

第十六章

深度學習在電腦視覺的應用

內容

- 16.1 前言
- 16.2 深度學習機的基本學習機制
- 16.3 三個著名的深度學習架構
- 16.4 語義分割應用
- 16.5 相機模組辨識應用
- 16.6 結論
- 16.7 作業

16.1 前言

- 介紹深度學習的基本原理、知名的深度學習架構、植基於深度學習的應用：語義分割 (Semantic Segmentation) 和相機模組辨識 (Camera Model Identification)。

16.2 深度學習機的基本學習機制

- 介紹深度學習機器如何利用以 loss function 減少為目的，先按照前向式 (Forward) 修正方式，然後按照反向式 (backward) 修正相關參數權重值方式，重複這種往返學習機制。最終將各個參數的權重值學習起來，並使 loss function 減少到夠小值。

範例1：給定兩個輸入 $I_1 = 0.5$ 和 $I_2 = 0.4$ ；理想輸出 $T_1 = 0.1$ 和 $T_2 = 0.9$ ，共有四個神經元 (Neuron)，神經元內含 sigmoid 激活函數。

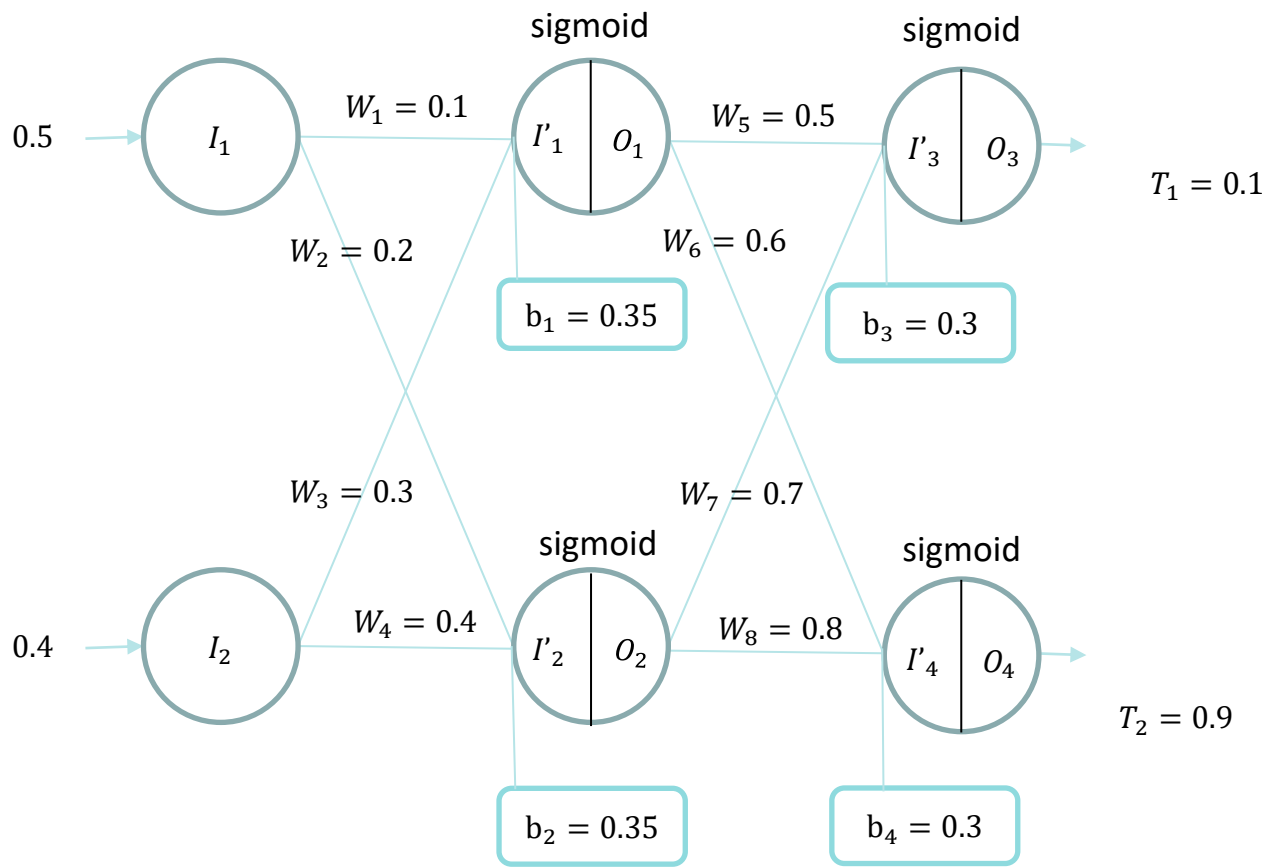


圖 16.2.1 反向傳播學習的模擬例子

- 神經元內含 sigmoid 激活函數 (參見圖 16.2.2) , 函數形式為

$$O_i = \frac{1}{1 + e^{-I_i}}$$

(16.2.1)

