Pretrained Model and Transfer Learning

預訓練模型與轉移學習

接下來，我們會下載網路上預先訓練的模型（它們的訓練都是以龐大的資料集為基礎），並學習如何利用相關領域專家的研究成果。我們可以把這些預先訓練的模型看成是一段定義好的程式，這個程式能讓使用者輸入資料，並傳回處理完的輸出結果。該程式的特性是由模型的**架構**和**訓練資料集**所決定的，而模型已用大量的**輸入 - 輸出對**（input-output pairs）來訓練過，並具有固定的輸出資料格式。現在我們只要使用這個現成的模型，便可以快速實作一個深度學習應用，省去自己訓練模型所需的時間。

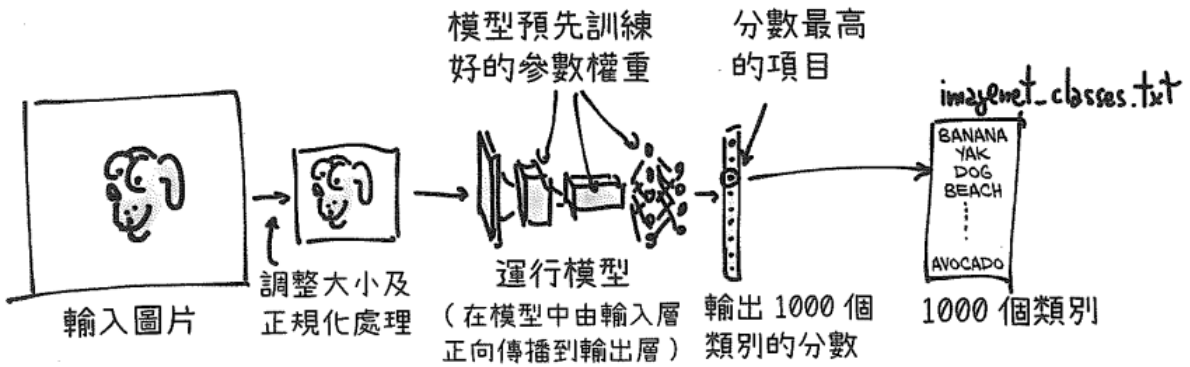
利用預先訓練的模型來辨識圖像中的物體

要運行一個『辨識物體』的模型。當研究人員在發表論文時，通常也會在網路上分享模型的程式碼。很多時候，程式碼中還會附帶訓練過的模型參數值，這些參數值是通過無數次的訓練得到的。透過使用這些模型，你可以省去很多訓練參數的時間。

我們現在要探索的模型已經在 ImageNet（http://imagenet.stanford.edu）的其中一個子資料集上訓練過很多次了。ImageNet 是一個非常龐大的資料集，裡面有超過 1400 萬張圖片，由史丹佛大學負責維護。該資料集中的所有圖片都由人工標籤過

ImageNet 和其他幾個公開資料集一樣，都起源於學術競賽。學術競賽向來是企業或研究單位的研究者們互相切磋的場合。其中針對 ImageNet 而舉辦的大型視覺辨識挑戰賽（ILSVRC），從 2010 年成立開始就受到廣大歡迎。這項比賽每年的任務都不同，如**圖片分類**（能將圖片分類到其所屬的類別）、**物體定位**（識別物體在圖片中的位置）、**物體辨識**（識別和標記圖片中的物體）、**場景分類**（對圖片中的場景進行分類）和**場景解析**（將圖片分割成不同語義類別的區塊，如牛、房子、乳酪、帽子）等。其中，圖片分類任務是從 1000 個類別中選出用來描述輸入圖片的標籤。輸出的結果按照**信心程度**排序，可以列出 **5 個**模型認為最符合圖片內容的標籤。

我們要把圖片輸入這個預先訓練的模型中。模型會對每張圖片生成對應的**預測標籤清單**（內含按照分數從高至低進行排列的標籤）。接下來便可以透過這份清單，檢查模型的預測結果。我們使用的模型可以精準地預測出大部分圖片的類別。



神經網路模型在圖片辨識中的**推論** (inference) 過程。

預訓練模型（Pretrained Model）是一種在大規模數據集上預先訓練的模型，這些模型通常會用於後續的特定任務中（例如通過遷移學習）。使用預訓練模型的好處是可以避免從零開始訓練，尤其當數據有限時，它可以顯著提升效果。

以下通過數學公式描述預訓練模型的概念及其應用：

1. 預訓練階段

首先，預訓練模型是在一個大的數據集 $D_{pre} = \{(x_{pre}, y_{pre})\}$ 上進行訓練，這裡的任務是源任務 $\mathcal{T}_{pre}$。我們的目標是學習一個模型 $f(x; \theta_{pre})$，並最小化損失函數：

預訓練資料

預訓練模型 $$\theta^*_{pre} = \arg\min_{\theta_{pre}} \frac{1}{N_{pre}} \sum_{i=1}^{N_{pre}} L(y_{pre_i}, f(x_{pre_i}; \theta_{pre}))$$

這個過程訓練出的模型參數 θ^*_{pre} 是基於源任務的最佳參數，它能夠很好地表徵數據的特徵。

 $x^* = \arg\min_x f(x)$ 找出 x^*，使得 $f(x)$ 在所有可能的 x 值中最小　argmin：函數最小值引數

2. 微調（Fine-tuning）階段

在使用預訓練模型時，我們將這個模型應用於目標任務 $\mathcal{T}_{target}$，其數據集為 $D_{target} = \{(x_{target}, y_{target})\}$。通常做法是使用預訓練模型 $f(x; \theta^*_{pre})$ 作為初始化，然後對目標任務進行微調。這可以通過下面的優化問題來表示：

$$\text{微調模型 } \theta^*_{target} = \arg\min_{\theta_{target}} \frac{1}{N_{target}} \sum_{i=1}^{N_{target}} L(y_{target_i}, f(x_{target_i}; \theta_{target}))$$

目標資料

此處，θ_{target} 的初始值是預訓練模型的參數 θ^*_{pre}，而不是從隨機初始化開始。這種微調的過程能夠在目標任務上快速收斂並提高性能。

3. 凍結部分層數

有時，我們會選擇凍結預訓練模型的某些層，只對部分層進行微調。例如，如果模型由多層組成，我們可以只微調最後幾層 θ_g，而保持前幾層的參數 θ_h 不變：

$$f(x;\theta) = g(h(x;\theta_h);\theta_g)$$

其中，θ_h 是在預訓練中學到的特徵提取層參數，而 θ_g 則是決策層的參數。在這種情況下，優化問題變為：

$$\theta_g^* = \arg\min_{\theta_g} \frac{1}{N_{target}} \sum_{i=1}^{N_{target}} L(y_{target_i}, f(x_{target_i}; \theta_h^*, \theta_g))$$

此時，θ_h^* 是固定的，只有 θ_g 參數在微調過程中被更新。

4. 特徵提取方法

另一種使用預訓練模型的方式是將它作為固定的特徵提取器，這種情況下我們不會更新預訓練模型的參數，只是利用其提取出的特徵作為輸入進行目標任務的訓練。這樣的特徵提取可以表示為：

