Day 19

深度學習與電腦視覺
學習馬拉松

cupoy 陪跑專家：楊哲寧

深度學習理論與實作

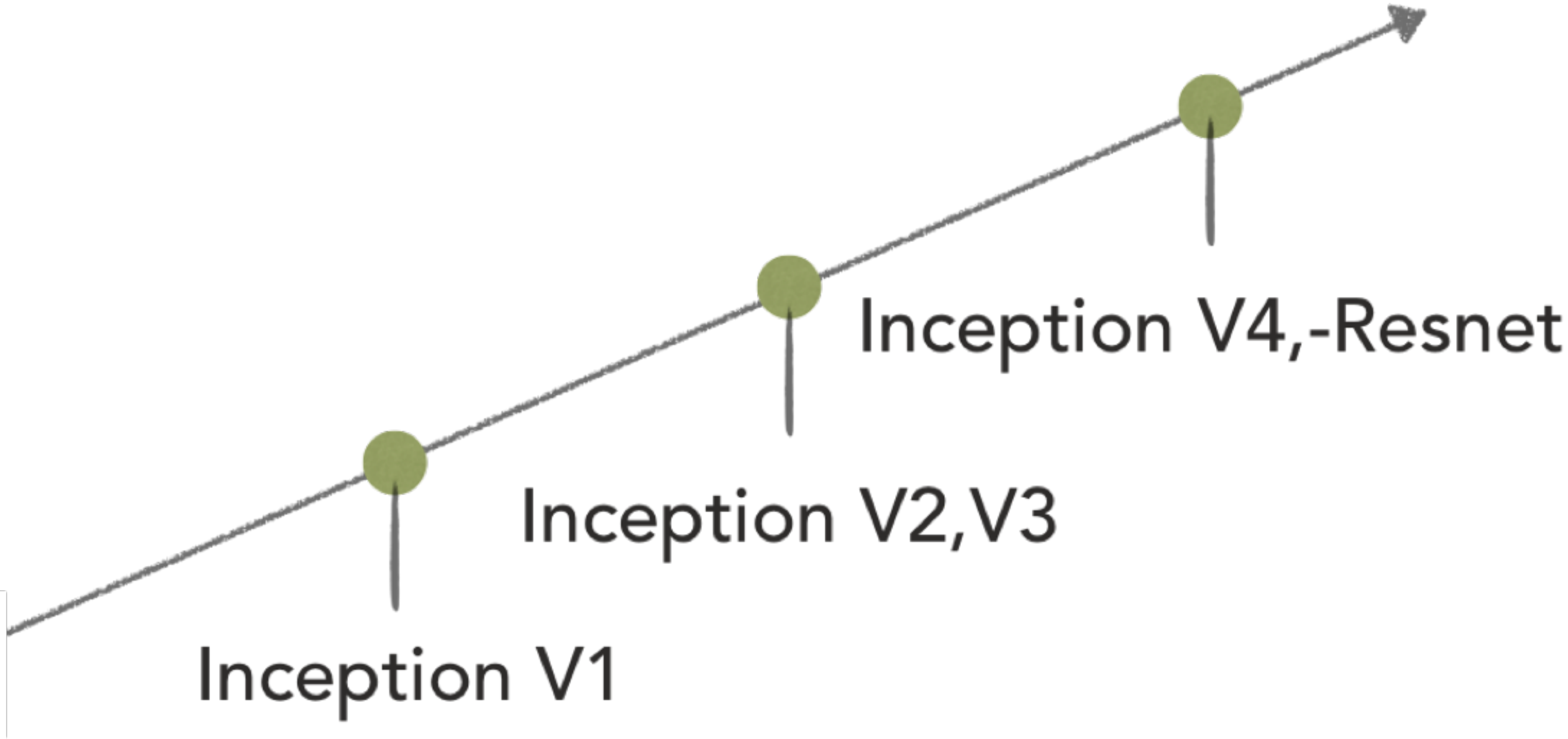
Classic CNN Backbone
(經典CNN框架)

