影響能力**。



深度學習Deep Learning

- 深度學習是機器學習(Machine learning)的一個分支，希望把資料透過**多個處理層(layer)**中的**線性或非線性轉換(linear or non-linear transform)**，自動抽取出足以**代表資料特性的特徵(feature)**。
- 在基礎的機器學習中，特徵通常是透過由**人力撰寫的演算法**產生出來的，需要經過各領域的專家對資料進行許多的分析及研究，了解資料的特性後，才能產生出有用、效果良好的特徵。這樣的過程就是**特徵工程(Feature engineering)**。

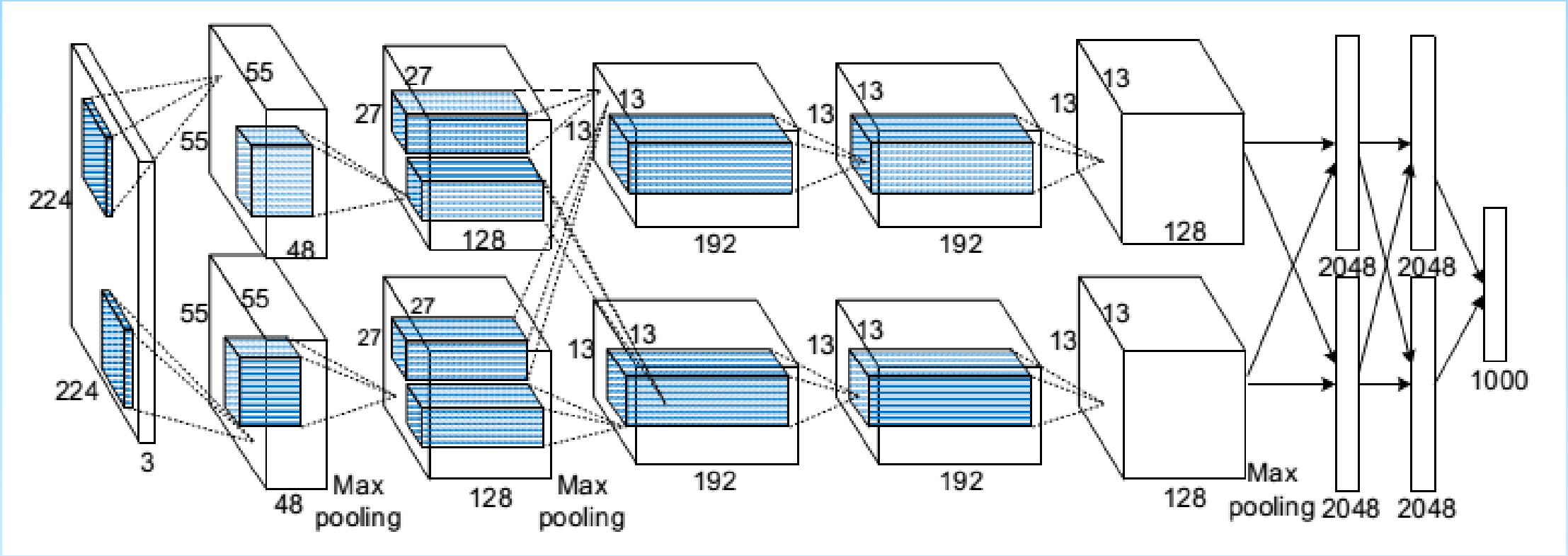
深度學習Deep Learning

- 深度學習具有**自動抽取特徵(feature extraction)**的能力，也被視為是一種特徵學習(Feature Learning, representation learning)，可以**取代專家的特徵工程所花費的時間**。
- 深度學習的訓練(Training)可以分為**三個步驟**：定義網路架構(define network structure)、定義學習目標(define learning target)、最後才是透過數值方法(Numerical method)進行訓練。

深度學習Deep Learning

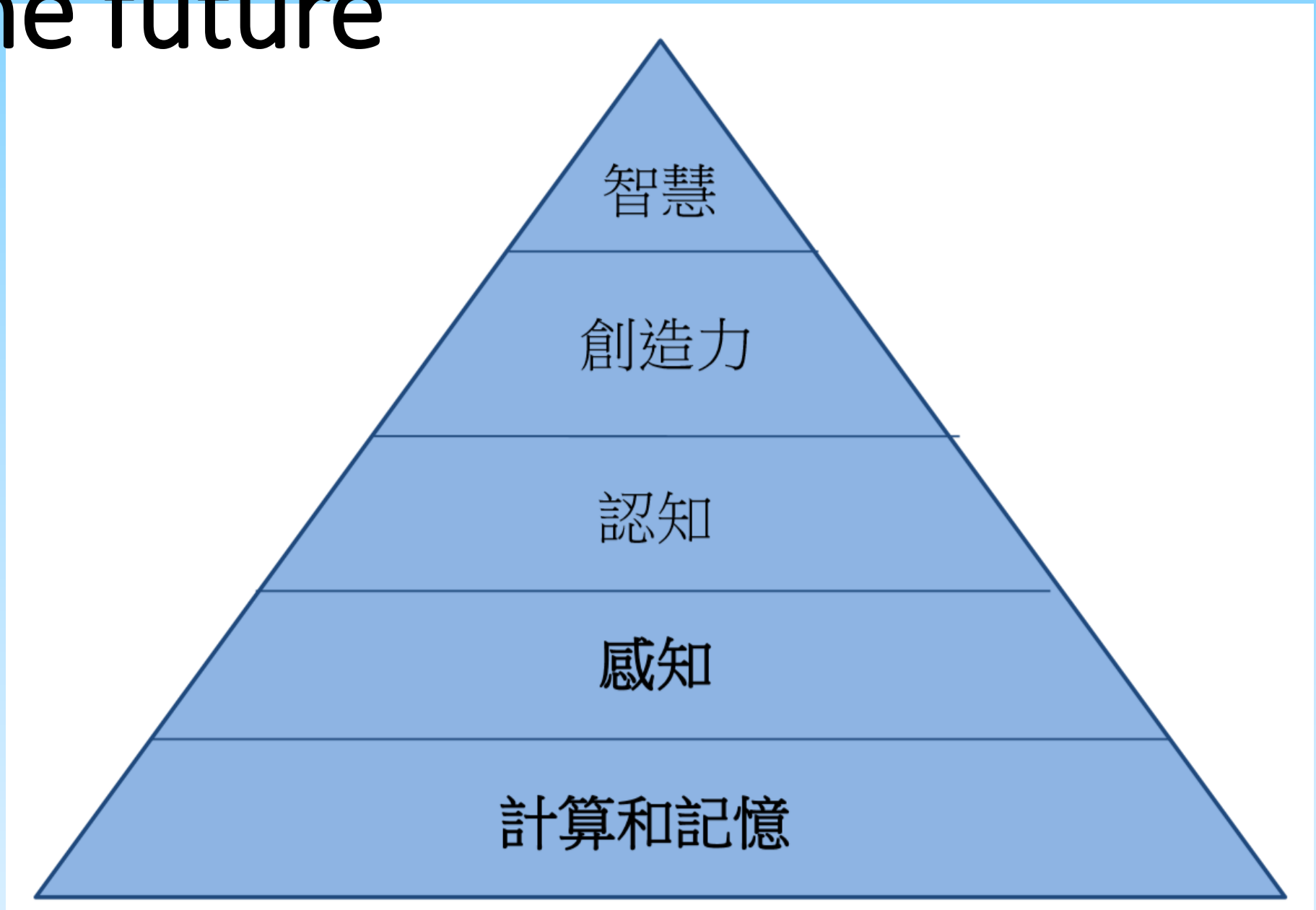
- **卷積神經網路(CNN)是最常見的深度學習網路**架構之一，因為網路架構中的卷積層(Convolutional layer)及池化層(Pooling layer)強化了模式辨識(Pattern recognition)及相鄰資料間的關係，使卷積神經網路應用在影像、聲音等訊號類型的資料型態能得到很好的效果。

深度學習Deep Learning



CNN model of AlexNet

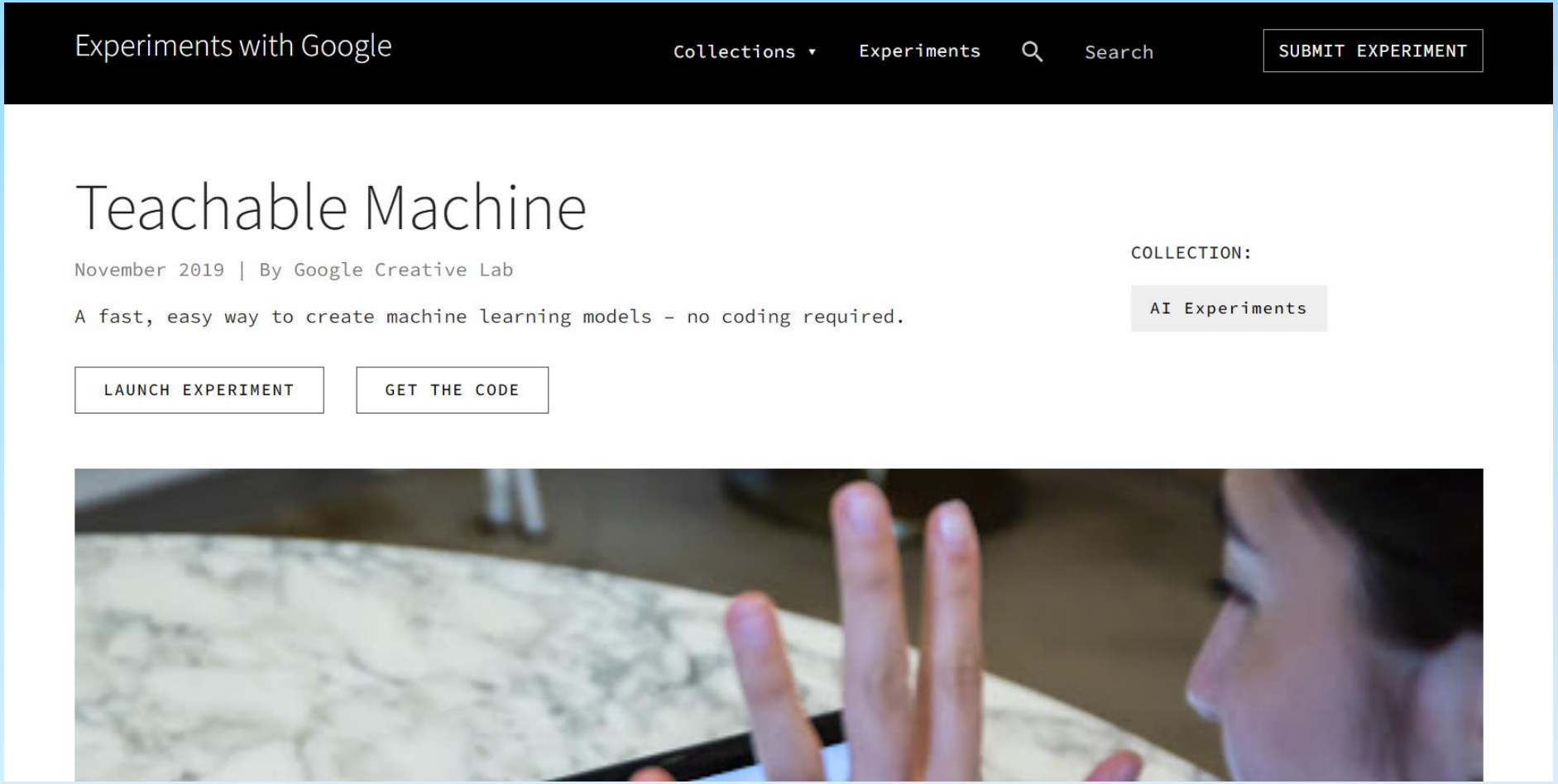
In the future



體驗AI辨識

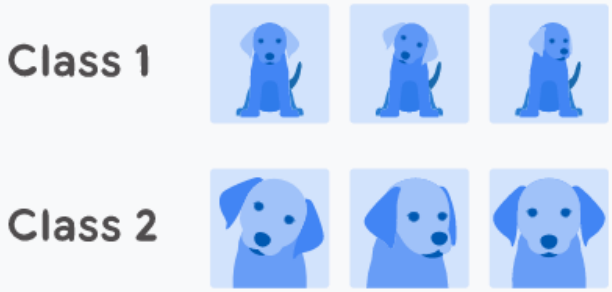
Teachable Machine

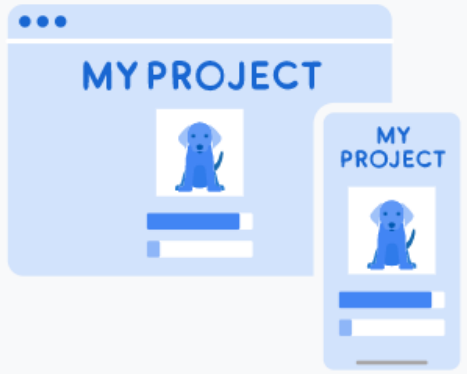
- https://experiments.withgoogle.com/teachable-machine



