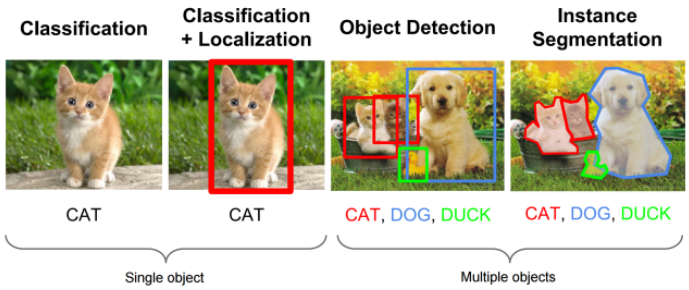
A simple way to collect your deep learning image dataset

Catch Zeng Follow
Feb 24 · 2 min read



Deep Learning has become the go-to method for solving many challenging problems. As we know, with enough training, a deep network can segment and identify the "key points" in the image.

> *If a very simple mechanism is large enough, it will have a magical effect.*

Therefore, this well-functioning deep learning **requires a lot of data**. The more training data, the better the accuracy of the model.

But where do we get all this data from? Well-annotated data can be both **expensive** and **time-consuming** to acquire. Hiring people to manually collect images and label them is **not efficient** at all. And, in the deep learning era, **data is very well arguably your most valuable resource**.

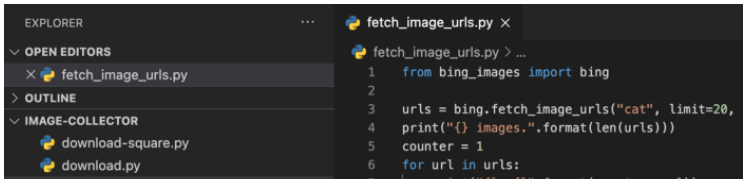
Here, I show a simple way to collect your deep learning image dataset.

The bing-images is a Python library to **fetch image URLs** and download using **multithreading** from Bing.com. It has the following features

- Support **file type** filters.
- Support Bing.com *filterui* filters.
- Download using **multithreading** and custom thread pool size.
- Support **purely** obtaining the image URLs.

Demo

Create a demo project, called *image-collector* here.



Install bing-images

Requirements

- Install Google Chrome Browser.
- Download `chromedriver` from here.
- Add `chromedriver` to PATH.

Fetch image URLs

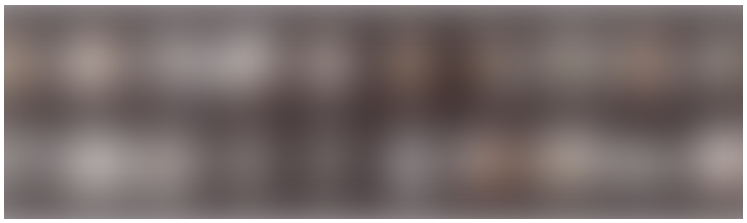
fetch_image_urls.*py*

Run

Download using multithreading

download.py

Run



Download square black-white images

download-square.py

The detailed code is at https://github.com/CatchZeng/bing_images.
See you!

Deep Learning | Image Collection | Bing Image Search | Bulk Image Downloader | Image Scraping

WRITTEN BY

Catch Zeng

Deep Learning and DevOps enthusiast

Follow

Analytics Vidhya

Analytics Vidhya is a community of Analytics and Data Science professionals. We are building the next-gen data science ecosystem https://www.analyticsvidhya.com

Follow

(Basics...
Srajan Rai

Mirco Milletari Ph.D. in Towards Data Science

Mohammad Sakib Mahmood in MLearning.ai

Priyansh Kedia in MLearning.ai