$$z = h(x; \theta^*_{pre})$$

然後，使用這些特徵 z 作為新的輸入來訓練一個簡單的分類器或回歸器 $g(z; \phi)$：

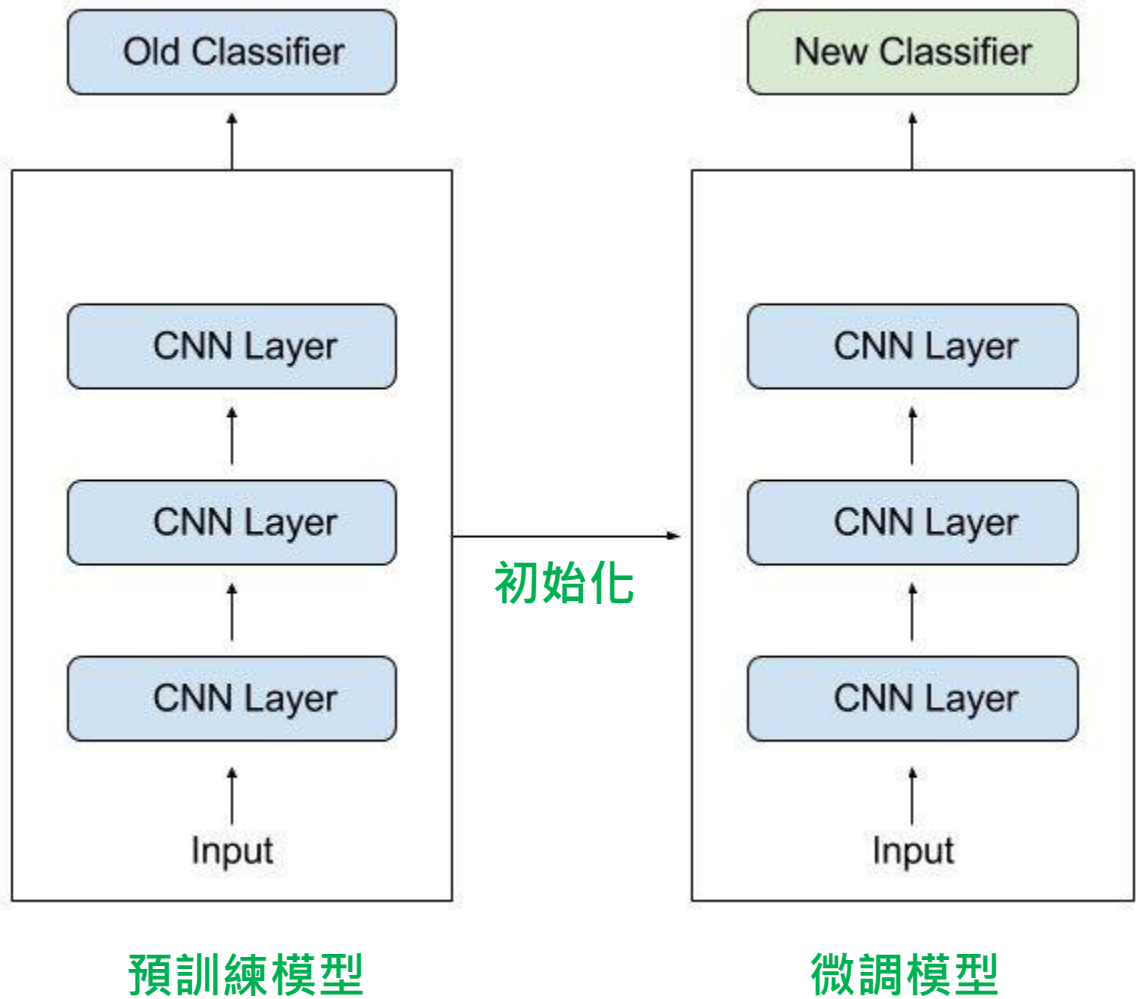
$$\phi^* = \arg\min_{\phi} \frac{1}{N_{target}} \sum_{i=1}^{N_{target}} L(y_{target_i}, g(z_i; \phi))$$

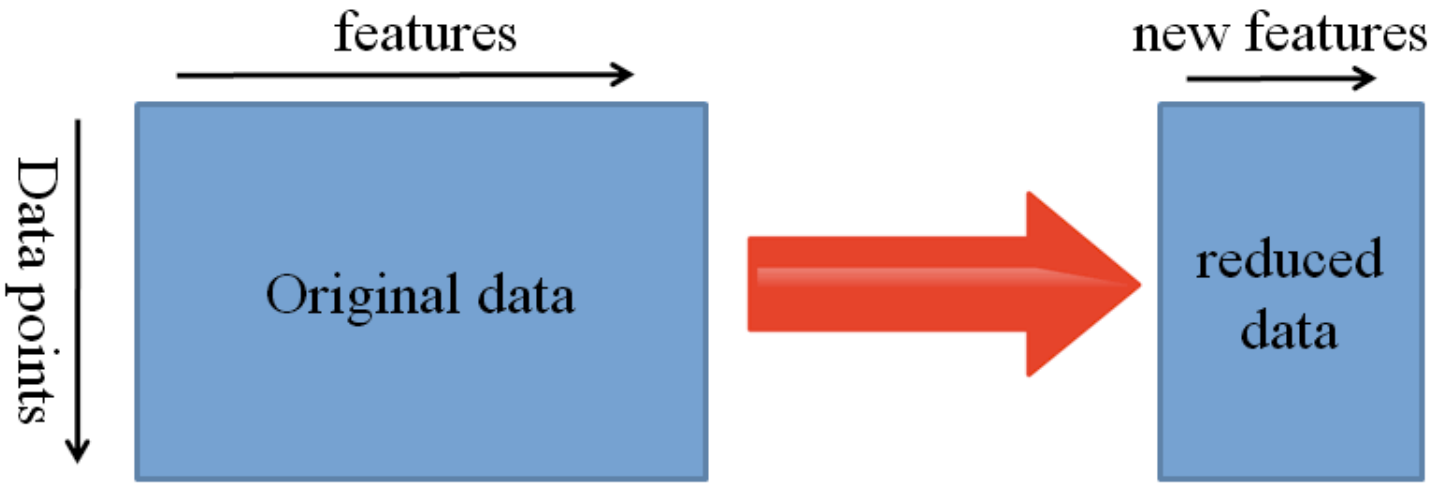
此處，預訓練模型的參數 θ^*_{pre} 是固定的，只有分類器 $g(\cdot; \phi)$ 的參數 ϕ 會被更新。

5. 總結

預訓練模型的使用可分為兩種主要方式：

- **微調（Fine-tuning）**：在目標任務上進行模型參數的更新。
- **特徵提取（Feature Extraction）**：固定預訓練模型的參數，並利用提取出的特徵來進行下游任務的訓練。


Old Classifier
CNN Layer
CNN Layer
CNN Layer
Input
初始化
New Classifier
CNN Layer
CNN Layer
CNN Layer
Input
預訓練模型
微調模型


features
new features
Data points
Original data
reduced
data

提取特徵

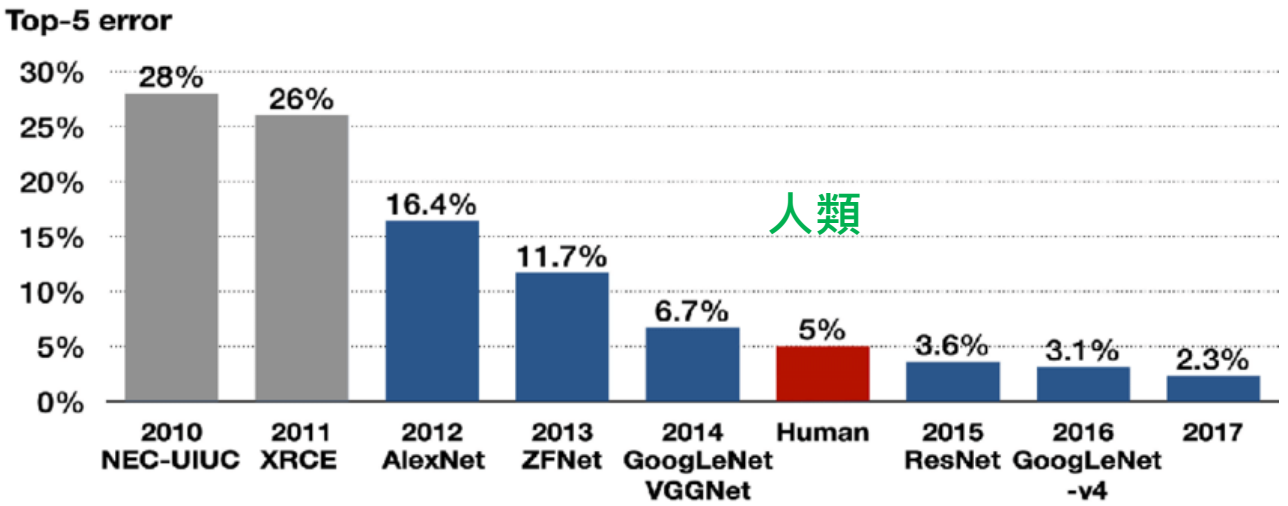
ImageNet專案是一個大型視覺資料庫，用於視覺目標辨識軟體研究。該專案已手動注釋了1400多萬張圖像，以指出圖片中的對象，並在至少100萬張圖像中提供了邊框。ImageNet包含2萬多個典型類別，例如「氣球」或「草莓」，每一類包含數百張圖像。儘管實際圖像不歸ImageNet所有，但可以直接從ImageNet免費獲得標註的第三方圖像URL。2010年以來，ImageNet專案每年舉辦一次軟體競賽，即ImageNet大規模視覺辨識挑戰賽(ILSVRC)。挑戰賽使用1000個「整理」後的非重疊類，軟體程式比賽正確分類和檢測目標及場景。

