

【新興領域：3月焦點5】你是我的眼-淺談電腦視覺辨識市場

關鍵字：人工智慧、視覺辨識、電腦視覺

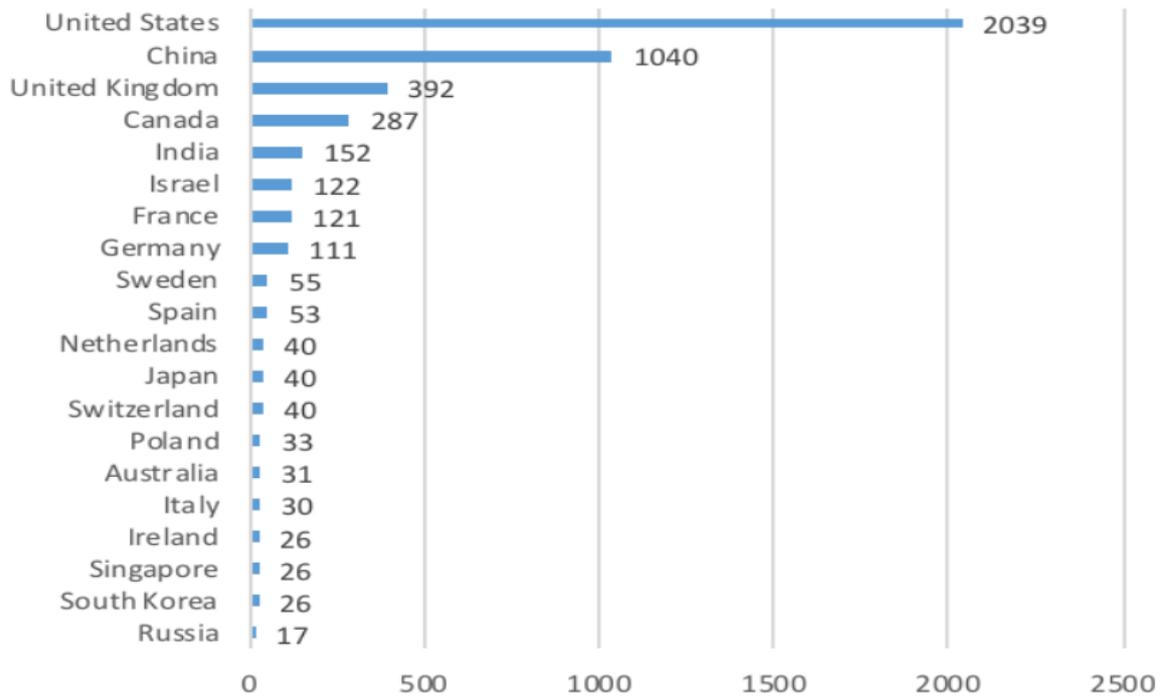
日期：2019.3.15



近年火熱的議題當屬人工智慧了，而前陣子的 AI 獨角獸-商湯科技即在視覺辨識中佔有一席之地。究竟什麼是視覺辨識技術呢？市場規模有多大？重要性？本文將由市場切入，對全球相關技術企業作一概要掃描。簡單介紹歷史再到淺談技術層面，之後對於商業上的應用層面以及富有趣味的應用做一介紹，最後則是整理臺灣相關視覺辨識企業。

根據維基百科上的定義，電腦視覺辨識是一門研究如何使機器「看」的科學，更進一步的說，就是指用攝影機和電腦代替人眼對目標進行辨識、跟蹤和測量等機器視覺，並進一步做圖像處理，用電腦處理成為更適合人眼觀察或傳送給儀器檢測的圖像。看似複雜的技術，但現今在日常生活的應用已經十分普遍了，舉凡手機解鎖、刷臉打卡上下班、工業自動化的應用、身分認證、自駕車感測器等。顯示了其應用領域層面的廣泛，也代表著富有高商業價值，因此對於教「電腦學會看一張圖片」這種技術不再只處於學術研究階段，而成功地進入日常生活當中。

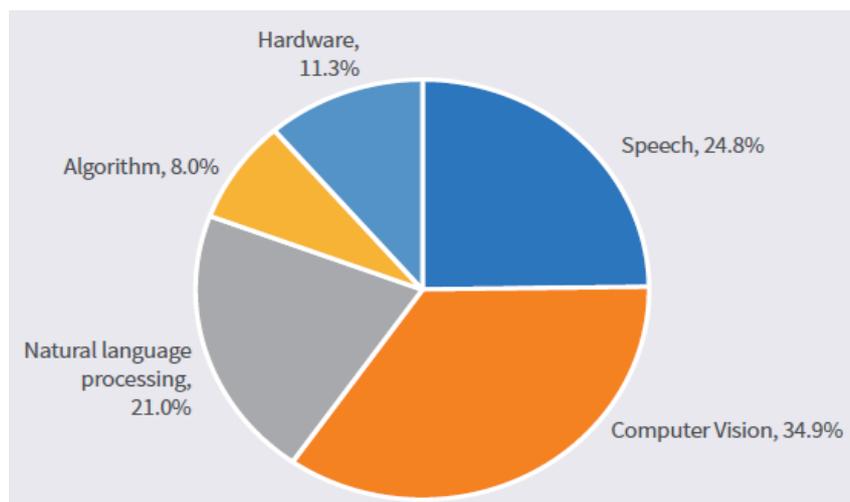
AI 市場概況



圖片來源：中國信通院

圖 5-1、全球 AI 企業數量(截止至 2018 上半年)