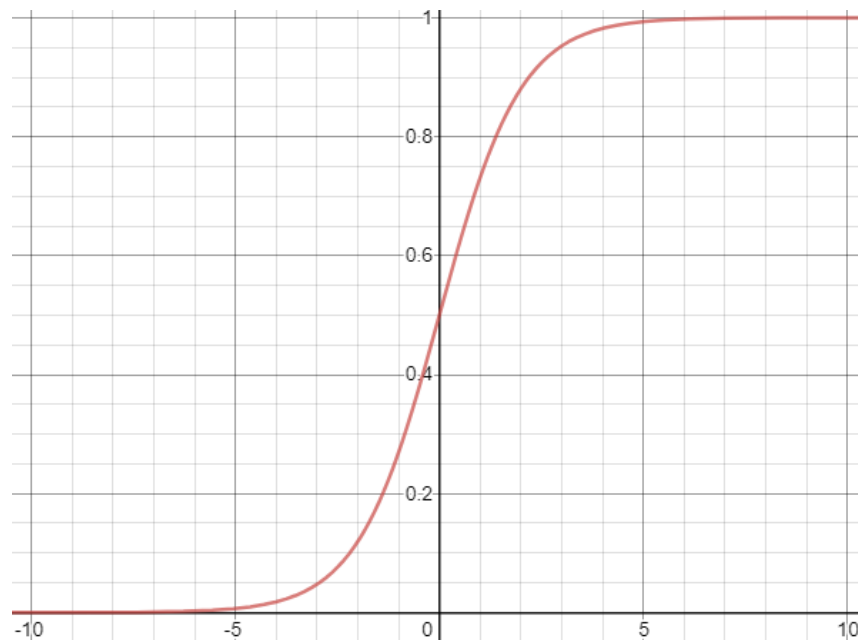


圖 16.2.2 sigmoid 激活函數

- 先模擬正向傳播。第一步結束時，我們得到

$$\begin{aligned} I'_1 &= W_1 * I_1 + W_3 * I_2 + b_1 \\ &= 0.1 * 0.5 + 0.3 * 0.4 + 0.35 \\ &= 0.05 + 0.12 + 0.35 \\ &= 0.52 \end{aligned}$$

(16.2.2)

- I_1 值經過式 (16.2.1) 激化函數作用，得到 O_1 值為

$$\begin{aligned} O_1 &= \frac{1}{1 + e^{-I_1}} \\ &= \frac{1}{1 + 0.59452054797} \\ &= 0.627147766313 \end{aligned}$$

(16.2.3)

同理，可求出 $I_2 = 0.61$ 和 $O_2 = 0.647940802081$ 、
 $I_3 = 1.06713244461$ 、 $O_3 = 0.744051203698$ 、
 $I_4 = 1.19464130145$ 、 $O_4 = 0.767570129409$ 。