IMAGENET

14,197

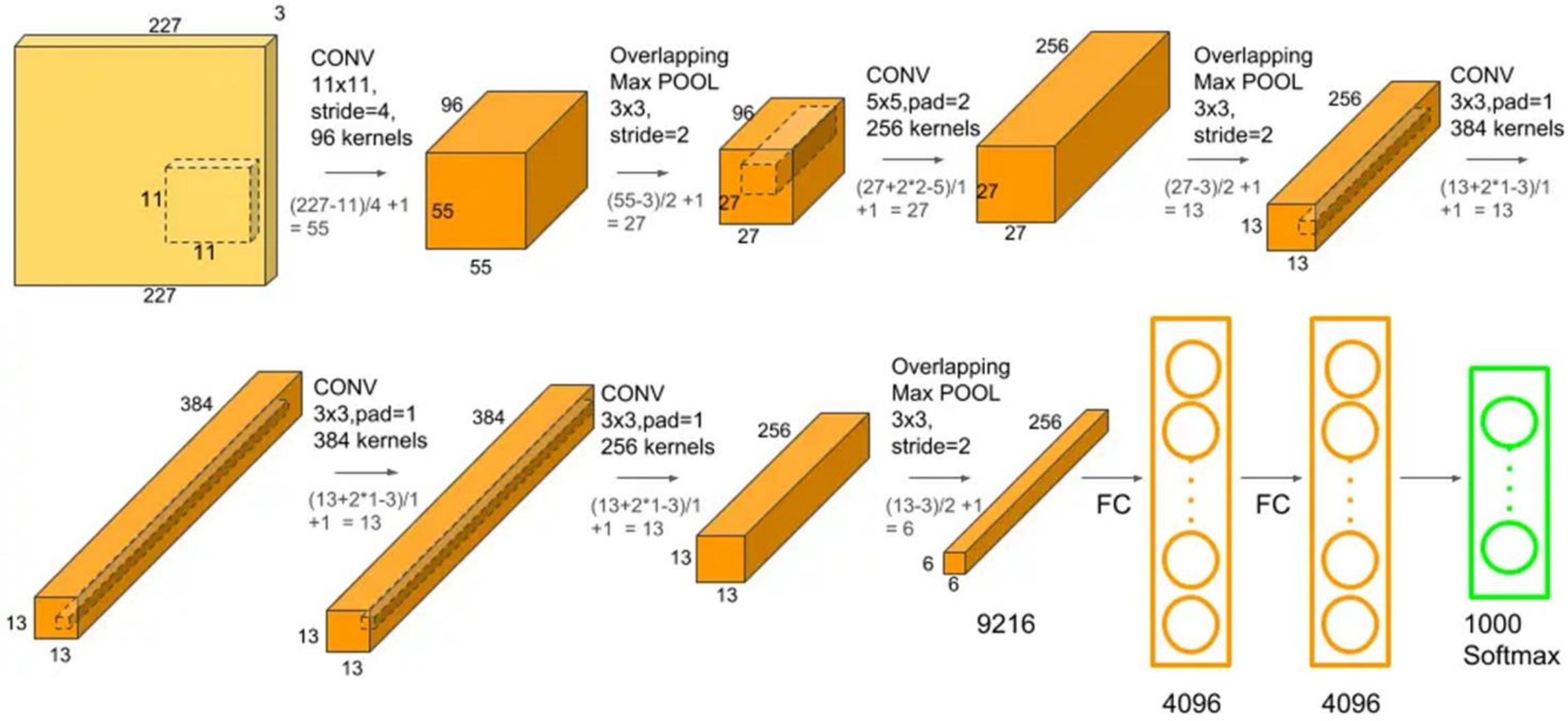
Home Download Challenges About

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)



比較項目	ImageNet	CIFAR-10
資料集大小	超過 1,400 萬張圖像	60,000 張圖像
圖像解析度	不固定，通常較高解析度 (224x224 或更大)	固定為 32x32 像素
圖像色彩	RGB 彩色圖像	RGB 彩色圖像
類別數量	1,000 類別（或更高，依據子集而定）	10 個類別
每類別圖像數量	每個類別約數千張圖像	每個類別約 6,000 張圖像
圖像來源	來自網路，包含各種自然、合成和物品圖像	日常生活中的物體圖像
類別內容	動物、物體、場景等廣泛類別	10 個基本類別：飛機、汽車、鳥、貓、鹿、狗、青蛙、馬、船、卡車
資料集應用	用於大規模的圖像分類、檢測、分割等任務	用於圖像分類任務
訓練與測試比例	訓練集包含約 1,280 萬張圖像，測試集約 5 萬張圖像	訓練集 50,000 張，測試集 10,000 張
模型訓練難度	訓練難度高，計算資源需求大	訓練相對簡單，適合快速驗證模型
常見模型	ResNet、VGG、Inception、EfficientNet	LeNet、VGG、ResNet

卷積神經網路之張量形狀(資料結構)變化



VGG（Visual Geometry Group）是一種深度卷積神經網路架構，最初由牛津大學的Visual Geometry Group提出，主要用於圖像分類任務。VGG 網路以其使用固定大小的 3x3 卷積核和層數加深為特點，並且在 ImageNet 圖像分類挑戰中取得了優秀的表現。

