INVIGOUR ENERGY
OIL & GAS CONSULTANCY

Superman vs Bird (or whatever) for Dummies

Lesson 1 Exercise (way overdue)

By Zaim – A noob

Why Superman and bird?

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

It's a Bird... It's a Plane... It's Superman

Original Broadway Cast Recording

Music	Charles Strouse
Lyrics	Lee Adams
Book	David Newman Robert Benton
Basis	*Superman* by Jerry Siegel and Joe Shuster
Productions	1966 Broadway 1975 ABC TV special 2007 Los Angeles Concert 2010 Dallas 2013 New York City Encores! 2014 London 2015 West End 2016 Germany

Contents

Contents

Dummies

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

- Non IT background
- Little coding experience
- May not even know how to transfer files in terminals
- Didn’t know the difference between Machine Learning and Deep Learning

Dummies

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

- Non IT background
- Little coding experience
- May not even know how to transfer files in terminals
- Didn’t know the difference between Machine Learning and Deep Learning

- Lots of logistics nightmare

Contents

General steps

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

cnnsuperbird

cnn fast ai using superman bird data set I downloaded from google

General steps

1. Download superman and bird images from google using someone's python scripts
2. Delete bad images
3. Start formatting the filenames using a simple python script I wrote
4. Separate the files into training and validation folders
5. upload to this github
6. clone back to my google cloud
7. run the jupyter notebook
8. experiment with batch size since the images are not a lot
9. also experiment with data augmentation, and the settings for the learn including annealing and differential learning rate

https://github.com/zaimawang/cnnsuperbird/blob/master/README.md

In the process I learned to create my own github

Contents

Challenge : Download Images

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

- Manually?

Challenge : Download Images

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

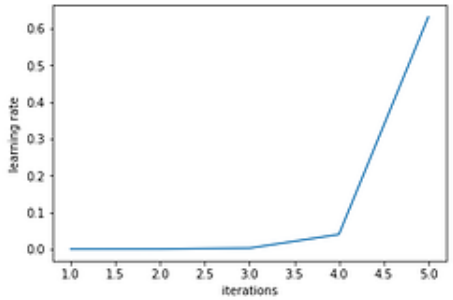
- Issues with small sample size

Wiki: Lesson 1

Part 1 v2

3. I used a batch size of 15. Still my learning rate schedule doesn't work

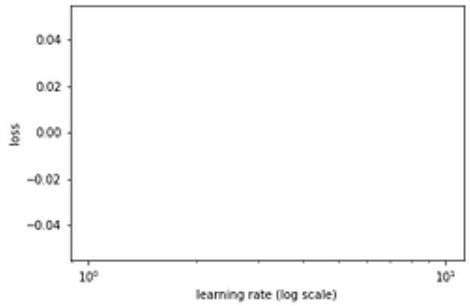
```
In [85]: learn.sched.plot_lr()
```



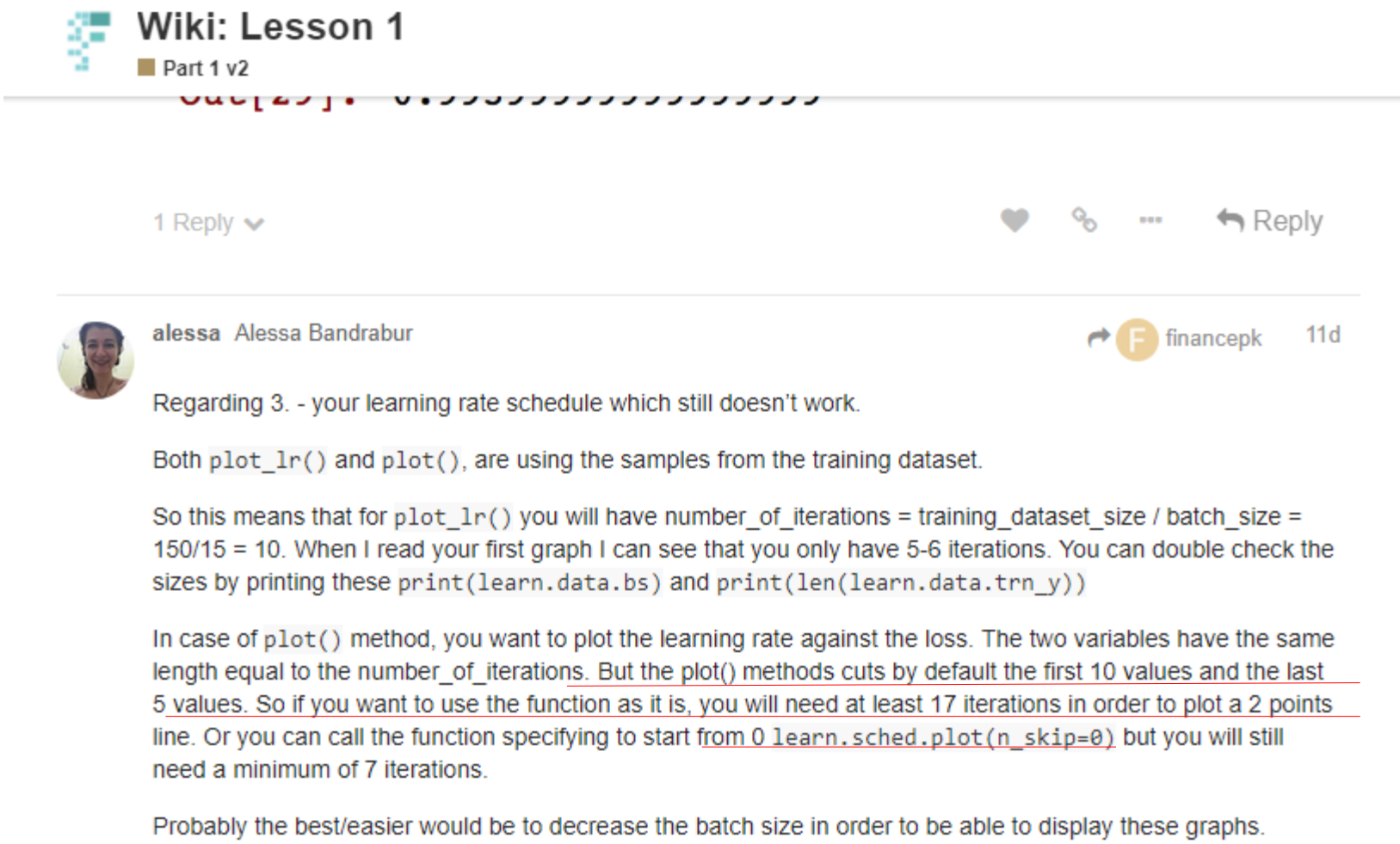
Note that in the previous plot *iteration* is one iteration (or *minibatch*) of SGD. In one epoch there a

We can see the plot of loss versus learning rate to see where our loss stops decreasing:

```
In [86]: learn.sched.plot()
```



Issue with small number of images, try minimum of 50



Wiki: Lesson 1

Part 1 v2

1 Reply

Reply

alessa Alessa Bandrabur

financepk 11d

Regarding 3. - your learning rate schedule which still doesn't work.