- 接下來模擬反向傳播

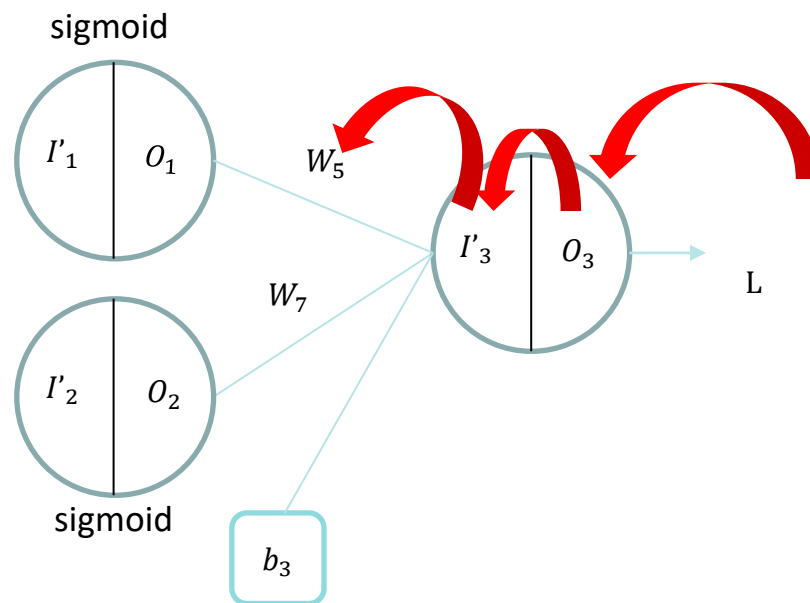


圖 16.2.3 反向傳播示意圖

- 計算總誤差

$$L = \frac{1}{2} (T_1 - O_3)^2 + \frac{1}{2} (T_2 - O_4)^2 \quad (16.2.4)$$

- 以更新 W_5 為例子，得到 $\frac{\partial L}{\partial W_5}$ 值為

$$\frac{\partial L}{\partial W_5} = \frac{\partial L}{\partial O_3} \frac{\partial O_3}{\partial I'_3} \frac{\partial I'_3}{\partial W_5} \quad (16.2.5)$$

- 將公式 (16.2.5) 拆開分別計算

$$\begin{aligned}\frac{\partial L}{\partial O_3} &= 2 * \frac{1}{2} (T_1 - O_3)^1 * -1 + 0 \\ &= 0.644051203698\end{aligned}$$

- 利用作業1的解答，我們得到

$$\begin{aligned}\frac{\partial O_3}{\partial I'_3} &= O_3 * (1 - O_3) \\ &= 0.190439009974\end{aligned}$$

$$\begin{aligned}\frac{\partial I'_3}{\partial W_5} &= 1 * O_1 + 0 + 0 \\ &= 0.627147766313\end{aligned}$$

- 利用以上三個偏微的結果，式 (16.2.5)可改寫為

$$\frac{\partial L}{\partial W_5} = -(T_1 - O_3) * O_3 * (1 - O_3) * O_1$$

(16.2.6)

- 假設 η 代表學習速率且假設 $\eta = 0.4$ ，則 W_5 的權重值更動為

$$\begin{aligned} W_5 &= W_5 - \eta * \frac{\partial L}{\partial W_5} \\ &= 0.469231510058 \end{aligned}$$

(16.2.7)

16.3 三個著名的深度學習架構

- CNN中的基本運算元件
- 卷積層 (Convolutional Layer: CONV) 與池化層 (Pooling Layer: POOL)：影像特徵 (Feature) 的提取
- 全連接層 (Fully Connected Layer: FC)：把先前提取的特徵去做分類。

