我爱机器学习

机器学习干货站

Show and Tell：谷歌在TensorFlow上开源图像描述系统

我爱机器学习(52ml.net) 2016年9月23日 0

选自Google Research blog

作者：Chris Shallue, Software Engineer, Google Brain Team

参与：李亚洲

原文链接：Show and Tell：谷歌在TensorFlow上开源图像描述系统

2014 年，Google Brain 团队的研究科学家训练了一个自动准确描述图像内容的机器学习系统。后来对该系统的进一步开发使其赢得了微软 COCO 2015 图像描述挑战赛的并列冠军，这项比赛是为了对比出准确描述图像的最佳算法。

今天，该团队在 TensorFlow 上开源最新版本的图像描述系统。此次公开的版本相比于原始版本包含了对图像描述系统计算机视觉组件的极大改进，可更快速的进行训练，并产出更精细、准确的描述。这些改进在论文 Show and Tell: Lessons learned from the 2015 MSCOCO Image Captioning Challenge 中有详尽的概述与分析。

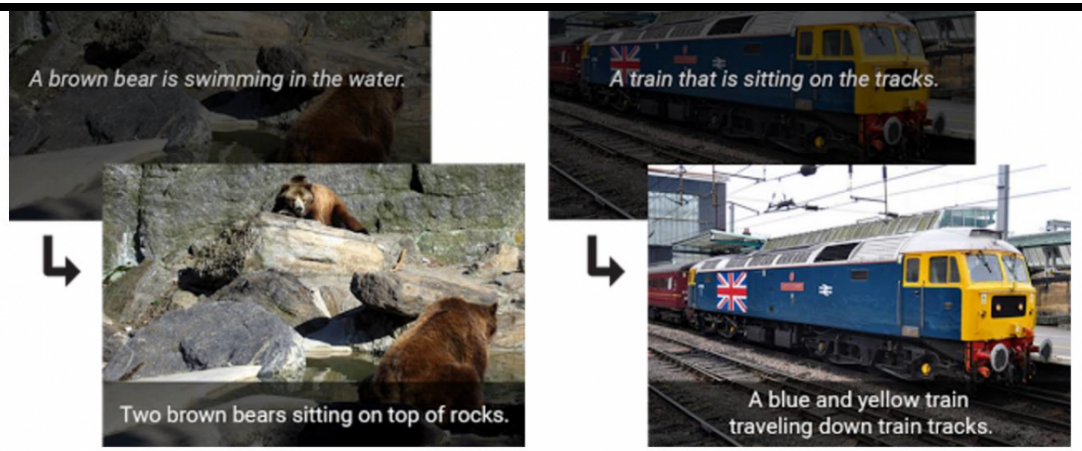
我们系统生成的对图片的自动描述

在 2014 年的系统中，我们使用 Inception V1 图像分类模型初始设定图像编码器，它可产生对识别图像中不同物体有帮助的编码。当时，它是可用的最好的图像模型，在 ImageNet 2012 图像分类任务基准上获得了 89.6% 的准确率。在 2015 年，我们使用最新的 Inception V2 取代了 V1，V2 模型当时在同样的分类任务上取得了 91.8 % 的成绩。在此视觉组件上的进步使得我们的描述系统在 BLEU-4 标准（该标准普遍使用于机器翻译中，用来评估机器生成的句子的质量）上的准确率上升了 2 个点，这也是它能在微软图像描述挑战赛中取得成功的关键因素。

今天开放的版本使用 Inception V3 初始设定图像编码器，V3 在 ImageNet 分类任务上取得了 93.9% 的准确率。使用更好的视觉模型初始设定图像编码器使得该图像描述系统有更好的能力识别图像中的不同物体，生成更细节、更准确的描述。相比于图像挑战赛中使用的那代系统，最新的系统在 BLEU-4 标准上的准确率又有了 2 个点的改进。

对视觉组件的另一个关键改进是对图像模型的精调。在该系统中，图像编码器是由对图像中物体分类的模型进行初始化的，此次精调就解决了这个问题，因为图像描述系统的目标是使用由图像模型生成的编码描述图像中的物体。例如，一个图像分类模型可以告诉你图像中有狗、草地和飞盘，但自然描述也可以告诉你草的颜色、狗与飞盘的关联。

