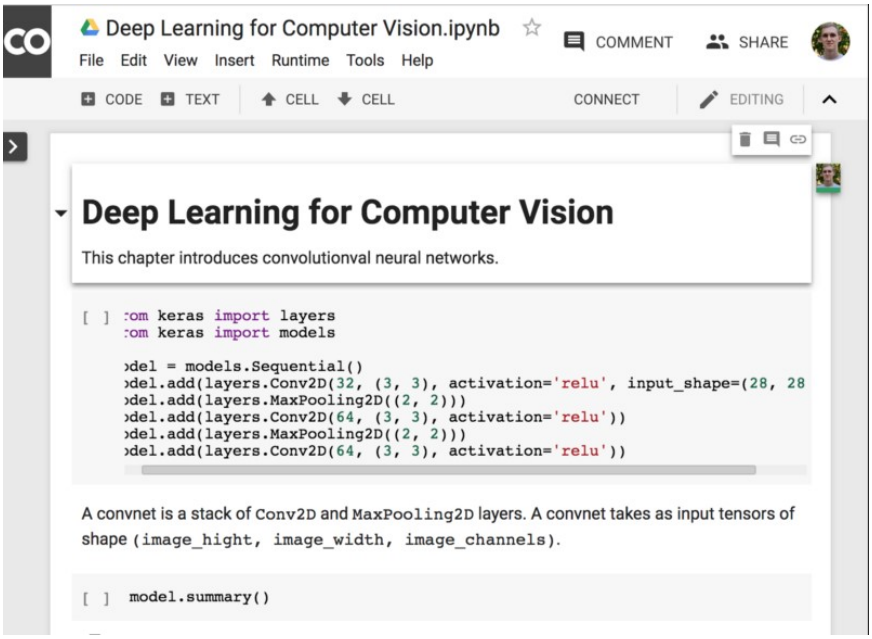
How to Upload Your Own Dataset into Google Colab

Yura Istomin Follow
May 14, 2018 · 3 min read



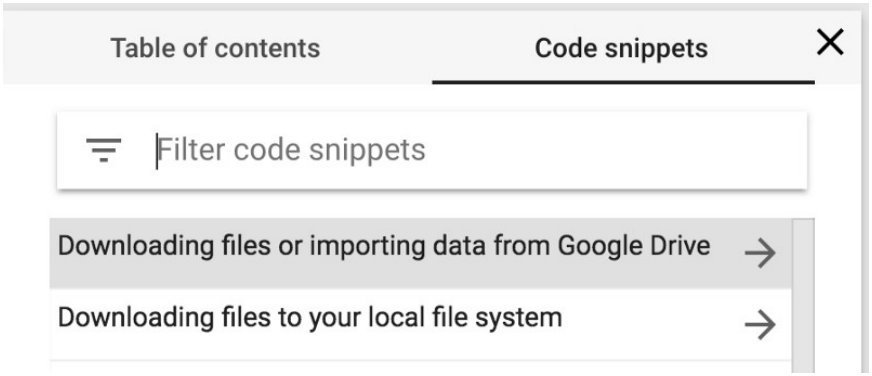
Deep Learning with Google Colaboratory

Not so long ago I started using Colab for working with Deep Learning. I don't really like its interface, but I love its **GPU**! If you're a student or you aren't sure whether you need Deep Learning or not, but you wanna try it without paying money for expensive GPU on AWS, GCP or other cloud platforms, then you are in the right place!

Briefly, Google Colab is a Jupyter Notebook with free GPU. More details —https://colab.research.google.com. Use it for training your nets, but please—don't mine Bitcoin there!