Both `plot_lr()` and `plot()`, are using the samples from the training dataset.

So this means that for `plot_lr()` you will have number_of_iterations = training_dataset_size / batch_size = 150/15 = 10. When I read your first graph I can see that you only have 5-6 iterations. You can double check the sizes by printing these `print(learn.data.bs)` and `print(len(learn.data.trn_y))`

In case of `plot()` method, you want to plot the learning rate against the loss. The two variables have the same length equal to the number_of_iterations. But the plot() methods cuts by default the first 10 values and the last 5 values. So if you want to use the function as it is, you will need at least 17 iterations in order to plot a 2 points line. Or you can call the function specifying to start from 0 `learn.sched.plot(n_skip=0)` but you will still need a minimum of 7 iterations.

Probably the best/easier would be to decrease the batch size in order to be able to display these graphs.

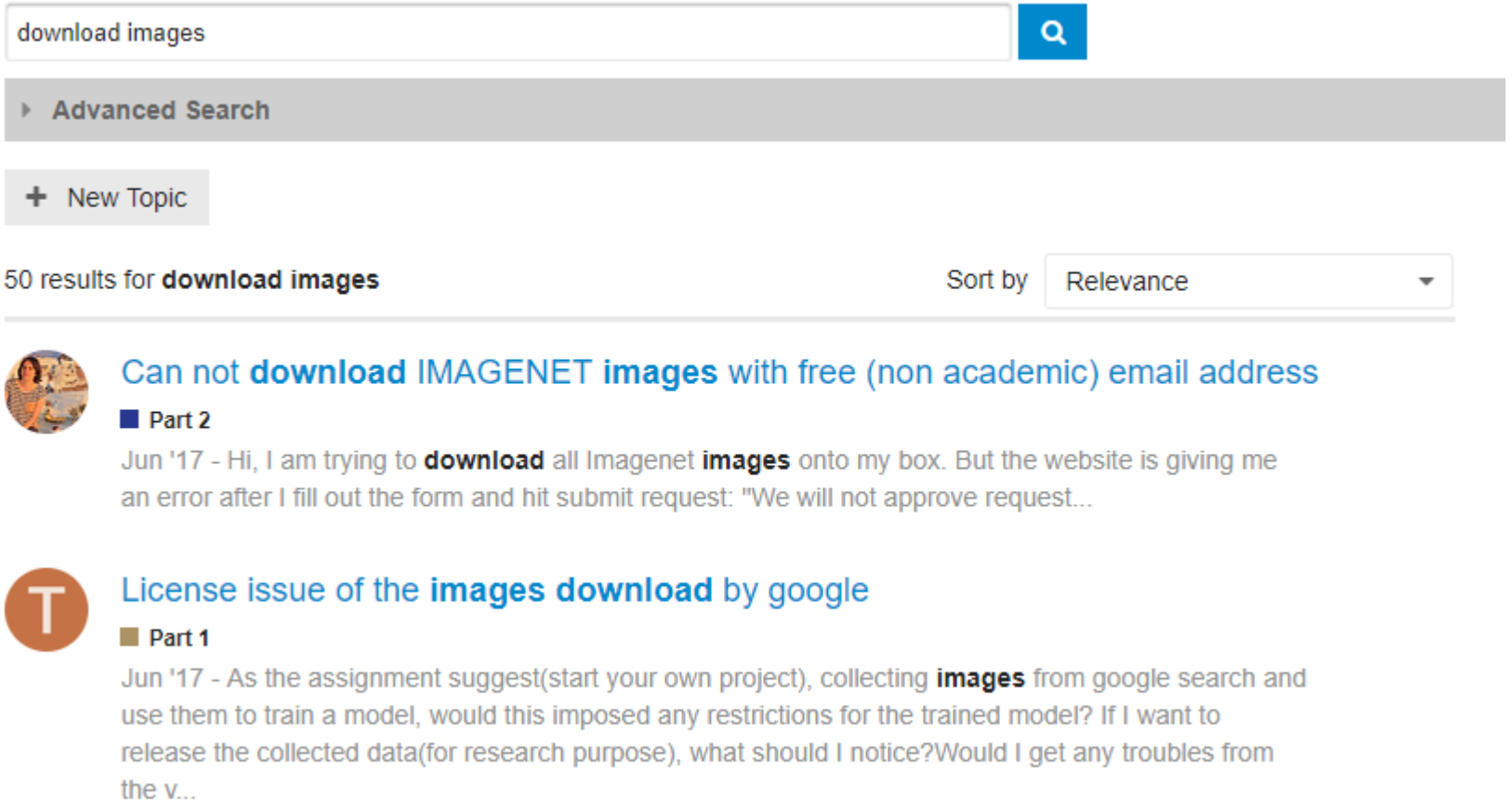
bs = 3

Contents

Download images : How? (1) Search forum.fast.ai

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

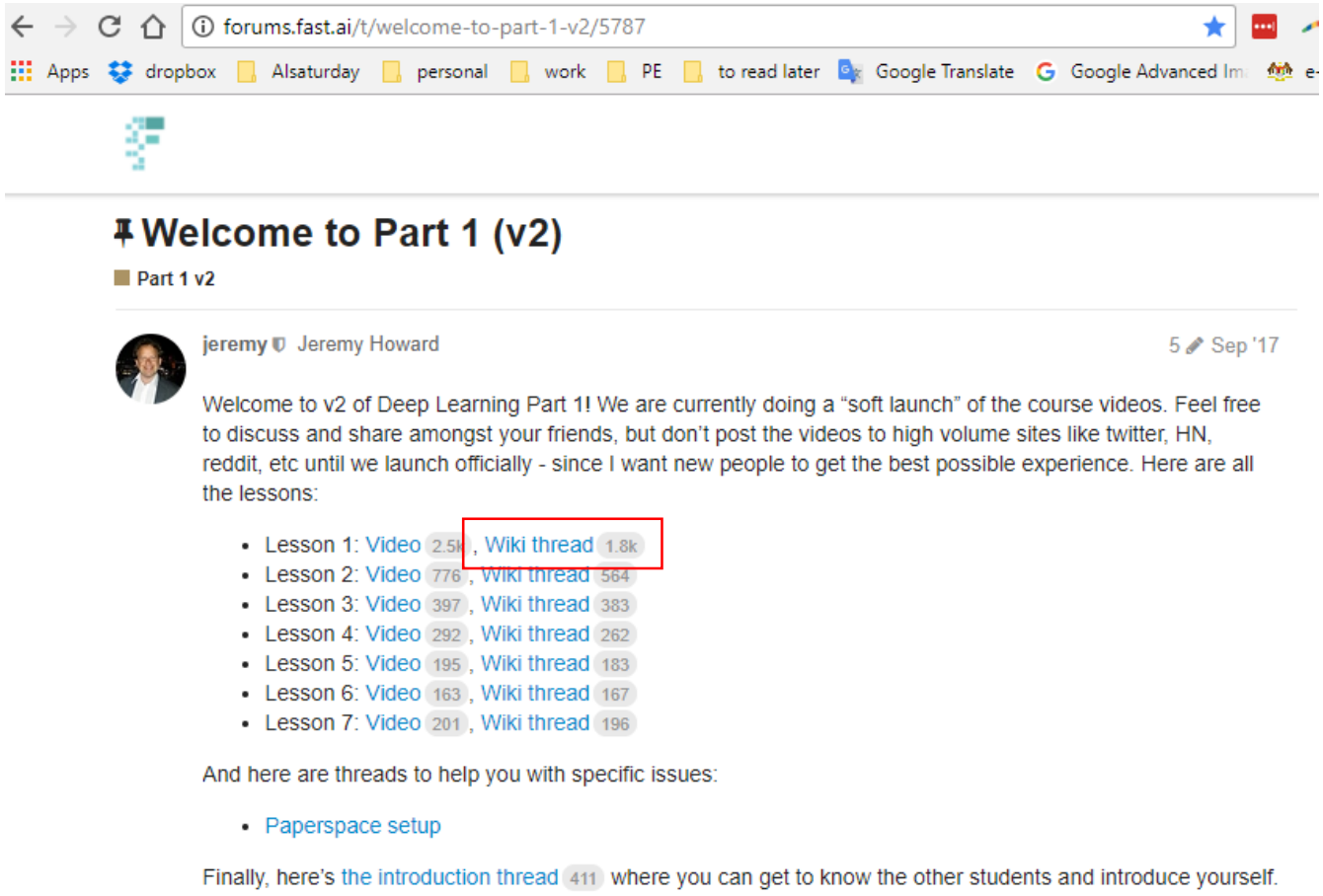
