

# SA123

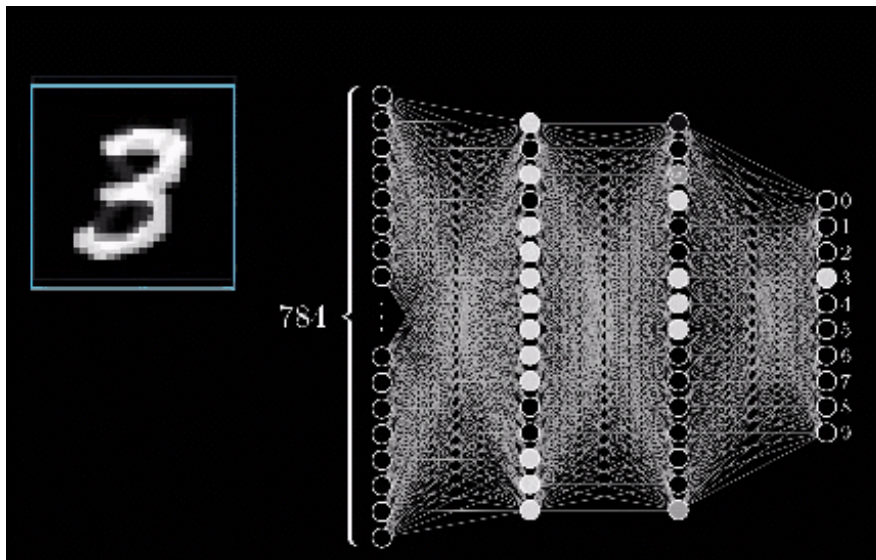
[首頁](#)

## 圖解10大CNN網路架構，通俗易懂！

🕒 2022年03月15日    👤 作者: 機器學習演算法與Python學習

點選機器學習算法與Python學習，選擇加星標

精彩内容不迷路

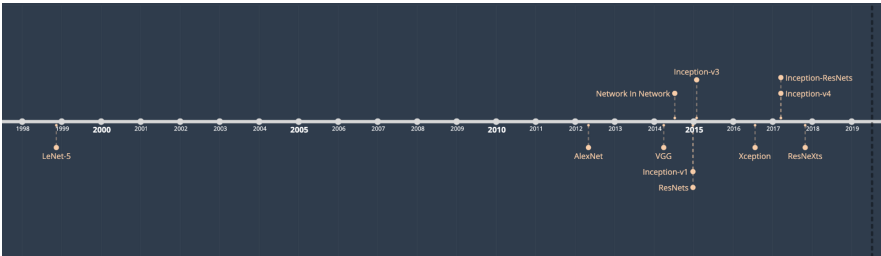


作者 / Raimi Karim, 出品 / AI科技大本營  
(ID: rgznai100)

本文精心選取了 10 個 CNN 體系結構的詳細圖解進行講述。由作者精心挑選。這些圖解展示了整個模型的精華，無需去逐個瀏覽那些 Softmax 層。除了這些示意圖，作者還提供了一些註釋，闡述了它們是如何不斷演變的——卷積層從 5 到 50 個、從普通的卷積層到卷積模組、從 2~3 tower 到 32 tower、卷積核從  $7 \times 7$  到  $5 \times 5$ 。

所謂“常見”，是指這些模型的預訓練權重通常被深度學習庫（如TensorFlow和PyTorch）所共享，提供給開發者使用，這些模型通常會在課堂上講授。其中一些模

型已經在競賽（如ILSVRC ImageNet大規模影象識別挑戰）中取得了成功。



將要討論的10個架構與相應的論文釋出時間

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
VGG16	528 MB	0.713	0.901	138,357,544	23
InceptionV3	92 MB	0.779	0.937	23,851,784	159
ResNet50	98 MB	0.749	0.921	25,636,712	-
Xception	88 MB	0.790	0.945	22,910,480	126
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
ResNeXt50	96 MB	0.777	0.938	25,097,128	-

The top-1 and top-5 accuracy refers to the model's performance on the ImageNet validation dataset.

Depth refers to the topological depth of the network. This includes activation layers, batch normalization layers etc.

