



Issued 5/2/18

Facebook is Using Your Instagram Photos to Train its Image Recognition

In the race to continue building more sophisticated AI deep learning models, Facebook has a secret weapon, billions of images on Instagram. In research that the company is presenting today at F8, Facebook details how it took what amounted to billions of public Instagram photos that had been annotated by users with hashtags and has used that data to train their own image recognition models. They relied on hundreds of GPUs running around the clock to parse through the data, but they were ultimately left with deep learning models that beat industry benchmarks, the best of which achieved 85.4 percent accuracy on ImageNet. When you're operating at this scale -- the largest of the tests used 3.5 billion Instagram images spanning 17k hashtags -- even Facebook doesn't have the resources to closely supervise the data. While other image recognition benchmarks may rely on millions photos that human beings have pored through and annotated personally, Facebook had to find methods to clean up what users had submitted that they could do at scale. The privacy implications here are interesting. On one hand, Facebook is only using what amounts to public data (no private accounts), but when a user posts an Instagram photo how aware are they that they're also contributing to a database that's training deep learning models for a tech mega-corp? These are the questions of 2018, but they're also issues that Facebook is undoubtedly growing more sensitive to out of self-preservation.