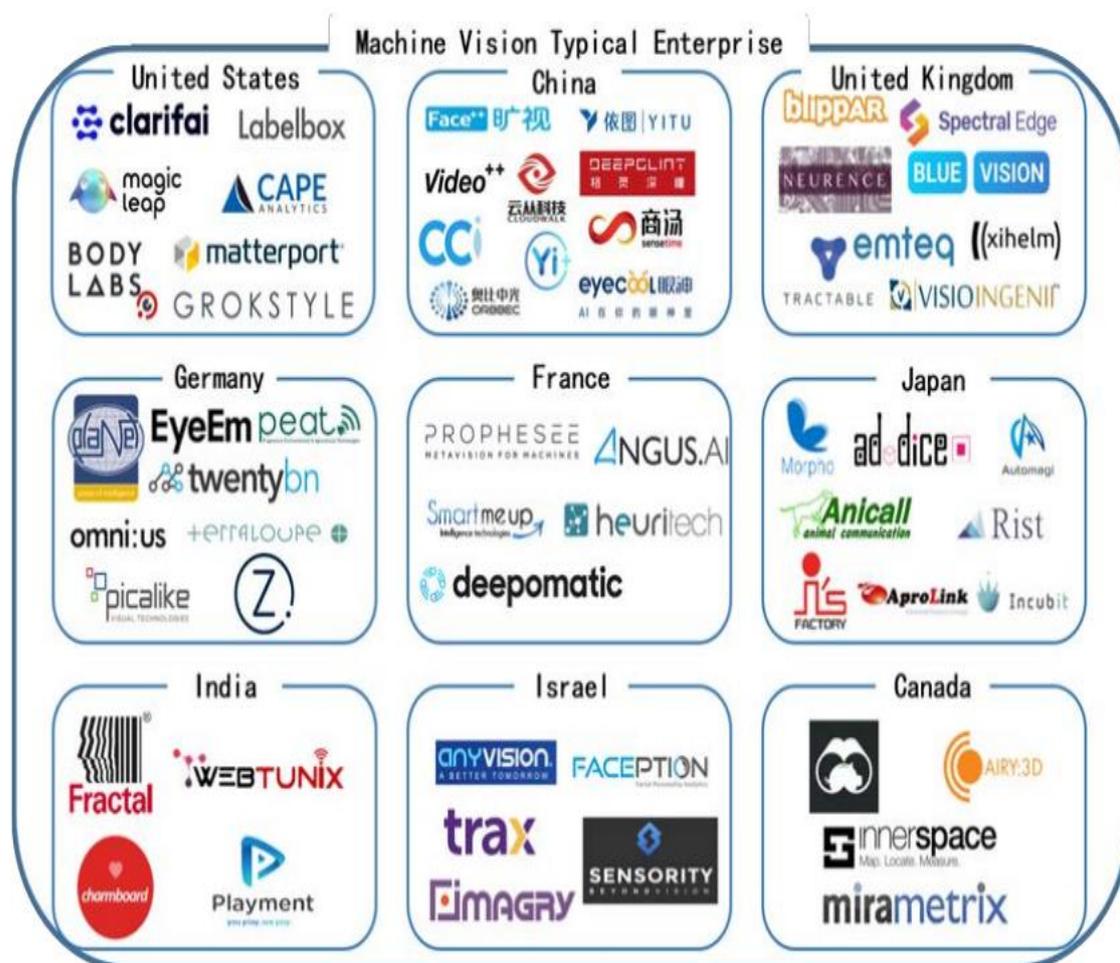
根據中國信通院報告截至 2018 上半年為止，全球共有 4,998 家 AI 企業。而美國與中國佔據近六成比例，美國共 2,039 家、中國 1,040 家。一般在討論全球 AI 實力時總是以美國、中國為主。但值得一提的是，英國的 AI 企業數量在全球排名第三也不惶多讓，且以開發 AlphaGo 聞名的 DeepMind 公司本身也是一家英國 AI 公司。



圖片來源：中國信通院

圖 5-2、中國 AI 市場結構

圖 5-2 為中國 AI 市場結構，電腦視覺市場份額高達 34.9%，常見應用如生物辨識、圖片影片的識別。在近兩年，深度學習演算法的技術突破將帶動視覺辨識、語音更廣泛的應用，但由於技術的到位需要對應的硬體設備方能發揮綜效，因此相關硬體、半導體 AI 晶片也在 AI 市場結構扮演重要的一角。