6個網路架構的預訓練權重可以在 Keras 中獲得，參見 [https://keras.io/applications/?source=post\\_page](https://keras.io/applications/?source=post_page)

寫這篇文章的原因在於沒有多少部落格和文章提到這些緊湊的結構圖解。因此，作者決定自己寫一篇作為參考。出於這個目的，作者閱讀了本文提到的論文和程式碼（絕大部分是 TensorFlow 和 Keras ），得到了這些成果。這裡還要特別指出，這些 CNN 網路結構的來源五花八門——計算機硬體效能的提高、ImageNet 競賽、解決特定問題、新想法等等。一位在 Google 工作的研究員 Christian Szegedy 曾經提到：

“這個程序絕大多數不只是由於更強大的硬體、更大的資料集和更大的模型，更是一系列新想法、演算法和網路結構的改進”。

現在讓我們來看看這些“巨獸”般的網路架構是如何逐漸演變的。

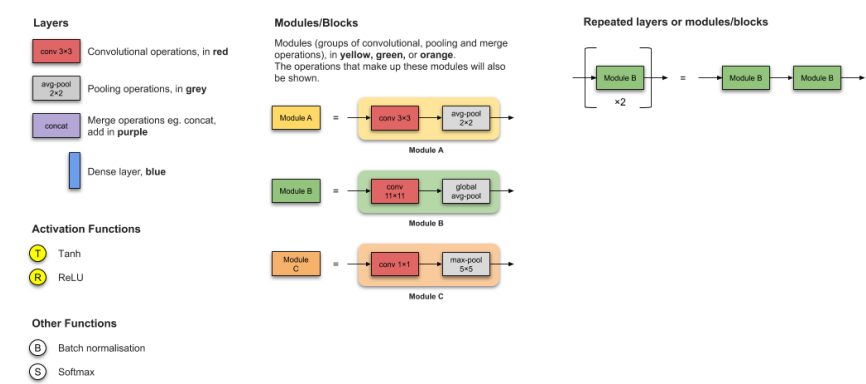
【作者注】對視覺化的註釋：請注意，在這些示意圖中，作者略去了一些資訊，如卷積過濾器的數量、Padding、Stride、Dropout 和flatten操作。

目錄（按發表時間排序）

1. LeNet-5

- 2. AlexNet
- 3. VGG-16
- 4. Inception-v1
- 5. Inception-v3
- 6. ResNet-50
- 7. Xception
- 8. Inception-v4
- 9. Inception-ResNets
- 10. ResNeXt-50

圖例



1. LeNet-5 (1998)

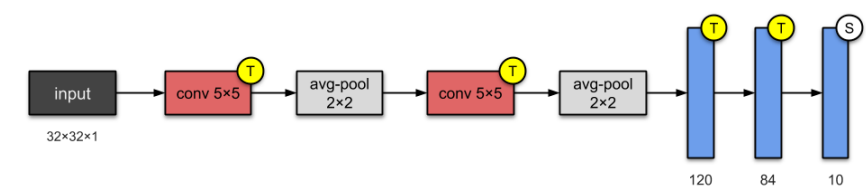


圖1：LeNet-5 網路結構

LeNet-5 一個最簡單的網路架構。它有 2 個卷積層和 3 個全連線層（總共 5 層，這種命名方式在神經網路中很常見，這個數字代表卷積層和全連線層的總和）。Average-Pooling 層，我們現在稱之為亞取樣層，有一些可訓練的權重（現在設計 CNN 網路時已經不常見了）。這個網路架構有大約 6 萬個引數。

有哪些創新？

這個網路架構已經成為標準的“模板”：堆疊式卷積和池化層，以一個或多個全連線層作為網路的末端。

相關論著

- 論文: Gradient-Based Learning Applied to Document Recognition  
連結: [http://yann.lecun.com/exdb/publis/index.html?source=post\\_page](http://yann.lecun.com/exdb/publis/index.html?source=post_page)
- 作者: Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner
- 發表在: Proceedings of the IEEE (1998)

## 2. AlexNet (2012)

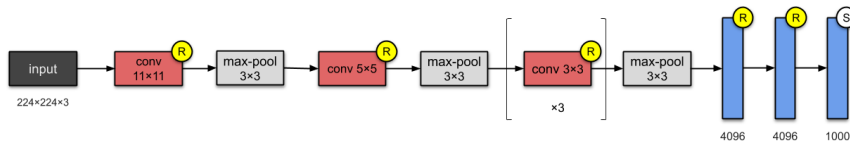


圖2: AlexNet 網路結構

AlexNet 有 60 M 個引數，共有 8 層：5 個卷積層和 3 個全連線層。AlexNet 只是在 LeNet-5 中堆疊了更多的層。在該論文發表時，論文作者指出他們的網路架構是“當前最大的 ImageNet 子集卷積神經網路之一”。