Teachable Machine

如何使用這項工具？





1 收集樣本

收集範例，並將範例分為你想讓電腦學會的類別。

影片：收集樣本

2 訓練模型

訓練模型並立即進行測試，看看它是否能將新範例正確分類。

影片：訓練模型

3 匯出模型

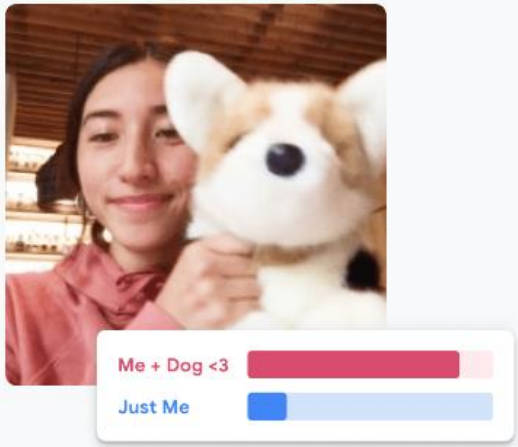
Export your model for your projects: sites, apps, and more. You can download your model or host it online.

影片：匯出模型

Teachable Machine

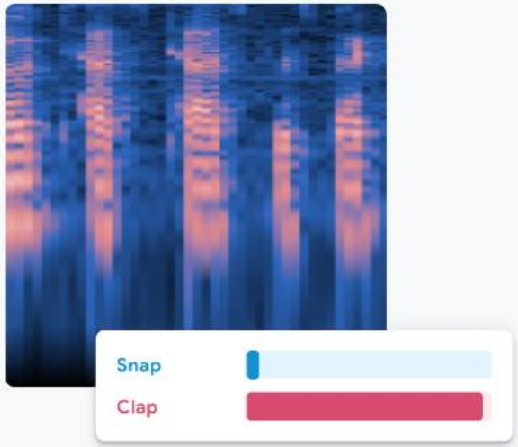
我可以使用哪些內容訓練模型？

Teachable Machine極具彈性，可處理現有檔案或即時擷取範例，想怎麼用，就怎麼用。你甚至可以選擇只在裝置上離線使用，不必擔心電腦上的任何網路攝影機或麥克風資料外流。