How? Search download imags at forums.fast.ai



Download images : How? (2) Use lessons wiki

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

http://forums.fast.ai/t/welcome-to-part-1-v2/5787



How to download images.
I found a python script

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

Wiki: Lesson 1

Part 1 v2

ecdrid Aditya

financepk 22d

Use this to download images(it works like a charm…
Tried and tested)

github.com 49

hardikvasa/google-images-download

google-images-download - Python Script to download hundreds of images from 'Google Images'. It is a ready-to-run code!

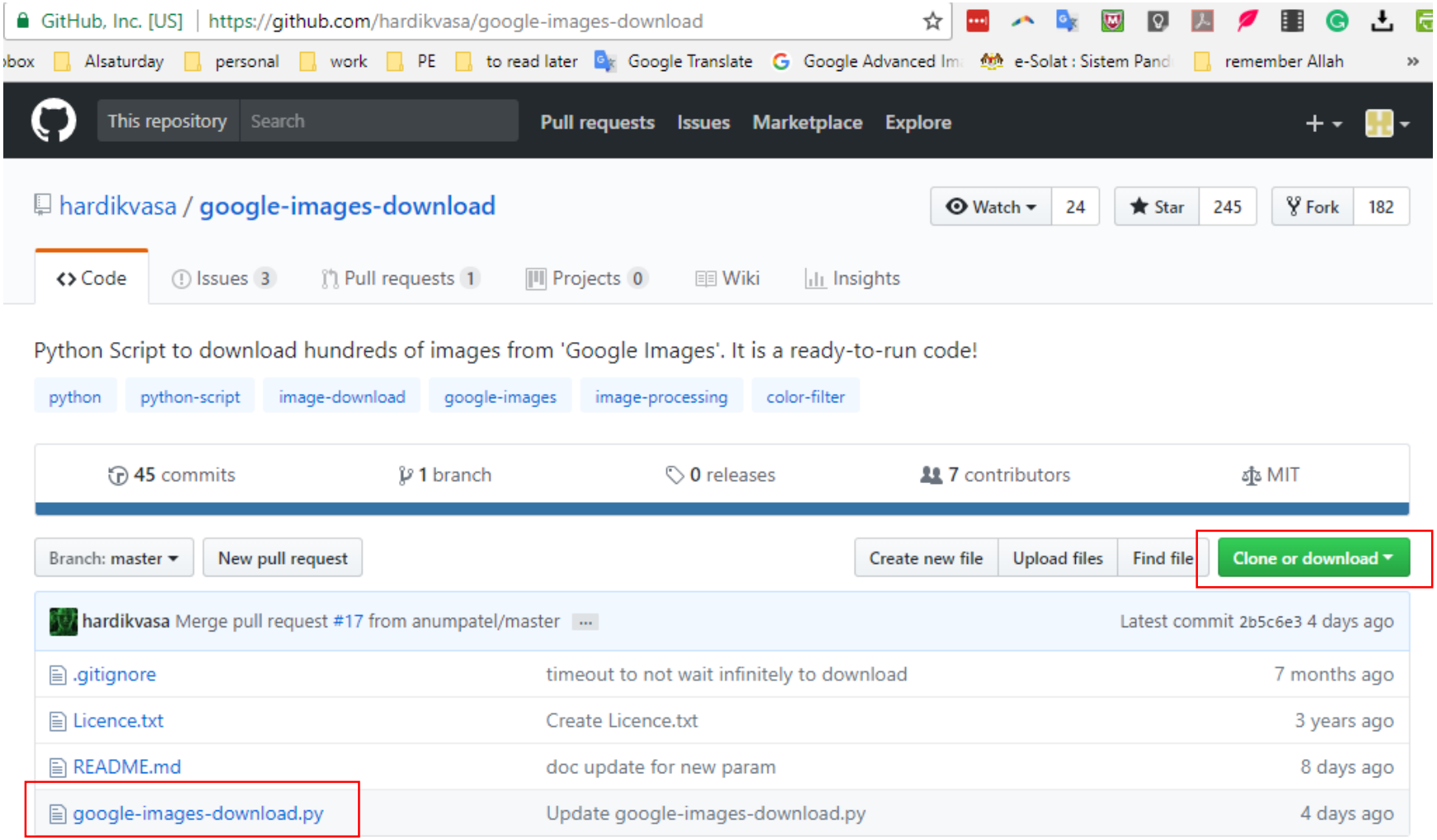
5 Likes Reply

Contents

Downloading images : Python Program

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

https://github.com/hardikvasa/google-images-download



Run the image downloader

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

Google Cloud SDK Shell

Go to the directory where the python file is

```
Welcome to the Google Cloud SDK! Run "gcloud -h" to get the list of available commands.
---

C:\Users\ACER\AppData\Local\Google\Cloud SDK>cd ../../../../Downloads/google-images-download-master

C:\Users\ACER\Downloads\google-images-download-master>
```