### 有哪些創新？

1. 他們的網路架構是首個採用 ReLU 作為啟用函式的 CNN ；
2. 在 CNN 中採用交織池化（Overlapping pooling）。

### 相關論著

- 論文:  
ImageNet Classification with Deep Convolutional Neural Networks

連結:

[https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks?source=post\\_pag](https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks?source=post_pag)

e

- 作者: Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton. University of Toronto, Canada.
- 發表在: NeurIPS 2012

### 3. VGG-16 (2014)

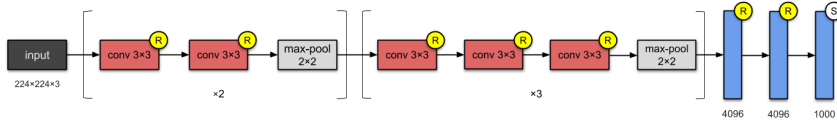


圖3: VGG-16 網路結構

你應該已經注意到，CNN 開始變得越來越深。這是因為提高深度神經網路效能最直接的方法就是增加它們的規模（Szegedy et. al.）。視覺幾何研究小組（VGG）的研究人員發明了 VGG-16，擁有 13 個卷積層和 3 個全連線層，繼承了 AlexNet 的 ReLU 傳統。它由 138 M 個變數組成，要佔用 500 MB 儲存空間。他們也設計了一個更深的版本 VGG-19。

#### 有哪些創新？

1. 正如他們在論文摘要中所提到的，該論文的貢獻是設計更深的網路（大約是 AlexNet 深度的兩倍）。

#### 相關論著

- 論文:

Very Deep Convolutional Networks for Large-Scale Image Recognition

連結: [https://arxiv.org/abs/1409.1556?](https://arxiv.org/abs/1409.1556?source=post_page)  
source=post\_page

- 作者: Karen Simonyan, Andrew Zisserman. University of Oxford, UK.
- 發表在 arXiv preprint, 2014

## 4. Inception-v1 (2014)

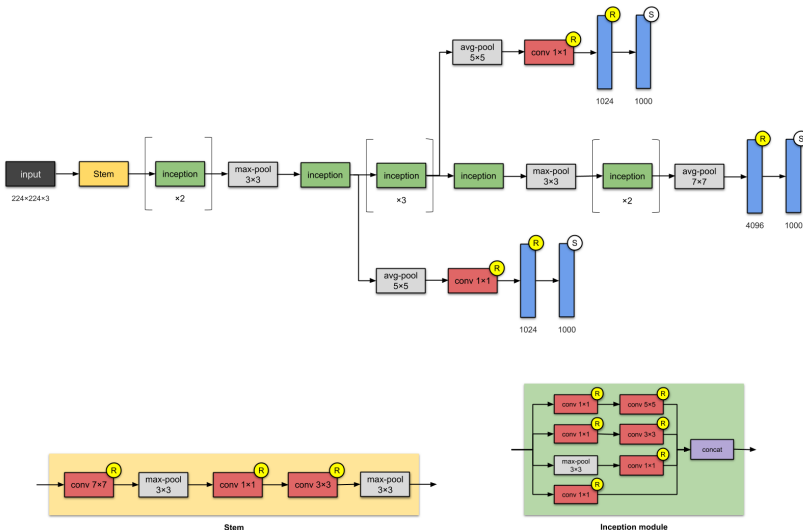


圖 4: Inception-v1 網路結構. 這個 CNN 有兩個輔助網路

(在推斷時被丟棄)，網路結構基於論文中的圖3。

這個 22 層的網路架構有 5 M 引數，被稱之為 Inception-v1。在這個架構中，大量應用了 Network in Network 方法（參見附錄），實現方法是採用 Inception Module。模組的架構設計是透過對稀疏結構預估進行研究完成。