VGG 的特點：

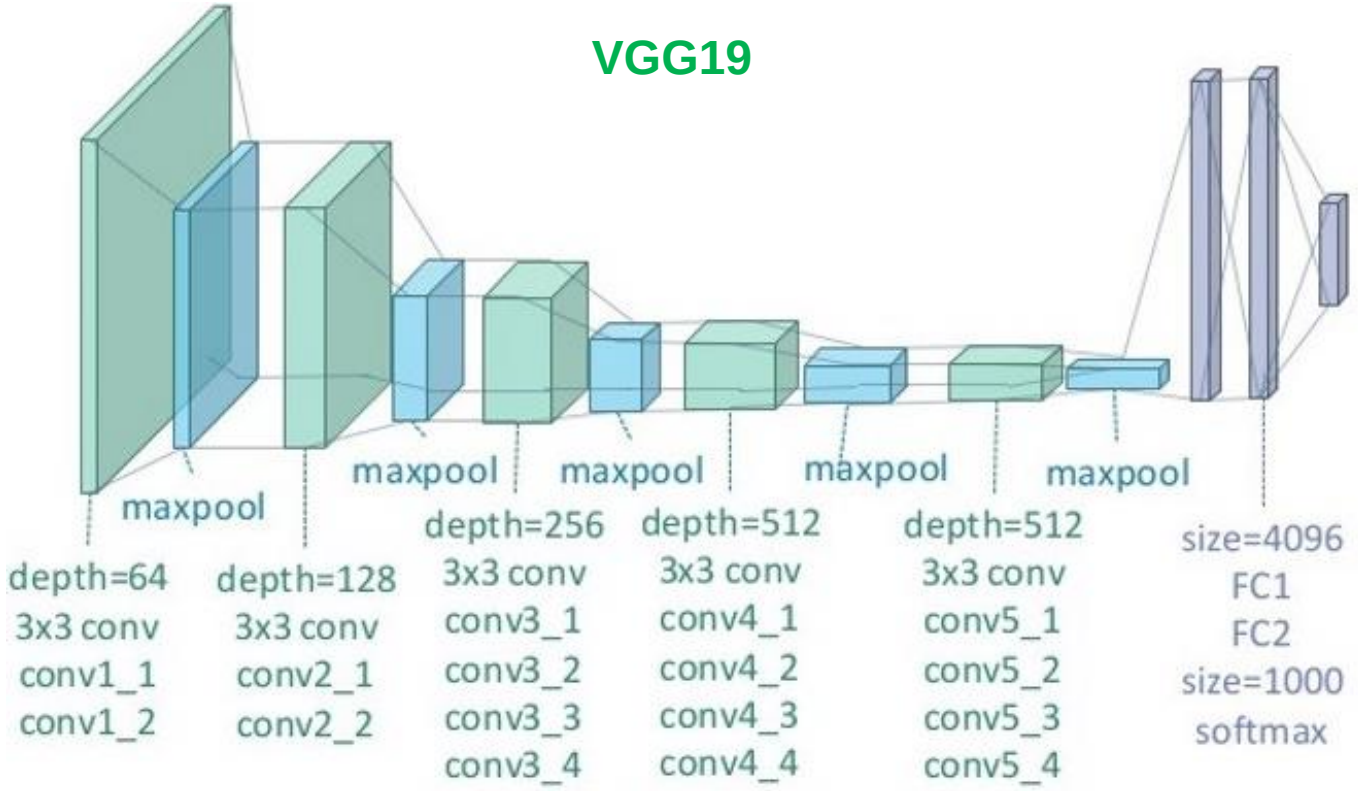
1. **深層網路**：VGG 網路通常有 16 層或 19 層深度（如 VGG16、VGG19），網路越深，學習到的特徵越豐富。
2. **小卷積核**：所有卷積層均使用 3x3 大小的卷積核，這使得網路能夠捕捉更細緻的局部特徵。
3. **池化層**：使用 2x2 的最大池化層來減少特徵圖的尺寸，從而在保持信息的同時降低計算成本。
4. **全連接層**：最後使用多個全連接層，並接上一個 softmax 層進行分類。

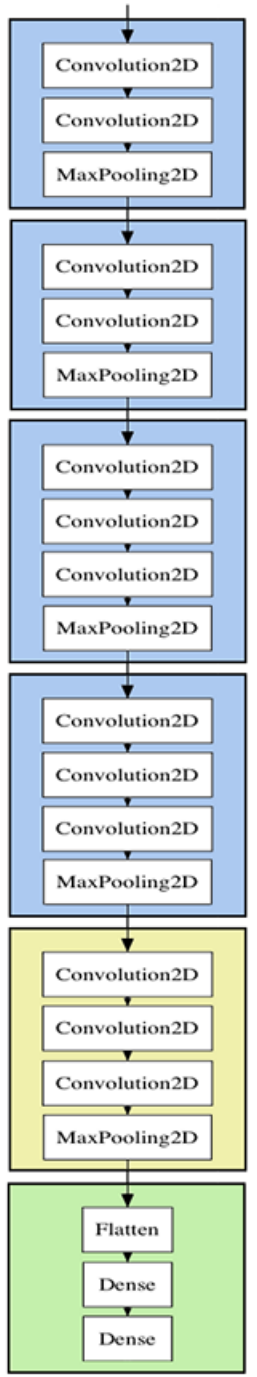
關於VGG

由於，這些模型使用了大量資料作訓練，且使用非常多層的處理，例如VGG使用ImageNet100萬張圖片，共1000種類別，幾乎涵蓋日常生活看到的事物，例如動物、交通工具...等，訓練出來的模型，就變成一種『通用解決方案』(Generic Solution)，如果要辨識照片內事物屬於這1000類，例如貓、狗、大象等，就可以直接拿VGG模型來用了，反之，如果，要辨識的內容不屬於1000類，也可以換掉input卷積層，只利用中間層萃取的特徵，這種能力稱為『Transfer Learning』

VGG16/VGG19 模型結構

VGG 是英國牛津大學Visual Geometry Group的縮寫，主要貢獻是使用更多的隱藏層，大量的圖片訓練，提高準確率至90%。VGG16/VGG19分別為16層(13個卷積層及3個全連接層)與19層(16個卷積層及3個全連接層)，結構圖如右。