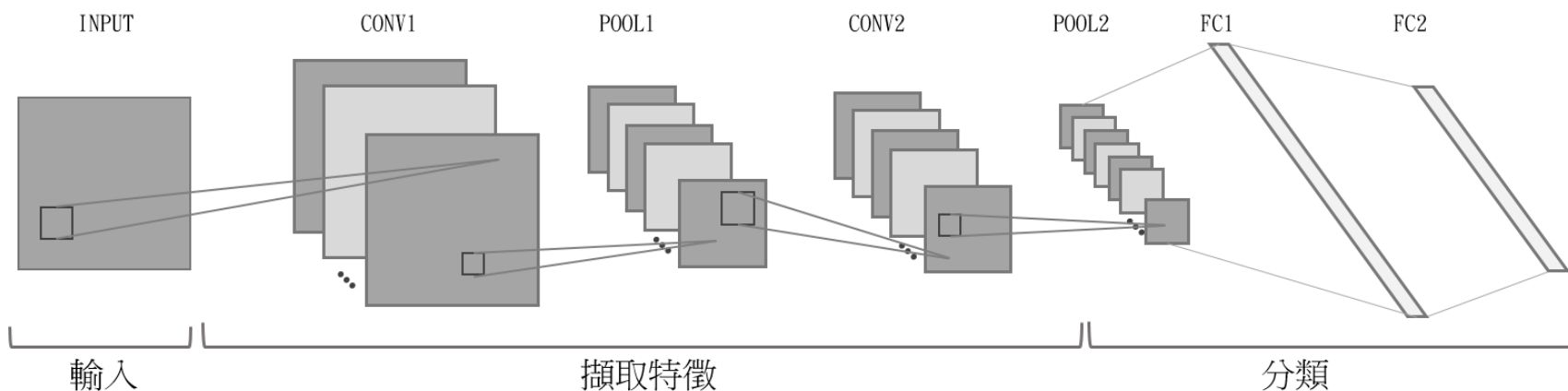


圖 16.3.1 卷積神經網路結構圖

■ 範例2：卷積的運算方式

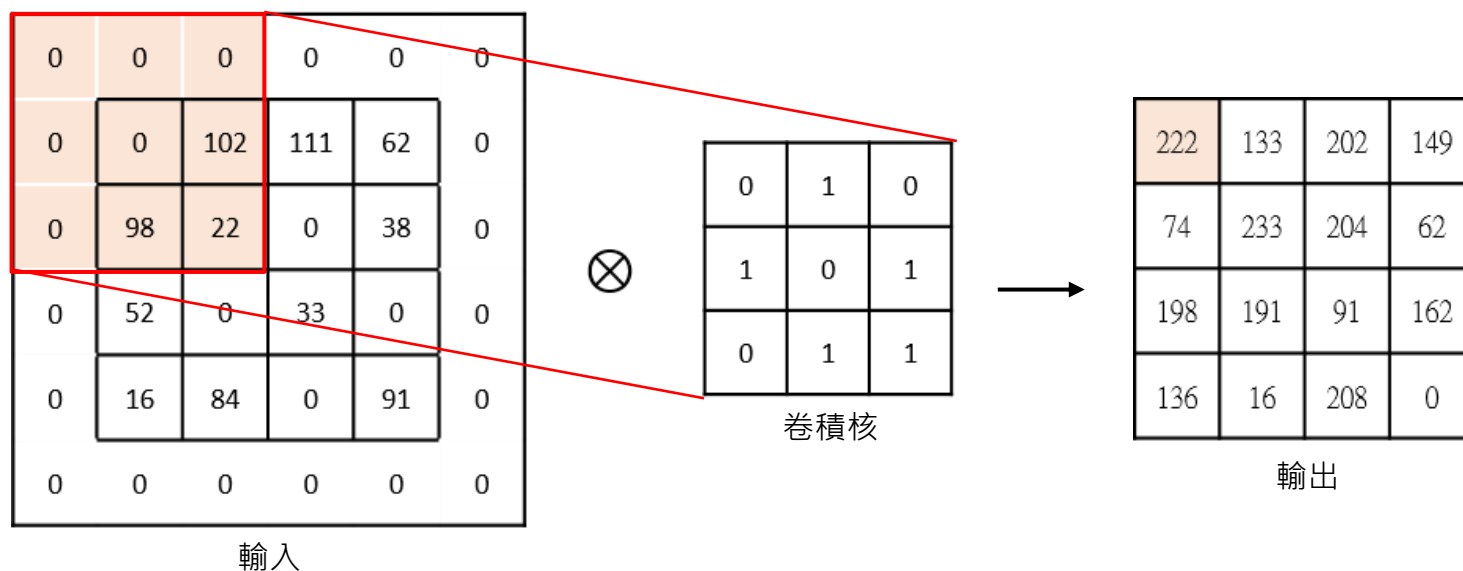


圖 16.3.2 卷積運算的範例

- 響應值222 (response) 的計算過程為
$$(0 * 0) + (0 * 1) + (0 * 0) +$$
$$(0 * 1) + (0 * 0) + (102 * 1) +$$
$$(0 * 0) + (98 * 1) + (22 * 1) = 222$$

(16.3.1)