重要知識點

了解InceptionV1-V3的演進

Inception

InceptionV1-V4以及 Inception-ResNet 的演進



InceptionV1

CUPOY

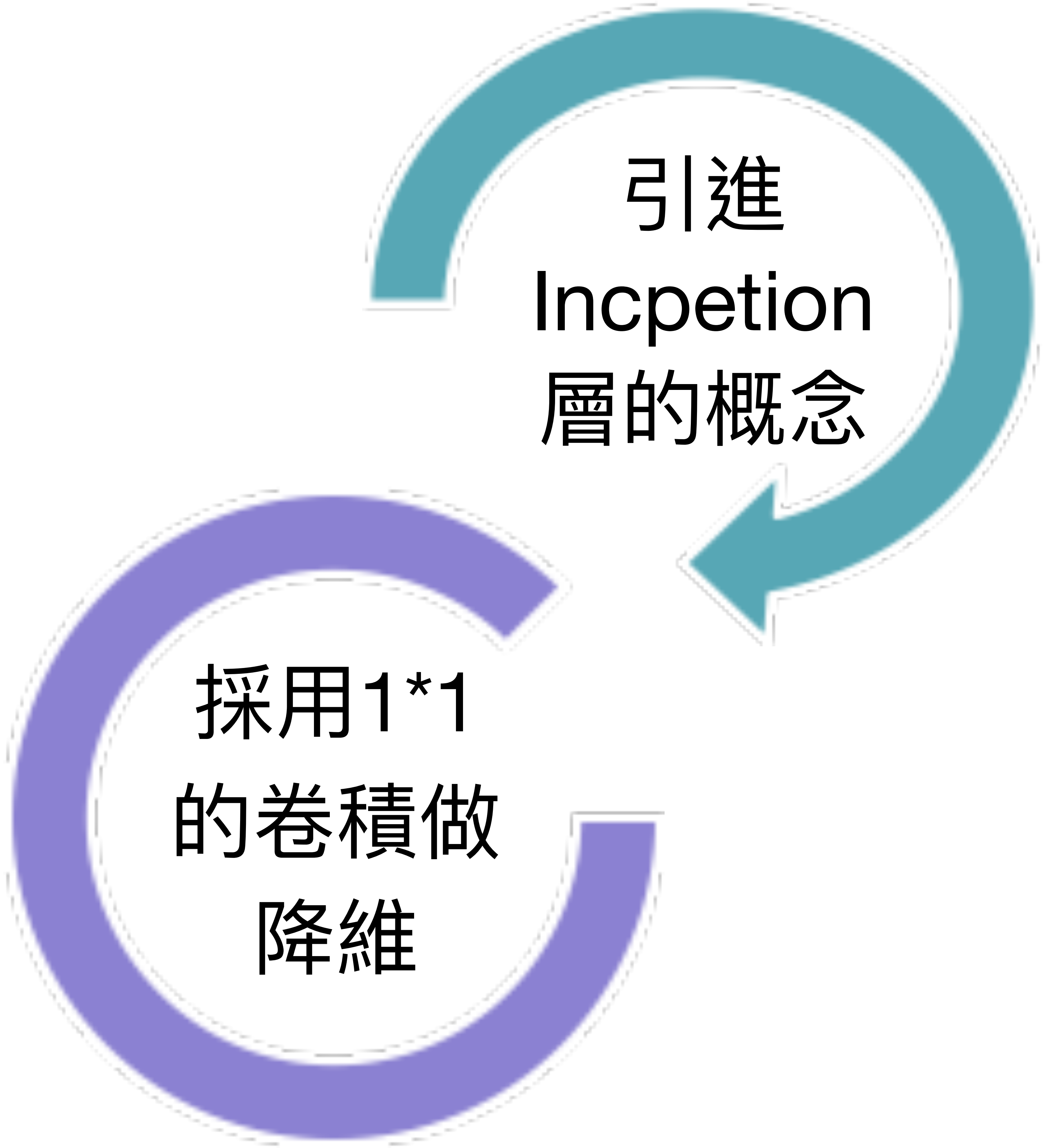
下表展示了幾個經典架構，其中 Inception 與 Vgg 分別為 2014 年 ImageNet 的冠亞軍

Model	AlexNet	Vgg	InceptionV1	ResNet
推出年	2012	2014	2014	2015
層數	8	19	22	152
卷積層數	5	16	21	151
卷積核大小	11,5,3	3	7,1,3,5	7,1,3,5
Inception(NIN)	-	-	+	-
全連接層	3	3	1	1
全連接層大小	4096*2,1000	4096*2,1000	1000	1000
Top-5 error	15.3%	7.3%	6.7%	3.57%

