

# 基于 TensorFlow 预训练模型快速、精准的图像分类器

曹大有, 胥 帅

(汉江师范学院 计算机科学系, 湖北 十堰 442000)

[摘 要] 图像分类是根据图像信息中所反映的不同特征, 把不同类别的目标区分开来的图像处理方法, 它利用计算机对图像进行定量分析, 把图像或图像中的每个像元或区域划归为若干个类别中的某一种, 以代替人的视觉判读. TensorFlow 是 Google 基于 DistBelief 进行研发的新一代人工智能学习系统, TensorFlow 是将复杂的数据结构传输至人工智能神经网络中进行分析 and 处理过程的系统. 根据 TensorFlow 的预训练模型建立了一个低成本、快速、精准的图像分类器, 根据实验结果, 准确率达 98%.

[关键词] 图像分类; 深度学习; 人工智能; 卷积神经网络

[doi] 10.19575/j.cnki.cn42-1892/g4.2017.03.006

[中图分类号] TP311

[文献标识码] A

[文章编号] 2096-3734(2017)03-0027-06

## 0 引 言

Artificial Intelligence, 也就是人工智能, 就像长生不老和星际漫游一样, 是人类最美好的梦想之一. 虽然计算机技术已经取得了长足的进步, 但是到目前为止, 还没有一台电脑能产生“自我”的意识.

但是自 2006 年以来, 机器学习领域, 取得了突破性的进展. 图灵试验, 至少不是那么可望而不可及了. 至于技术手段, 不仅仅依赖于云计算对大数据的并行处理能力, 而且依赖于算法. 这个算法就是 Deep Learning. 借助于 Deep Learning 算法, 人类终于找到了如何处理“抽象概念”这个亘古难题的方法.

深度学习的概念源于人工神经网络的研究. 包含多隐层的多层感知器就是一种深度学习结构. 深度学习通过组合低层特

征形成更加抽象的高层表示属性类别或特征, 以发现数据的分布式特征表示<sup>[1]</sup>.

深度学习的概念由 Hinton 等人于 2006 年提出. 基于深度置信网络 (DBN) 提出非监督贪心逐层训练算法, 为解决深层结构相关的优化难题带来希望, 随后提出多层自动编码器深层结构. 此外 Lecun 等人提出的卷积神经网络是第一个真正多层结构学习算法, 它利用空间相对关系减少参数数目以提高训练性能<sup>[1]</sup>.

深度学习是机器学习研究中的一个新的领域, 其动机在于建立、模拟人脑进行分析学习的神经网络, 它模仿人脑的机制来解释数据, 例如图像, 声音和文本<sup>[2]</sup>.

TensorFlow 是谷歌基于 DistBelief 进行研发的新一代人工智能学习系统, 其命名来源于本身的运行原理. Tensor (张量) 意味着 N 维数组, Flow (流) 意味着基于数

[收稿日期] 2017-03-10

[基金项目] 湖北省教育厅 2015 年度人文社会科学研究重点项目《大数据时代大学生个性化学习技术的研究》(15D137); 汉江师范学院 2014 年度教学研究项目《大数据时代大学生个性化学习技术的研究》(2014003)

[作者简介] 曹大有 (1964-), 男, 湖北十堰人, 汉江师范学院计算机科学系教授, 主要从事软件开发技术、机器学习、深度学习研究.

据流图的计算, TensorFlow 为张量从流图的一端流动到另一端计算过程. TensorFlow 是将复杂的数据结构传输至人工智能神经网络中进行分析 and 处理过程的系统.

TensorFlow 可被用于语音识别或图像识别等多项机器深度学习领域, 对 2011 年开发的深度学习基础架构 DistBelief 进行了各方面的改进, 它可在小到一部智能手机、大到数千台数据中心服务器的各种设备上运行. TensorFlow 完全开源, 任何人都可以用.

## 1 问题的提出

计算机视觉 (Computer Vision) 是一个持续流行了多年的问题, 原因是在: 探索、影像内容的理解、医学应用和地图识别等领域有太多的应用. 大家都有一个愿景: 让计算机能够像人一样去“看”一张图片, 甚至“读懂”一张图片.