依此類推，共得 16 個響應值 (也稱特徵值)，這些響應值構成輸出的特徵圖 (Feature Map)。

- 池化層用於保留重要特徵，且把特徵圖縮小至原本的四分之一，達到降低卷積網路運算量的效果。

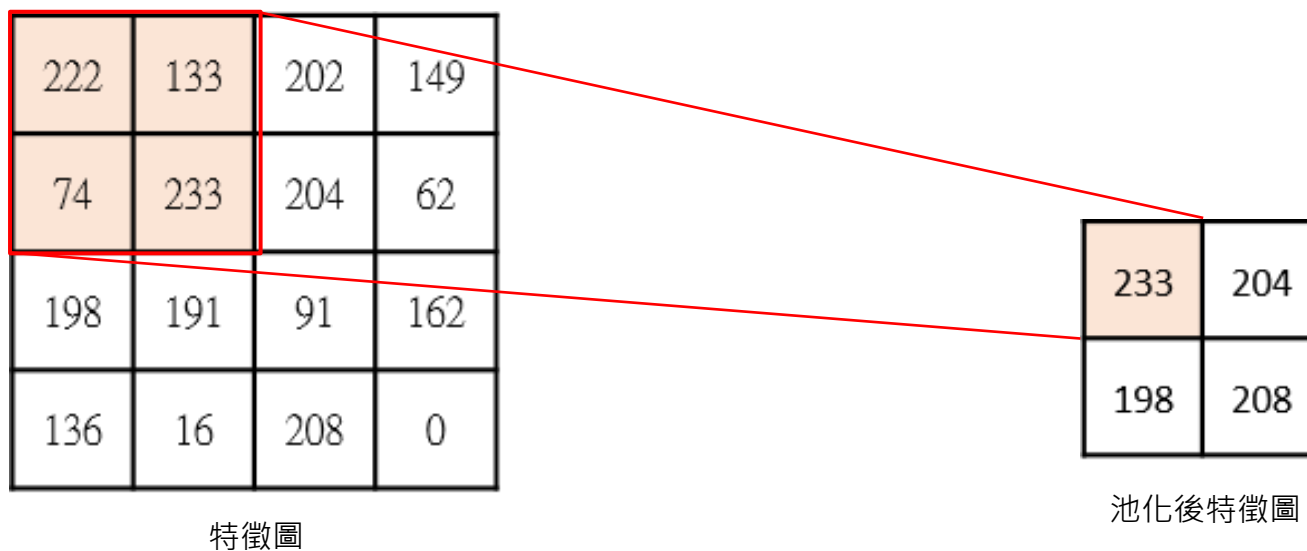


圖 16.3.3 池化運算

- 全連接層會把池化層運算完的結果平坦化 (Flatten) 後，連上全連接層神經網路。

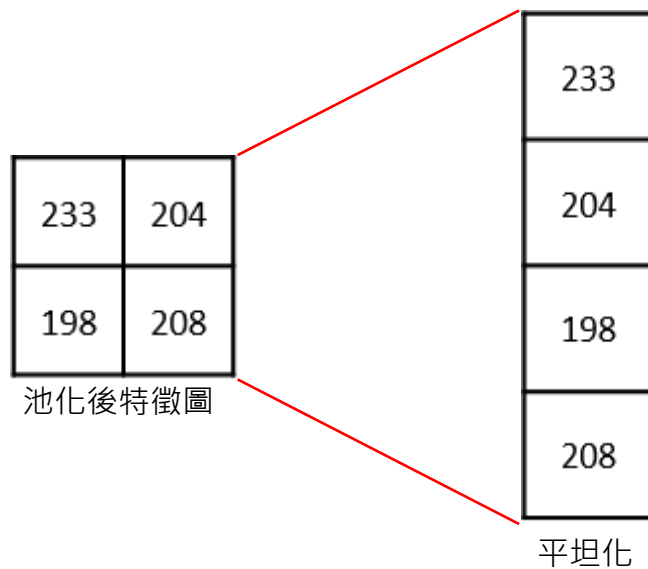


圖 16.3.4 平坦化

- AlexNet：由 Alex Krizhevsky 所提出，該架構在2012年的 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 比賽中取得了第一名。