Conv block 1: frozen 凍結

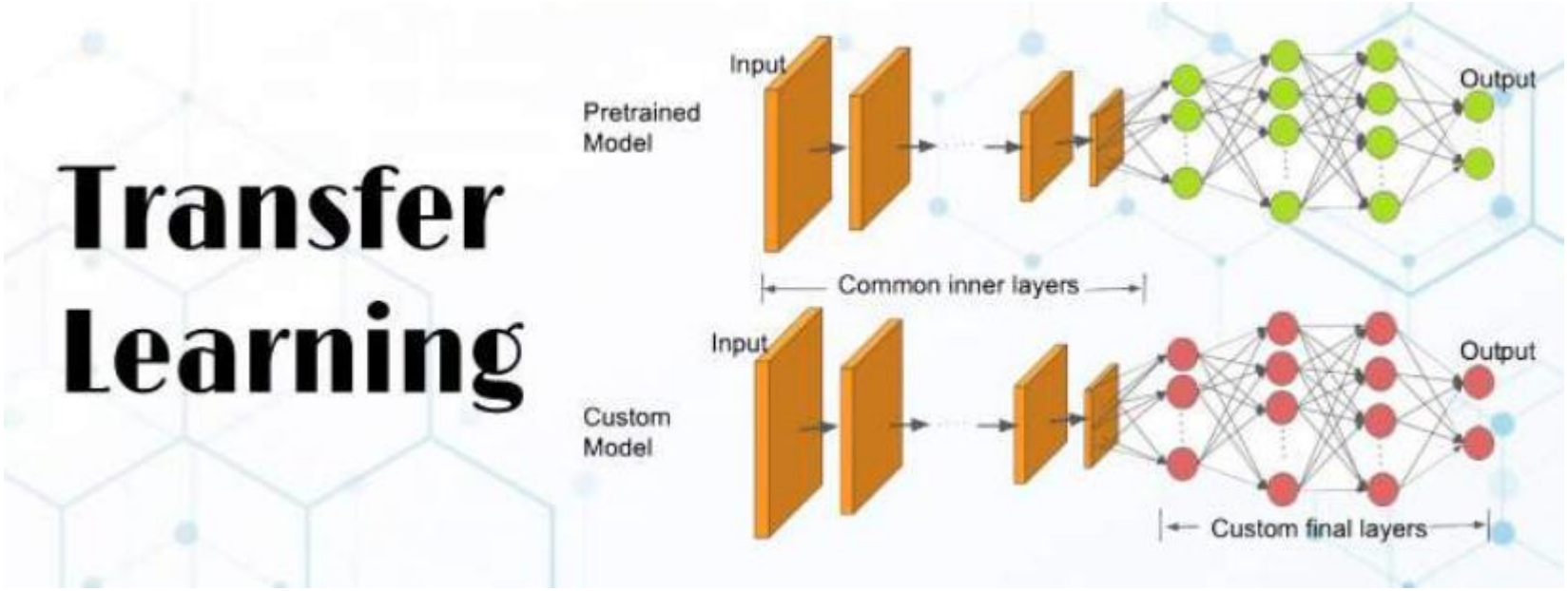
Conv block 2: frozen 凍結

Conv block 3: frozen 凍結

Conv block 4: frozen 凍結

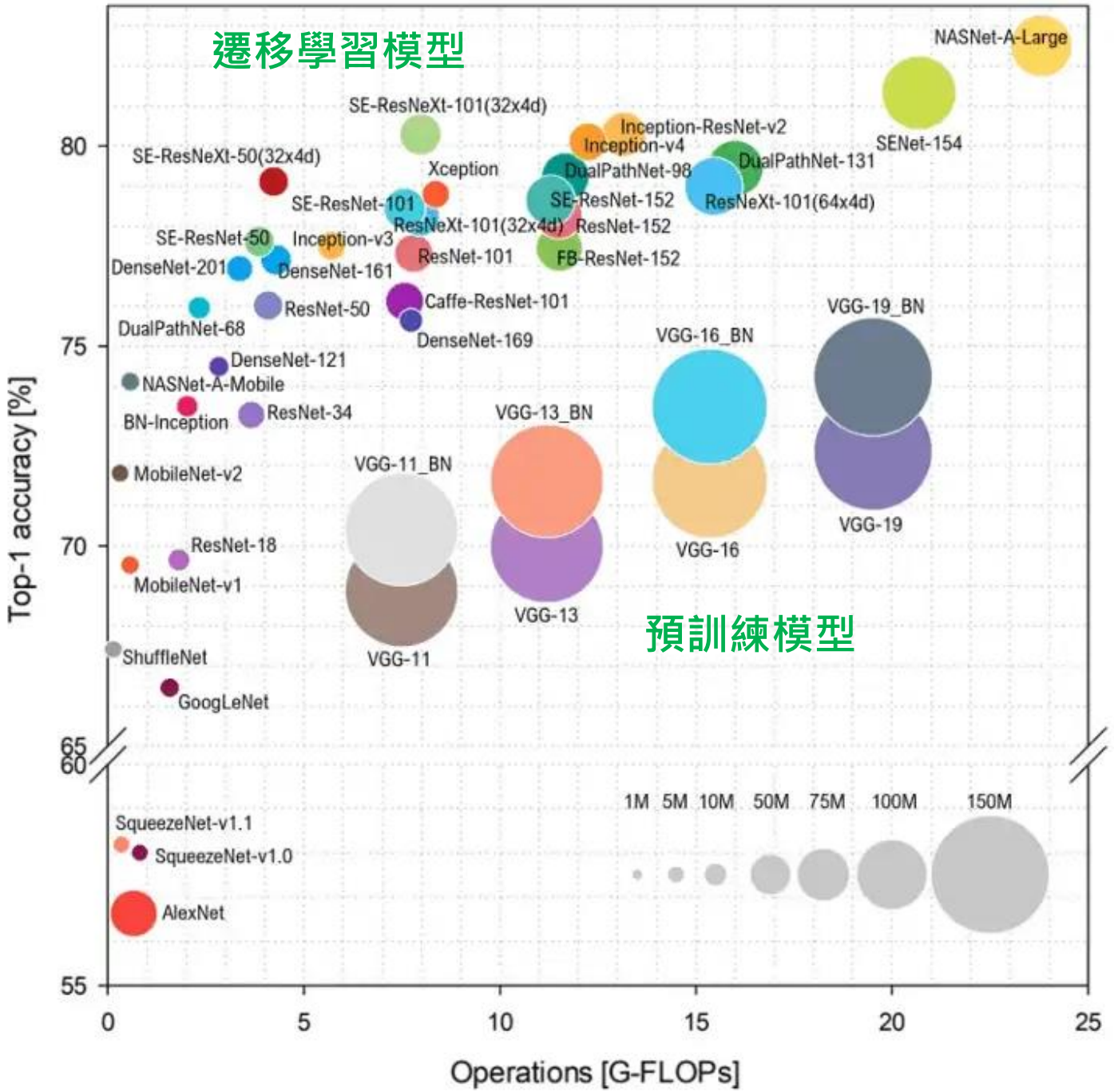
We fine-tune Conv block 5 微調

We fine-tune our own fully-connected classifier 微調



轉移學習就是把已經訓練好的模型、參數，轉移至另外的一個新模型上使得我們不需要從零開始，重新訓練一個新模型。是用於解決數據標記困難、數據取得不易等問題的重要解決手段。

正確率



指數(次方)

exponent

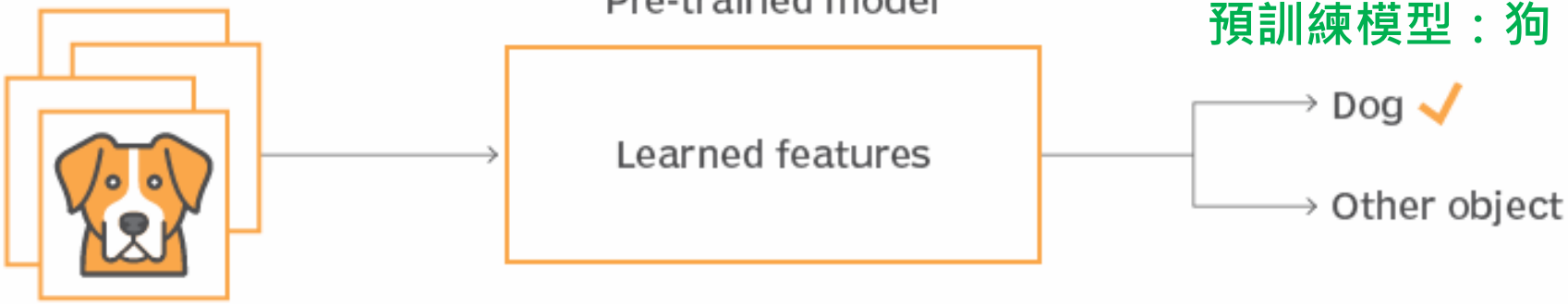
$$1.2345 = \underbrace{12345}_{\text{mantissa}} \times 10^{\overbrace{-4}^{\text{exponent}}}$$

尾數(精確度)