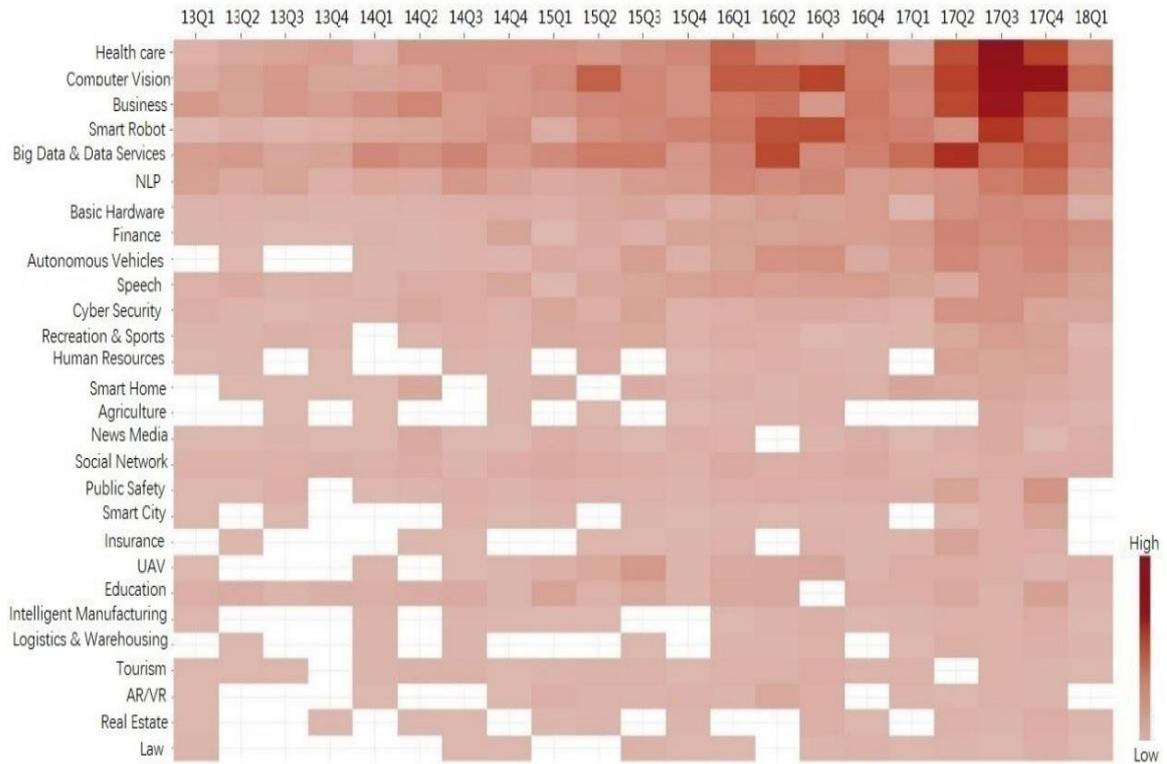


圖片來源：中國信通院

圖 5-3、全球電腦視覺企業

與傳統視覺辨識相比，AI 視覺辨識有快速、高準確性的優點。視覺辨識技術應用在企業上包括圖片或影片的辨識、分類、偵測等，在生活中常見於影像監控、自駕車、臉部辨別、醫療診斷、工業自動系統等。2017 年視覺辨識市場規模僅 23.7 億美元，但隨著近年技術的突破與應用的廣泛，2023 年市場可望成長至 253.2 億美元。

以美國與中國視覺辨識企業為例，Cape Analytics 是一家依據衛星圖像數據，為房屋估價，而且讓保險公司進行準確且快速的報價。中國則有港中大教授創辦的 AI 獨角獸-商湯科技，以及應用視覺辨識技術在金融、交通、醫療等領域的依圖科技。



圖片來源：中國信通院

圖 5-4、全球 AI 領域投資熱力圖(2013-2018Q1)

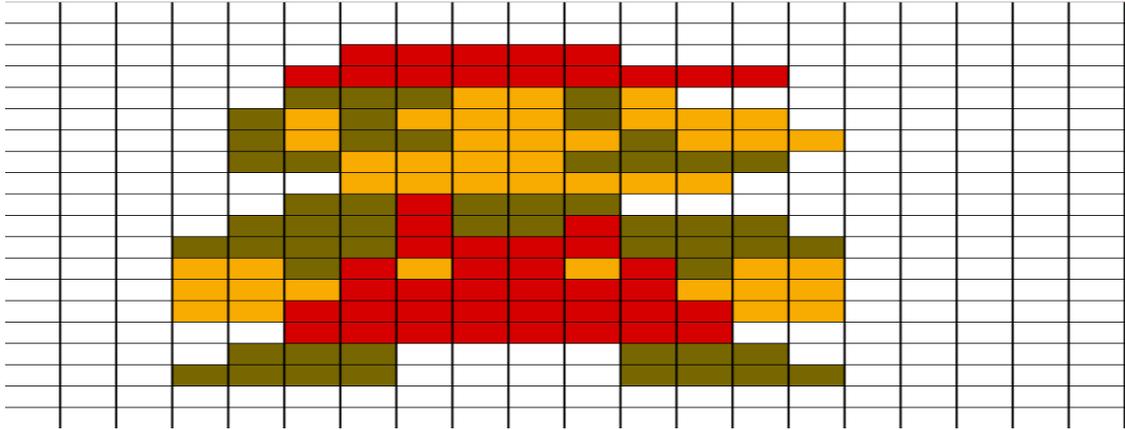
以每季 AI 領域投資繪製成熱力圖，可以觀察到電腦視覺與醫療照護在近年都有熱度上升的趨勢，而其餘領域的投資則無一明顯趨勢，較為分散。

