**Sample Code & 作業內容**

請參考範例程式碼Day67-Keras\_Dataset\_Introduce，請嘗試改用CIFAR100，作業請提交Day67-Keras\_Dataset\_HW

[檢視範例](https://ai100-2.cupoy.com/samplecodelist/D67)

**參考資料**

**延伸閱讀**

* **Keras: The Python Deep Learning library**

[**https://github.com/keras-team/keras/**](https://github.com/keras-team/keras/)

* Keras dataset

[**https://keras.io/datasets/**](https://keras.io/datasets/)

* Predicting Boston House Prices

[**https://www.kaggle.com/sagarnildass/predicting-boston-house-prices**](https://www.kaggle.com/sagarnildass/predicting-boston-house-prices)

**其餘公開數據集介紹**

**Imagenet**

Imagenet數據集有1400多萬幅圖片，涵蓋2萬多個類別；其中有超過百萬的圖片有明確的類別標註和圖像中物體位置的標註，具體信息如下：

1）Total number of non-empty synsets : 21841

2）Total number of images: 14,197,122

3）Number of images with bounding box annotations: 1,034,908

4）Number of synsets with SIFT features: 1000

5）Number of images with SIFT features: 1.2 million

Imagenet數據集是目前深度學習圖像領域應用得非常多的一個領域，關於圖像分類、定位、檢測等研究工作大多基於此數據集展開。Imagenet數據集文檔詳細，有專門的團隊維護，使用非常方便，在計算機視覺領域研究論文中應用非常廣，幾乎成為了目前深度學習圖像領域算法性能檢驗的“標準”數據集。數據集大小：~1TB（ILSVRC2016比賽全部數據）  
下載地址：  
[**http://www.image-net.org/about-stats**](http://www.image-net.org/about-stats)

**COCO**

COCO(Common Objects in Context)是一個新的圖像識別、分割和圖像語義數據集，它有如下特點：

1）Object segmentation

2）Recognition in Context

3）Multiple objects per image

4）More than 300,000 images

5）More than 2 Million instances

6）80 object categories

7）5 captions per image

8）Keypoints on 100,000 people

COCO數據集由微軟贊助，其對於圖像的標註信息不僅有類別、位置信息，還有對圖像的語義文本描述，COCO數據集的開源使得近兩三年來圖像分割語義理解取得了巨大的進展，也幾乎成為了圖像語義理解算法性能評價的“標準”數據集。

Google開源的開源了圖說生成模型show and tell就是在此數據集上測試的，想玩的可以下下來試試。數據集大小：~40GB   
下載地址：[**http://mscoco.org/**](http://mscoco.org/)