对于一张输入的图片, 要判定它属于给定的一些标签、类别中的哪一个. 看似很简单的问题, 却一直是计算机视觉的一个核心问题, 应用场景也很多. 它的重要性还体现在其他的一些计算机视觉的问题 (比如说物体定位和识别、图像内容分割等) 都可以基于它去完成.

图像识别的难点在于看似很直接, 实际上会有下面这样一些困难: 视角不同, 每个事物旋转或者侧视最后的构图都完全不同; 尺寸大小不统一, 相同内容的图片也可大可小; 变形, 很多东西处于特殊的情形下, 会有特殊的摆放和形状; 光影等干扰、幻象; 背景干扰; 同类内的差异 (比如椅子有靠椅、吧椅、餐椅、躺椅等等).

识别的途径简单说来, 也就下面三步. 输入: 我们给定  $K$  个类别的  $N$  张图片, 作为计算机学习的训练集; 学习: 让计算机逐张图片地“观察”和“学习”; 评估: 就像我们

上学要考试检测一样, 我们还得考考计算机学得如何, 于是我们给定一些计算机不知道类别的图片让它判别, 然后再比对我们已知的正确答案.

最常用的方法是: 最近邻分类器 (Nearest Neighbor Classifier) 和  $K$  最近邻分类器 (KNN). 最近邻分类器想法就是: 既然我们现在有每个像素点的值, 那我们就根据输入图片的这些值, 计算和训练集中的图片距离, 找到最近的图片的类别, 作为它的类别, 想法很直接, 这就是“最近邻”的思想, 但这种做法在图像识别中, 其实效果并不是特别好.

最近邻的思想在 CIFAR 上得到的准确度为 38.6%, 我们知道 10 个类别, 我们随机猜测的话准确率差不多是  $1/10 = 10\%$ , 所以说还是有识别效果的, 但是显然这距离人的识别准确率 (94%) 实在是低太多了, 不那么实用.

KNN 是对最近邻思想的一个调整. 在使用最近邻分类器分类, 扫描 CIFAR 训练集的时候, 会发现有时候不一定距离最近的和当前图片是同类, 但是最近的一些图片中有很多和当前图片是同类. 所以我们将最近邻扩展为最近的  $N$  个临近点, 然后统计一下这些点的类目分布, 取最多的那个类目作为自己的类别, 这就是 KNN 的思想.

KNN 也有两个不足: 一是  $K$  值的确定; 二是预测过程非常耗时, 因为要和全部训练集中的图片比对一遍. 用深度神经网络可以解决这些问题, 深度神经网络解决图像问题时训练是一个相对耗时间的过程, 但是识别的过程非常快.

深度学习是机器学习研究中的一个新的领域, 其动机在于建立、模拟人脑进行分析学习的神经网络, 它模仿人脑的机制来解释数据, 例如图像、声音和文本. 深度学

习是无监督学习的一种。

深度学习的概念源于人工神经网络的研究。含多隐层的多层感知器就是一种深度学习结构。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征,以发现数据的分布式特征表示。

目前研究人员正在使用的深度学习框架不尽相同,有 TensorFlow、Torch、Caffe、Theano、Deeplearning4j 等,这些深度学习框架被应用于计算机视觉、语音识别、自然语言处理与生物信息学等领域,并获取了极好的效果。

Google 不仅是大数据和云计算的领导者,在机器学习和深度学习上也有很好的实践和积累,在 2015 年年底开源了内部使用的深度学习框架 TensorFlow。

TensorFlow 是使用数据流图(Data Flow Graph)的形式进行计算。图中的节点代表数学运算,而图中的线条表示多维数据数组(Tensor)之间的交互。TensorFlow 灵活的架构可以部署在一个或多个 CPU、GPU 的台式以及服务器中,或者使用单一的 API 应用在移动设备中。TensorFlow 最初是由研究人员和 Google Brain 团队针对机器学习和深度神经网络进行研究所开发的,目前开源之后可以在几乎各种领域适用。