-k is keyword, -l is limit (how many)

```
C:\Users\ACER\Downloads\google-images-download-master>python google-images-download.py -k "superman" -l 120

Item no.: 1 --> Item name = superman
Evaluating...
Total Image Links = 100
Total time taken: 13.817165133633728 Seconds
Starting Download...
completed ====> 1. 250px-SupermanRoss.png
completed ====> 2. lead_960.jpg
completed ====> 3. 1330222e9a6df3b80d8c21ecffc0c600._SX1280_QL80_TTD_.jpg
completed ====> 4. GalleryChar_1900x900_MOS_52e05e3fe24a61.04593858.jpg
completed ====> 5. superman-returns.jpg
completed ====> 6. maxresdefault.jpg
completed ====> 7. dc-comics-batman-v-superman-superman-half-scale-polystone-statue-prime-1-902664-01.jpg
completed ====> 8. 71BDjMoIXYL._SL1500_.jpg
completed ====> 9. 170px-Superman_S_symbol.svg.png
completed ====> 10. latest.jpg
```

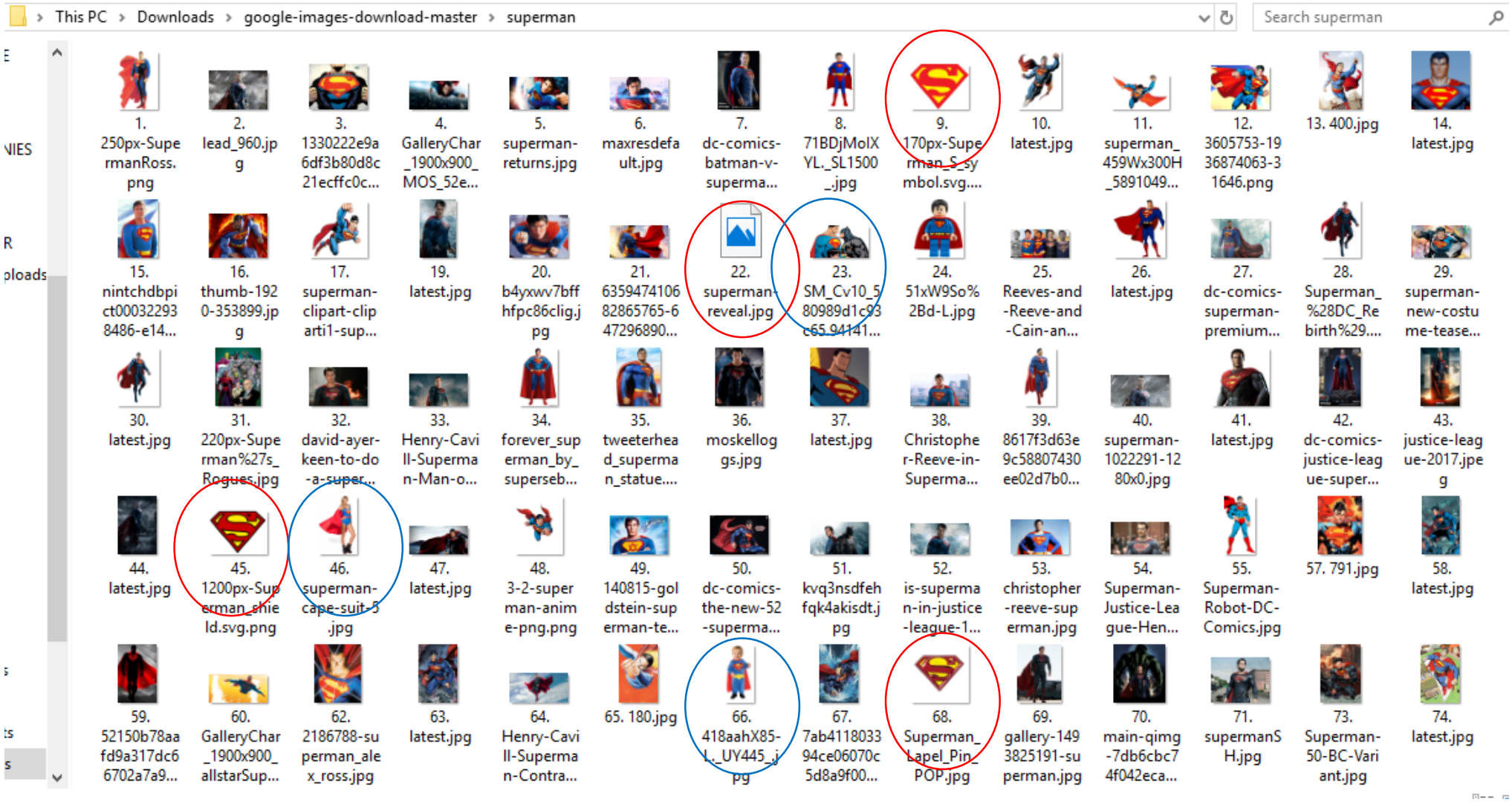
Somehow the limit is 100, google limit?