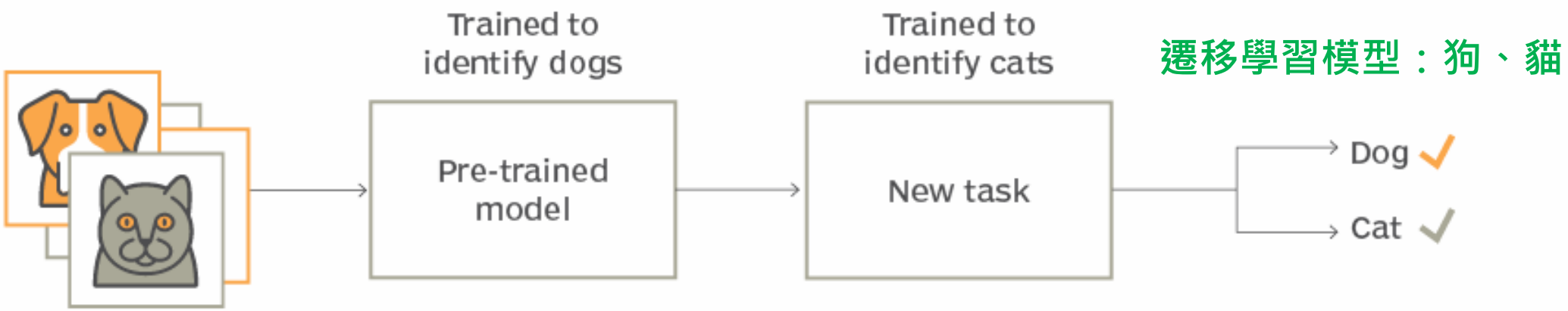
浮點：實數的近似值表示法

FLOPs = floating point operations

Training from scratch



Transfer learning

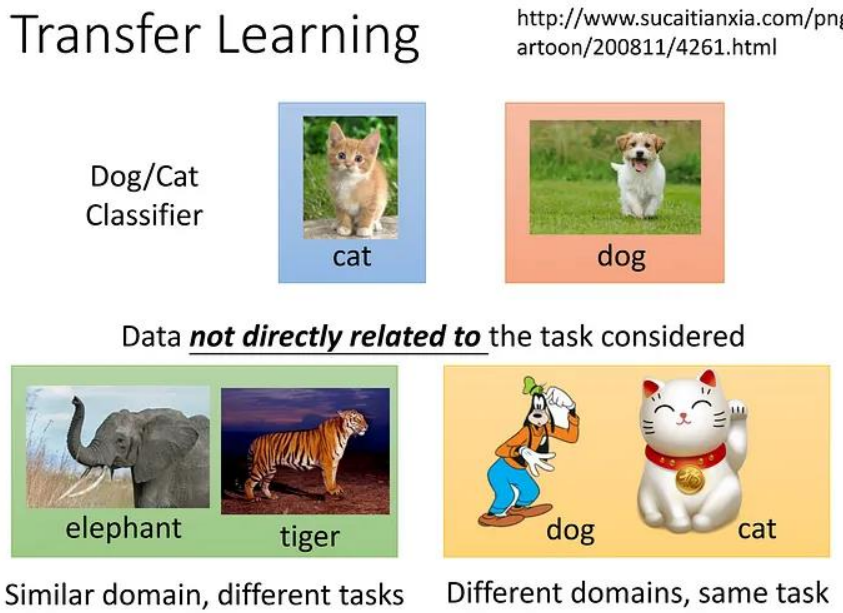


遷移學習 (transfer learning)

- 什麼是遷移學習呢？舉個例子，我們現在有一個可以判斷貓和狗圖片的分類器，但是我們想要一個可以判斷和貓狗沒有直接相關圖片的分類器，這個沒有直接相關的意思可能是下面兩種：

a. 是動物的照片，但不是貓和狗，可能是大象和老虎 (Similar domain, different tasks)

b. 是貓和狗的圖片，但不是真的貓狗，而是招財貓和高飛狗 (Different domains, same task)



- ✓ 那原本的貓狗分類器，是否對這個新的分類器有幫助？這就是遷移學習所要探討的目標
- ✓ 需要辨別的資料為目標資料 (Target data)，原本既有的資料則稱為來源資料 (Source data)

- 轉移學習就是把已經訓練好的模型、參數，轉移至另外的一個新模型上，使得我們不需要從零開始，重新訓練一個新model
- ✓ 舉例來說，你可以train好一個based on Cifar 10的CNN，然後把這個訓練出來的模型套用至其他影像辨識的數據上、甚至是使用這個模型成為一個特徵萃取機制

- Why Transfer Learning?

僅基於數據的角度來談的話有三點

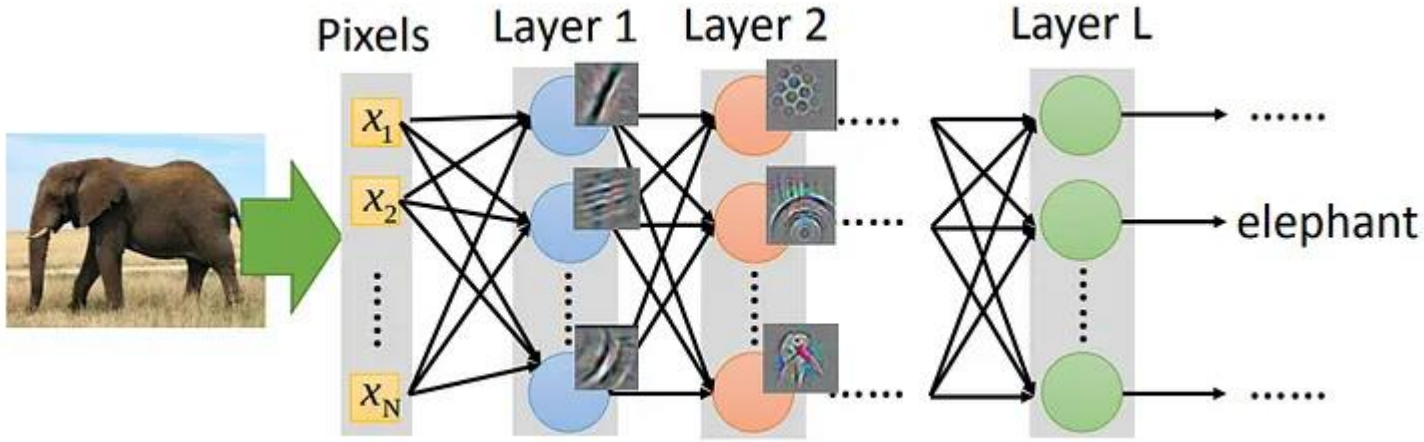
1. 收集數據比較困難
2. 標記(label)數據很耗時，很繁瑣，需要大量的人力
3. 訓練one2one的model，缺失了Machine Learning 中對於泛化性能的要求

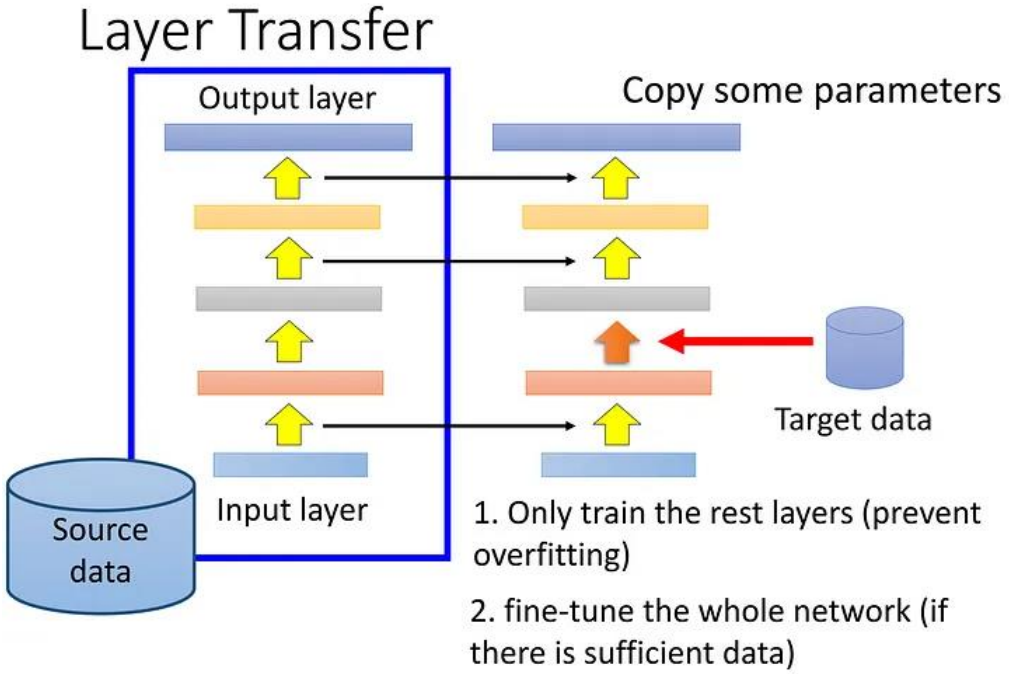
上面三點，就使得我們必須進行Transfer Learning

- 使用的方法叫作 Model fine-tuning，我來用上面的貓狗圖片舉個例子：

先將source data(貓狗圖片) 訓練出一個模型，使用這個訓練完的模型參數當作target data(象虎圖片)模型參數的初始值，繼續以target data(象虎圖片)訓練該已初始化的模型，得到新的模型。這個使用別人當作初始值，自己再做訓練的過程叫作 fine-tune

✓ 但是我們若直接使用target data fine-tune出新的模型的話，通常 結果會很差。原因是因為target data通常數量很少，原來的模型複雜度相對太高，所以訓練出來的結果很容易over-fitting。如果target data很多的話，直接重新訓練一個新的模型就可以了，並不需要遷移學習。這時候我們就需要 layer transfer了