視覺辨識技術探討

「一畫勝千言」，圖像能夠傳遞的訊息有時候遠勝於文字。我們每日見到的街道景象、建築物、人事物，透過大腦處理眼睛接受的影像而留存於腦海中。但是對於電腦來說，上述不過是一個個代表不同數字的矩陣，電腦無法簡易區別各圖片間的差異。電腦視覺技術是研究者透過設計機器，賦予機器能夠像人類一樣能夠對視覺接受到的視覺資訊理解。(以下圖 5-5 至圖 5-9 摘自 Deep Learning for Computer Vision)

圖像

在聊聊圖像辨識前，我們需要瞭解什麼是「圖片」。而圖片中最重要的元素稱為像素(Pixel，Picture + element)，是圖片顯示的基本單位，我們可以把圖片想成是一個烤肉架，像素就是烤肉架中每個方格中顏色與光的強度，也可以用欄與列的數量表示一組有像素個數的矩陣。



圖片來源：<https://reurl.cc/K40p9>

圖 5-5、像素示意圖

像素成像又可分為灰階影像(GrayScale)與 RGB 彩色像素(Color Pixel)，灰階影像是以數字 0-255 代表黑色到白色，數字越低越接近深色，如下圖所示。



圖 5-6、灰階像素示意圖

另一方面，RGB 彩色像素顧名思義就是由色彩三原色 Red、Green、Blue 所構成的。通常表示顏色的方式為三原色的數字高低(紅、綠、藍)，數值介於 0-255。例如 (255, 255, 255) 表示白色、(0,0,0) 黑色、(0, 0, 255) 深藍色。

既然我們知道了由三原色就可以組成彩色像素，那麼更進一步推論，一張照片也可以拆解成極端的三原色照片加總。所以只要拿到一張照片，都可以將它想成多維度的矩陣。(包含了照片的寬、高、色彩亮度，或者說像素的欄列個數與色彩維度)

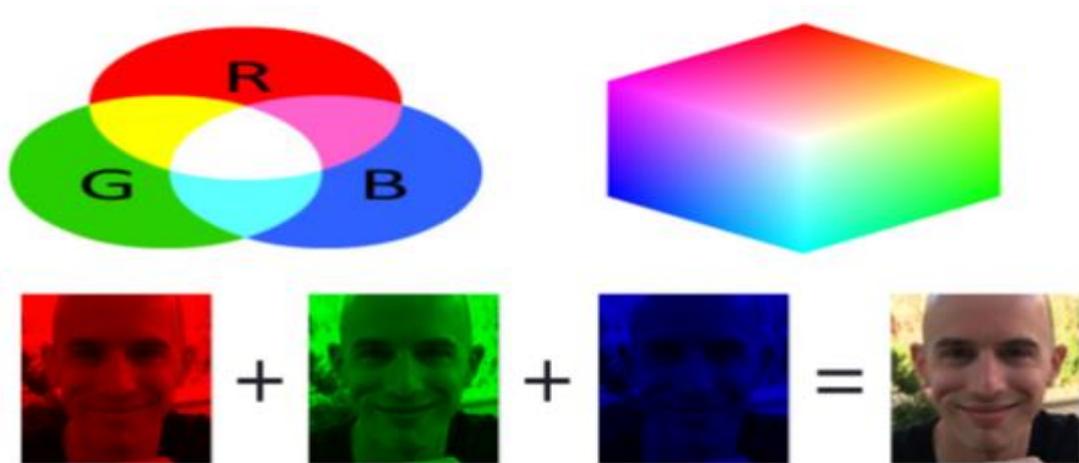


圖 5-7、RGB 色彩像素示意圖

圖像分類技術

介紹完圖片是怎麼構成的，就可以介紹圖像分類技術。圖像分類技術現今可在浪頭上，由於其廣泛的應用而受歡迎。簡單來說即是透過機器學習演算法與電腦視覺，去提取影像的特徵，而結果是可以簡單的標示出一張圖片裡面的物件，或者更進一步產生針對圖片作描述的句子。舉例來說，我們所擁有識別種類集合為{貓、狗、熊貓}，則欲正確標籤一張狗狗的照片，則電腦將會透過機率的方式呈現，如有 95%此照片是狗、4%為貓、1%為熊貓。

數據科學家又被稱為 21 世紀最性感的工作，前陣子熱門的大數據，在人工智慧中也扮演了不可或缺的角色。為了使圖片辨識的準確度提高，所需要的就是大量的圖片，使得電腦知道什麼樣的東西是狗？什麼是貓？柴犬跟秋田有什麼差別？而現今有許多開源的圖片資料庫，如能夠辨識手寫文字的 MNIST、包含飛機、汽車、船等的 CIFAR-10 等，使得圖片辨識技術的進展能夠透過網路串流、合作。