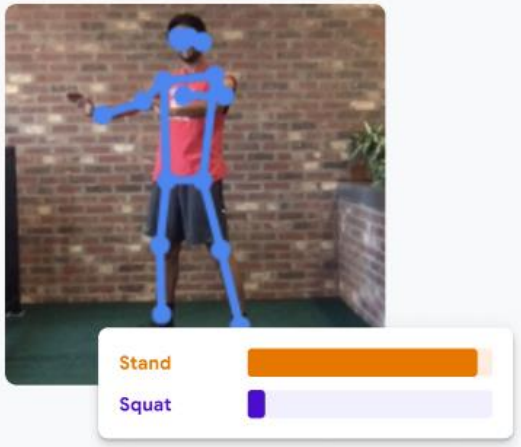
圖片

使用檔案或網路攝影機，訓練模型將圖片分類。



音訊

錄製簡短的音訊樣本，藉此訓練模型將音訊分類。



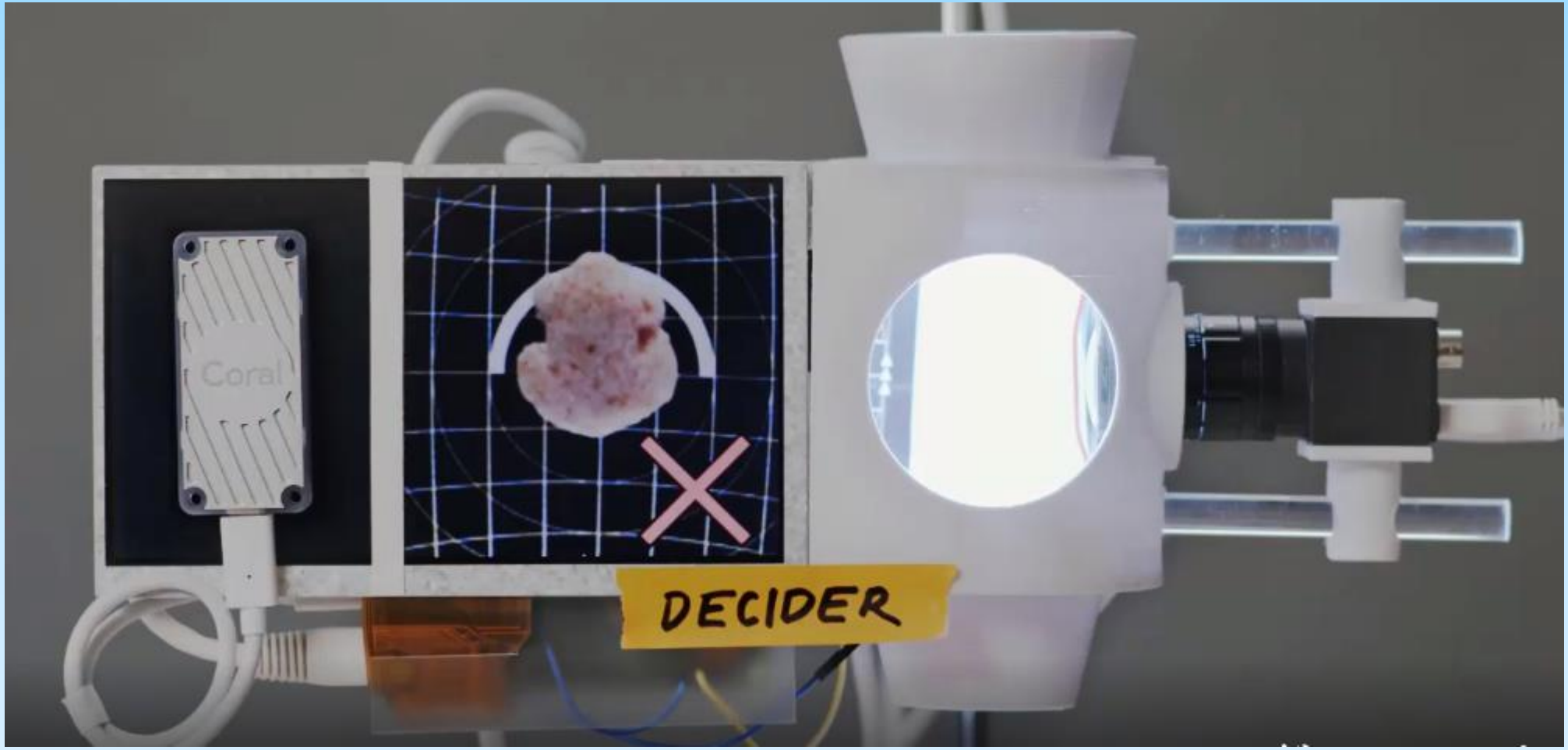
姿勢

使用檔案或在網路攝影機中擺姿勢，藉此訓練模型將身體姿勢分類。

https://experiments.withgoogle.com/tiny-sorter/view

https://coral.ai/projects/teachable-sorter#step-2-setup-pi-and-install-libraries

- https://youtu.be/ydzJPeeMiMI

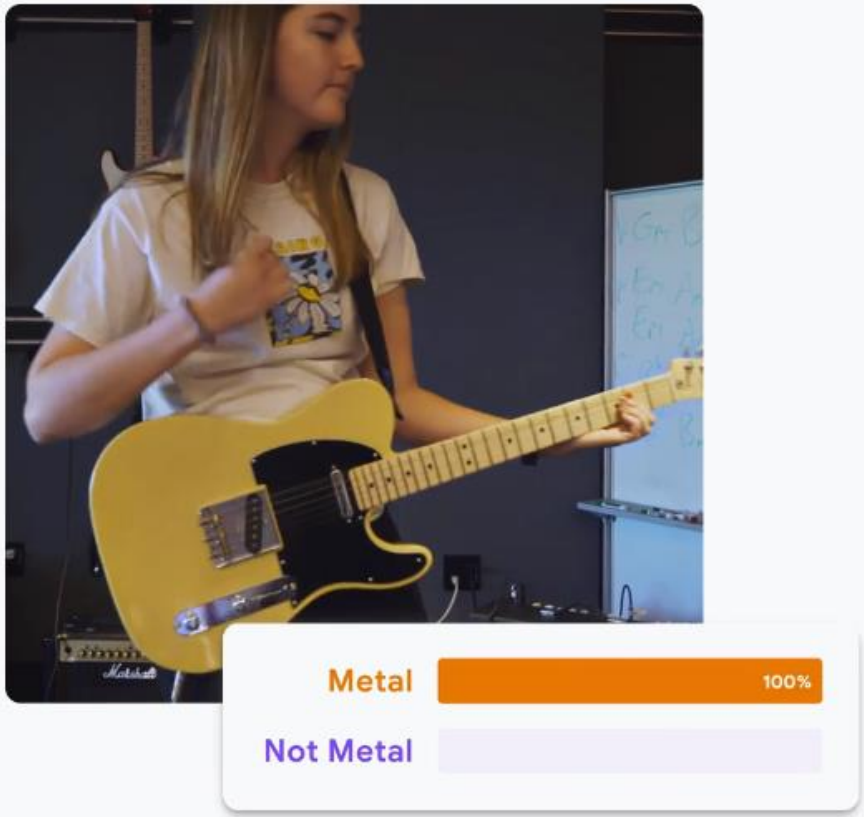


Teachable Machine

訓練電腦辨識你的圖片、音訊和姿勢。

輕鬆快速地建立機器學習模型，以便用於網站、應用程式和其他地方，不需要編寫程式或具備專業知識。

開始使用



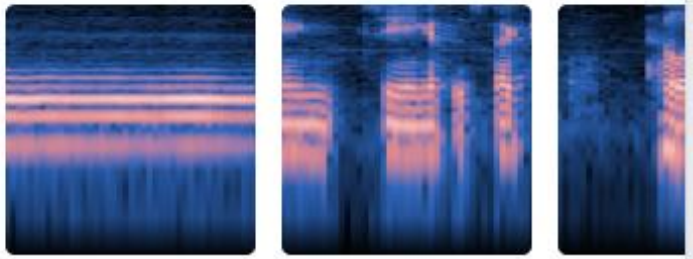
新增專案

從雲端硬碟開啟現有專案。

從檔案開啟現有專案。

圖片專案

以圖片 (使用現有檔案或透過網路攝影機拍攝圖片) 訓練模型。



音訊專案

以長度一秒的音訊 (使用現有檔案或透過麥克風錄音) 訓練模型。

姿勢專案

以圖片 (使用現有檔案或透過網路攝影機拍攝圖片) 訓練模型。

新增圖像專案

標準圖像模型

適用於大多數用途

224 x 224 像素的彩色圖像

可匯出至 TensorFlow、TFLite 和 TF.js

模型大小：約 5 MB

內嵌圖像模型

適用於微控制器

96 x 96 像素的灰階圖像

可匯出至 TFLite for Microcontrollers、TFLite 和 TF.js

模型大小：約 500 KB

查看支援這些模型的硬體

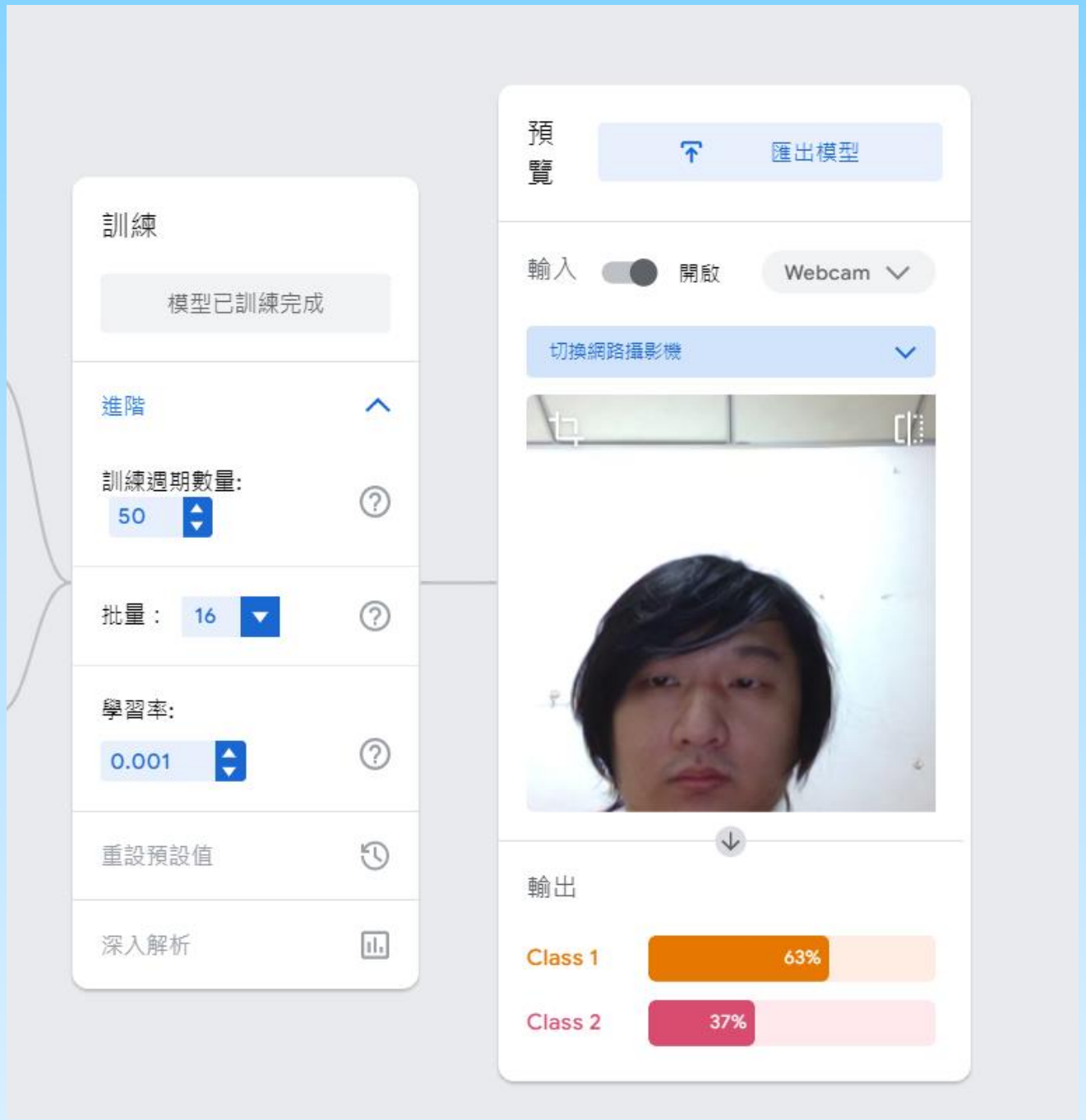
。

以圖片 (使用現有檔案或透過網路攝影機拍攝圖片) 訓練模型。

以長度一秒的音訊 (使用現有檔案或透過麥克風錄音) 訓練模型。

以圖片 (使用現有檔案或透過網路攝影機拍攝圖片) 訓練模型。


Class 1
57 個圖片樣本
網路攝影機
上傳
Class 2
97 個圖片樣本
網路攝影機
上傳
訓練
訓練中...
00:11 - 23 / 50
進階
預覽
匯出模型
必須先在左側訓練模型，才能在這裡預覽。


訓練
模型已訓練完成
進階
訓練週期數量:
50
批量：
16
學習率:
0.001
重設預設值
深入解析
預覽
匯出模型
輸入
開啟
Webcam
切換網路攝影機
輸出
Class 1
63%
Class 2
37%

輸入

切換網路

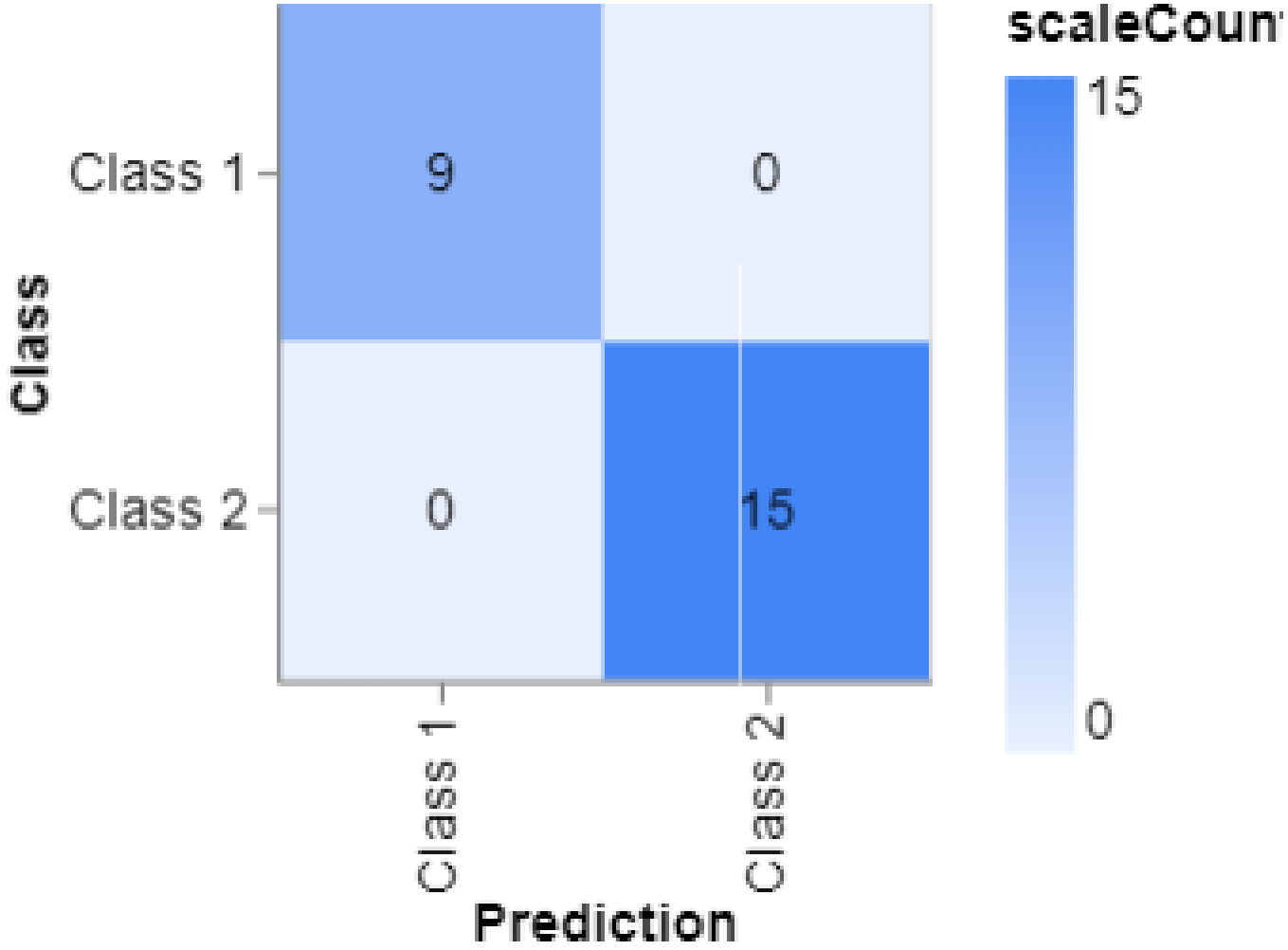
相關詞彙

各類別的準確率

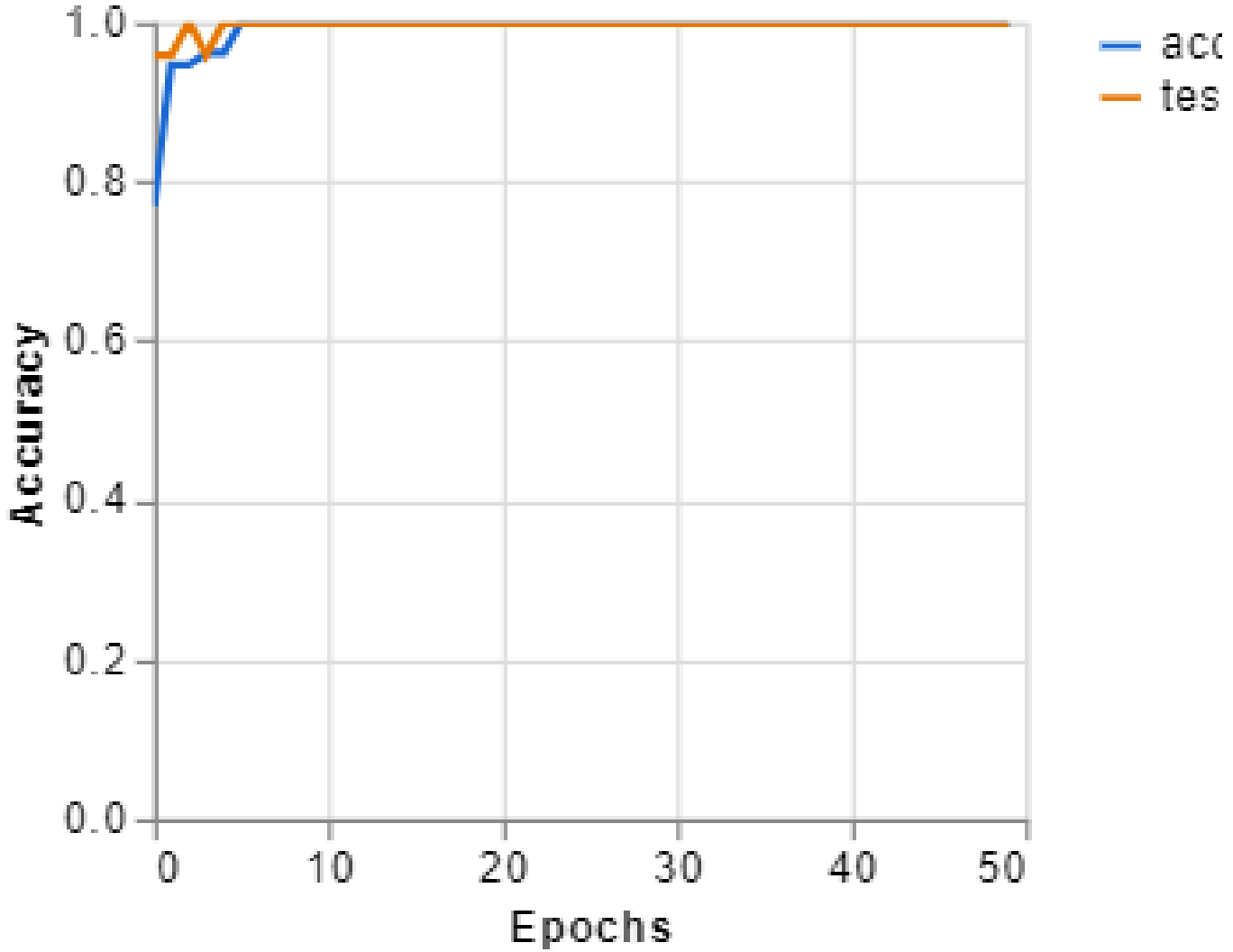
CLASS	ACCURACY	# SAMPLES
Class 1	1.00	9
Class 2	1.00	15

