Kaggle 比賽介紹 : 漁業監控

2017/04/24 R Ladies Event @ Dcard

By 智程

About me

- Linkernetworks 資料科學家
- Taipei Ethereum Meetup Co-organizer

這比賽在比什麼

大海茫茫 漁船作業監管不易

任務是分辨六個魚種

長鰭鮪
大目鮪
鬼頭刀
月魚
鯊魚
黃鰭鮪

ALB: Albacore tuna (*Thunnus alalunga*)

BET: Bigeye tuna (*Thunnus obesus*)

DOL: Dolphinfish, Mahi Mahi (Coryphaena hippurus)

LAG: Opah, Moonfish (*Lampris guttatus*)

SHARK: Various: Silky, Shortfin Mako

YFT: Yellowfin tuna (*Thunnus albacares*)

Fish images are not to scale with one another

學到什麼

Kernels 與論壇挖寶

1,355 Kernels

Finding BoatIDs
119 Votes / 2 months ago / Python

Using InceptionV3 features - SVM classifier
5 Votes / 3 days ago / Python

Fish detection
46 Votes / 60 days ago / Python

Fishy Keras [LB: 1.25267]
63 Votes / 2 months ago / Python

mxnet + xgboost simple solution
48 Votes / 2 months ago / Python

fish_keras_test
7 Votes / 17 days ago / Python

Notebook | Code | Comments (1) | Log | Versions (4) | Forks (3) | Fork Notebook

```
SHARK
YFT
```

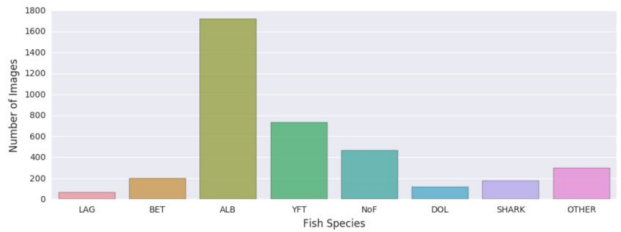
So there are 8 folders present inside the train folder, one for each species.

Now let us check the number of files present in each of these sub folders.

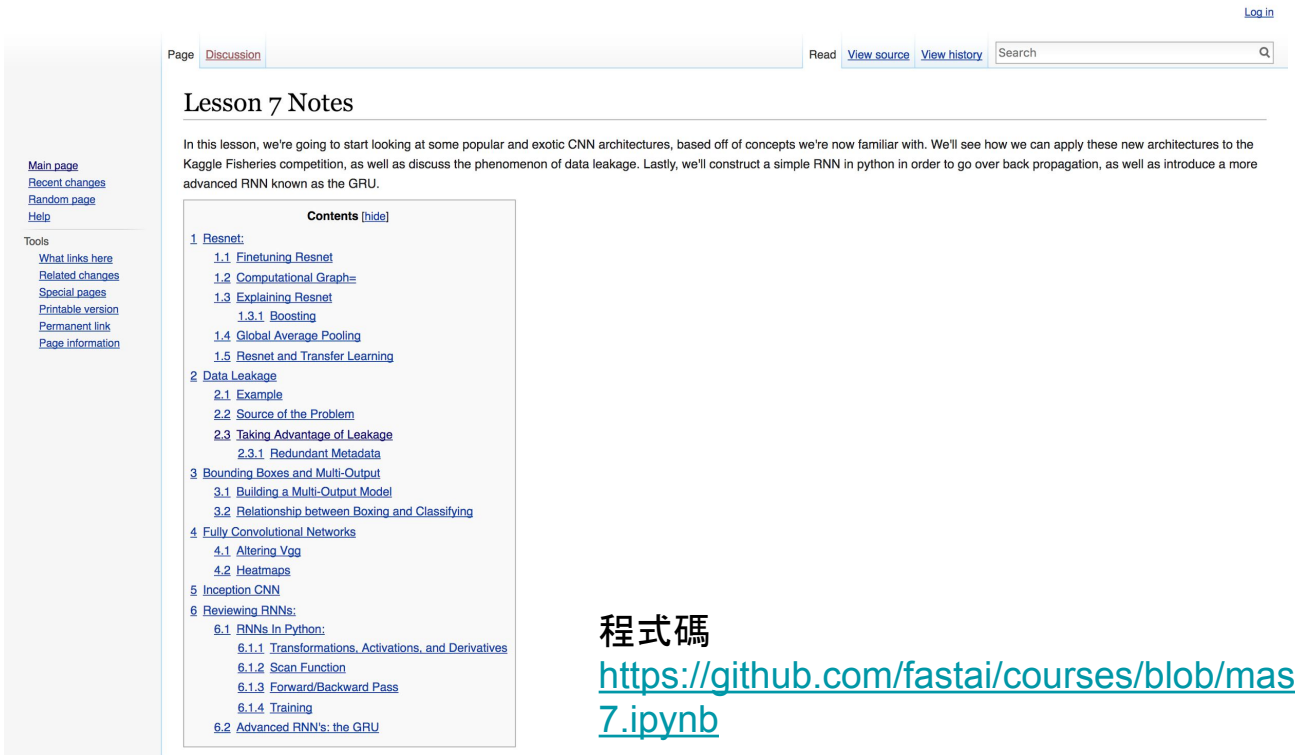
In [2]:
```
sub_folders = check_output(["ls", "../input/train/"]).decode("utf8").strip().split('\n')
count_dict = {}
for sub_folder in sub_folders:
    num_of_files = len(check_output(["ls", "../input/train/"+sub_folder]).decode("utf8").st
rip().split('\n'))
    print("Number of files for the species",sub_folder,":",num_of_files)
    count_dict[sub_folder] = num_of_files

plt.figure(figsize=(12,4))
sns.barplot(list(count_dict.keys()), list(count_dict.values()), alpha=0.8)
plt.xlabel('Fish Species', fontsize=12)
plt.ylabel('Number of Images', fontsize=12)
plt.show()
```

```
Number of files for the species ALB : 1719
Number of files for the species BET : 200
Number of files for the species DOL : 117
Number of files for the species LAG : 67
Number of files for the species NoF : 465
Number of files for the species OTHER : 299
Number of files for the species SHARK : 176
Number of files for the species YFT : 734
```