對於一隻貓，我們可以形容牠的毛色為何、毛摸起來的質地、爪子、眼珠的顏色等，但對於電腦來說就是一大堆數字了。那麼該如何使電腦知道貓的毛色為何、毛的質地、眼珠顏色呢？這時候就要採用特徵提取(Feature extraction)，特徵提取是使用演算法，得到一組組特徵向量(Feature vector)，也就是將下圖所示那一大堆數字變成能夠代表貓特徵的向量，試著讓電腦能夠瞭解我們眼中所描述的貓。



151	121	1	93	165	204	14	214	28	235	29	142	142	75	22	100	111	28	6	5
62	67	17	234	27	1	221	37	189	141	137	168	41	206	100	70	219	127	114	191
20	168	155	113	178	228	25	130	139	221	205	154	226	14	89	86	242	67	203	15
236	136	158	230	10	5	165	17	30	155	247	47	128	123	253	229	181	251	232	28
174	148	93	70	95	106	151	10	160	214	68	75	24	99	93	63	215	222	102	180
103	126	58	16	138	136	98	202	42	233	206	246	85	103	215	3	62	64	77	216
235	103	52	37	94	104	173	86	223	113	126	80	165	149	196	75	186	60	179	193
212	15	179	139	48	232	194	46	174	37	44	253	164	253	14	216	175	30	46	254
119	81	241	172	95	170	29	210	22	194	137	23	33	203	241	21	144	63	244	188
129	19	33	253	229	5	152	233	52	44	32	214	142	121	249	109	99	232	183	71
88	200	194	185	140	200	223	190	164	102	45	36	152	27	190	137	61	1	237	247
113	16	220	215	143	104	247	29	97	203	1	14	241	70	2	30	151	67	169	205
9	210	102	246	75	9	158	104	184	129	32	80	102	32	99	169	91	166	73	214
124	52	76	148	249	107	65	216	187	181	186	219	9	203	209	240	40	249	119	122
6	251	52	208	46	65	185	38	77	240	177	252	38	203	119	0	217	139	139	157
150	194	28	206	148	197	208	28	74	93	154	145	49	251	150	185	235	23	230	156
33	183	248	153	168	205	146	100	254	218	157	168	223	60	247	118	5	180	16	206
130	53	128	212	61	226	201	110	140	183	102	208	195	246	140	138	54	191	139	79
165	246	22	102	151	213	40	138	8	93	17	233	85	169	166	24	49	40	160	97
152	251	101	230	23	162	70	238	75	24	84	242	247	144	203	3	19	24	198	88
187	105	152	83	167	98	125	180	136	121	67	67	185	98	123	106	168	105	127	153
139	197	55	209	28	124	208	208	104	40	37	113	214	252	203	80	146	211	7	16
123	19	144	223	62	253	202	108	47	242	142	241	66	86	214	133	146	253	189	200
220	144	31	16	136	123	227	62	183	163	67	215	174	111	189	54	144	56	59	163

圖 5-8、人類看到的 vs 電腦看到的

前面這樣說明，看來圖示辨識好像十分容易。但其實圖像辨識的技術要解決很多挑戰，例如不同角度下所拍的狗狗圖片、大小不同但本質相同的俄羅斯娃娃要如何辨識、人臉被髮型遮住、背景燈光照明亮度等，這些問題都是演算法需要去克服的。

類神經網路(Artificial Neural Networks)是一門涉及認知神經與機器學習的科學，理論架構與生物神經網路相仿。簡單來說，如果是給電腦識別此圖是否為貓咪，那麼 Input 可拆解各項特徵(毛色、四肢、爪等)，再來給各項特徵適當的權重後得到一數值。並且透過類似門檻的 Step Function，將數值轉換成最後的 Output 來作判斷。也就是下圖的一連串過程目的在於得到 Output。(0 或者 1，例如判斷所輸入的圖片 Input 為貓咪為否)