混淆矩陣

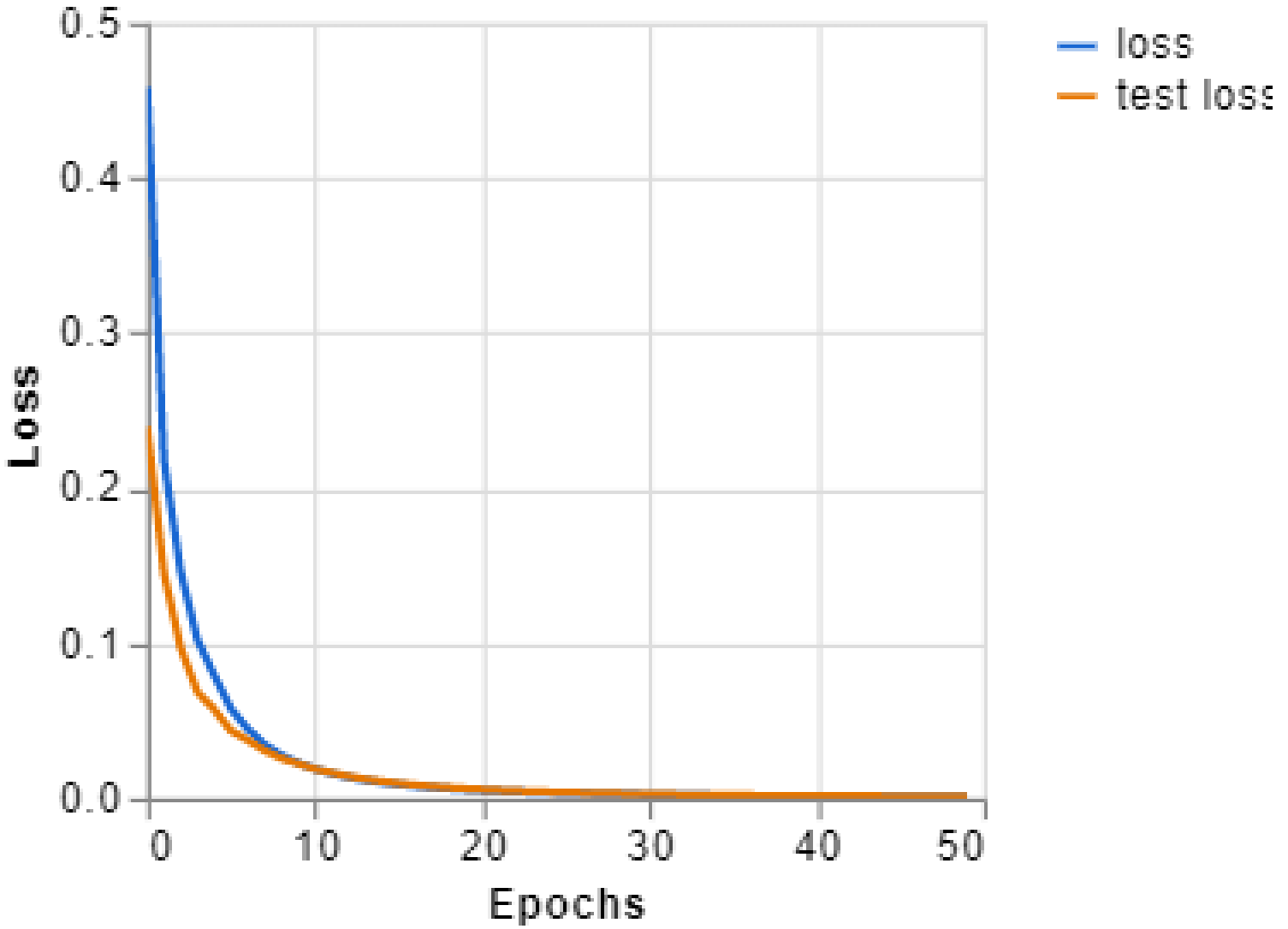
混淆矩陣


Class 1
Class 2
Class
9
0
0
15
Class 1
Class 2
Prediction
scaleCoun
15
0

每個訓練週期的準確率



每個訓練週期的損失



匯出模型以便在專案中使用。

Tensorflow.js | Tensorflow | Tensorflow Lite

https://teachablemachine.withgoogle.com/models/[...]

When you upload your model, Teachable Machine hosts it at this link. (FAQ: Who can use my model?)

透過以下程式碼片段使用模型：

Javascript | p5.js | 在 Github 提供

Learn more about how to use the code snippet on github.

複製

```
<div>Teachable Machine Image Model</div>
<button type="button" onclick="init()">Start</button>
<div id="webcam-container"></div>
<div id="label-container"></div>
<script src="https://cdn.jsdelivr.net/npm/@tensorflow/tfjs@1.3.1/dist/tf.min.js"></script>
<script src="https://cdn.jsdelivr.net/npm/@teachablemachine/image@0.8/dist/teachablemachine-
image.min.js"></script>
<script type="text/javascript">
    // More API functions here:
    // https://github.com/googlecreativelab/teachablemachine-community/tree/master/libraries/image

    // the link to your model provided by Teachable Machine export panel
    const URL = "./my_model/";

    let model, webcam, labelContainer, maxPredictions;

    // Load the image model and setup the webcam
    async function init() {
        const modelURL = URL + "model.json";
        const metadataURL = URL + "metadata.json";

        // load the model and metadata
```

匯出模型

關閉 Webcam

100%

匯出模型以便在專案中使用。

Tensorflow.js | Tensorflow | Tensorflow Lite

將模型轉換成 keras .h5 模型。請注意，轉換作業是在雲端進行，但系統不會上傳訓練資料，只會上傳訓練完成的模型。

透過以下程式碼片段使用模型：

Keras

在 Github 提供

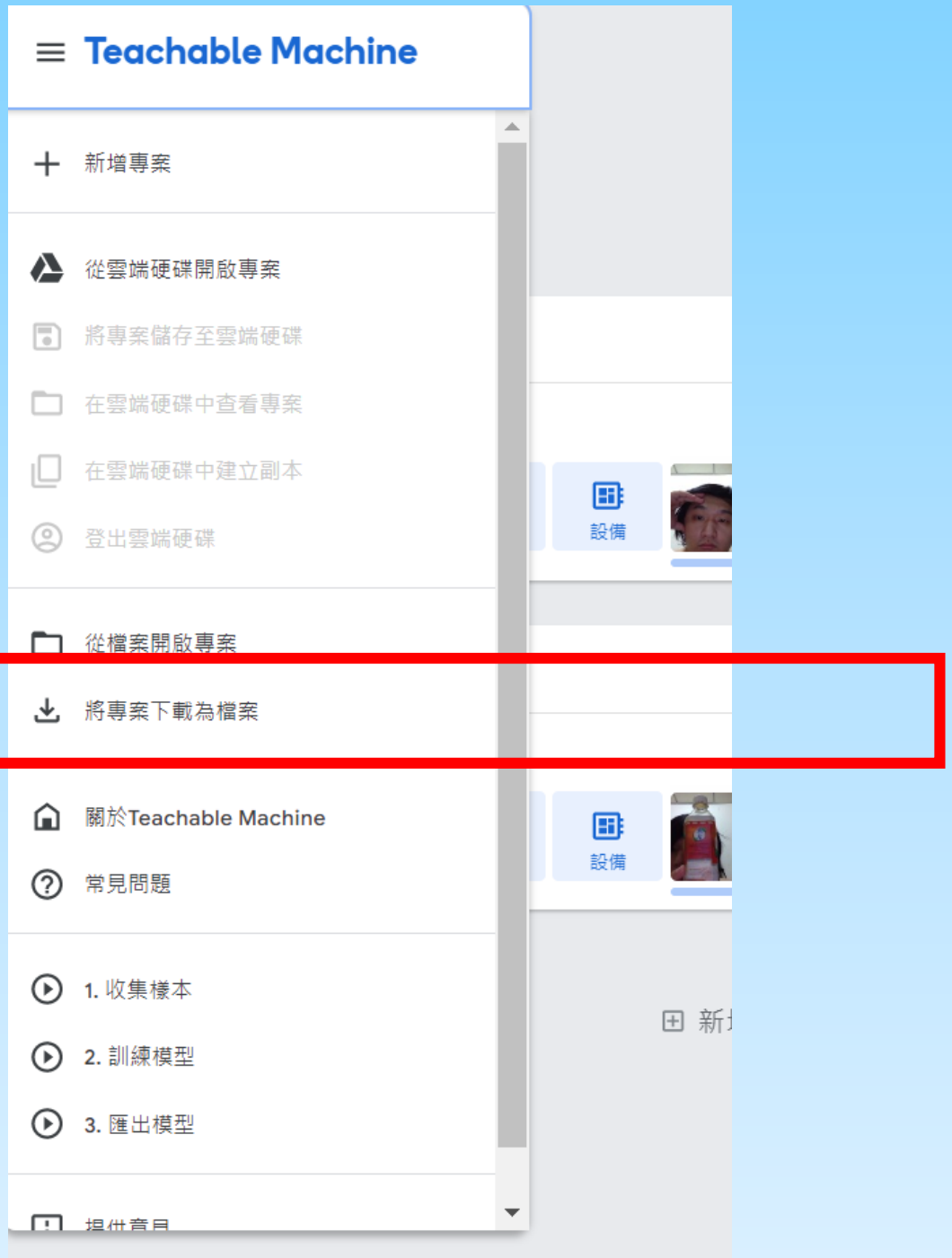
複製

```
from keras.models import load_model
from PIL import Image, ImageOps
import numpy as np

# Load the model
model = load_model('keras_model.h5')

# Create the array of the right shape to feed into the keras model
# The 'length' or number of images you can put into the array is
# determined by the first position in the shape tuple, in this case 1.
data = np.ndarray(shape=(1, 224, 224, 3), dtype=np.float32)
# Replace this with the path to your image
image = Image.open('<IMAGE_PATH>')
#resize the image to a 224x224 with the same strategy as in TM2:
#resizing the image to be at least 224x224 and then cropping from the center
size = (224, 224)
image = ImageOps.fit(image, size, Image.ANTIALIAS)

#turn the image into a numpy array
image_array = np.asarray(image)
# Normalize the image
normalized_image_array = (image_array.astype(np.float32) / 127.0) - 1
# Load the image into the array
data[0] = normalized_image_array
```