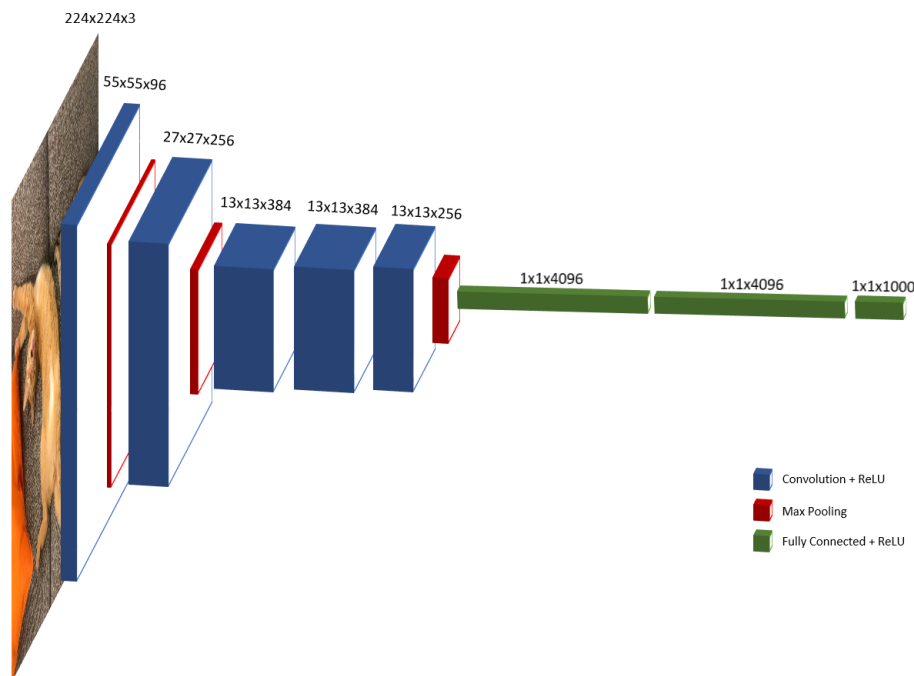


圖 16.3.5 AlexNet架構圖

- VGG-16 (Visual Geometry Group-16)：由 Karen Simonyan & Andrew Zisserman 所提出，該架構在2014年的 ILSVRC 比賽中拿到了第二名。