每個模組體現了 3 個思想：

1. 採用不同過濾器的並行卷積塔，然後進行堆疊，採用 1×1、3×3、5×5 卷積核，識別不同特徵，從而對其進行“聚類”。這個想法受到 Arora 等人的論文“*Provable bounds for learning some deep representations*”啟發，建議採用逐層構建的方式，這樣可以分析最後一層的相關統計，並把它們聚類到高相關的各單元組。
2. 1×1 卷積核用來進行維度裁減，以避免計算瓶頸。
3. 1×1 卷積核在一個卷積內增加了非線性。
4. 該論文作者也引入了兩個輔助分類器，以在分類器的最後階段擴大差異，增加了反向傳播的網格訊號，提供了額外的正則化。輔助網路（與輔助分類器相連的分支）在推理時被丟棄。

值得注意的是，「這個網路架構的主要成果是提高網路內部計算資源的利用率」。

### 作者注：

模組的命名 (*Stem*和*Inception*) 在這個版本的*Inception*網路架構中還沒有提出，直到後面一些版本即*Inception-v4*和*Inception-ResNet*中才正式使用。作者把這些加入到這裡是為了更容易進行比較。

## 有哪些創新？

1. 採用緊密模組/板塊構建網路。不採用堆疊卷積層的方法，而是堆疊由卷積層組成模組的方法。*Inception* 這一名字來自於科幻電影《盜夢空間》。

## 相關論著

- 論文:

Going Deeper with Convolutions

連結: [https://arxiv.org/abs/1409.4842?](https://arxiv.org/abs/1409.4842?source=post_page)  
[source=post\\_page](https://arxiv.org/abs/1409.4842?source=post_page)

- 作者: Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich.  
Google, University of Michigan, University of North Carolina
- 發表在: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

## 5. Inception-v3 (2015)



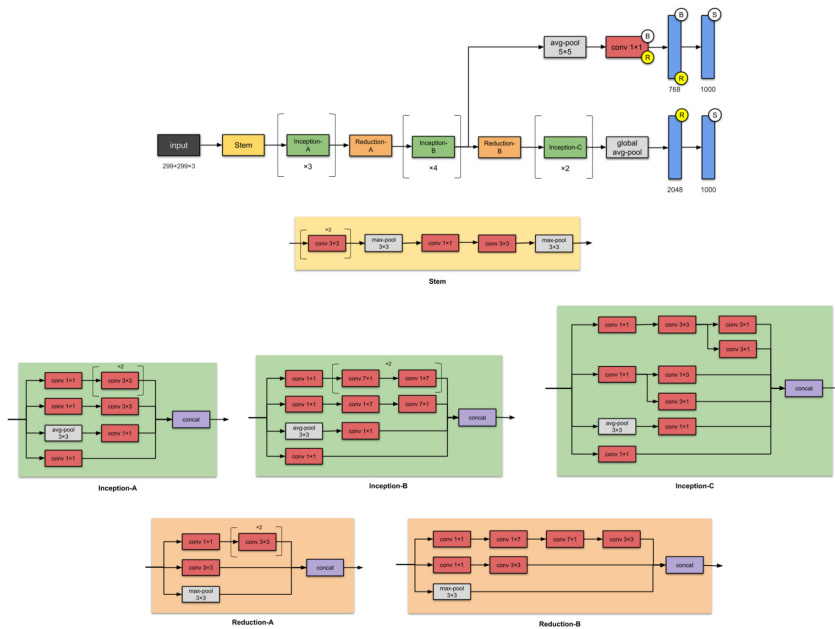


圖 5: Inception-v3 網路架構 這個 CNN 有兩個輔助網路（在推理時被丟棄）。注: 所有卷積層之後採用batch norm和 ReLU 啟用

Inception-v3 是 Inception-v1 的後續版本，有 24 M 個引數。Inception-v2 去哪裡了？別擔心，它只不過是 v3 的一個早期原型，因此與 v3 十分相似，但不常被使用。該論文作者在提出 Inception-v2 時，在上面做了很多實驗，並記錄了一些成功經驗。Inception 是這些成功經驗的結晶（如對最佳化器、損失函式的改進，在輔助網路中對輔助層增加批次正則等等）。

提出 Inception-v2 和 Inception-v3 的原因是要避免表示瓶頸（這意味著大幅度地降低了下一層的輸入維度），並透過採用分片方法提高了計算效率。

模組的命名（Stem 和 Inception）在這個版本的 Inception 網路架構中還沒有提出，直到後面一些版本即 Inception-v4 和 Inception-ResNet 中才正式使用。作者把這些加入到這裡是為了更容易進行比較。