Teachable Machine
新增專案
從雲端硬碟開啟專案
將專案儲存至雲端硬碟
在雲端硬碟中查看專案
在雲端硬碟中建立副本
登出雲端硬碟
從檔案開啟專案
將專案下載為檔案
關於Teachable Machine
常見問題
1. 收集樣本
2. 訓練模型
3. 匯出模型
設備
設備

水果甜度與感測器

水果甜度

水果甜度-表面有皺褶

水果甜度-深淺變化


香蕉放七天營養變化
大家最喜歡吃哪種呢?
香蕉放置一段時間後,顏色會一直變化,口味不同之外,營養素也大不同。
C:碳水化合物/S:總糖量/Fru:果糖/Fb:膳食纖維
第0天
第1天
第3天
第7天
竟然放久熱量更低!
熱量高
熱量低
熱量:90Kcal
C:23.3g
S:9.2g
Fru:2.8g
Fb:2g
VitA:44IU
VitC:11.5mg
熱量:87Kcal
C:22.2g
S:13g
Fru:4.3g
Fb:1.7g
VitA:46IU
VitC:6.4mg
熱量:84Kcal
C:21.3g
S:11.9g
Fru:4.5g
Fb:1.7g
VitA:0IU
VitC:4.8mg
熱量:68Kcal
C:17.1g
S:10.8g
Fru:4.5g
Fb:1.6g
VitA:0IU
VitC:4.3mg
每100g計算營養素
YouTube 營養師杯蓋
資料來源:食品成分資料庫

水果甜度-光澤變化

挑選櫻桃的4個重點

POINT 1

果梗硬

從櫻桃梗看是否新鮮

顏色鮮綠且形狀完整的梗蒂，表示新鮮且沒有遭到碰撞

POINT 2

硬度、彈性

彈性佳、果肉結實的櫻桃是新鮮的象徵，水分也比較充足。

POINT 3

光澤潤

櫻桃的採收、包裝及運輸都必須在短時間內完成，維持表皮的光澤；若是儲運過程不佳、陳列過久，都會破壞光澤、顯得黯淡，甚至讓果肉脫水，導致表皮皺摺。

POINT 4

根蒂凹、顏色深

根蒂處形狀愈凹、顏色愈深表示甜度較高。

水果甜度-顏色變化

較成熟的蜜棗顏色會越黃，甜度較高但是口感鬆軟；反之，顏色越綠的蜜棗有較脆的口感，但是甜度就較低。前者因成熟度高，所以不耐久放，應盡早食用。
挑選時要注意果蒂周圍是否凹陷寬廣，光滑平順無高低不平的皺摺，若果蒂周圍突起者就代表品質較差。

水果甜度-科展

中華民國第 59 屆中小學科學展覽會

作品說明書

國小組　化學科

080203

揭開柳橙甜度的秘密

學校名稱：高雄市三民區東光國民小學

水果甜度-科展

(一) 甜度：

人對甜味的喜愛是出於本能；在沒有人工的化學甜味劑出現之前，糖是最主要的甜味來源。**甜度可以說是糖溶液所呈現出的甜味感覺程度。目前真正的甜度尚無法用儀器分析，而是靠官能品評來打分數。**

糖濃度愈高，甜度愈大，這是不爭的事實；但是在糖溶液中若有其他溶質的存在，也會影響甜味的感覺，例如糖中有酸、水果抹鹽、果蔬汁加少量的鹽，均有助於甜度的加強。因此**甜度的大小必須在僅由糖提供甜味的情況下，才能說它與糖度的多寡有關。**又糖的種類不同時，甜度也不一樣，例如**葡萄糖不如蔗糖及果糖甜。溫度不同時，甜度的表現也會不同，例如葡萄糖與果糖在冷水中的甜度比熱水中甜度大。**水果及果汁飲料中的糖，往往不是單純的一種糖，因此甜度與糖度之間還有一些差距。

水果甜度-科展

(二) 糖度：

目前國人通稱的糖度，和蔗糖工業上所用的糖度略有不同。**在果實、果汁、食品品質分析上所稱的糖度通常是以 Brix 度數表示，指的是用比重計法測定樣本溶液的比重。Brix 度數雖然習慣上看成是含糖重量百分率(一般果汁中存在的糖類如葡萄糖、果糖與蔗糖的比重或折射率相似。)**然而實際上所代表的意義，應該是**指果汁所含溶解物概略重量百分率。**一般的水果及水果產品，**若不含不溶性物質(如果汁)時，其可溶固形物含量即以糖度折射計法所得到的讀數表示**

水果甜度-大學研討會

SEAIT2017 第 9 屆企業架構與資訊科技研討會

水果成熟度分級檢測系統之設計

張正弘 老師	李權容 學生	藍家翔 學生
德明科大	德明科大	德明科大
資訊科技科	資訊科技科	資訊科技科
lion@takming.edu.tw	penda@gmail.com	penda@gmail.com

水果甜度-大學研討會

摘要

目前台灣的水果自動化分級技術雖然已發展完全，但成本仍居高不下，一台桌上型的機器就要價95萬元，更不要提價格高達3000萬元的線上型分級器，所以若是我們能夠做出價格低廉，機動性高，檢測成果穩定的水果自動化分級器，相信能在農產的市場裡打出一片天。水果自動化分級器在農產的市場裡絕對是不可或缺的，一直以來消費者在購買水果時的共同疑問就是水果夠不夠甜，而這種主觀又抽象的問題也總是得不到準確的答案，要是我們能夠讓水果廠商都購買一台水果自動化分級器的話，有了統一且標準的甜度指標，對消費者來說相信也會更有說服力。

本研究的貢獻於果農，方便水果商或農民可以不用再以拍打或者是其他方式來辨別水果成熟程度、水果種類或者是甜度等等，便於分類、分級。我們主要以影像處理方式及電壓晶體辨別水果色澤及瑕疵、檢測水果甜度、判斷大小、重量辨別。在這個科技時代以機器對水果實施非破壞性個體檢驗，除可提高品質、增加產品經濟價值、確保果農收益外，將有助於提升消費者購買信賴，藉由機器分析數據，診斷果農栽培管理方法的妥適性，除了可降低人成本外更能提高水果品質。

水果甜度-大學研討會

4.實驗設計、過程及結果

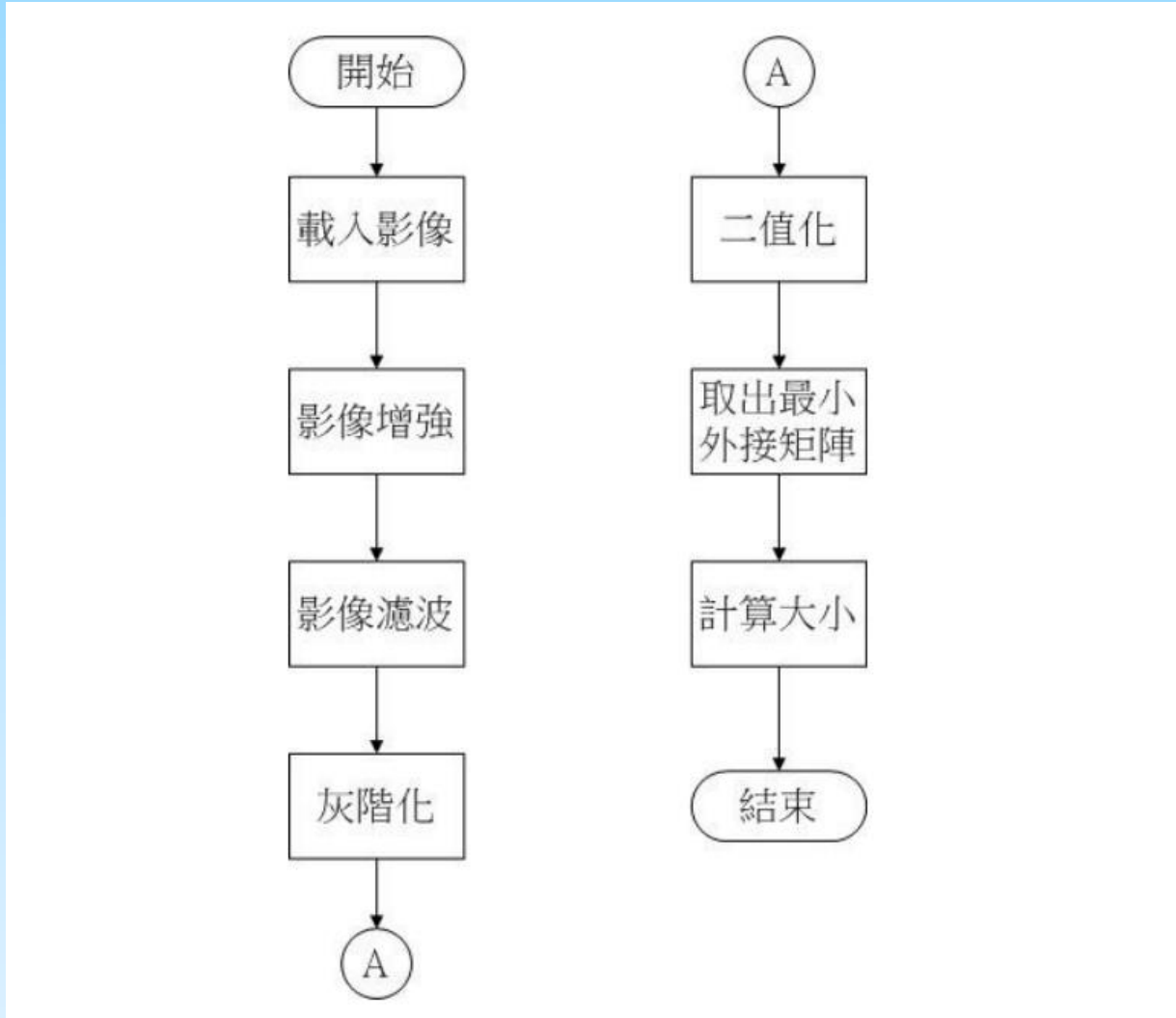
初期實驗的目標是找出水果的相對大小，並藉以分類水果的級數，一般是越大的水果等級較高。我們先載入靜態的水果原始彩色影像，此處選擇檸檬為實驗的水果，處理的過程包含先進行影像的前處理，前處理的步驟為影像的增強及濾波，待影像品質改善後，先將影像灰階化，再將灰階轉二值化，二值化的閥值採用Otsu's的方法決定。未來也可以在二值化步驟後標示出水果的瑕疵處，接著就可以計算水果的大小。圖9中為處理的流程圖，並附上部分處理過程的程式碼。