Repeat for “bird”, “cats” or “dogs” or whatever

Contents

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

Delete non-relevant data



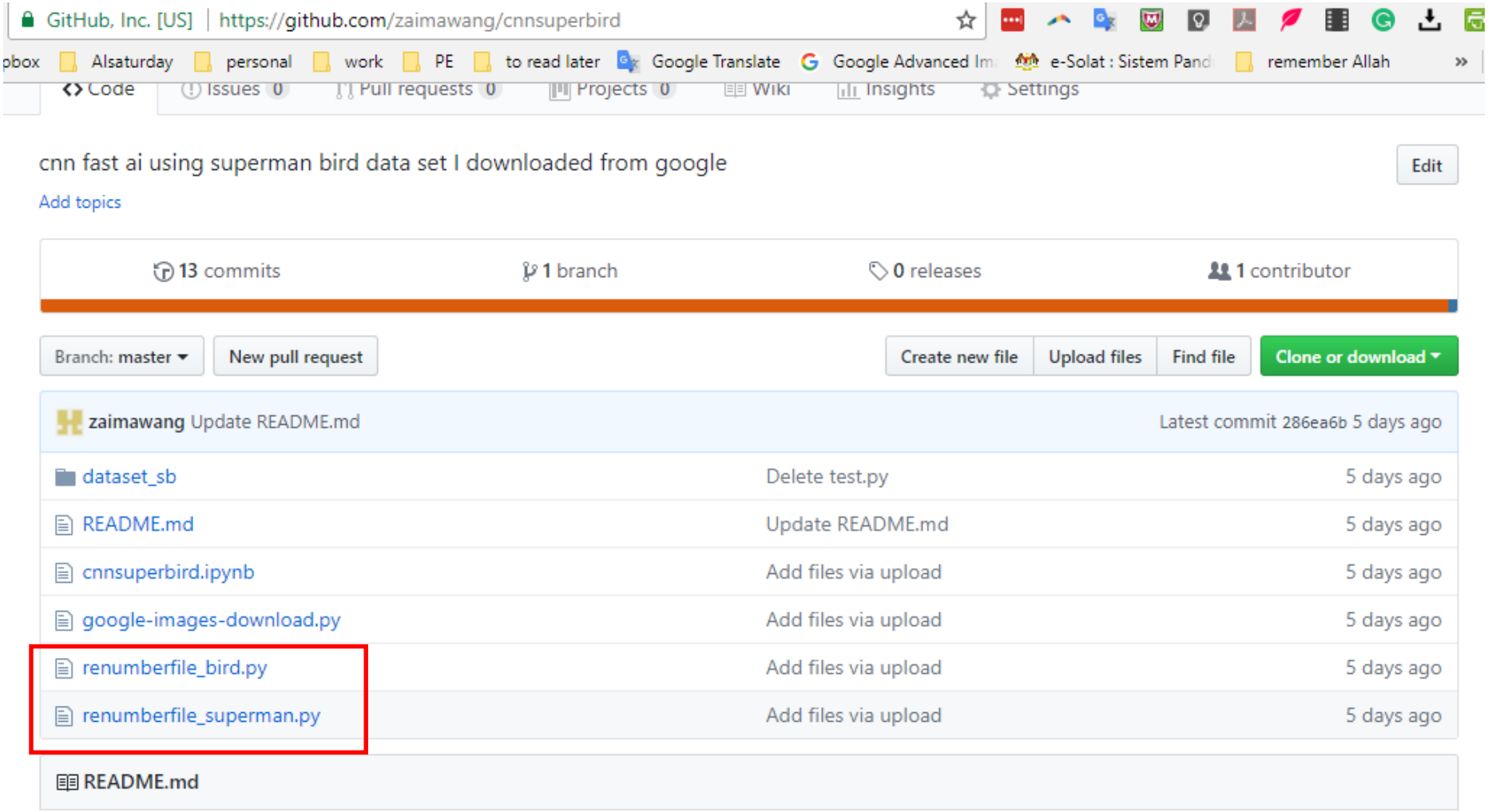
Repeat for bird

Contents

Renaming Files? Use my python script

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

https://github.com/zaimawang/cnnsuperbird



Set your path and filename (superman/bird)

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

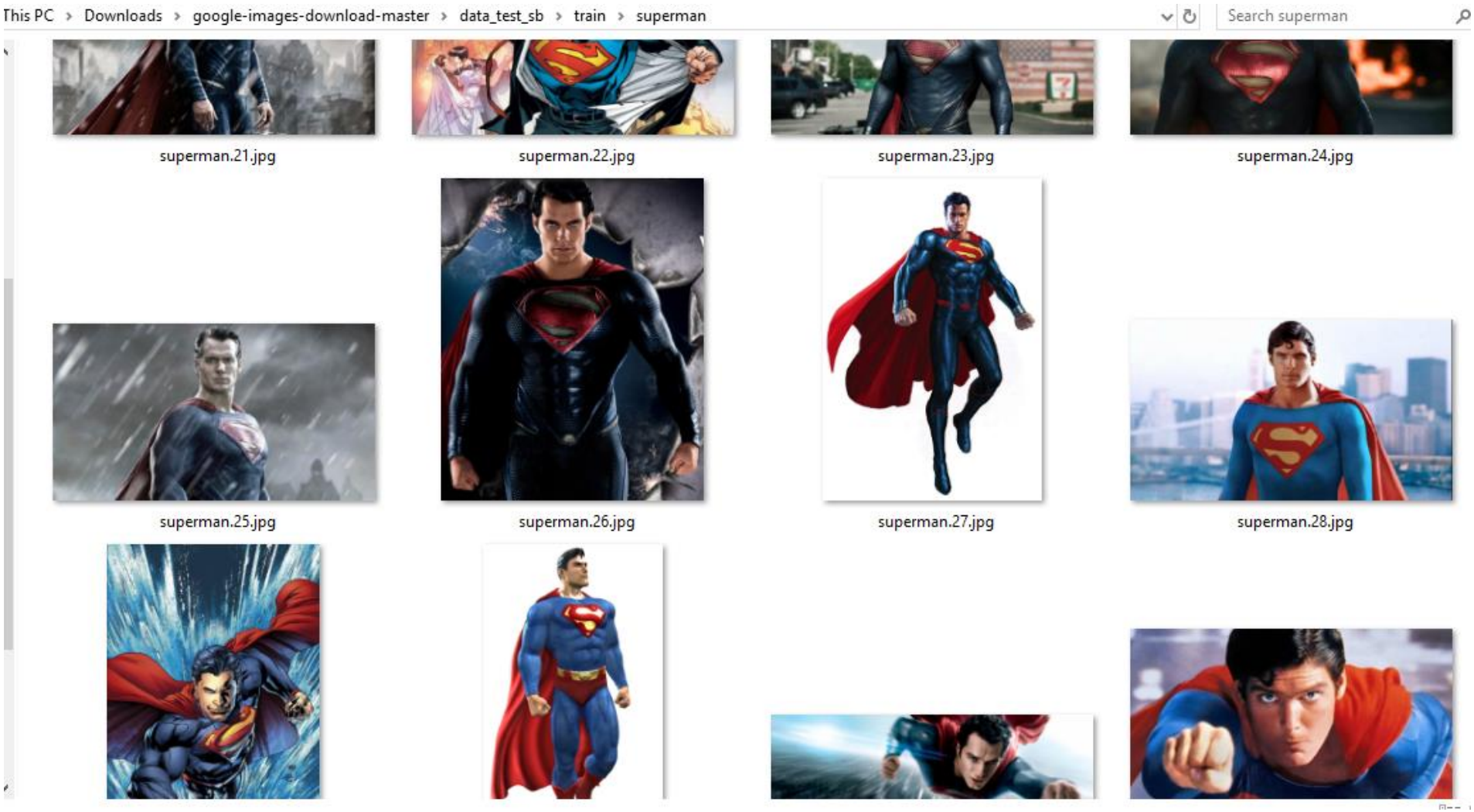
renumberfile_superman.py | google-images-download.py

```
1   # os is a library that gives us the ability to make OS changes
2   import os
3
4
5   def increment_file_numbers(directory, string, integer):
6       # iterate over every file name in the directory
7       print (directory)
8       file_number = startcount
9       for file_name in os.listdir(directory):
10          file_number = file_number + 1
11          extension = ".jpg"
12          file_name2 = objectname+"."
13          # example format : superman.01.jpg
14          new_file_name = '%s/%s%02d%s' % (directory, file_name2, file_number, extension)
15          old_file_name = '%s/%s' % (directory, file_name)
16          print (old_file_name)
17          print (new_file_name)
18          # rename the file!
19          os.rename(old_file_name, new_file_name)
20
21
22  # This is the path to the files
23  # C:\Users\ACER\Downloads\google-images-download-master\superman
24  PATH = os.path.abspath('/Users/ACER/Downloads/google-images-download-master/superman')
25  objectname = "superman"
26  startcount = 0
27
28  # Let's rename the files
29  increment_file_numbers(PATH, objectname, startcount)
30
```