輸入層是圖案的所有像素點，第1層 layer 將會是判斷最基礎的圖形，例如橫線，直線，非常小的圓圈之類的，第2層就會是判斷稍微複雜一點點的圖形，第3層就是再更複雜一點點的圖形...以此類推。由此可知，我們可以了解到只要是判別圖片的CNN，前面幾層layer跟要判斷的目標沒有關係，不管是要判斷貓，狗，鳥，大象，老虎，都不是像素等級的微小圖形，但是前面幾層layer卻也極為重要，因為不管是甚麼圖形，都是由像素級的圖形拼湊而成。因此我們可以將前面幾層layer保留，代換掉後面幾層的 layer 之後重新訓練模型的參數，便可以拿到一個新的並且符合需求的模型了，而且只訓練後面幾層，參數也較少，較不會發生over-fitting

層遷移的典型做法：

1. 凍結前幾層，訓練後幾層：

- 將預訓練模型的前幾層凍結，只訓練新添加的層，這樣可以保持模型已經學到的通用特徵，而專注於新任務的專門特徵。
- 公式：

$$\theta_{new} = \theta_{\text{pretrained}}^{(1:k)} \cup \theta_{\text{new}}^{(k+1:n)}$$

其中：

- $\theta_{\text{pretrained}}^{(1:k)}$ 是前 k 層預訓練模型的權重，這部分的權重保持不變（凍結）。
- $\theta_{\text{new}}^{(k+1:n)}$ 是從第 $k+1$ 到第 n 層的新的模型參數，這部分將在新任務上進行訓練。

2. 部分凍結，部分微調：

- 將部分層的權重凍結，部分層進行微調。這適用於新任務與原始任務有較大差異時，允許模型適應新資料的特徵。
- 公式：

$$\theta_{\text{fine-tuned}} = \theta_{\text{pretrained}}^{(1:k)} + \Delta\theta^{(1:k)}$$

其中：

- $\Delta\theta^{(1:k)}$ 是在新任務上對前 k 層的微調，這部分是新資料學習到的增量。

3. 全面微調：

- 將整個預訓練模型解凍，在新資料集上進行全面微調。適用於新任務與原始任務差異非常大時。
- 公式：

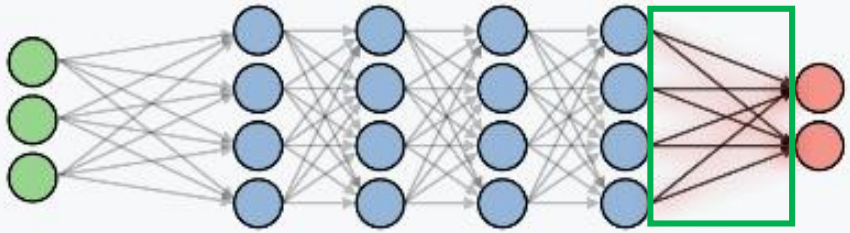
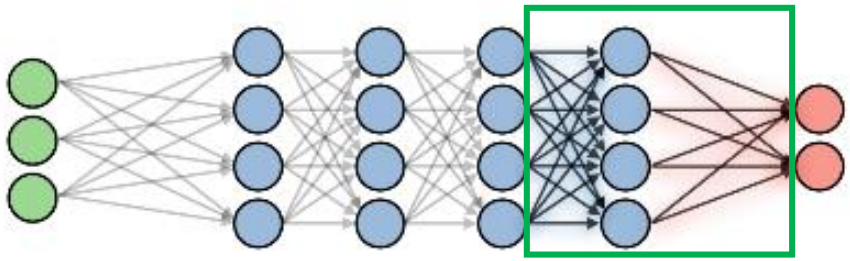
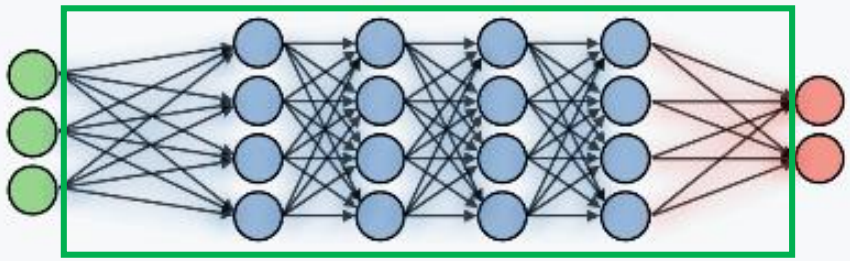
$$\theta_{\text{fine-tuned}} = \theta_{\text{pretrained}} + \Delta\theta$$

全部參數 θ 都會根據新資料進行調整。

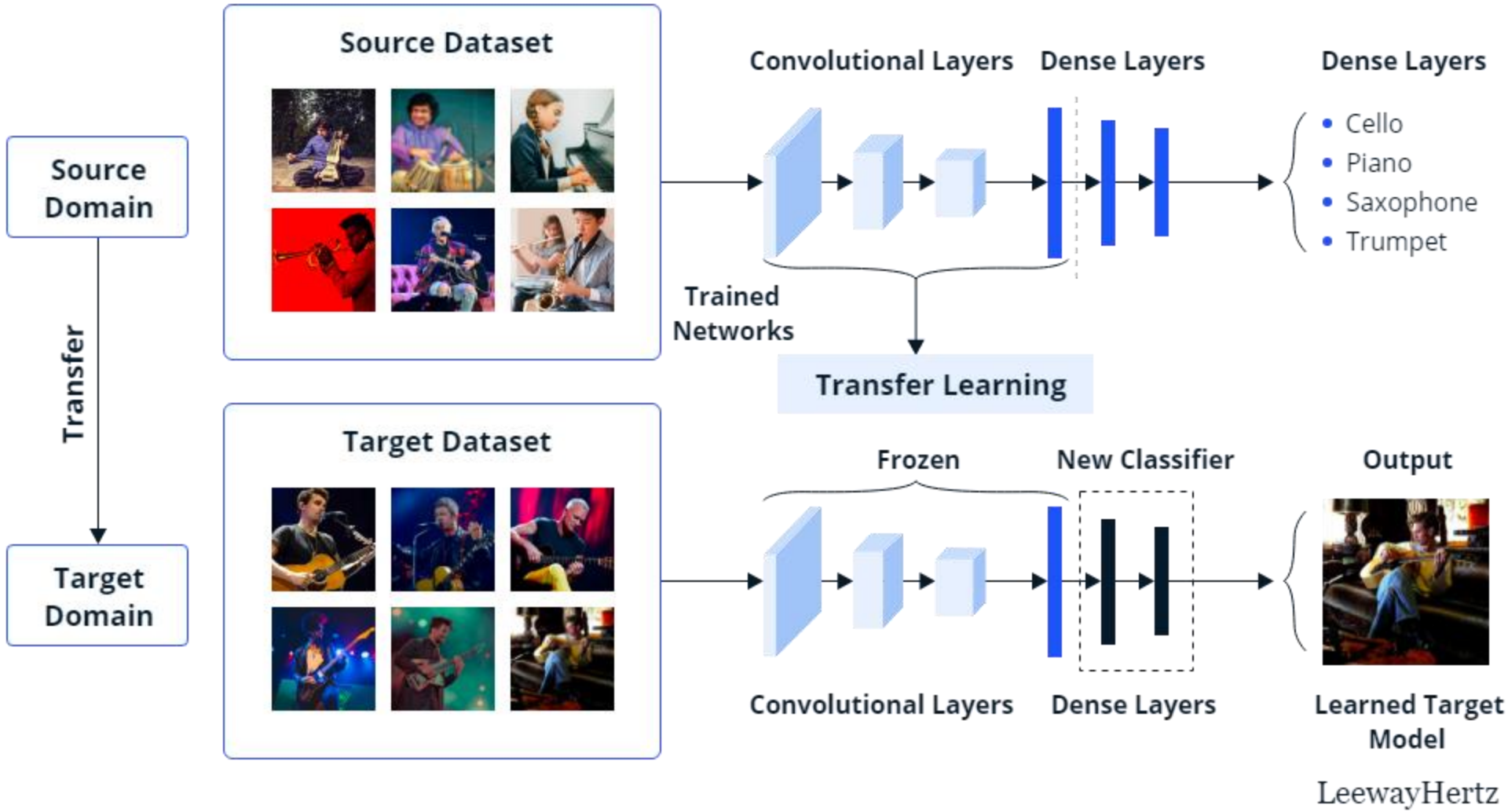
只訓練輸出層

只訓練後幾層與輸出層

只有初始化(不用隨機)