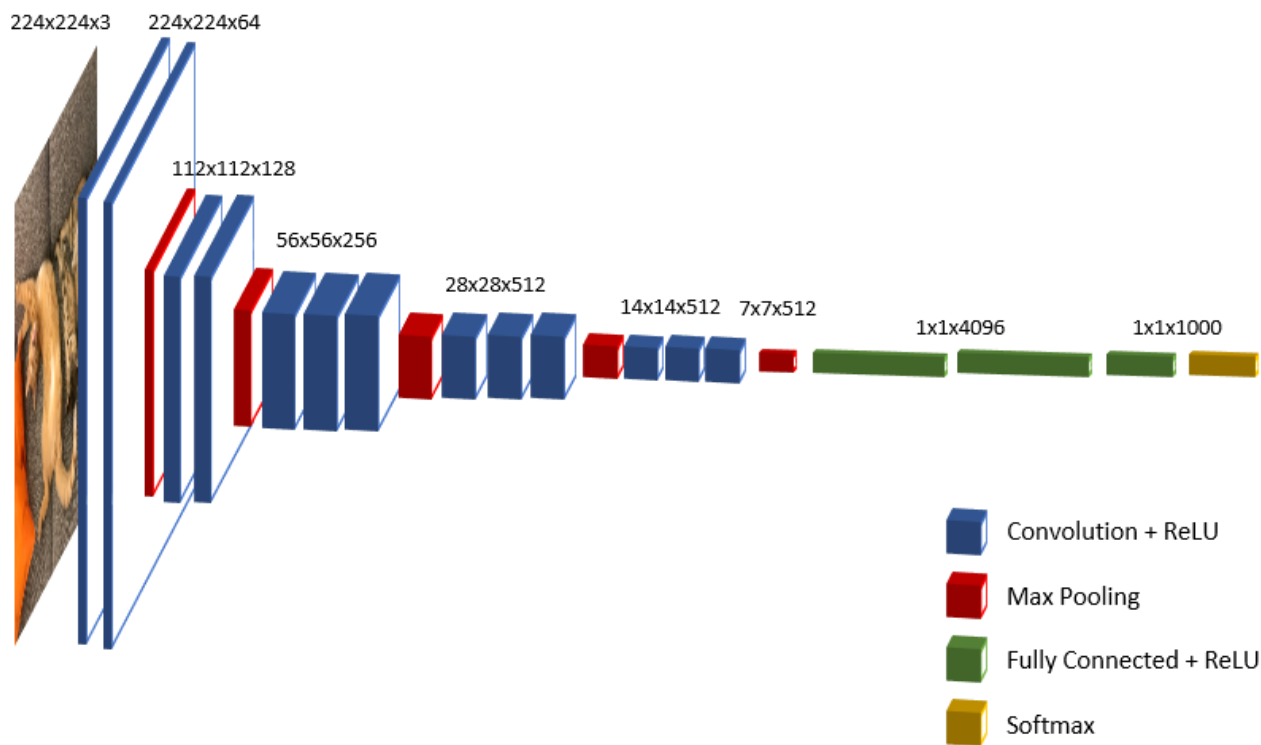


圖 16.3.6 VGG-16架構圖

- FCN：由 Jonathan Long 等人所提出，最初是用於語義分割 (Semantic Segmentation) 的應用。

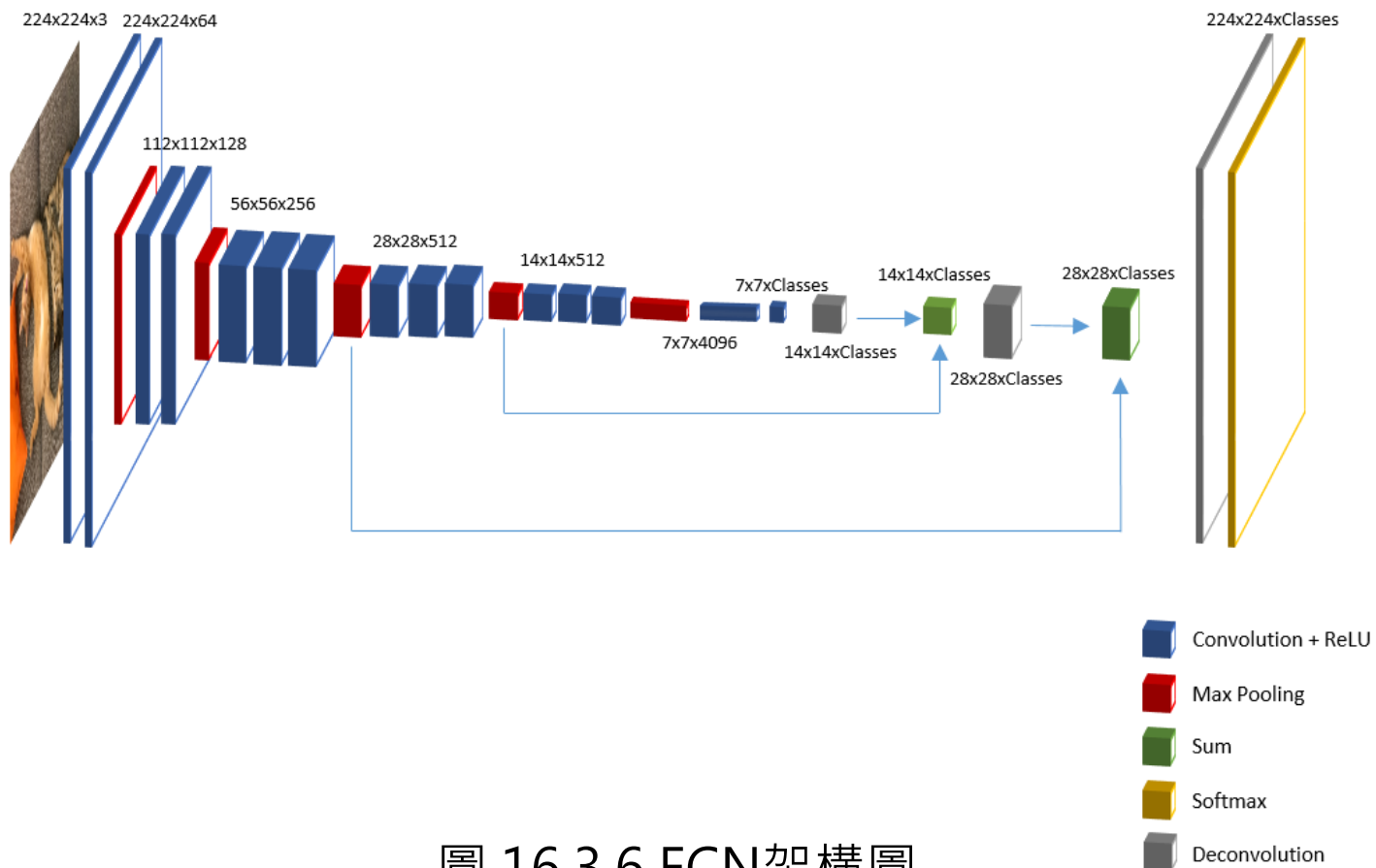


圖 16.3.6 FCN架構圖

16.4 語義分割應用

- 語義分割 (Semantic Segmentation) 在電腦視覺與影像處理中，主要是將圖像中的不同物件予以切割出來，對許多的應用是相當重要的工作。
- 兩種常用的訓練資料集：CamVid Dataset、GTA5 Dataset。



圖 16.4.1 CamVid圖像標記示意圖



圖 16.4.2 GTA5 圖像標記示意圖

- 適應場景 (Domain Adaptation)：真實資料集與合成資料集之間的落差，會將GTA5 Dataset 的合成圖像使用風格轉換 (Style Transfer) 轉換為類似真實世界風格的合成圖像。