## 有哪些創新？

1. 首先採用批次正則化（為了簡化，上圖中未反映這一點）的設計者之一。

## 與之前的Inception-v1版本相比，有哪些改進？



1. 把 $n \times n$  卷積分解成不對稱的卷積 $1 \times n$  and  $n \times 1$  卷積。
2. 把  $5 \times 5$  卷積分解成 2 個  $3 \times 3$  卷積操作
3. 把  $7 \times 7$  卷積替換成一系列  $3 \times 3$  卷積。

## 相關論著

- 論文:

Rethinking the Inception Architecture for Computer Vision

連結: [https://arxiv.org/abs/1512.00567?](https://arxiv.org/abs/1512.00567?source=post_page)

- 作者: Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna. Google, University College London
- 發表在: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

## 6. ResNet-50 (2015)

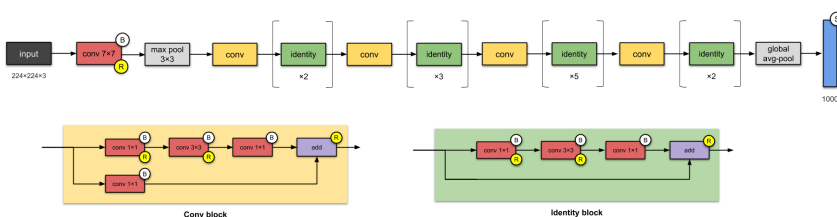


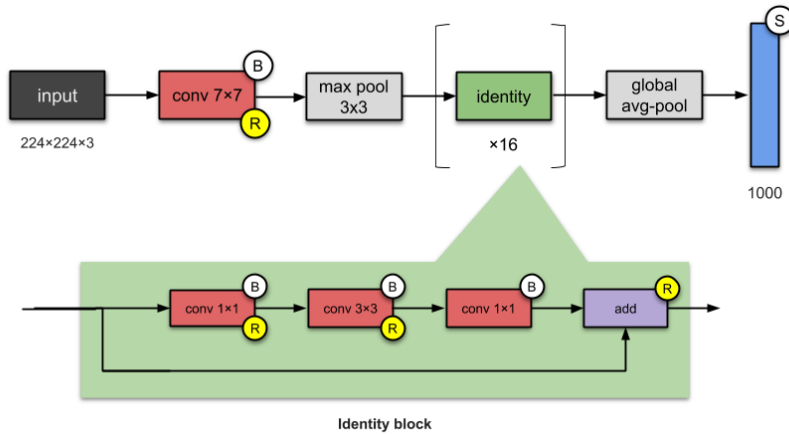
圖 6: ResNet-50 網路架構

對，這就是對文章開頭問題的回答。

在上面的幾個 CNN 中，我們只是看到設計時增加了層數，就達到了更好的效能。但是，“網路深度不斷增加，精度達到了飽和（這一點並不奇怪），因此網路效能開始快速下降”。微軟研究院的專家在構建更深的網路時，採用 ResNet（殘差網路）解決這個問題，可以不再採用全連線的方式。

ResNet 是最早採用批次正則化（Ioffe 和 Szegey 在2015 提交給 ICML 的批次正則化論文）的網路之一。上圖中是 ResNet-50 的網路架構，採用了 26 M 個引數。

ResNet 的基本建構單元是 conv 層和識別板塊。因為它們看上去很相似，你可以把 ResNet 簡化成下圖：



## 有哪些創新？

1. 大量避免使用全連線（他們不是第一個這麼做的）。
2. 設計更深的 CNN 網路（最大可達 152 層），而不必損失網路的生成能力。
3. 首先採用批次正則化的網路架構之一。

## 相關論著

### 論文：

Deep Residual Learning for Image Recognition

連結: [https://arxiv.org/abs/1512.03385?](https://arxiv.org/abs/1512.03385?source=post_page)  
source=post\_page

- 作者: Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Microsoft
- 發表在: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