參考來源： itread

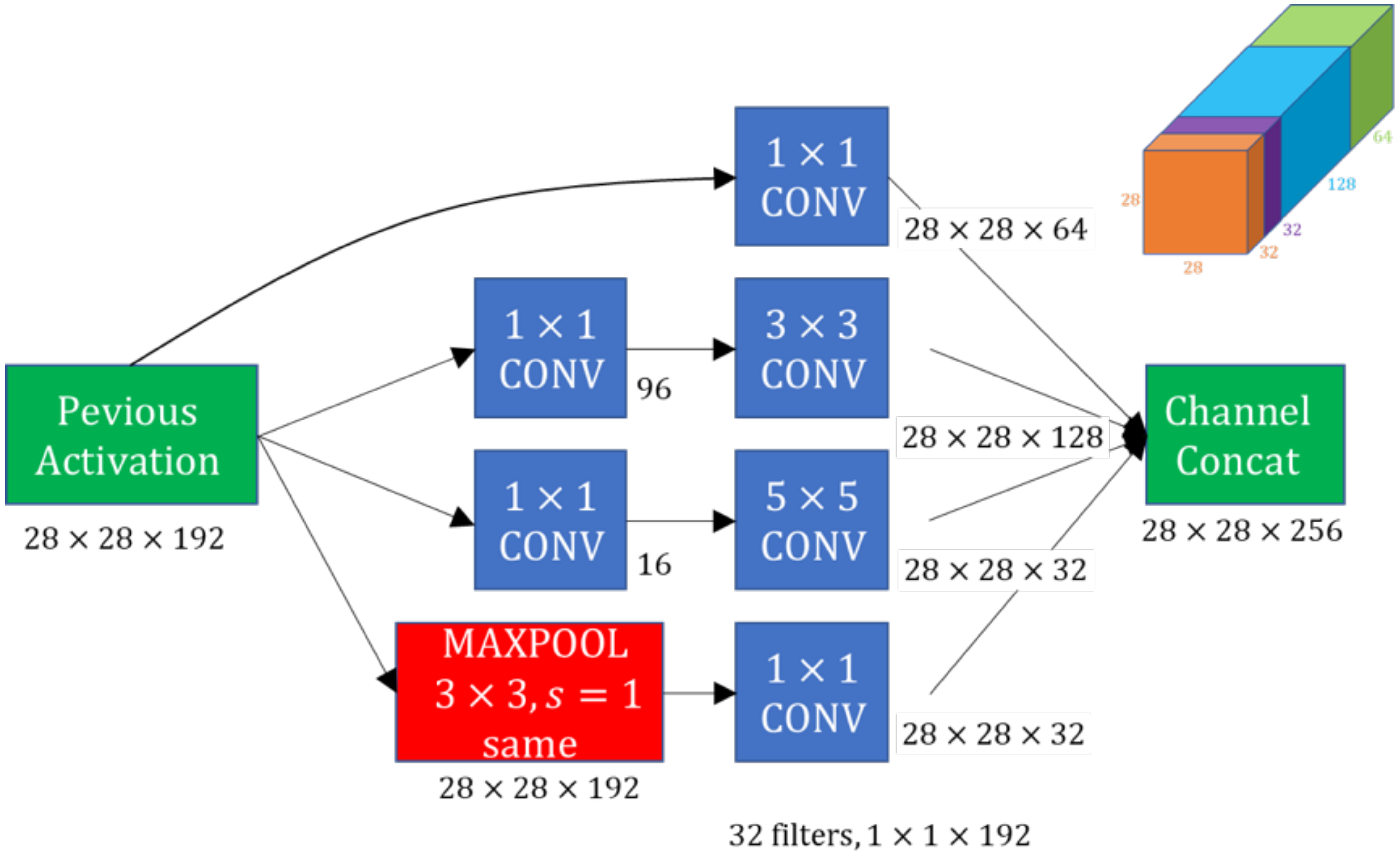
InceptionV1

InceptionV1 兩大重點



InceptionV1-Inception Layer

Inception Layer:：目的在於結合不同 Receptive Field的特徵圖。



Inception Layer

參考來源： DataHacker

InceptionV1-Inception Layer

- 由於信息位置的巨大差異，為卷積操作選擇合適的卷積核大小就比較困難

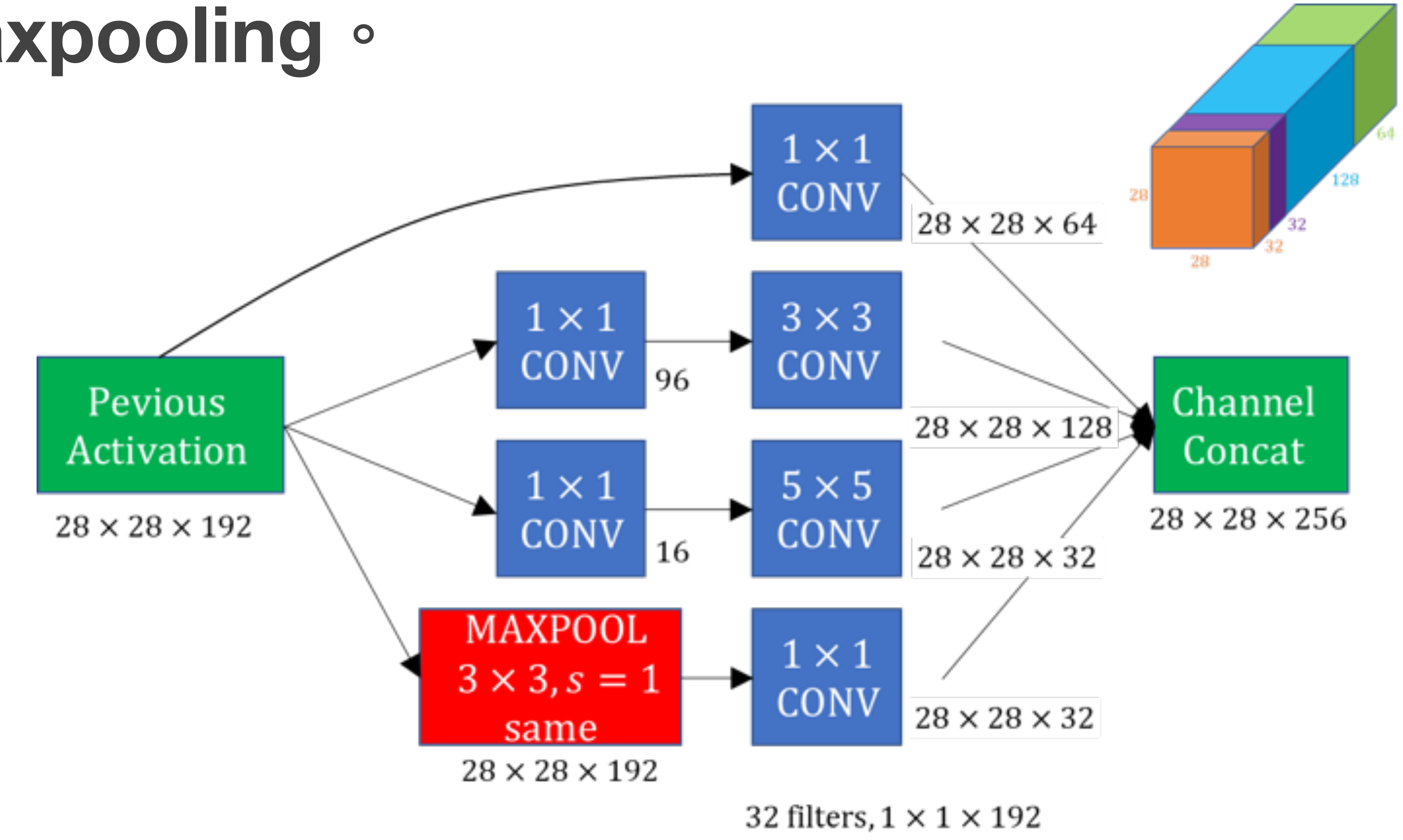
信息分佈更**全局性的圖**像偏好較**大**的卷積核

信息分佈比較**局部的圖**像偏好較**小**的卷積核

- Inception概念：結合不同大小的 Kernels借以獲得不同尺度的訊息。

InceptionV1-Inception Layer

圖中展示了經典的 Inception架構，接在 Feature Maps 後一共有**四條分支**，其中三條先經過 **1*1 kernel 的壓縮**，這樣做的意義主要是為了控制輸出 **Channels 的深度**，並同時能增加模型的**非線性**，一條則是先通過 3*3 kernel 的 **Maxpooling**。