Edit path and filename

```
C:\Users\ACER\Downloads\google-images-download-master>python renumberfile_superman.py
```

Files have been renamed and renumbered



Move half to a new folder called **train/superman** and another half to **valid/superman**

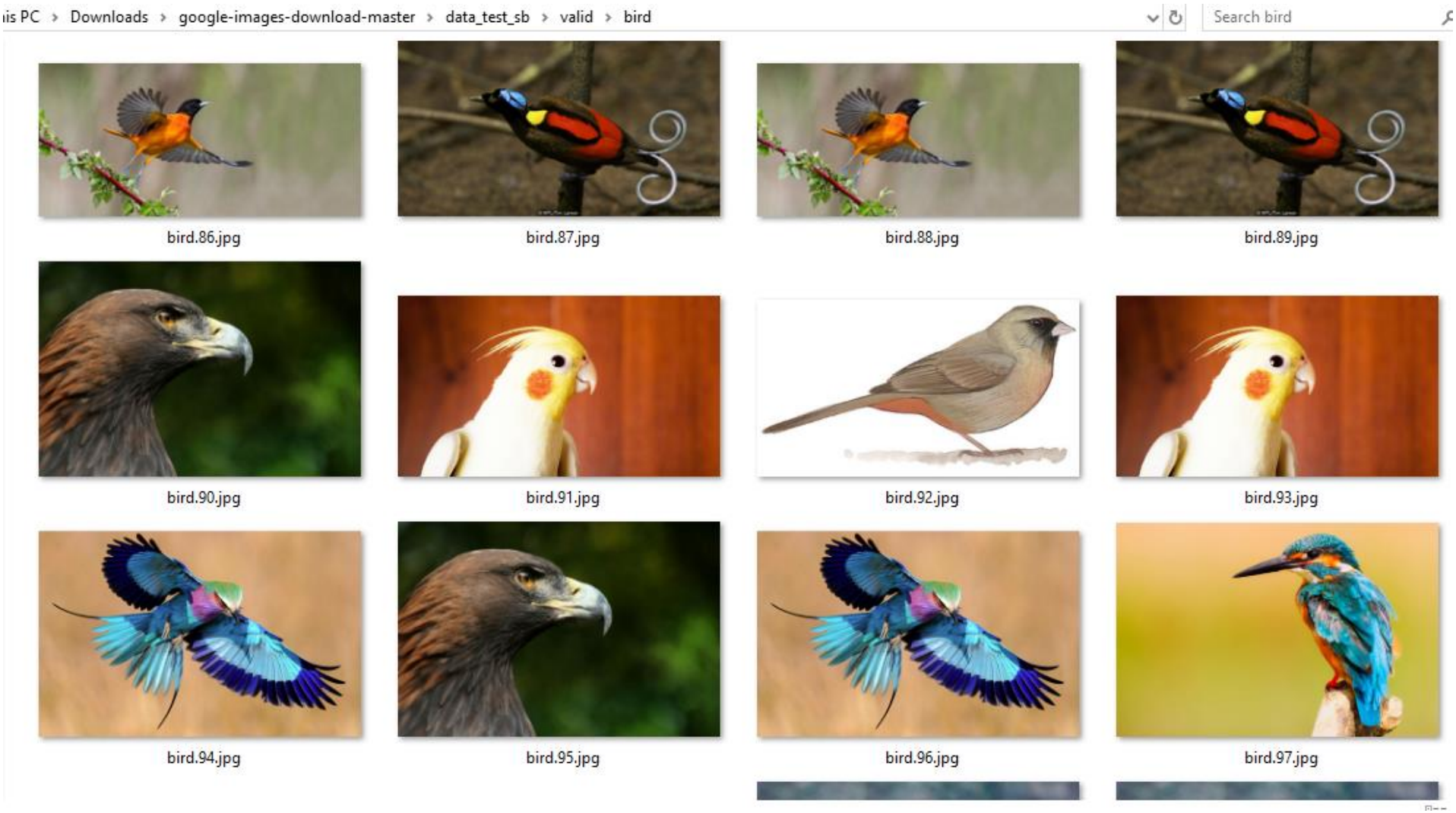
Folder structure

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

- Your Ipythonnotebook.ipynb sits in the same folder as superbirddata folder
- superbirddata folder
  - Train
    - superman
    - Bird
  - Valid
    - superman
    - Bird

Similarly for bird do the same

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

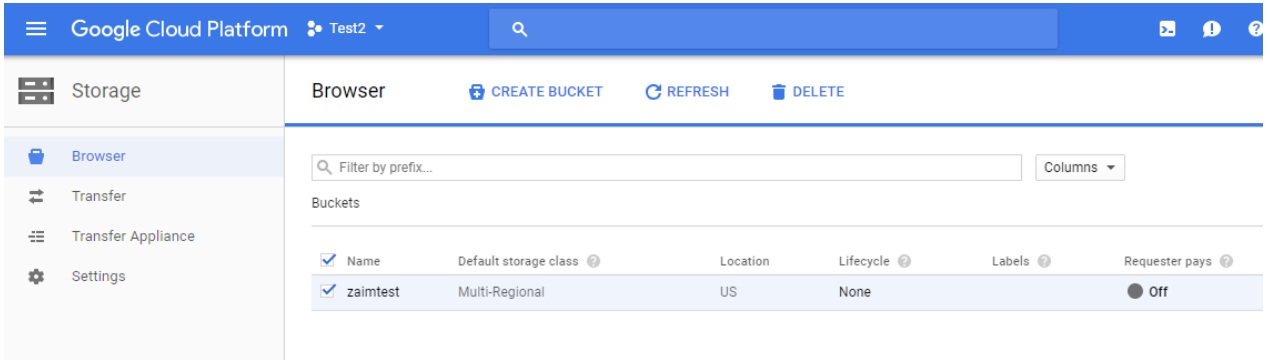


Contents

File transfer method 1 : (bucket and gsutil)

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

- But images now in pc but need to have it in the cloud
- There are a few ways to do it
  - Upload files to google cloud BUCKET, but create a bucket first (hint: search for bucket)