## 7. Xception (2016)

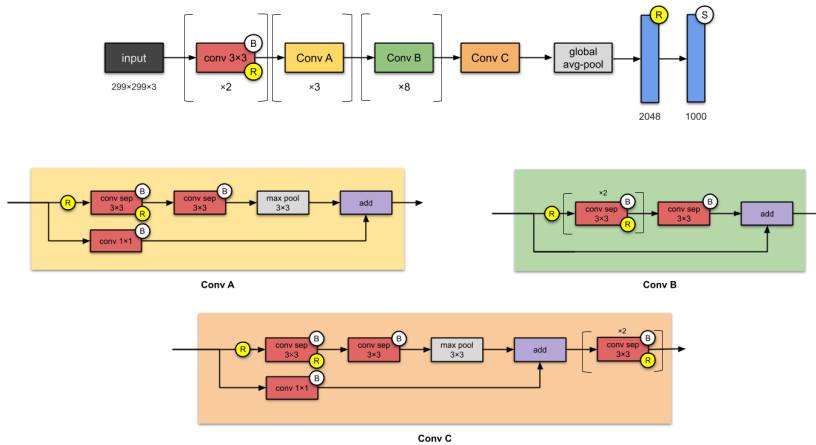


圖 7: Xception 網路架構 注：深度可分解卷積用 conv sep 指代

Xception 是 Inception 網路結構的一個應用，其中 Inception 模組被替換成深度可分解卷積。它和 Inception-v1 有大致相當數量的引數（23M）。Xcption 採用了對 *eXtreme* 的 Inception 假設：

- 首先，交叉通道（或交叉特徵圖）相關性可以被  $1 \times 1$  卷積探測到。
- 因此，每個通道的空間相關性可以透過常規的  $3 \times 3$  或  $5 \times 5$  卷積探測到。

把這個思想推到極致，意味著對每個通道進行  $1 \times 1$  卷積，對每個輸出進行  $3 \times 3$  卷積。這等同於將 Inception 模組替換成深度可分解卷積。

### 有哪些創新？

1. 完全基於深度可分解卷積層，引入 CNN。

### 相關論著

- 論文：

Xception: Deep Learning with Depthwise Separable Convolutions

連結: [https://arxiv.org/abs/1610.02357?](https://arxiv.org/abs/1610.02357?source=post_page)

source=post\_page

- 作者: François Chollet. Google.
- 發表在: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

## 8. Inception-v4 (2016)

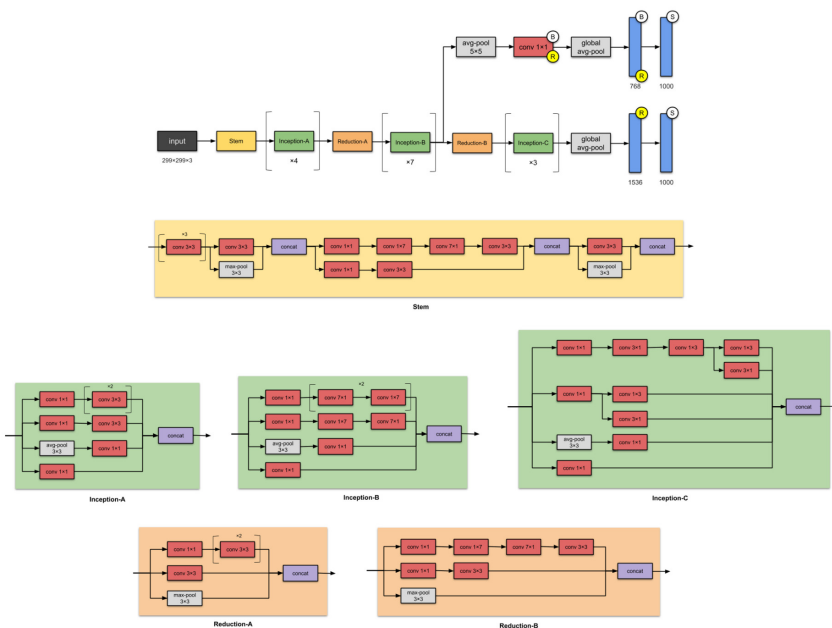


圖 8: Inception-v4 網路架構 這個 CNN 有兩個輔助網路 (在推理時會被丟棄)。注: 所有卷積層採用batch norm 和 ReLU 啟用。

Google 的研究人員再次提出了 Inception-v4 (43M個引數)。這是對 Inception-v3 的一次改進，主要的差別是 Stem 組和對 Inception-C 模組的小改動。該論文的作者也“對每種網格大小的 Inception 板塊增加了 Uniform 選擇”。他們也提到採用“殘差連線可以大幅度提高訓練速度”。

