Section – Loading data into Fastai library

Source: fastai/nbs/04_data.external.ipynb

full link (https://github.com/fastai/fastai/blob/master/nbs/04_data.external.ipynb)

* Has URLs, download(http fetch, request), untar_data(URLs.NAME)

* Has fastai defined URLs for PLANET_SAMPLE, PLANET_TINY (downloads from fastai acct at s3.aws bucket)

* Planet zip file full link: 'https://s3.amazonaws.com/fast-ai-sample/planet_sample.tgz'

* def download_url(url, dest, ... chunk_size=xx)

  Use to download image from any web link.

Section – show image in notebook, 3 in a row as a batch

A)

From PIL import Image

im = Image.open(fname)

plt.imshow(im);

B)

After opening 3 images (save to memory) in Jupyter notebook,

Show_batch(3, img_size=10, 4)

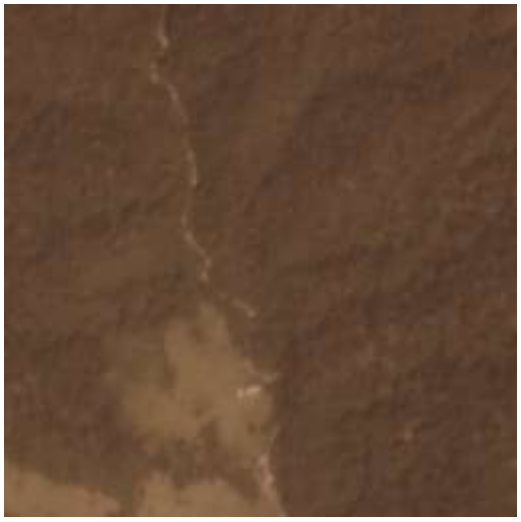
Section – Chp 19, loading image dataset

chp 19 - os.walk, and glob -- use to iterate over folder image objects/files. Loads into fastai dataset.

chp 3 - pathlib, Path class -- use to change/specify folder path in os.

Section – Ideas, Broad

- Roads seem to have cars in a row and headlights (2 images).
  Can we use this for later parts, goal 2 – extending project to recent years and new areas?
  (add image)



- Infrared channel – tif file format: Scikit-Image library
  Use this to strip out 4th channel from original tif files. Use 4th channel by itself to differentiate road and river. Also try re-creating jpg images using a different formula to get brighter jpg images.

- Labels – simplify: Only use 2-5 labels (primary forest, road, agriculture, river, haze)
  JHoward often talks about simplifying data to get better results and faster training.
  I've started a sample with only a few labels.
  (add link folder)
  http://github.com/JennEYoon//...
  \geo-ml\rainforest\tests\planet-jpg

```
fname, label_one
train_0_rbg.jpg, haze
train_1_rgb.jpg, agriculture
train_2_rgb.jpg, primaryforest
train_3_rgb.jpg, primaryforest
train_4_rgb.jpg, road
train_5_rgb.jpg, haze
train_6_rgb.jpg, agriculture
train_7_rgb.jpg, haze
train_8_rgb.jpg, agriculture
train_9_rgb.jpg, road
```

- Train model on smaller dataset (500 or 1000 instead of 40,000)
- JHoward also talks about using a smaller dataset to improve model or speed up training.

Maybe select for road, river, agriculture, and primary forest labels only – to bump up occurrence of our main categories.

Section – Jupyter command notes:

From Jupyter notebook, fastbook cell 1:

```python
!pip install –Uqq fastbook
# U is update, bash command
# qq is quiet 2nd level, bash command.  Show only critical error messages
```