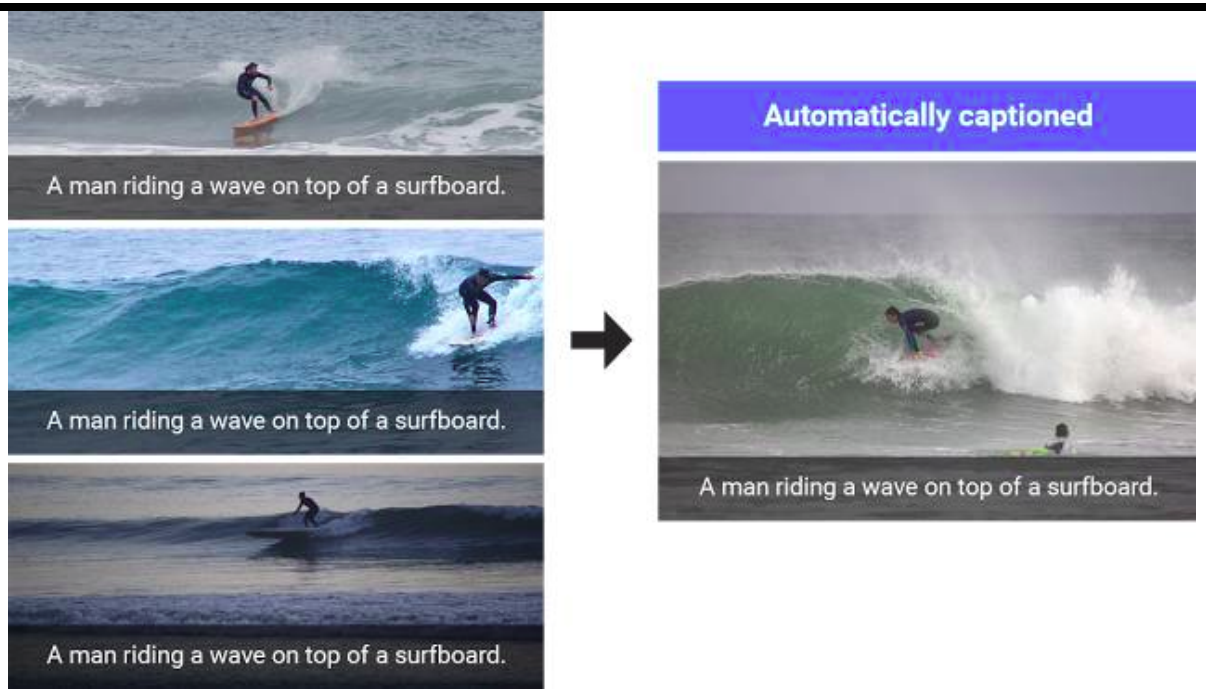
在精调阶段，通过在人类生成的图像描述数据上联合训练视觉和语言组件对图像描述系统进行了改进。这使得该系统能从图像中迁移出对生成描述有帮助但对物体分类没必要的信息。特别的，在精调之后，该系统能更准确的描述物体的颜色了。更重要的是，精调阶段必须发生在语言组件学会生成描述之后，否则随机初始化语言组件的噪声会不可逆的破坏视觉组件。