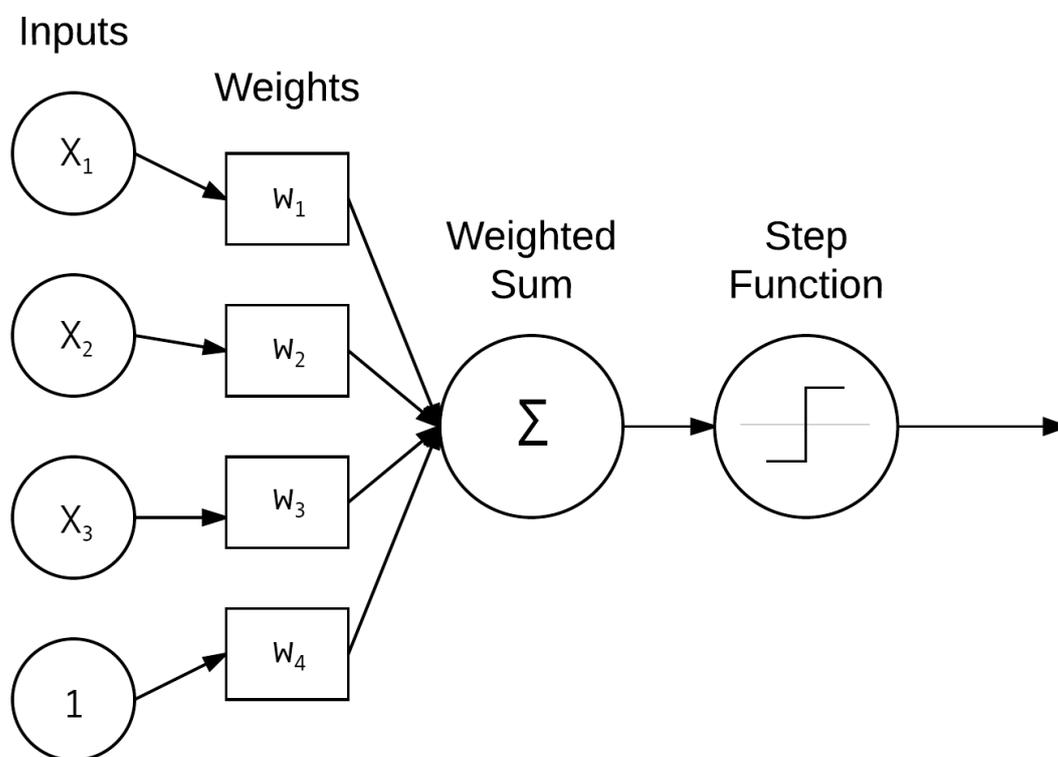


圖 5-9、簡易類神經網路架構

視覺辨識歷史

早期臉部視覺技術還被當作是科幻小說的題材，但現今隨著智慧型手機臉部解鎖等技術商業化應用，對於大眾而言，這項技術早已是稀鬆平常了。但你知道嗎？這項技術最早可追溯於 1950 年代唷。以下將依年代做視覺辨識發展歷程介紹。

1. 1957 年感知機的發明

其實早在 1943 年，人工神經網路(Neural Network，NN)的模型概念就被心理學家們所提出了，也開創了對於神經網路的研究。而直到 1957 年，由美國神經學家 Frank Rosenblatt 提出感知機(Perceptron machine)，此階段的技術僅能進行簡單的辨識，如區分出三角形與正方形的區別。

2. 1959 年 MIT AI 實驗室的設立

在著名的達特茅斯會議後，MIT 設立了 AI 實驗室，研究人工智慧議題。1966 年的 7 月，AI 實驗室共同創辦人 Marvin Minsky 給了他指導學生一個[任務](#)「使相機連結上電腦，並且

使得電腦描述出它透過相機所看到的事物。」當然技術對於 50 年前而言太早了，涉及諸多技術直到今天都還在發展，但是此階段的發展也對於視覺辨識的演進貢獻莫大。

3. 1970 年代商業化應用的出現

光學字元辨識(Optical Character Recognition)是一項對於文字圖像進行辨識處理的技術。1974 年推出了基於 ORC 技術的 Kurzweil Computer 產品，可以辨識各種形式字體，這也是將視覺辨識技術商業化的一個轉折點。

4. 1980 年代 EigenFace 技術

1987 年 Sirovich 與 Kirby 提出特徵臉(Eigenface)進行人臉辨識的技術，該技術被認為是第一種有效的人臉辨識方法，且具有高效率優勢、計算速度非常快，短時間就可以處理大量人臉。但是，特徵臉在應用時有個問題，就是在不同的光照條件和成像角度時，會導致識別率大幅下降。