總之，值得注意的是，論文提到 Inception-v4 因為模型規模增大效果更好。

**與之前的Inceptio-v3版本相比，有哪些改進？**

1. 改變了 Stem 模組。
2. 增加了更多的 Inception 模組。

3. 採用了 Inception-v3 的 Uniform 選擇，意味著在每個模組中採用了相同數量的過濾器。

## 相關論著

### 論文:

Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning

連結: [https://arxiv.org/abs/1602.07261?](https://arxiv.org/abs/1602.07261?source=post_page)  
source=post\_page

- 作者: Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alex Alemi. Google.
- 發表在: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence

## 9. Inception-ResNet-V2 (2016)

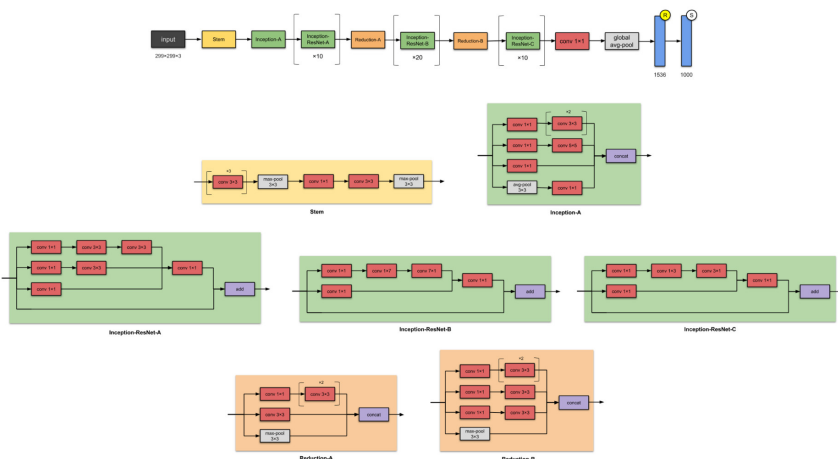


圖 9: Inception-ResNet-V2 網路結構。注：所有卷積層採用batch norm和 ReLU 啟用。

在提出 Inception-v4 的同一篇論文中，作者們也提出了 Inception-ResNet: Inception-ResNet-v1 和 nception-ResNet-v2 網路系列，v2 系列有 56M 個引數。

**與之前的Inception-v3版本相比，有哪些改進？**

1. 把 Inception 模組轉換為殘差 Inception 模組。

2. 增加了更多的 Inception 模組。
3. 在 Stem 模組後面，增加了一種新型的 Inception 模組 (Inception-A) 。

## 相關論著

### 論文:

Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning

連結: [https://arxiv.org/abs/1602.07261?](https://arxiv.org/abs/1602.07261?source=post_page)  
source=post\_page

- 作者: Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alex Alemi. Google
- 發表在: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence

## 10. ResNeXt-50 (2017)

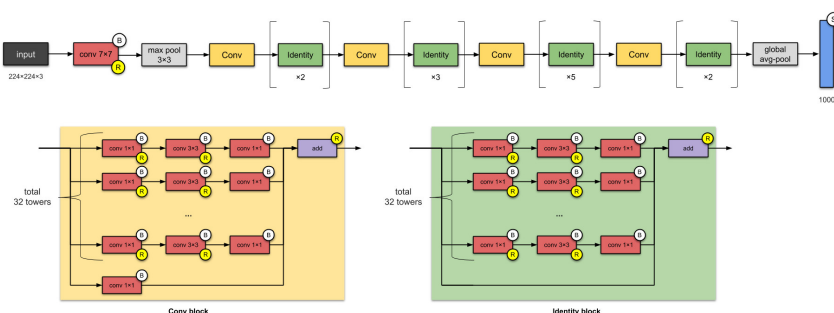


圖 10: ResNeXt 網路架構

如果你想起了 ResNet，對，它們是相關的。ResNeXt 有 25 M 個引數(ResNet-50 有 25.5M 個)。它們之間的差別是 ResNeXt 在每個模組上增加了並行塔/分支、路徑。上圖總計有 32 個 tower。

### 有哪些創新？

1. 在一個模組中增加了並行塔的數量（基數）。

## 相關論著

- 論文:

Aggregated Residual Transformations for Deep  
Neural Networks

連結: [https://arxiv.org/abs/1611.05431?](https://arxiv.org/abs/1611.05431?source=post_page)  
[source=post\\_page](https://arxiv.org/abs/1611.05431?source=post_page)