圖 16.4.3 GTA5風格轉換示意圖

16.5 相機模組辨識應用

介紹如何在給定的全彩影像中，辨識出其相機模組。相機模組的識別一直是相當具有挑戰性的研究議題，該議題於身分驗證，以及相片竄改的偵測有很重要的應用。

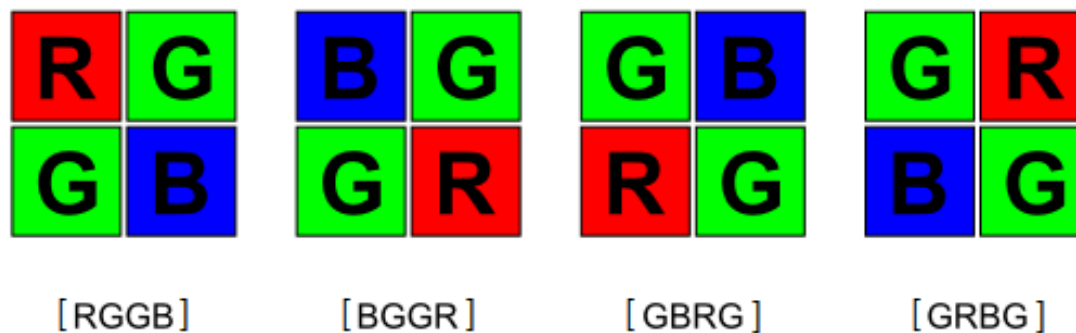


圖 16.5.1 Bayer CFA之四個模組

- Bondi 等人於2015年提出了以深度學習為主的相機模組識別演算法，其中納入識別的主流相機品牌有18個相機模組。

品牌	型號
CANON	Ixus70
Casio	EX-Z150
Fuji	FinePixJ50
Kodak	M1063
Nikon	CoolPixS710
Nikon	D200
Nikon	D70
Olympus	mju-1050SW
Panasonic	DMC-FZ50
Pentax	OptioA40
Praktica	DCZ5.9
Ricoh Caplio	GX100
Rollei	RCP-7325XS
Samsung	L74wide
Samsung	NV15
SONY	DSC-H50
SONY	DSC-T77
SONY	DSC-W170

表 16.5.2 相機品牌與其對應型號

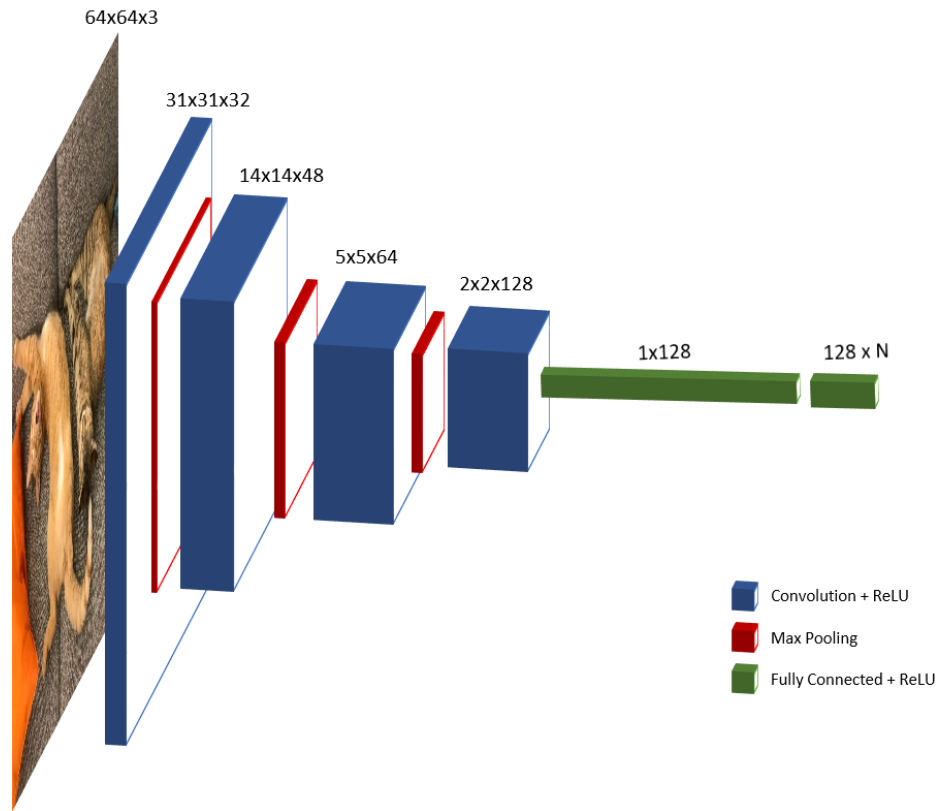


圖 16.5.3 相機模組識別演算法使用的深度學習架構