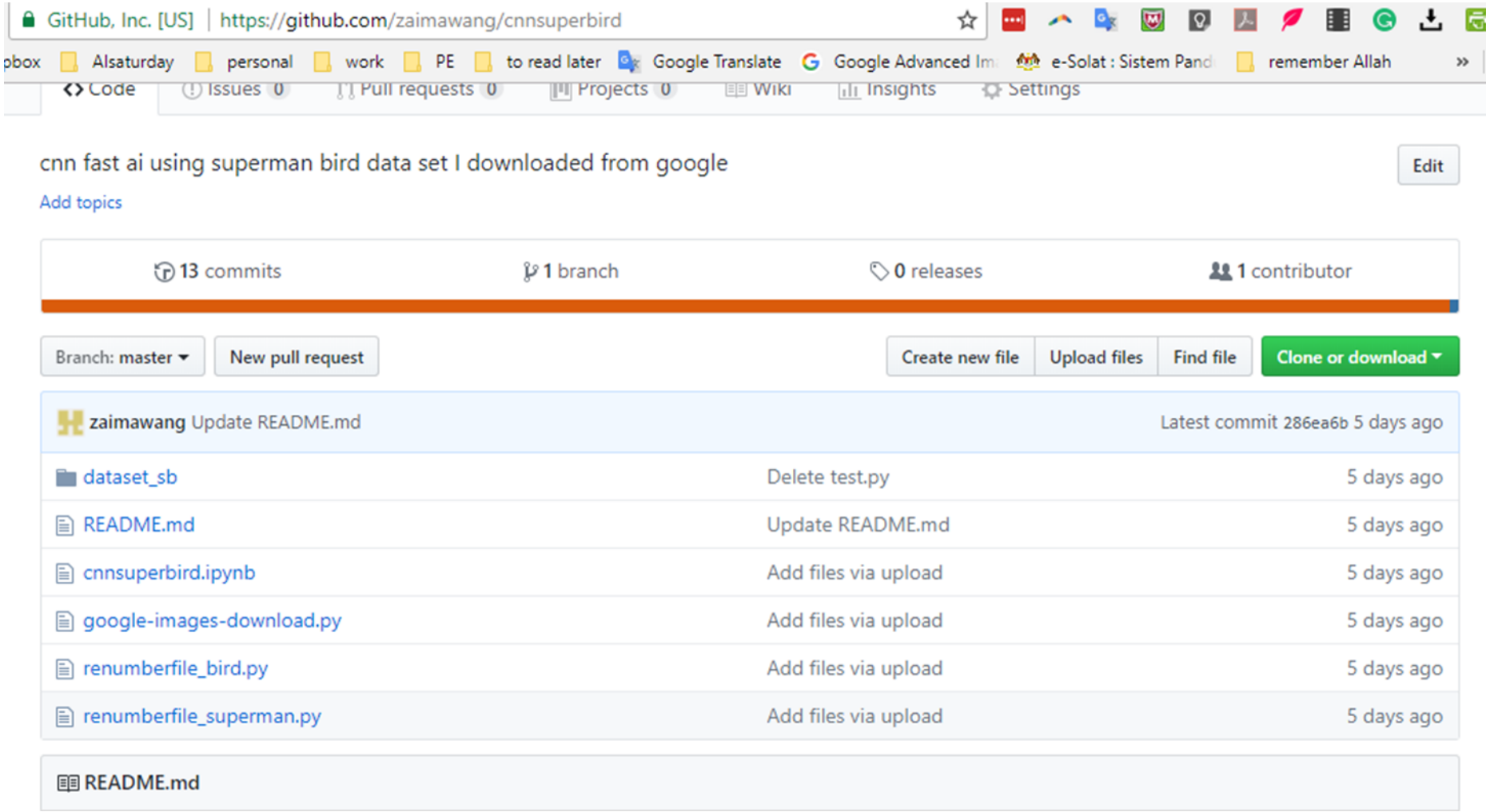
  - AND GSUTIL https://cloud.google.com/storage/docs/gsutil/commands/cp

```
(fastai) ACER@fastai-instance-1:~/fastai/courses/dl1$ gsutil cp -r gs://zaimtest .
```

Your jupyter notebook doesn't have access to the bucket, that is why we gsutil the the bucket to where jupyter notebook can see it

File transfer method 2 : (put in github and then git clone)

INVIGOUR ENERGY
OIL & GAS CONSULTANCY



Git cloning

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

ACER@fastai-instance-1: ~

```
(fastai) ACER@fastai-instance-1:~$ git clone https://github.com/zaimawang/cnnsuperbird
Cloning into 'cnnsuperbird'...
remote: Counting objects: 223, done.
remote: Compressing objects: 100% (9/9), done.
remote: Total 223 (delta 3), reused 0 (delta 0), pack-reused 214
Receiving objects: 100% (223/223), 56.08 MiB | 12.67 MiB/s, done.
Resolving deltas: 100% (12/12), done.
Checking connectivity... done.
(fastai) ACER@fastai-instance-1:~$ ls
anaconda3  cnnsuperbird  data  downloads  fastai
(fastai) ACER@fastai-instance-1:~$
```

Normally you want to git clone under fast/courses/dl1

File transfer method 3: Use your jupyter notebook

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

https://stackoverflow.com/questions/34734714/ipython-jupyter-uploading-folder

1 You could upload a zip file, then start a notebook and use the zipfile module from Python to extract it. – Thomas K Jan 12 '16 at 12:01

add a comment

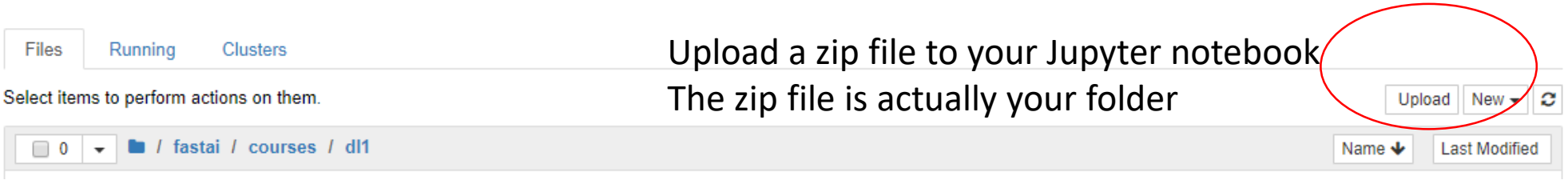
1 Answer

active oldest votes

Convert it into a single Zip file and upload that. to unzip the folder use the code down bellow

5

```
import zipfile as zf
files = zf.ZipFile("ZippedFolder.zip", 'r')
files.extractall('directory to extract')
files.close()
```



Upload a zip file to your Jupyter notebook
The zip file is actually your folder

A few other methods not discussed here

Contents

Run your jupyter notebook

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

jupyter cnnsuperbird Last Checkpoint: Last Sunday at 9:16 PM (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Run Code

In [1]:
```
# Put these at the top of every notebook, to get automatic reloading and inline plotting
%reload_ext autoreload
%autoreload 2
%matplotlib inline
```

