Create custom deep learning datasets in minutes [Day 14 of 17]

"Adrian at PyImageSearch" <adrian@pyimagesearch.com>
收件人:navicester@163.com
时　间:2021-2-28 11:00:00
附　件:

I have a confession to make.

I wasn't always the suave, cool, got-it-all-together guy I am today.

When I was a kid (and possibly well into my teenage years), I was what one might charitably call *a bit obsessed* with Pokemon, the mythical, animal-like creatures from the popular *(in many circles)* video game/TV/movie franchise.

Fifteen years later...I'm pretty much the same. I'm still not cool, and it's unlikely that anyone would describe me as "suave". And my Pokemon obsession is *still* strong enough that I occasionally write blog posts using Pokemon as an example, *just for funsies.*

So today, we're going to start building a Pokedex, which is used to catalogue and explore the different species of Pokemon you might encounter on your travels.

Talk about solving real-world problems! 😉

But seriously, building a Pokedex is a great way to explore how simple it is to create your own image datasets, especially if you leverage existing APIs and packages.

One of my personal favorites is the Bing Image Search API. While I'm not a big Microsoft/Windows guy, I have to admit, their Bing Image Search API is *awesome* for building deep learning datasets.

There's also another way you can leverage the power of search engines to create your own datasets...

In my many years of advanced study, I came across a small search engine company called Google *(you might have heard of them)*, and detailed a method you can use to leverage to power of Google Images to build your own custom datasets.

If you picked up on my throwing a little shade at Google, it's because they deprecated their image search API, so the process of using Google Images is time consuming and tedious. Bing's solution is more elegant and more enjoyable to use.

As a dedicated student of computer vision and deep learning, you'll want to explore both methods before you make any decisions about which to use.

Be sure to get your dataset ready soon (it's okay if you choose *not* to center yours on the world of Pokemon as I did — instead, choose something that's meaningful to you), because tomorrow, you'll learn how to train a Keras + deep learning model on the data!

Adrian Rosebrock
Chief PyImageSearcher

P.S. I used "minutes" as the time measurement in the subject of this email. From the time it takes you to read the post, download the code, and start downloading your dataset, it will not take you longer than 20-30 minutes.

However.

The actual process of downloading the images and saving them to disk may take anywhere from minutes, to hours, to days depending on (1) how many images you intend on downloading and (2) the speed of your internet connection.

Make sure you start downloading your dataset now so you'll be ready for tomorrow's lesson!