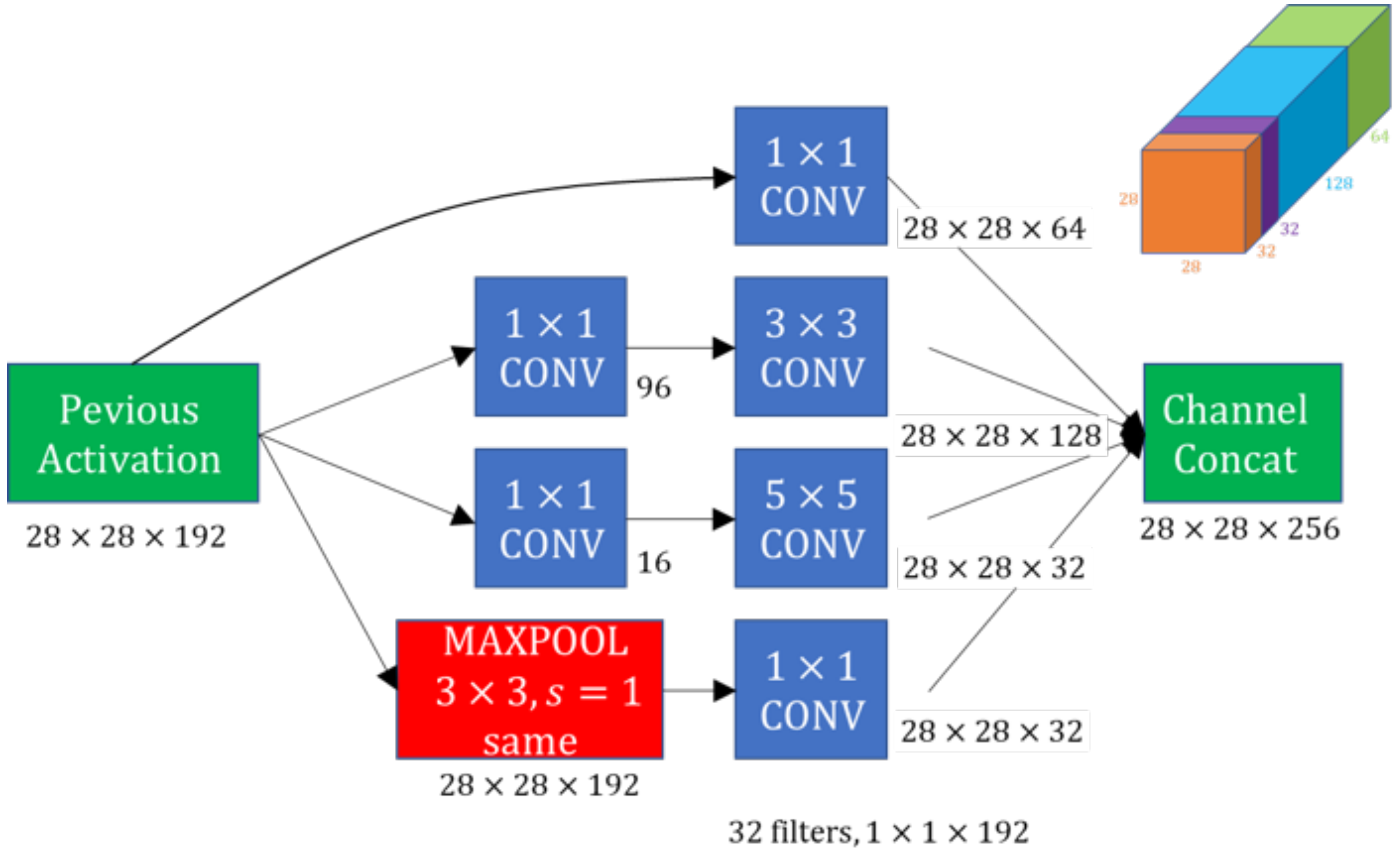


Inception Layer

參考來源： DataHacker

InceptionV1-Inception Layer

為了確保輸出 Feature Map 在長寬上擁有一樣尺寸，我們就要借用 **Padding** 技巧，1*1 kernel 輸出大小與輸入相同，而 3*3、5*5 kernel 則分別設定補邊值為**1**、**2**，在 tensorflow、Keras 中最快的方式就是設定 **padding=same**，就能在步伐為1時確保輸出尺寸維持相同。