- 作者: Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, Kaiming He. University of California San Diego, Facebook Research
- 發表在: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

## 附錄: Network In Network (2014)

我們注意到，在一個卷積中，畫素的值是過濾器中權重和當前滑動視窗的線性組合。考慮一個只有 1 個隱藏層的 mini 神經網路。這是他們為什麼把它叫做 Mlpconv 的原因。我們要處理的網路就是一個只有 1 個隱藏層的（卷積神經）網路。

Mlpconv 的思想和 $1 \times 1$  卷積核密切相關，成為 Inception 網路架構的主要特徵。

### 有哪些創新？

1. MLP 卷積層， $1 \times 1$  卷積。
2. 全域性平均池化（對每個特徵圖進行平均，把結果向量反饋到 Softmax 層）。

## 相關論著

- 論文:

Network In Network



連結: [https://arxiv.org/abs/1312.4400?  
source=post\\_page](https://arxiv.org/abs/1312.4400?source=post_page)

- 作者: Min Lin, Qiang Chen, Shuicheng Yan. National University of Singapore
- 發表在: arXiv preprint, 2013

下面列出了可以让你對神經網路視覺化的相關資源:

- Netron
- TensorBoard API  
  
by TensorFlow
- plot\_model API  
  
by Keras
- pytorchviz package

### 參考文獻

作者在文中使用了提出這些 CNN 網路結構的論文作為參考文獻。除了這些論文，作者列出了一些本文中的其它參考文獻:

- [https://github.com/tensorflow/models/tree/master/research/slim/nets\(github.com/tensorflow\)](https://github.com/tensorflow/models/tree/master/research/slim/nets(github.com/tensorflow))
- Implementation of deep learning models from the Keras team([github.com/keras-team](https://github.com/keras-team))
- Lecture Notes on Convolutional Neural Network Architectures: from LeNet to ResNet([slazebni.cs.illinois.edu](https://slazebni.cs.illinois.edu))
- Review: NIN — Network In Network (Image Classification)([towardsdatascience.com](https://towardsdatascience.com))

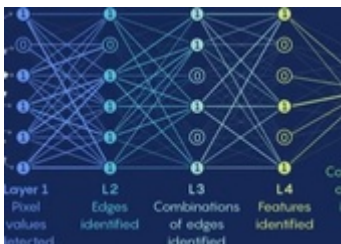


你點的每個“在看”，我都認真當成了AI

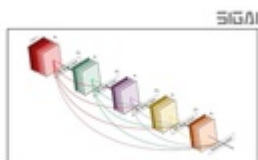
## 相關文章



大量總 | 一文學會八篇經典  
CNN論文



圖解十大 CNN 架構



基於內容的影像檢索技術綜  
述-CNN方法



【附論文】白話CNN經典模  
型：GoogLeNet（從  
Inception v1到v4的演進）



## 從LeNet到GoogLeNet： 逐層詳解，看卷積神經網路 的進化

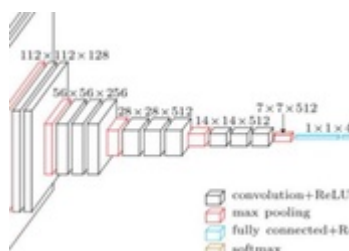
## CNN結構演變總結（一）經典模型

√ 结构演变总结（一）经典模型

## MobileNet系列之 MobileNet\_v2

NetV2: Inverted Residuals and Linear Bottlenecks

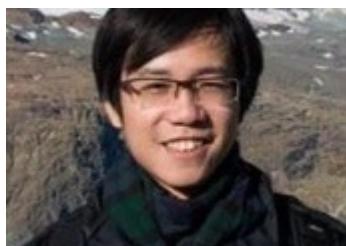
Andrew Howard, Menglong Zhu, Andrey Zhmoginov,  
Google Inc.,  
{adler, howarda, menglong, azhmogin, lichen}@google.com



## 一份對Keras中深度學習模型進行微調的全面指南（第二部分）

## 乾貨 | ImageNet冠軍模型 SE-Net詳解

IMAGENET



## CVPR 引用量最高的10篇論文！何愷明ResNet登頂，YOLO佔據兩席！

Copyright © 2022 | Powered by sa123.cc