水果的大小計算可以由三個方式來進行

a. 直接由二值化影像來計算圖中的白色點的數目，超過某的數字或比例及可以定義為大的水果。

b. 求取外接最小矩形框，並讀取矩形的長、寬資料，以長、寬中的最大值為水果大小的判斷數值，外接矩形的圖示如圖13。

水果甜度-大學研討會



水果甜度-大學研討會

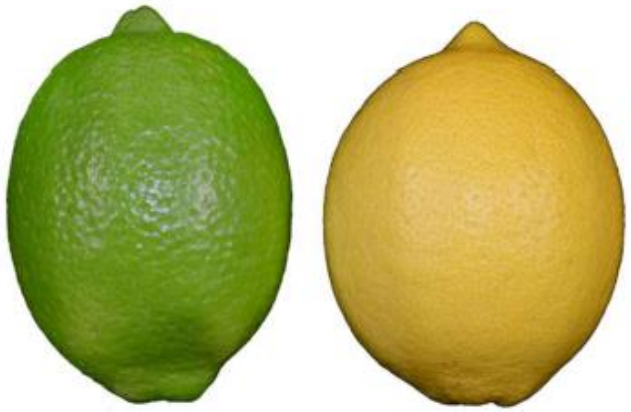


圖10 檸檬原圖

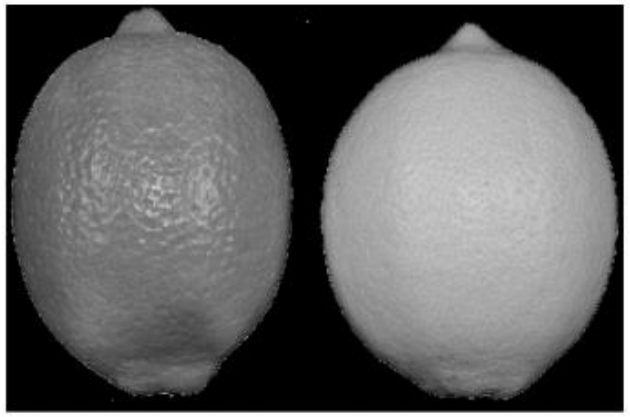


圖11 灰階化後的檸檬圖

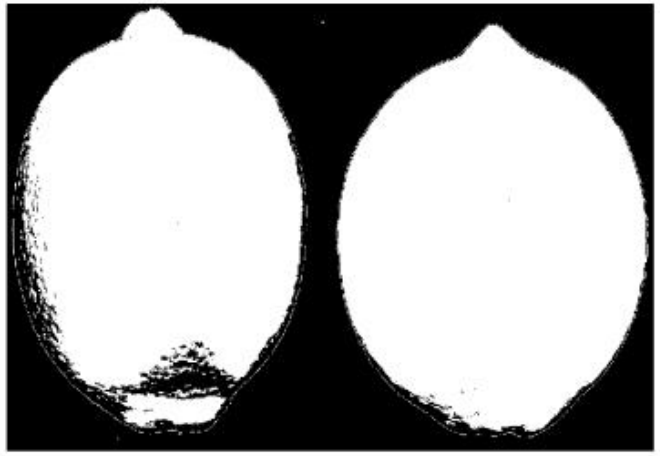


圖12 二值化後的圖

圖13 外接最小矩形框

水果甜度-大學研討會

5.討論及未來工作

本論文依據水果的影像特性進行水果大小的判定及瑕疵的檢測，依前章的實驗結果，目前所做出的結果已經可以達到初步判讀大小的數據。但是水果影像存在著非常多的變異性，例如拍攝的角度及水果本身的光澤造成反光等，處理過後的影像必定存在非常多的雜訊，此雜訊會影響二值化的結果，並改變外接矩形的大小。未來將繼續研究影像增強及影像濾波等技術，增強影像對比，消除雜訊，並考慮以形態學濾波進行二值化的進接觸，以達到最佳影像。未來也希望可以繼續做出瑕疵檢測效果及水果品種分類。

Prompt

如何不讓生成式文本" 一本正經胡說八道"

交待背景

說明需求

補充或排除

提問流程建議

概念對齊 → 創意碰撞 → 內容擴寫

如何不讓生成式文本”一本正經胡說八道”

抽絲剝繭

保持存疑

反向驗證

美國著作權局（USCO）

機器生成的圖像無法受到著作權保護

正確清楚的語意

- 隨便的形容

cat in meeting →

會議中有貓

貓正在會議

貓在會議中

貓會開會

正確清楚的語意

cat in meeting

正確清楚的語意

- 更實際的形容

3 cats chatting →

3隻貓聚一起(聊天)

3隻貓在聊天(開會)

正確清楚的語意

3 cats chatting

各式風格關鍵字使用

繪圖「風格」	A stylized Cyberpunk（賽博龐克）、A stylized Cthulhu Mythos（克蘇魯神話）、迪士尼（Disney）、皮克斯（Pixar animation）
場景「風格」	realism（現實主義）、surrealism（超現實主義）、anti-utopia（反烏托邦）、印象派風格（Impressionism）、漫畫風格（Cartoon Style）、Photorealist (真實感)
場景「物件」	ruins（廢墟）、city（城市）、street（街道）、universe（宇宙）
使用某藝術家的「畫風」	Miyazaki Hayao（宮崎駿）、Shinkai Makoto（新海誠）、Pablo Picasso（畢卡索）、Vincent Van Gogh（梵谷）
使用某遊戲的「畫風」	botw（曠野之息）、Pokémon（寶可夢）、The Elder Scrolls（上古卷軸）

圖片調整燈光&視角

Composition
視角

closeup view（特寫鏡頭）
Wide-angle view（廣角鏡頭）
A bird's-eye view（鳥瞰）

Lighting
燈光

Soft light（柔光）
Hard light（硬光）
Cold light（冷光）

燈光範例

- Strawberry birthday Cake, closeup view, Soft light
- （草莓生日蛋糕、特寫鏡頭、柔光）

視角+光線範例

- ruins, warm light
- （廢墟，暖光）

視角+光線範例

- ruins, Wide-angle view, Cold light
- （廢墟，廣角鏡頭，冷光）

再增加畫面鏡頭感

「攝影＋對焦」讓圖片有商業攝影的感覺

- photography（攝影）、cinematic（電影）、Long Shot (遠景)
- in focus（對焦）、depth of field（景深）

實際範例

- Strawberry Cake, photography, closeup view, in focus, Soft light
- （草莓蛋糕、攝影、特寫鏡頭、對焦、柔光）

實際範例

- Church ruins, cinematic, Wide-angle view, depth of field, Warm light
- （教堂廢墟，電影，廣角鏡頭，景深，暖光）

實際範例

指定色票

- SVG Blue Haven logo #A9A9A9, #89CFF0, #FF6600

實際範例

插畫形式

- vector art, cat fishing in a pond, with a rod and reel in its paws. Muted colors with earth tones and simple lines.

Muted colors 啞色(低飽和色調)

Earth tones 大地色系

實際範例

等軸測投影
Isometric projection

Isometric cutaway illustration of empty house for 15 year old little girl, color pink mattress, super cute, no girl.

實際範例

線圖風格

one line drawing Taipei 101

產品設計

product design

Google pixel 7, pro rococo style, product design.

用手繪實現真實

S__170352893-02.jpg

生成式AI 產生影片

Google One

Gemini

大學生使用 Google AI Pro 版 Gemini 可以免付費 1 年

~~每月 $650~~ 前 12 個月每月 $0

取得學生優惠

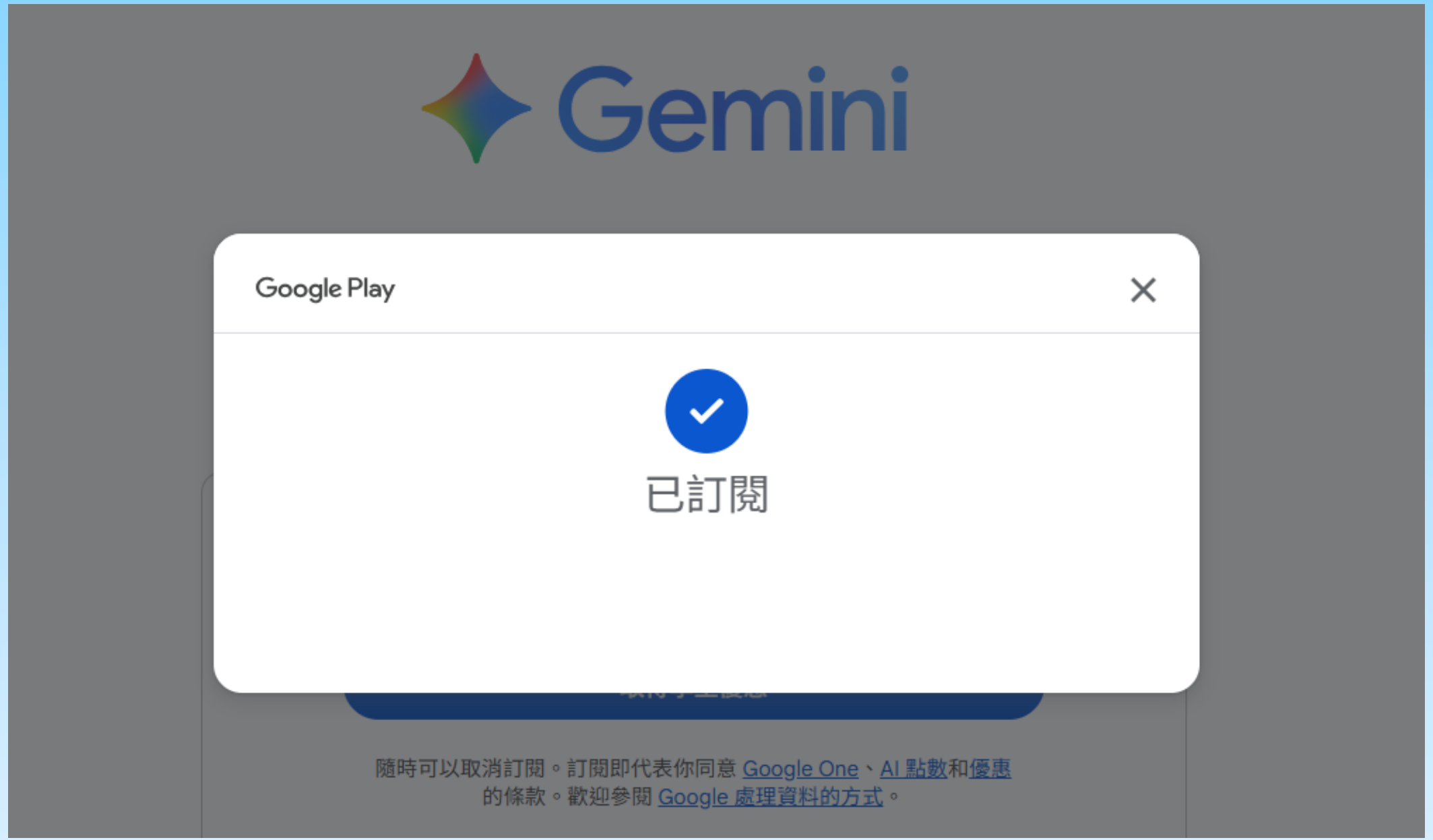
隨時可以取消訂閱。訂閱即代表你同意 Google One、AI 點數和優惠的條款。歡迎參閱 Google 處理資料的方式。

精選 Gemini 福利

功課輔導和考前準備
分析整本教科書 (最多 1,500 頁)

寫作輔助
生成初稿、精修論點及調整構想

生成影片
利用 Veo 3 技術，將簡單的文字轉換成精彩影片


Gemini
Google Play
已訂閱
隨時可以取消訂閱。訂閱即代表你同意 Google One、AI 點數和優惠
的條款。歡迎參閱 Google 處理資料的方式。