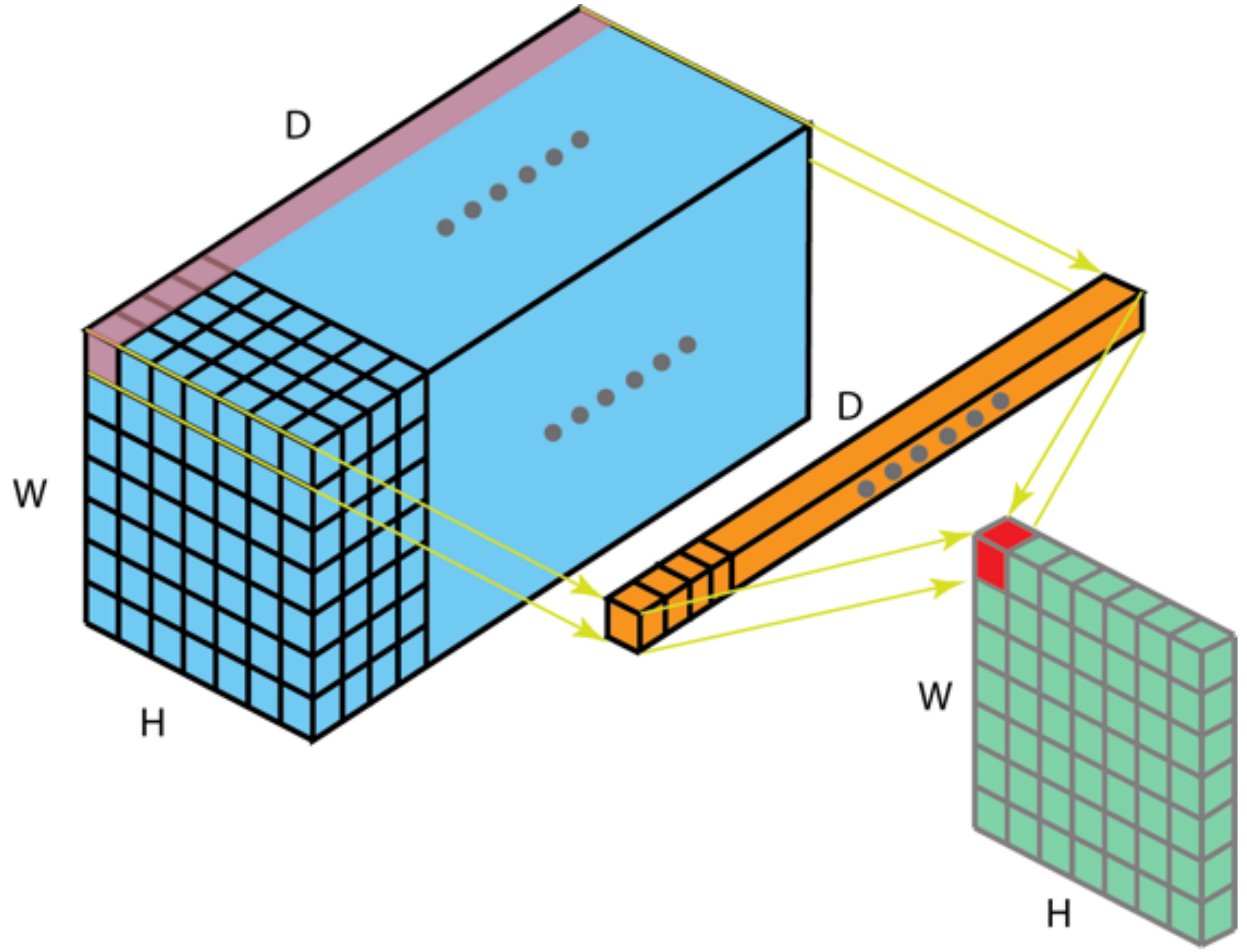
Inception Layer

參考來源： DataHacker

InceptionV1- 1*1 Convolution

1*1 kernel 的**壓縮**其實就是一般的卷積，然而它的好處在於能用相當少的參數量，達到壓縮特徵圖深度的目的，舉個例子來說，當輸入 Feature Map 為 (batch_size,14,14,192)，要將其壓縮為(batch_size,14,14,n)，我們只需要 1*1*192*n+n 個參數量，當然同樣事情也可以用 3*3 kernel 達成，但參數量就會變為 3*3*192*n+n。