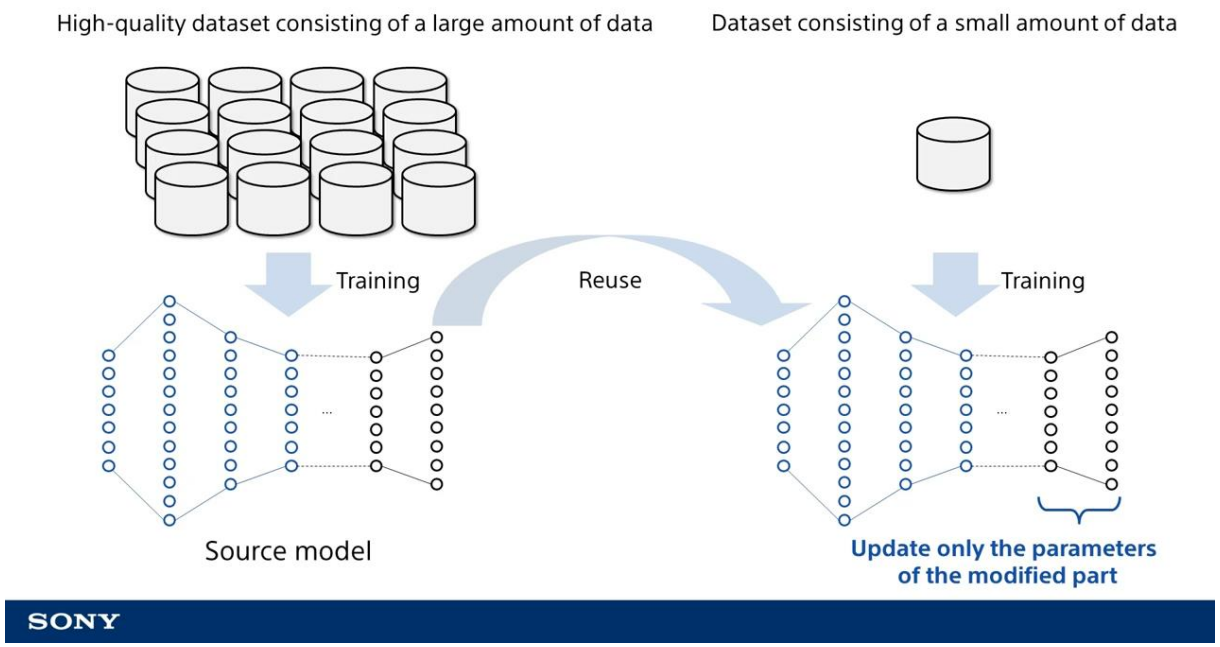
Training size	Illustration	Explanation
Small		Freezes all layers, trains weights on softmax
Medium		Freezes most layers, trains weights on last layers and softmax
Large		Trains weights on layers and softmax by initializing weights on pre-trained ones

softmax：柔性最大化函數


Source Domain
Transfer
Target Domain
Source Dataset
Target Dataset
Convolutional Layers
Dense Layers
Dense Layers
Cello
Piano
Saxophone
Trumpet
Trained Networks
Transfer Learning
Frozen
New Classifier
Output
Convolutional Layers
Dense Layers
Learned Target Model
LeewayHertz

Transfer Learning

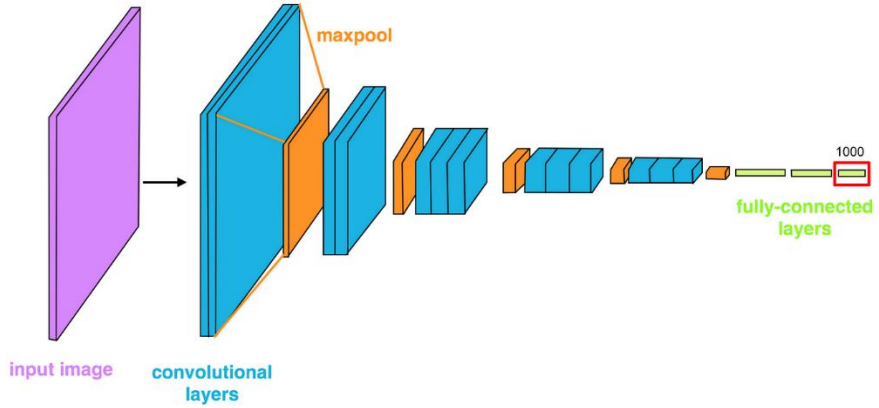
11:34



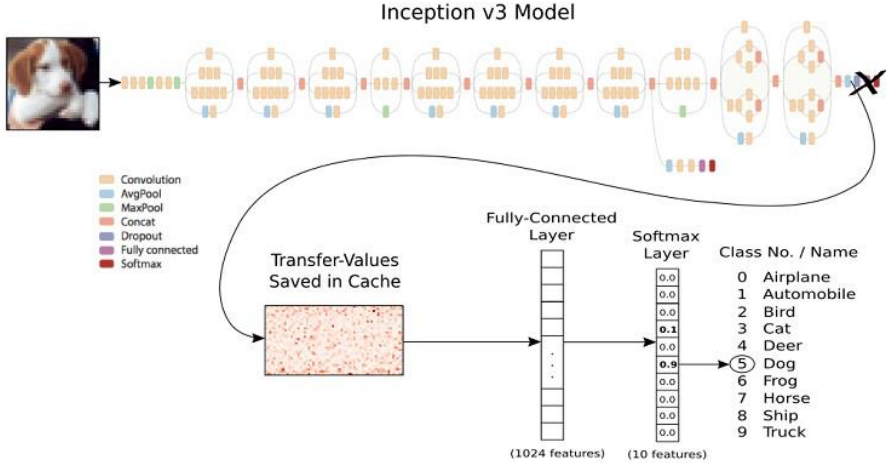
Types of Pretrained Model

- **VGG-16**
  - 牛津大學的Visual Geometry Group (VGG)提出。這個模型的名字中的 "16" 代表它擁有 16 層帶有可學習參數的網絡層
- **InceptionV3**
  - 由 Google 提出的深度卷積神經網絡模型，這個模型參與了2014年ImageNet大規模圖像識別競賽 (ILSVRC) 並取得了成功。
- **ResNet**
  - 微軟研究院的KaimingHe 等人在2015 年提出的一種深度卷積神經網絡架構，並在 ImageNet 大規模圖像識別競賽 (ILSVRC) 中取得了傑出表現，最大特點是解決了傳統深層神經網絡在增加層數時會出現的**梯度消失問題**

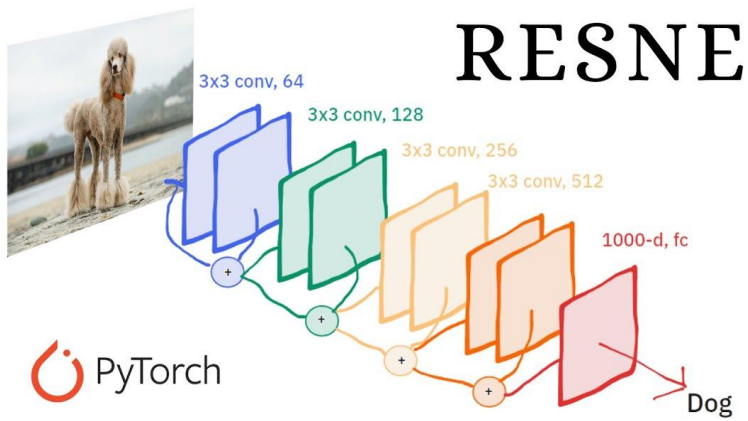
VGG-16 Architecture



Inception v3 Model



RESNET



Generative Neural Network