In [2]:
```
# This file contains all the main external libs we'll use
from fastai.imports import *
from fastai.transforms import *
from fastai.conv_learner import *
from fastai.model import *
from fastai.dataset import *
from fastai.sgdr import *
from fastai.plots import *
```

In [6]:
```
PATH = "cnnsuperbird/dataset_sb/"
```

Set your path where you have your pictures

Set small batch size due to small sample image

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

jupyter cnnsuperbird Last Checkpoint: Last Sunday at 9:16 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run Code

In [12]:
```
# Uncomment the below if you need to reset your precomputed activations
# !rm -rf {PATH}tmp
```

In [13]:
```
bs=5
```

In [14]:
```
tfms = tfms_from_model(resnet34, sz, aug_tfms=transforms_side_on, max_zoom=1.1)
```

In [15]:
```
arch=resnet34
data = ImageClassifierData.from_paths(PATH, bs, tfms)
```

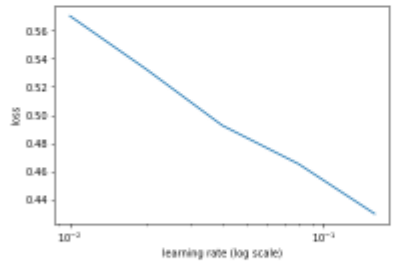
In [16]:
```
learn = ConvLearner.pretrained(arch, data, precompute=True)
```

In [22]:
```
learn.lr_find()
```

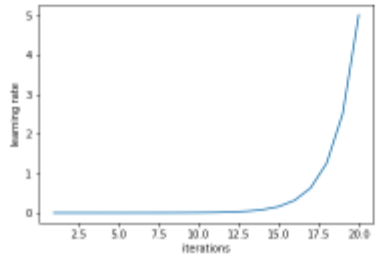
Epoch 100% 1/1 [00:03<00:00, 3.11s/it]

```
[  0.       0.51294  669.55872    0.91   ]
```

In [23]:
```
learn.sched.plot()
```



In [25]:
```
learn.sched.plot_lr()
```



Why Superman and bird is not exactly the best choice

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

- Images are too different?
- Easy for cnn to distinguish

Improving accuracy

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

In [26]:
```
learn = ConvLearner.pretrained(arch, data, precompute=True)
learn.fit(0.01, 4)
```

Epoch 100% 4/4 [00:00<00:00, 6.94it/s]

```
[ 0.       0.34434  0.83409  0.99   ]
[ 1.       0.15373  0.83279  0.99   ]
[ 2.       0.09719  0.82931  0.99   ]
[ 3.       0.09749  0.83868  0.99   ]
```

In [28]:
```
learn.unfreeze()
```

In [29]:
```
lr=np.array([0.0001,0.001,0.01])
```

In [32]:
```
learn.fit(lr,3,cycle_len=1,cycle_mult=1)
```

Epoch 100% 3/3 [00:10<00:00, 3.49s/it]

```
[ 0.       0.2933   0.82867  0.99   ]
[ 1.       0.18966  0.8387   0.99   ]
[ 2.       0.16883  0.81245  1.     ]
```

In [33]:
```
learn.save('model')
```

In [34]:
```
learn.load('model')
```

Correctly Classified

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

In [44]:
```
# 1. A few correct labels at random
plot_val_with_title(rand_by_correct(True), "Correctly classified")
```

```
Correctly classified
```

Most uncertain prediction

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

```
In [166]: most_uncertain = np.argsort(np.abs(probs -0.5))[:4]
          plot_val_with_title(most_uncertain, "Most uncertain predictions")
```

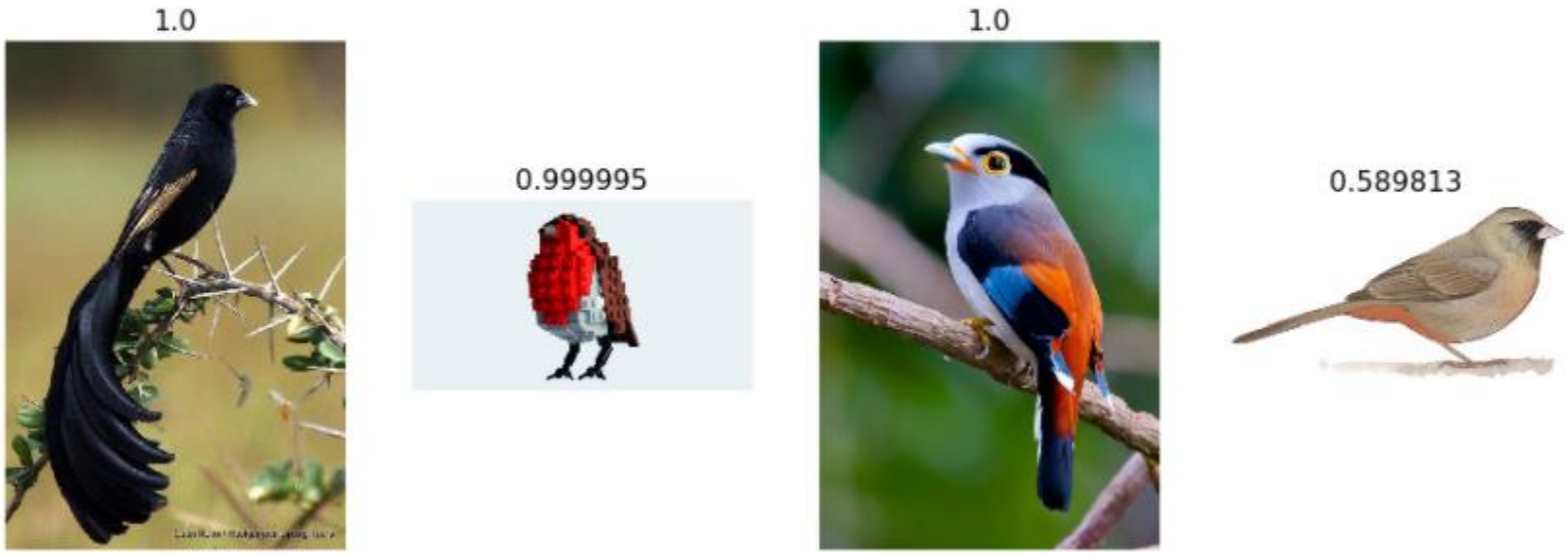
```
Most uncertain predictions
```

Not as accurate model (95%)

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

In [57]:
```
# 2. A few incorrect labels at random
plot_val_with_title(rand_by_correct(False), "Incorrectly classified")
```

```
Incorrectly classified
```



Summary: Some of the things to try

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

- Batch size
- data augmentation / image transformation
- Differential learning rate
- Multi length cycle, TTA
- Add another category?

Where you can find this file?

INVIGOUR ENERGY
OIL & GAS CONSULTANCY

- https://github.com/zaimawang/cnnsuperbird/share_cnn_superbird.pdf

THANK YOU