Kernel 參數 Bias

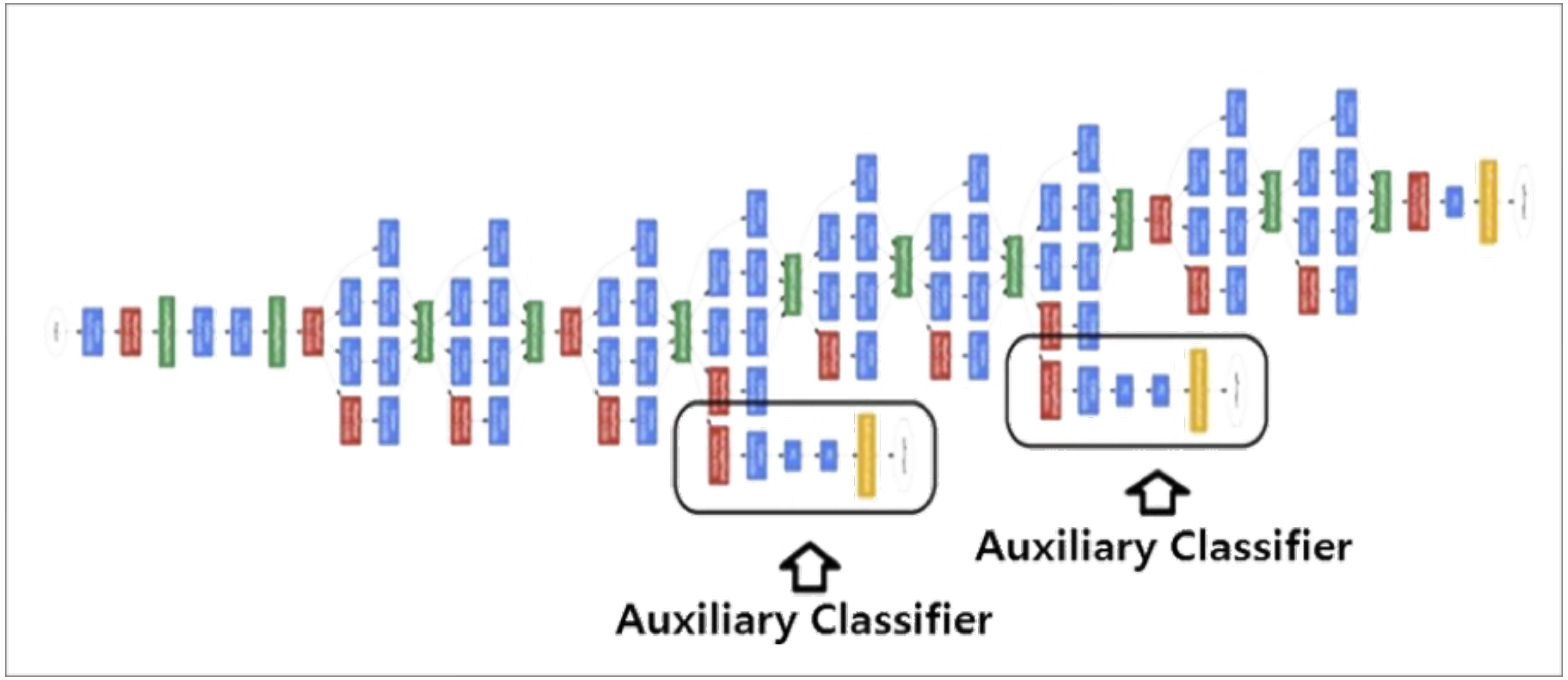


1*1 Convolution

參考來源： TowardsDataScience

InceptionV1- Auxiliary classifiers

Auxiliary classifiers 在之後的文獻中比上少看見，原論文中提到其存在的目的在於通過**中間層的分類損失**，監督神經網路，最後會按一個較小的權重（0.3）加到最終分類結果中，這樣相當於做了模型融合，同時給網絡增加了反向傳播的梯度信號，也提供了**額外的正則化**。



參考來源： Medium

InceptionV2

InceptionV2 的重點就在於引入了『Batch Normalization』，之前已有獨立篇章介紹，在這就不多加贅述。

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
Parameters to be learned: γ, β

Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

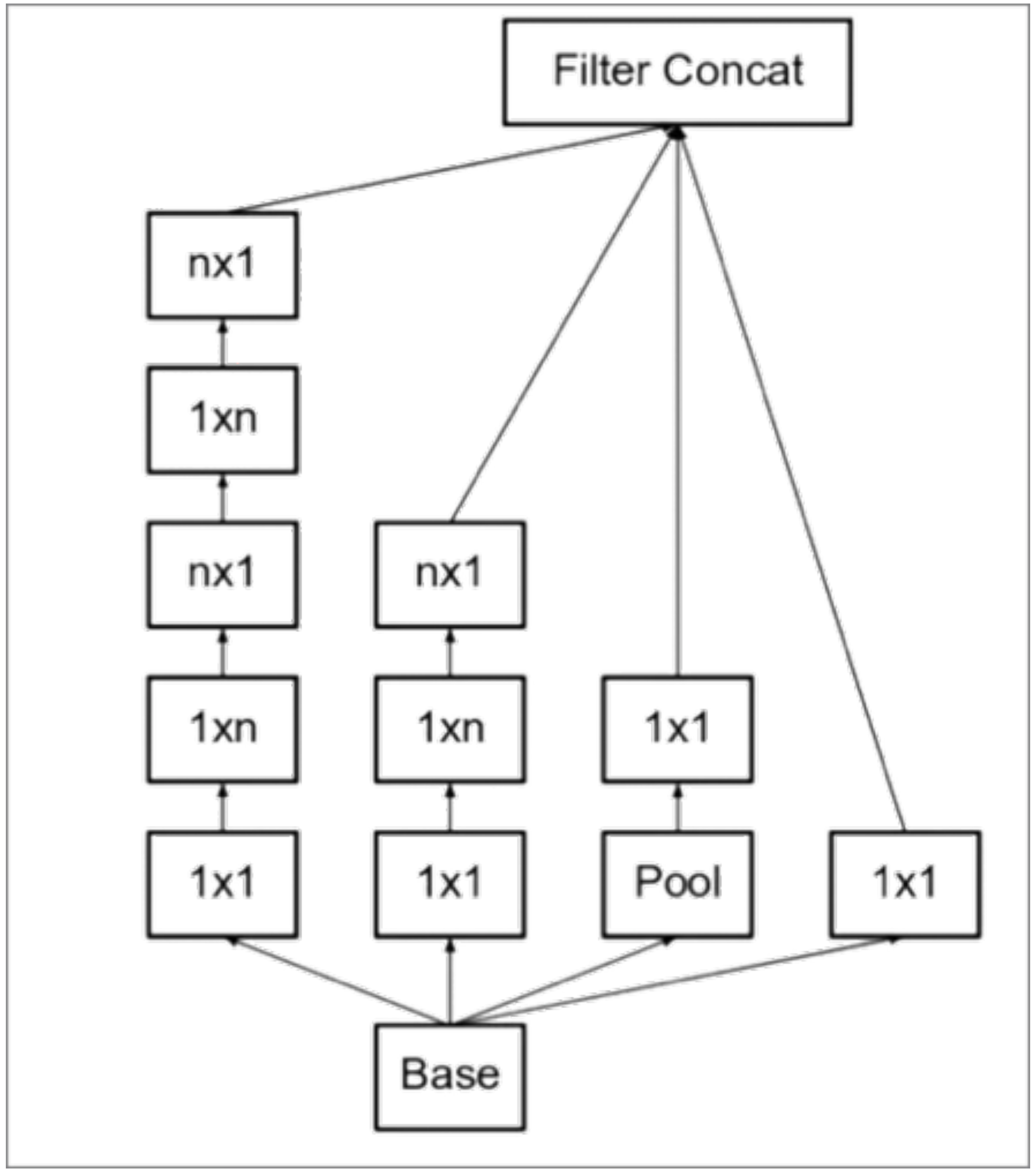
$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma \widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

參考來源： TowardsDataScience

InceptionV3

- 將**n*n**的卷積核以**1*n+n*1**的卷積核取代
- 將1個卷積拆成2個卷積，使得網絡深度更深，增加了網絡的非線性（每增加一層都要經過Activation Function） 。
- 如原本3*3的 Kernel 就會被拆解為1*3 + 3*1，因此觀看右圖我們可發現 **Inception Block 架構變得更深**。
- 然而原文也提及，此種結構放在神經網路前**幾層效果並不好**，建議放到較深的層數中使用。



參考來源： Cornell University

參考資料

CUPOY

Network In Network

Min Lin, Qiang Chen, Shuicheng Yan

(Submitted on 16 Dec 2013 (v1), last revised 4 Mar 2014 (this version, v3))

We propose a novel deep network structure called "Network In Network" (NIN) to enhance model discriminability for local patches within the receptive field. The conventional convolutional layer uses linear filters followed by a nonlinear activation function to scan the input. Instead, we build micro neural networks with more complex structures to abstract the data within the receptive field. We instantiate the micro neural network with a multilayer perceptron, which is a potent function approximator. The feature maps are obtained by sliding the micro networks over the input in a similar manner as CNN; they are then fed into the next layer. Deep NIN can be implemented by stacking mutiple of the above described structure. With enhanced local modeling via the micro network, we are able to utilize global average pooling over feature maps in the classification layer, which is easier to interpret and less prone to overfitting than traditional fully connected layers. We demonstrated the state-of-the-art classification performances with NIN on CIFAR-10 and CIFAR-100, and reasonable performances on SVHN and MNIST datasets.