左图：更好的图像模型使其能生成更详细、更准确的描述；右图：在精调图像模型之后，图像生成系统更能准确的描述物体颜色。

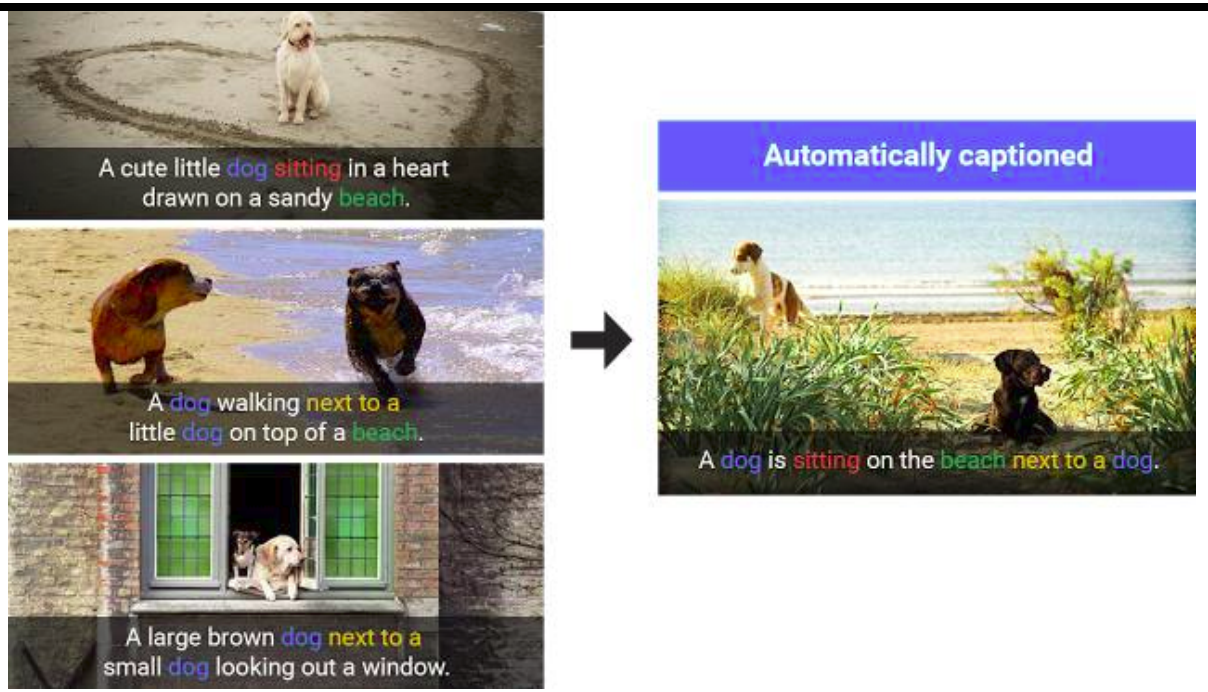
直到最近，我们的图像描述系统才被部署到 DistBelief 软件框架中。今天公开的在 TensorFlow 中的实现取得了同样等级的准确率和极大的速度进步：在英伟达 K 20 GPU 上，相比于在 DistBelief 中每训练一步的 3 秒时间，在 TensorFlow 框架中每训练一步的时间降到了 0.7 秒，意味着总体训练时间只是之前的 25%。

关于该系统的一个天然问题是它是否能够生成之前从未看到过的环境与互动的新型描述。该系统是通过千百张人类描述过的图像进行训练的，在看到类似于之前的场景时它总会重复使用人类的描述。



在看到类似于之前的场景时，该系统总是重复使用人类的描述

所以，它真的理解每张图像中的物体和物体之间的关联吗？或者说它总是在从训练数据中反刍图像的描述？无比令人激动的是，我们的模型真的开发出了在全新场景上生成准确的图像描述的能力，表明对图像中物体与环境的更深的理解。此外，它也学习如何用自然口音的英语表达知识，尽管除了阅读人类描述之外它没受过额外的语言训练。



我们的模型使用从训练集类似场景中学到的概念，生成全新的图像描述。

我们希望在 TensorFlow 中开源的这一模型能推进图像描述研究与应用，也使得更多刚兴趣的人能进行学习。想要训练自己的图像描述系统以及获取更多关于该神经网络架构的细节，浏览该模型的代码主页：https://github.com/tensorflow/models/tree/master/im2txt。

虽然我们的系统使用的是 Inception V3 图像分类模型，你也可以尝试使用最新发布的 Inception-ResNet-v2 模型训练该系统，看看结果是否会更好。

> **欢迎加入我爱机器学习QQ14群：336582044**

微信扫一扫，关注我爱机器学习公众号

微博：我爱机器学习

分类 深度学习 科技速递

标签 Google Image Caption TensorFlow

上一篇文章
提升深度学习模型的表现，你需要这20个技巧（附论文）

下一篇文章
深度学习漫游指南：强化学习概览

或许您还喜欢这些文章

综合

Google Influential Papers for 2013

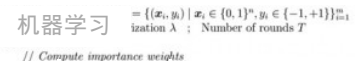
机器学习

Google develops computer vision accurate enough to solve its own CAPTCHAs

机器学习

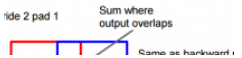
Up Next: Retrieval Methods for Large Scale Related Video Suggestion

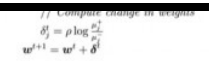
机器学习



机器学习

深度学习





Sibyl: Google’s system for Large Scale Machine Learning

为什么机器学习广泛的用在谷歌的广告系统中，而很少用在搜索排序中？

[导读]Image Classification and Segmentation with Tensorflow and TF-Slim

发表评论

评论

姓名 *

电子邮件 *

站点

发表评论