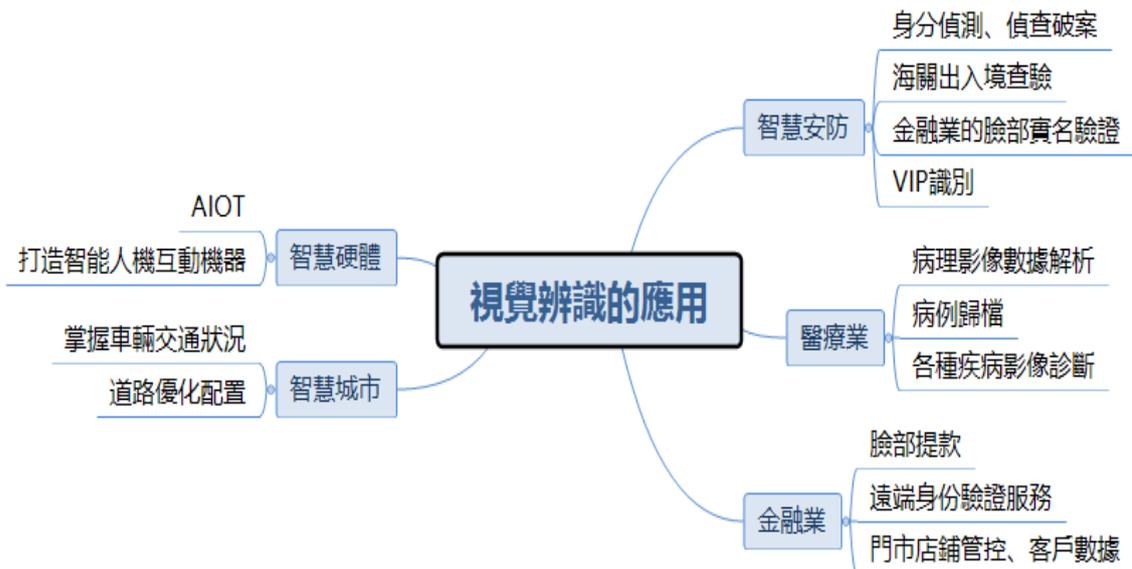
5. 1990-2010 年

此階段視覺辨識技術持續進展，主要是圖像資料庫的設立，如犯罪防治上的應用，但此時技術還未完全成熟。

6. 2010-現在

此階段的應用就為人所知了，如社交平台 FaceBook 上臉部標籤、無人商店識別客戶身分，一直到 iPhoneX 的解鎖等，在現今這個階段視覺辨識的應用已經十分普遍了。

視覺辨識應用



資料來源：Findit 整理繪製

圖 5-10、視覺辨識的應用

視覺辨識的應用最常見的莫過於人臉辨識了，除了在 iPhoneX 上的人臉解鎖外，該技術還可運用在身分證上。例如到中國銀行開戶時，櫃員會核對身分證與本人臉部相符合，或者網路訂票也藉由臉部辨識進行實名制。醫療業上對於疾病的診斷也是一常見的應用，醫生在判斷病人的病歷時，往往會參考 AI 對於該病患所診斷預測的疾病，另外人工智慧對於病歷檔案的整理歸納也能發揮其功效。在安防上，新聞時有所聞通緝犯觀看巨星演唱會而被抓到的例子，或者藉由視覺辨識管理人口出入境，AI 技術套用在治安是具有其效果的。

除了上述較為常見的產業面應用外，下表整理出較為有趣的 AI 視覺辨識案例。

表 5-1、具特色的視覺辨識應用整理

名稱	應用內容描述	網址
犬類分辨	能夠準確辨識秋田與柴犬	https://gcp.expert/automi/
城市車流解決方案	監控車流並彈性調整交通號誌，改善了 62% 的堵車狀況。	https://www.bnext.com.tw/article/51602/ai-smart-city-traffic-fisheye
產學界合作利用 AI 養蝦	AI 技術養蝦，判斷適當的餵食點，提升育成率至七成。	https://www.bnext.com.tw/article/51594/ai-shrimp-sun-yat-sen-university
牛臉辨識技術	美國農場與 Cainthus 合作，能夠迅速的辨認每一隻牛的身分，給予每頭牛專門的照顧、提高生產效率。	https://meet.bnext.com.tw/articles/view/44364?utm_source=dable
魚塭養殖	AI 數據管控水質狀況、減少用電成本。	https://buzzorange.com/techorange/2019/01/24/ai-fishpond/
神像識別	辨識神像的表情，判斷不同神像的年齡與情緒。	https://buzzorange.com/techorange/2019/02/22/find-your-budda/
打麻將	研究者對麻將進行數學和 AI 方面的研究，嘗試回答當前 14 張牌的牌面到底有多好，以及該打出哪一張牌？	https://www.jiqizhixin.com/articles/2019-03-12-2
AI 生成 4k 慢動作影片	在標準影片的基礎上，以令人難以想像的準確度，生成新的「假畫面」，進而變成流暢且清晰的慢動作影片。	https://buzzorange.com/techorange/2018/07/30/nvidia-ai-4k/
威利在哪裡	能夠以 AI 辨識出圖片中的威利，並且透過機械手機在 4.5 秒內找到威利。	https://buzzorange.com/techorange/2018/08/10/lets-ruin-where-is-willy/

生成對抗網路的各種應用	此技術可以應用在創造出人臉、變換圖片場景、改變圖片內容等。	http://bit.ly/2T0O7tg
-------------	-------------------------------	---