Comments: 10 pages, 4 figures, for iclr2014
Subjects: **Neural and Evolutionary Computing (cs.NE)**; Computer Vision and Pattern Recognition (cs.CV); Machine Learning (cs.LG)
Cite as: arXiv:1312.4400 [cs.NE]
(or arXiv:1312.4400v3 [cs.NE] for this version)

Bibliographic data

[Enable Bibex(What is Bibex?)]

Submission history

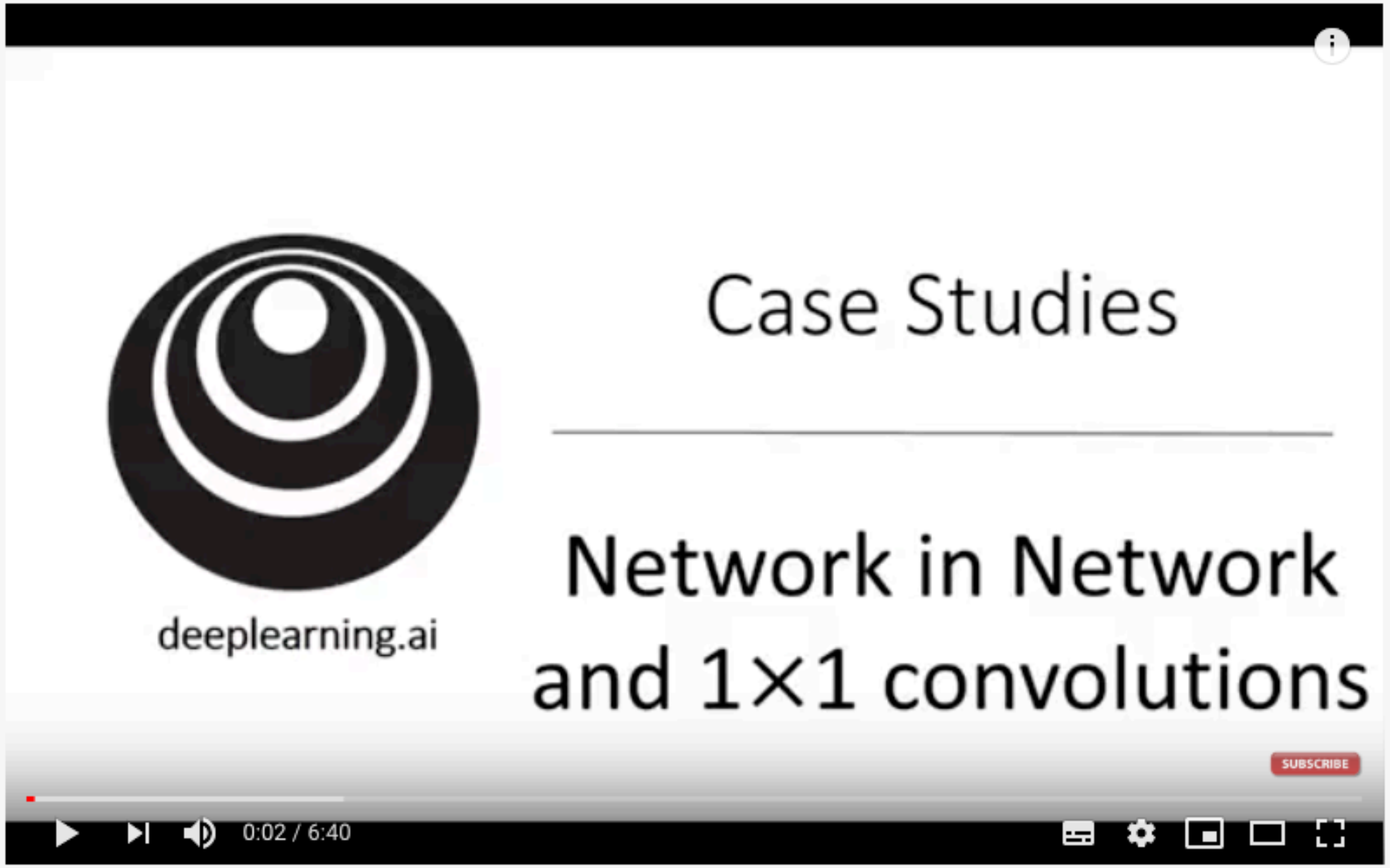
From: Min Lin [view email]
[v1] Mon, 16 Dec 2013 15:34:13 UTC (501 KB)
[v2] Wed, 18 Dec 2013 09:30:27 UTC (509 KB)
[v3] Tue, 4 Mar 2014 05:15:42 UTC (445 KB)

Network in Network

首次提出1*1 convolution降維

連結

參考資料



Neural Networks - Networks in Networks and 1x1 Convolutions

吳恩達：1x1 Convolutions

連結

解題時間 Let’s Crack It

請跳出 PDF 至官網 Sample Code & 作業開始解題