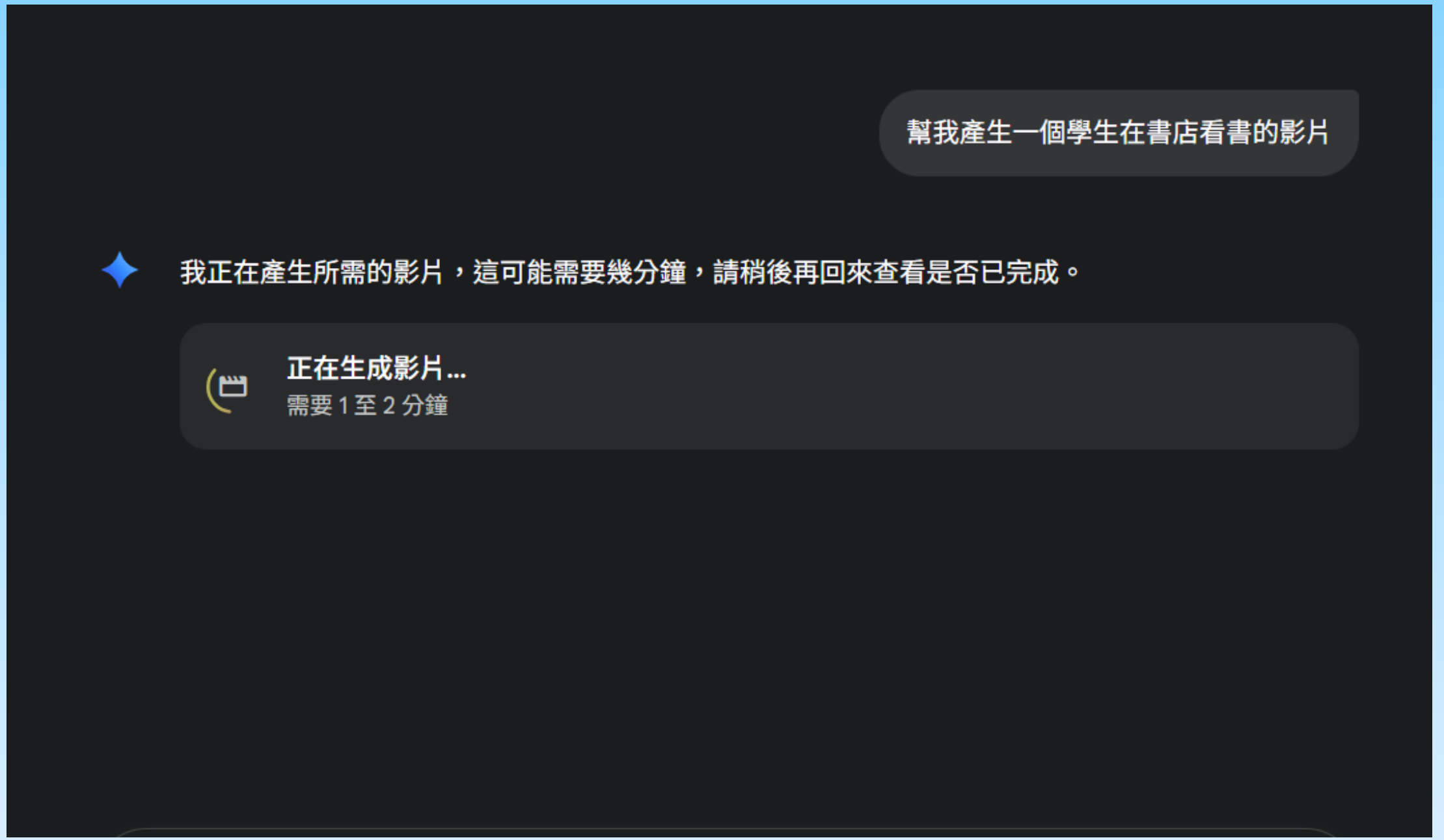
Gemini

你已訂閱 Google AI Pro

盡情使用 Gemini、NotebookLM 等服務

Google AI Pro 福利

- 使用 Google 最強模型的實用功能和進階版功能
- 使用 Veo 模型製作優質影片
- 可運用 Gemini 的更多功能，包括 Deep Research
- NotebookLM 的語音摘要和來源數量增為 5 倍
- Gmail 和 Google 文件等服務內建 Gemini
- 享有總計 2 TB 的儲存空間和更多進階版福利


幫我產生一個學生在書店看書的影片
我正在產生所需的影片，這可能需要幾分鐘，請稍後再回來查看是否已完成。
正在生成影片...
需要 1 至 2 分鐘

幫我產生一個學生在書店看書的影片

顯示思路

您的影片已準備就緒！

Vidu

- Vidu 於 2025 年第 3 季釋出 Q1 版本，正式開放 Reference-to-Video（圖像參照生成影片）功能，讓創作者得以上傳多張角色或場景圖片，並透過一則文字描述指令生成風格一致的短片。此更新結合原有 Image-to-Video 與多圖融合能力，使生成片段在不同視角與鏡位切換時，仍可維持角色樣貌、背景元素及敘事邏輯的一致性，大幅提升影片連貫度與實用性。

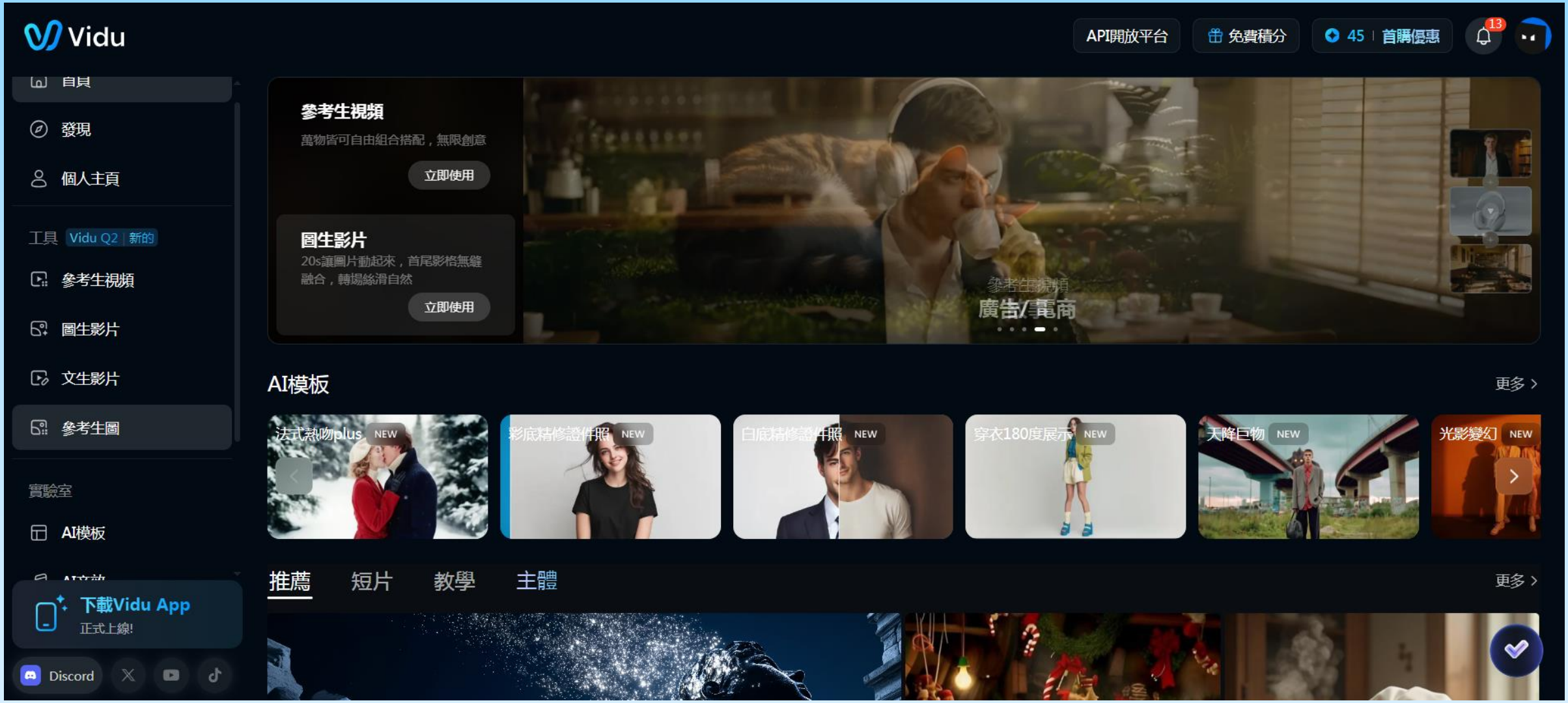
Vidu

- 1｜支援多圖參照（Multi-Reference Input）
- 使用者可上傳最多 7 張參照圖，涵蓋角色、場景與道具，Vidu 會根據圖片中之特徵建立視覺模組，並於影片中自動還原相關細節。這種模組化設計可靈活重組，擴展不同組合應用。
- 2｜語意驅動畫面生成（Semantic Understanding）
- 即使用戶未提供所有畫面元素（例如道具或背景），只要於 Prompt 中提及，系統亦可自動生成對應畫面。這顯示其具備語意補全能力，可根據描述合理推理畫面場景。
- 3｜角色與畫面一致性提升（Temporal Coherence）
- 影片中的角色表現、服裝、體態將根據所選參照圖保持一致。即使出現鏡位轉換、視角切換或多角色互動，Vidu 亦可維持角色外觀與動作邏輯的穩定，解決傳統生成影片中角色「跳格」、「變形」等問題。
- 此三大特點令 Reference-to-Video 成為目前市面上少有同時兼顧「靈活構圖」與「穩定敘事」的生成方案，無需分鏡圖亦可完成具敘事邏輯的視覺片段。

Vidu

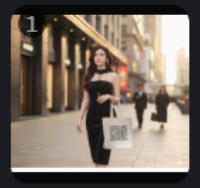

免费获得积分
45
免费版 | 积分明细 >
购买积分
订阅套餐
待完成
历史任务
下载App
下载Vidu App获得额外积分，随时随地创作不停
+20
已完成
每日登录
登录即领积分，限时福利，不容错过！
+5
已完成
新人礼包
新人注册，免费赠送
+20
已完成
品牌资讯
订阅我们的品牌资讯，获取产品更新、使用技巧和特别优惠
+20
已完成

Vidu


Vidu
API開放平台
免費積分
45 | 首購優惠
13
發現
個人主頁
工具 Vidu Q2 | 新的
參考生視頻
圖生影片
文生影片
參考生圖
實驗室
AI模板
下載Vidu App
正式上線!
Discord
參考生視頻
萬物皆可自由組合搭配，無限創意
立即使用
圖生影片
20s讓圖片動起來，首尾影格無縫融合，轉場絲滑自然
立即使用
廣告/電商
AI模板
更多
法式熱吻plus NEW
彩底精修證件照 NEW
白底精修證件照 NEW
穿衣180度展示 NEW
天降巨物 NEW
光影變幻 NEW
推薦
短片
教學
主體
更多

首頁 | AI圖片 | API開放平台 | 免費積分 | 45 | 首購優

參考生視頻 | 圖生影片 | 文生影片



+ 第2幀

時長: 5秒

請讓模特兒拿起這個提包，並且走在路上

專業模式

使用教程 | 試用範例

模型 Vidu Q2

清晰度 1080p

創作 0

進行中 1

創作任務提交成功，內容生成中...