臺灣視覺辨識企業介紹

隨著人工智慧技術的進展，催生了不少家以 AI 為技術的新創企業。而臺灣在當中的布局也是少不了的，以下介紹以人工智慧的圖形辨識為主要技術之企業。

表 5-2、臺灣視覺辨識企業整理

企業名稱	成立時間	企業介紹	主要商品/服務	獲投
 Viscovery (創意引晴)	2008	Viscovery 透過自行開發專利算法，自動分析並標籤大量的圖片與影片，將圖片與影片中視覺內容轉化成結構化、多維度、含有高度商業價值 Visual Big Data (視覺大數據)。VisionAI 服務廣泛應用於電子商務、智慧零售、媒體廣告、工業 4.0 等領域。	<ul style="list-style-type: none"> • Visual Recognition - 自動化標示人臉、物件、場景、商標、動作、文字、情緒、抽象概念...等。 • Visual Search- 透過輸入視覺化的關鍵字，精確搜尋圖片和影片畫面中的實際內容。 • VisualBig Data- 透過視覺分析後產生的影音大數據，可加以計算、歸納、總結、和探索影音內容的價值。 • Custom Recognition Model- 為企業開發客製化辨識模組，滿足特定的視覺辨識需求。 	<ol style="list-style-type: none"> 1. 2013 年獲得 Pinehurst Advisors 的 30 萬美元種子資金。 2. 2015 年獲投 A 輪 500 萬美元。 3. 2016 年獲中華開發工業銀行(CDIB) 領投，GD1 Fund II、漢鼎亞太、起點創業營(軟銀中國早期基金)等 1,000 萬美元。
 大猩猩科技公司 (Gorilla Technology Group)	2000	即時影像分析和物聯網技術的全球領導廠商。開發各種以智慧影像分析為中心的内容管理應用，包括智慧零售、智能學校、智慧企業、智能監控和廣播媒體。大猩猩應用機器學習和深度學習算法，實現了數字流處理和全面視頻分析的自動化，擴大商機創造增值服務	<ul style="list-style-type: none"> • 智慧零售-以影片及感測器分析顧客行為及購物動向，並轉換為運營管理的參考資訊，有效提升店內運營績效。 • 智慧企業-結合人臉辨識的員工考勤、訪客管理及教育訓練的雲端管理平台，可有效掌握員工出缺勤狀況及人員進入紀錄，並利用線上課程強化員工的職務技能。 • 智慧監控-藉由視頻分析獲取的資訊進行即時環境監控以及重大事件的調查采證，以有效提升場域安全。 	<ol style="list-style-type: none"> 1. 2015 年獲得澳洲電信 Telstra 的 C 輪投資。 2. 2018 年獲得日本 SBI 集團(軟銀投資前身)的 D 輪 1,500 萬美元

			<ul style="list-style-type: none"> 智慧媒體-藉由數位媒體管理平台所提供的視頻管理及製作流程，提升視頻協同製作的效率以及數位資料的跨區存取。 	
 盛星科技 (Astra)	2016	<p>主要產品為 MSP 平台，結合監視器，以人臉影像辨識技術，協助客戶可以掌握不同時間點的客流數量、不重複訪客與重複訪客人數等資料。</p>	<ul style="list-style-type: none"> Astra 提供大型電信商或服務供應商 MSP 平台，透過電信商的網絡及業務管道，對中小型企業、店家提供影像串流與人臉辨識應用服務，讓店家得已掌握店內不同時間點的客流數量，不重複訪客與重複訪客人數等資料。 	<ol style="list-style-type: none"> 2017 年獲產業局研發補助。 2017 年獲得日本、東南亞地區的創投基金 KK Fund、East Ventures 等投資
 盾心科技 (UMBO CV)	2014	<p>以開發智慧安防為主的人工智慧公司，創立自動監控識別系統，給予智慧性的安全防護，目前已應用在倉庫、零售商店、街道等。</p>	<ul style="list-style-type: none"> 智慧安控攝影機 Umbo SmartDome 雲端平台 Aqua SmartCloud，透過監控畫面學習和辨識影像中的物件與事件，系統可以即時通報負責單位，爭取更多救援時間。 Light 事件偵測-透過錄影時間與影像空間環境中擷取大量資訊進行解讀與分析，即時辨識異常影像與可疑的行為活動，大幅防止犯罪發生。 	<ol style="list-style-type: none"> 2016 年獲得 Appworks 領投，群聯電子、緯創資通以及 Mesh Ventures 的種子輪 9,200 萬台幣。 2017 年獲 CDIB Venture Capital Corp 領投，Substance Capital、AppWorks、Mesh Ventures 跟投 680 萬美元。
 思凱睿克 (SkyREC)	2013	<p>提供「實體零售 Google Analytics」，以雲端服務的形式，讓零售業者能以最客觀的、最迅速的資訊來了解商店狀況，藉此制定最佳的店內行銷策略，以達到最高效的商店營運。</p>	<ul style="list-style-type: none"> 店內外客流量及入店轉換率 業績轉換率(提袋率) 店內熱點區域 員工排班優化 人氣商品分析 顧客來訪頻率 顧客來訪時間 顧客店內動線 促銷成效分析 績效預測分析 	<ol style="list-style-type: none"> 近幾年獲得聯發科技、義美、友達光電種子輪投資，金額未揭露。

資料來源：Findit 整理

結語

提到以人工智慧為商業模式的企業，十之八九都與視覺辨識脫不了關係，可以說是目前應用最廣的技術了。但仔細端看視覺辨識技術之企業，目前仍以辨識圖像為主，也就是使得電腦得以判斷圖片是什麼？並且試著能夠對於圖片進行分類、描述等。但視覺辨識中也有創造出圖片的技術，如近來火熱的創造人臉技術等，可惜的是，其商業應用還不像單純視覺辨識一樣廣泛。但隨著人工智慧議題熱門，相關領域產官學的多方合作，我們仍值得期待未來 AI 所帶來的商業化革命。

劉聖元 (台灣經濟研究院研六所助理研究員)

參考資料

《Deep Learning for Computer Vision》· Dr.Adrian RoseBrock

《China AI development report 2018》

《World AI industry Development Blue Book 2018》

數位時代

科技報橘