Google Neural Net application able to place photo location better than humans

February 25 2016, by Bob Yirka

Top: GeoGuessr panorama, Bottom: Ground truth location (yellow), human guess (green), PlaNet guess (blue). Credit: arXiv:1602.05314 [cs.CV]

(Tech Xplore)—A trio of researchers at Google, led by Tobias Weyand has developed a deep-learning machine that is capable of beating humans at identifying where a photograph was taken, using only pixel information. In the paper they have uploaded to the *arXiv* preprint server, the team describes how they built their application, called PlaNet, how it works, and how well it compared to humans doing the same task.

At first glance it might seem impossible to figure out where a photograph was taken without any more information than is shown in the picture, but in many cases, people are able to do it anyway—they use situation awareness, such as the weather shown, plants in the picture, objects and any number of other elements that set off cues in the brain.

But, could a computer do the same thing? That is what the researchers with this new effort sought to find out.

The team didn't try to replicate the way humans identify photo locations, instead, they placed a grid over a [virtual world](), dividing the planet into squares of different sizes, depending on how many pictures are generated in certain areas in the real world. More people take pictures in New York City, for example, than in Indianapolis Indiana. Next, they fed their system millions of stored images with geolocation information attached to them. After that, the neural network took over, forming relationships between pixels in images with the places where the photos were taken. They finished by validating the network using several million more pictures with geolocation attached.

After that, the next thing to do was test to see how well the application actually worked. To do that they fed the system 2.3 million images from Flickr, which all had geotags, but the system wasn't given access to the geotags, instead, it had to figure out the locations by itself. The team found that the system could correctly guess down to street level 3.6 percent of the time. That number improved to 10.1 percent at city level, 28.4 percent at country level, and 48 percent at continent level.

Taking their test one step further, the team had their system compete with ten humans in an online game that consisted of guessing where a photograph was taken. PlaNet beat the humans 28 times during 50 rounds and demonstrated a much lower localization error, which showed the application is already better at the task than us humans. If the team adds an ability to pick out elements in pictures such as toys, trees, skin color, etc. and relate them to locations, the way we humans do, there is no telling how good it could get.

  More information: — www.geoguessr.com/

— PlaNet - Photo Geolocation with Convolutional Neural Networks, arXiv:1602.05314 [cs.CV] arxiv.org/abs/1602.05314

Abstract

Is it possible to build a system to determine the location where a photo was taken using just its pixels? In general, the problem seems exceptionally difficult: it is trivial to construct situations where no location can be inferred. Yet images often contain informative cues such as landmarks, weather patterns, vegetation, road markings, and architectural details, which in combination may allow one to determine an approximate location and occasionally an exact location. Websites such as GeoGuessr and View from your Window suggest that humans are relatively good at integrating these cues to geolocate images, especially en-masse. In computer vision, the photo geolocation problem is usually approached using image retrieval methods. In contrast, we pose the problem as one of classification by subdividing the surface of the earth into thousands of multi-scale geographic cells, and train a deep network using millions of geotagged images. While previous approaches only recognize landmarks or perform approximate matching using global image descriptors, our model is able to use and integrate multiple visible cues. We show that the resulting model, called PlaNet, outperforms previous approaches and even attains superhuman levels of accuracy in some cases. Moreover, we extend our model to photo albums by combining it with a long short-term memory (LSTM) architecture. By learning to exploit temporal coherence to geolocate uncertain photos, we demonstrate that this model achieves a 50% performance improvement over the single-image model.

Citation: Google Neural Net application able to place photo location better than humans (2016, February 25) retrieved 25 April 2024 from https://techxplore.com/news/2016-02-google-neural-

net-application-photo.html