圖生影片 2025年11月18日22:27

請讓模特兒拿起這個提包，並且走在路上

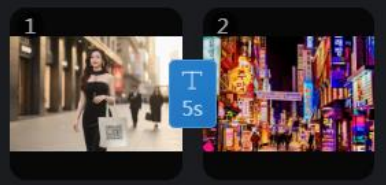
生成中

進行中

請讓模特兒拿起這個提包，並且走在路上

首页 | AI图片 | API开放平台 | 免费积分 | 45 | 首购优惠

参考生视频 | 图生视频 | 文生视频



+ 第3帧

第1帧 - 第2帧: 5s

將第一張照片的人物，於第二張照片作為背景，走在街道上

专业模式

使用教程 | 试用样例

模型 Vidu Q2

清晰度 1080p

创作 10

进行中

奖励合计60积

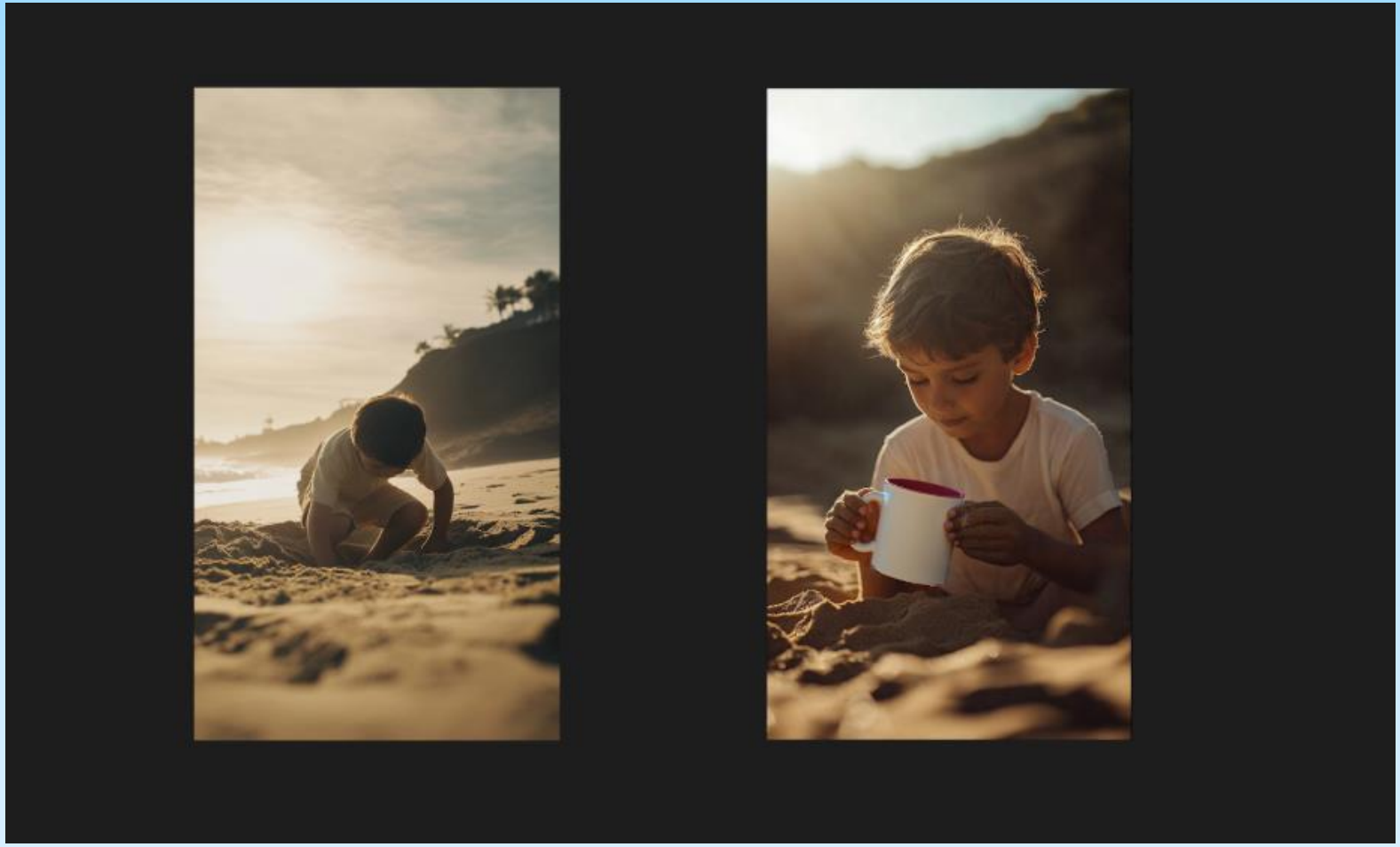
做同款 | 投稿

課堂測試

- 將選一張照片作為產品宣傳照，使一張照片動起來。
- 另外，使用兩張照片 ，作場景切換。

Step 1：構思影片故事與畫面

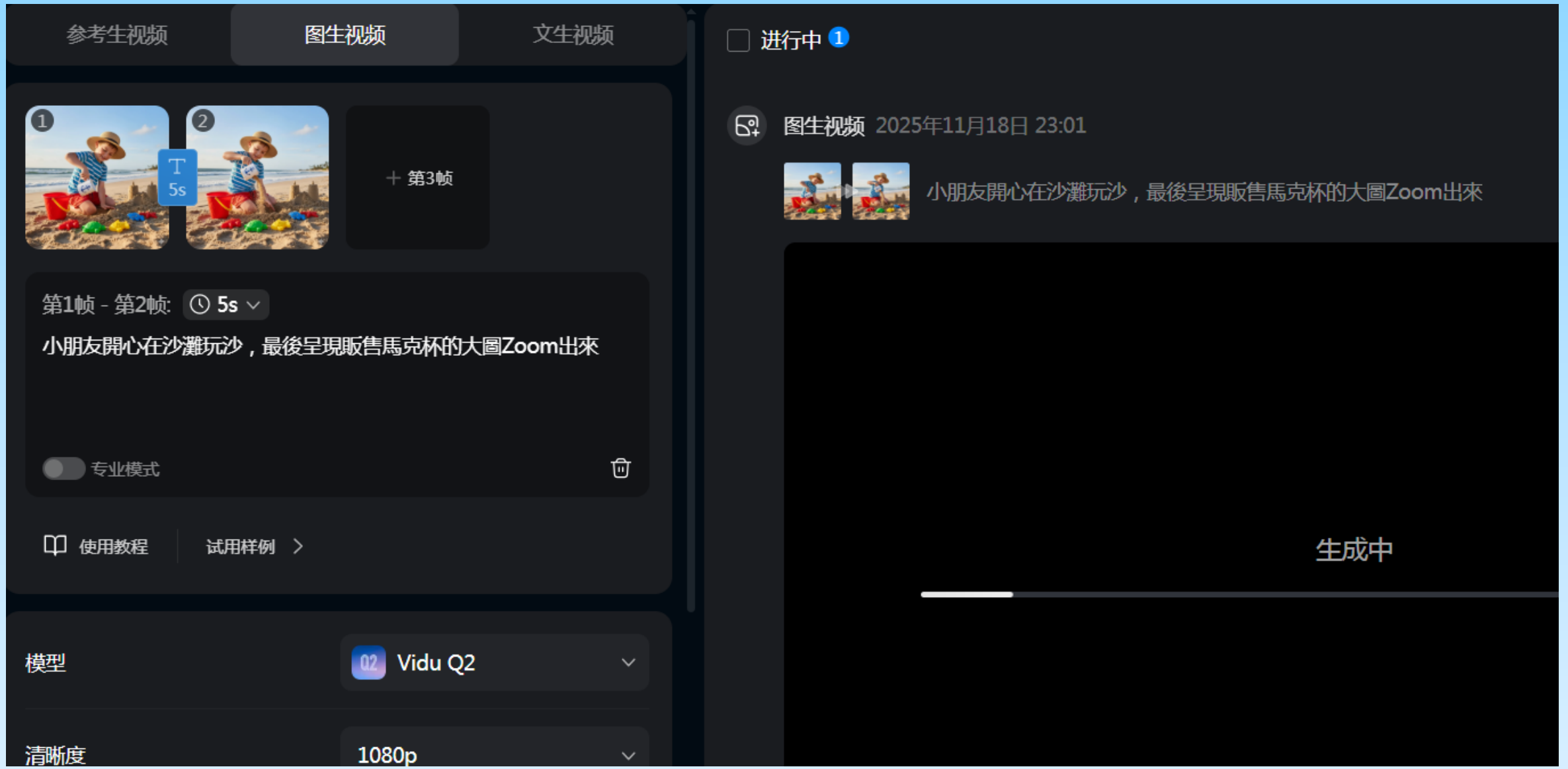
- 假設我們想製作一支「小男孩拿馬克杯挖沙坑」的短片，這就需要以下圖片素材：



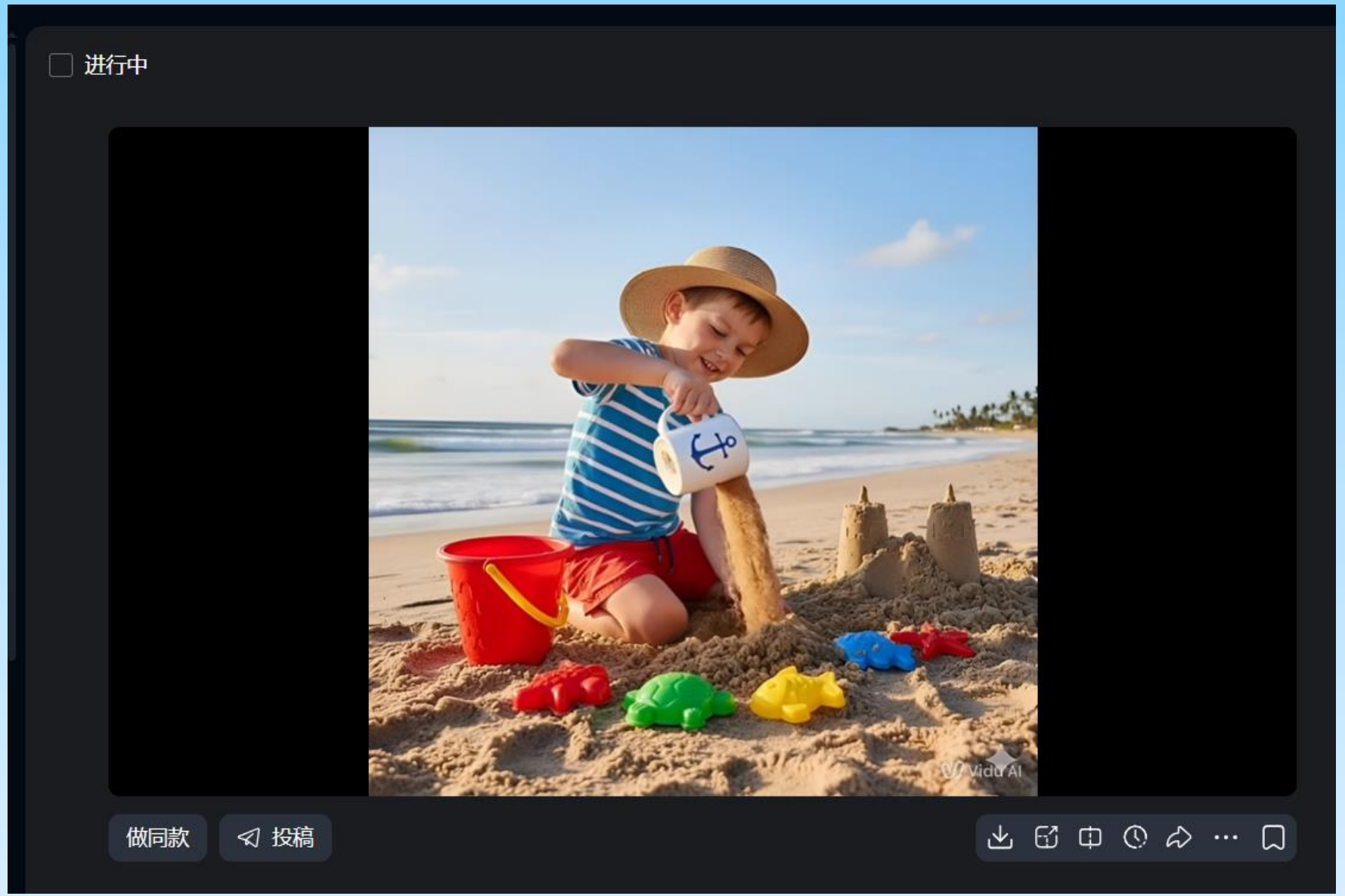
Step 2：使用 Gemini 生成圖片

- 假設我們想製作一支「小男孩拿馬克杯挖沙坑」的短片，這就需要以下圖片素材：

Step 3：進入 Vidu 製作影片



完成




非常感謝