When you do Deep Learning in Google Colab, you need a training data. There are two ways to upload it into the Colab: download your dataset to the Google Drive and then import it into your notebook, or just upload the data directly from your local machine. Colab has **Code**

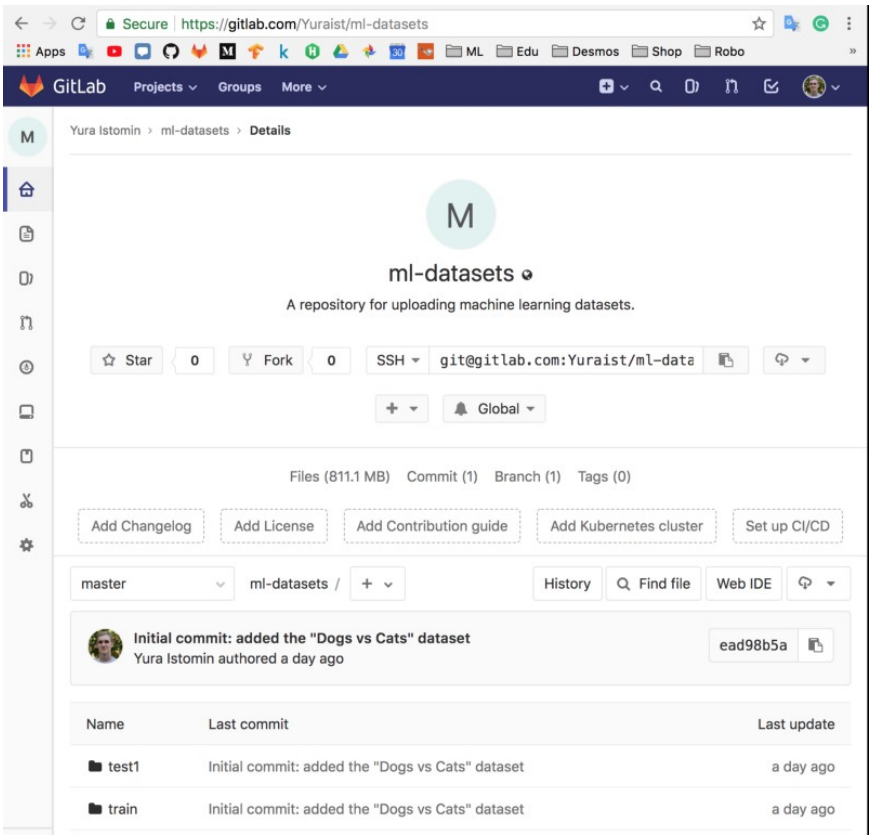
snippets for this tasks.



I tried to use both. But neither of this solutions work the way I want them to. If you want to use Google Drive for big image dataset (i.e. 'Dogs vs Cats' by Kaggle), you should upload zips with images and then unzip them into the Drive. It takes a lot of time! And even if you do it, perhaps, you'll tinker with importing the data directly into the notebook.

Thus, you could import data directly from the local system. But it doesn't work well too.

I was thinking how to import my data simply and easily. And I decided to do this using **git**. I downloaded the required dataset on my computer. Then I created a new repository on GitLab (you can use GitHub, or Bitbucket or etc.) and uploaded my files into the repo.



After that I opened the Colab notebook and cloned repository into it. Now I have all the necessary data. This solution works clearly and I can get my files with one line of code!

```
[ ]  ! git clone https://gitlab.com/Yuraist/ml-datasets.git

     Cloning into 'ml-datasets'...
     remote: Counting objects: 37396, done.
     remote: Compressing objects: 100% (37394/37394), done.
     remote: Total 37396 (delta 2), reused 37396 (delta 2)
     Receiving objects: 100% (37396/37396), 810.11 MiB | 35.20 MiB/s, dor
     Resolving deltas: 100% (2/2), done.
     Checking out files: 100% (37500/37500), done.
```

```
[33] ! ls

     cats_and_dogs  datalab  ml-datasets
```

+ form

If you want to know more about Jupyter Notebook (the same as Colab), you can read this cool tutorial—Getting Started With Jupyter Notebook for Python

Thanks for reading!