挖到 Fast.ai 第七門課



Log in

Page Discussion | Read View source View history | Search

Lesson 7 Notes

In this lesson, we're going to start looking at some popular and exotic CNN architectures, based off of concepts we're now familiar with. We'll see how we can apply these new architectures to the Kaggle Fisheries competition, as well as discuss the phenomenon of data leakage. Lastly, we'll construct a simple RNN in python in order to go over back propagation, as well as introduce a more advanced RNN known as the GRU.

Main page
Recent changes
Random page
Help

Tools
What links here
Related changes
Special pages
Printable version
Permanent link
Page information

Contents [hide]

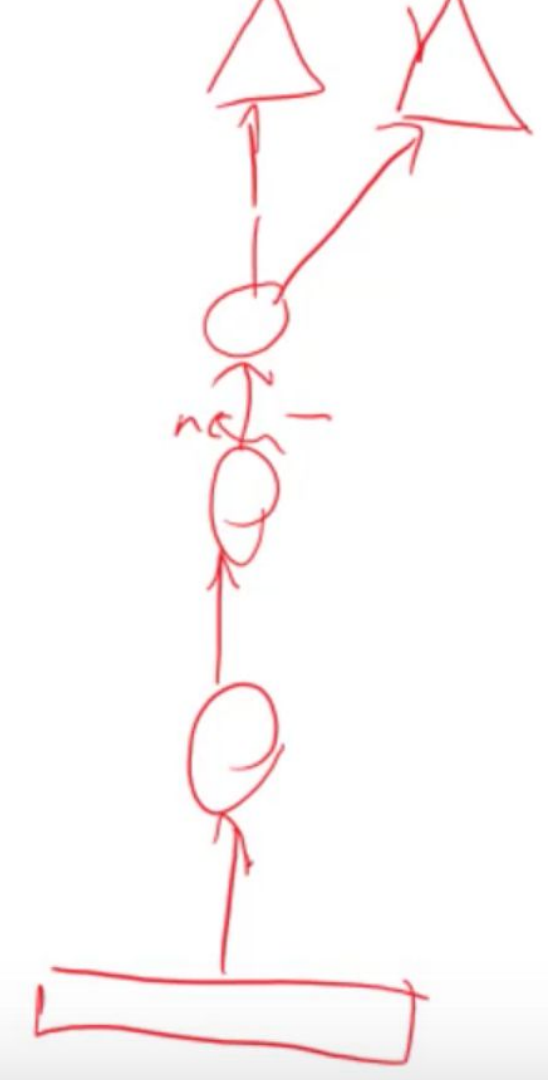
程式碼
https://github.com/fastai/courses/blob/master/deeplearning1/nbs/lesson7.ipynb

觀念：Data Leakage

- 你的模型會偷吃步
  - 模型會無所不用其極提高準確度。
  - 例如：同一台船可能都抓同一種魚，因此模型可能會辨識船的資訊來猜測魚種
- 千萬不要偷吃步
  - 例如：某船在捕捉非法魚種時，模型認船不認魚就會猜錯了。

技巧：Bounding Boxes and Multi-Output

- 論壇有人提供 Bbox 資料。
- 模型同時要輸出分類與 bbox，會有暗示中間層的效果



技巧: Heatmap 技

- 把神經網路中間層視覺化
- 可確認模型看著該看的東西

延伸議題

台灣是遠洋漁業的大玩家

「根據「中西太平洋漁業委員會」(WCPFC) 2014年的統計，在中西太平洋水域的鰹魚、黃鰭鮪(黃鰭金槍魚)、大目鮪(大目金槍魚)漁獲量，台灣以27.3萬公噸排名第三，僅輸給美國、韓國。其中價值最高的大目鮪漁獲量更高達8476噸，位居世界第二，堪稱太平洋鮪魚的霸主。」

參照到文章吧

Chih-Cheng Liang
chihchengliang@gmail.com
Jan 22 · 2 min read

Kaggle 比賽介紹：漁業監控

比賽頁面影片截圖：辨識漁船上作業的魚種示意圖

大約 2015 年的時候，我就意識到除非能夠認真放下手邊的工作一陣子，才有