Data Flow Graph: 使用有向图的节点和边共同描述数学计算。Graph 中的 Nodes 代表数学操作,也可以表示数据输入输出的端点。边表示节点之间的关系,传递操作之间互相使用的多位数组(Tensors),Tensor 在 Graph 中流动——这也就是 TensorFlow 名字的由来。一旦节点相连的边传来了数据流,节点就被分配到计算设备上异步的(节点间)、并行的(节点内)执行。

TensorFlow 有以下特点:

**机动性:** TensorFlow 并不只是一个规则的 Neural Network 库,事实上如果你可以将你的计算表示成 Data Flow Graph 的形式,就可以使用 TensorFlow。用户构建 Graph,写内层循环代码驱动计算,TensorFlow 可以帮助装配子图。定义新的操作只需要写一个 Python 函数,如果缺少底层的数据操作,需要写一些 C++ 代码定义操作。

**可适性强:** 可以应用在不同设备上,CPU、GPU、移动设备、云平台等。

**自动差分:** TensorFlow 的自动差分能力对很多基于 Graph 的机器学习算法有益。

**多种编程语言可选:** TensorFlow 很容易使用,有 Python 接口和 C++ 接口。其他语言可以使用 SWIG 工具使用接口。(SWIG—Simplified Wrapper and Interface Generator,是一个非常优秀的开源工具,支持将 C/C++ 代码与任何主流脚本语言相集成。)

**最优化表现:** 充分利用硬件资源,TensorFlow 可以将 Graph 的不同计算单元分配到不同设备执行,使用 TensorFlow 处理副本。

与 Caffe、Theano、Torch、MXNet 等框架相比,TensorFlow 在 Github 上 Fork 数和 Star 数都是最多的,而且在图形分类、音频处理、推荐系统和自然语言处理等场景下都有丰富的应用。最近流行的 Keras 框架底层默认使用 TensorFlow,著名的斯坦福 CS231n 课程使用 TensorFlow 作为授课和作业的编程语言,国内外多本 TensorFlow 书籍已经在筹备或者发售中,AlphaGo 开发团队 Deepmind 也计划将神经网络应用迁移到 TensorFlow 中,这无不印证了 TensorFlow 在业界的流行程度。

TensorFlow 不仅在 Github 开放了源

代码,在《TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems》论文中也介绍了系统框架的设计与实现,其中测试过 200 节点规模的训练集群也是其他分布式深度学习框架所不能媲美的。Google 还在《Wide & Deep Learning for Recommender Systems》和《The YouTube Video Recommendation System》论文中介绍了 Google Play 应用商店和 YouTube 视频推荐的算法模型,还提供了基于 TensorFlow 的代码实例,使用 TensorFlow 任何人都可以在 ImageNet 或 Kaggle 竞赛中得到接近 State of the art 的好成绩。

TensorFlow 的流行让深度学习门槛变得越来越低,只要你有 Python 和机器学习基础,入门和使用神经网络模型变得非常简单。TensorFlow 支持 Python 和 C++

两种编程语言,再复杂的多层神经网络模型都可以用 Python 来实现。

但 Tensorflow 仍然是一个偏底层的库,实际使用时开发人员需要编写大量的代码,阅读他人的代码也很费劲,然而 Google 实际上也公开了大量对 Tensorflow 做了轻量级的封装的预训练模型。文中就利用 TensorFlow 预训练好的 Inception 模型,来建立一个低成本、快速、精准的图像分类器。

## 2 基于 Inception 模型的图像分类器

Inception 是用来训练 2012 年 ImageNet 的 Large Visual Recognition Challenge 数据集。这是计算机视觉领域的一类标准任务,模型要把整个图像集分为 1000 个类别,例如“斑马”、“达尔玛西亚狗”,和“洗碗机”。如下图所示,这里展示了一部分 AlexNet 的分类结果:

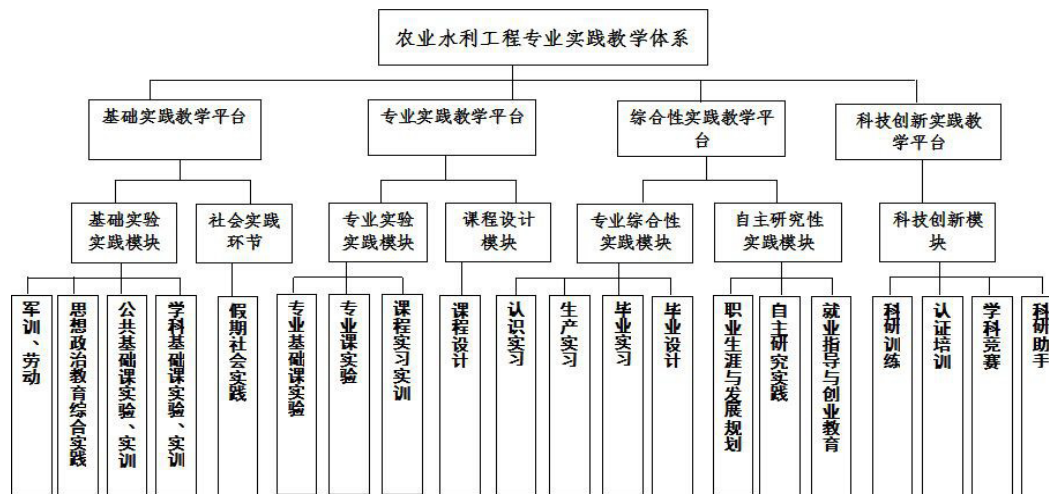


图 1 AlexNet 的分类图

下面我们就在 Ubuntu 上利用 Inception 模型来训练建立自己的图像分类器。整个训练过程可以分为四个主要步骤:

### 1) 收集训练用的图像。

我们从“百度图片”上搜索了 5 种动物 cuckoo(布谷鸟)、frog(青蛙)、lion(狮子)、

owl(猫头鹰)、sparrow(麻雀)的图片。为了使图像分类器有很好的泛化能力,为每类动物收集了在不同环境及地点各 100 张图片以建立训练图像集,且每种动物图像样本的大约 10% 是幼鸟图像,这样分类器是也可以找到幼鸟和成年鸟之间的不同之

处.当然若有一些扭曲的图像会更好,虽然这可能会降低我们的训练过程的速度.这里我们这所以不需要用更多的样本来训练模型,这是因为 TensorFlow 会用 Inception 模型之前训练特征检测器来重新训练一个新模型.

一旦有了合适的图像数量和类型,我们就在 TensorFlow 目录下创建一个文件夹:

```
$ cd ~/tensorflow
$ mkdir My_file && cd My_file
&& mkdir animal_photos && cd animal_photos
```

```
$ mkdir cuckoo frog lion owl sparrow
```

然后将每类动物的图像集移动到相应的文件夹中,该脚本接受 PNG、JPG、GIF 和 TIF 类型图像.

2)使用 TensorFlow 和 Inception 模型来训练一个计算图/模型.

然后使用克隆的 TensorFlow 源码中内置 python 脚本训练了一个新模型和相关标签.我们用来重新训练的原始计算图 Inception 模型是谷歌研究人员花费了两个星期在一台包含八个 NVidia Tesla K40 GPU 的计算机上训练得到的.过程如下:

```
$ cd ~/tensorflow
$ python tensorflow/examples/
image_retraining/retrain.py \
--bottleneck_dir My_file/
bottlenecks \
--model_dir My_file/inception \
--output_graph My_file/retrained_
graph.pb \
--output_labels My_file/retrained_
labels.txt \
--image_dir My_file/animal_photos
```

以上过程会先下载 Inception 模型到 My\_file/inception 文件夹下.在我们的没有 GPU 支持的笔记本上(配置为:8G 内存、

Inter(R) Core(TM) i5-3230M CPU@ 2.60GHz \* 4),用时不到 10 分钟,经过 3999 步训练模型.以下是训练结果:

Final test accuracy = 98%

我们的 TensorFlow 模型的最终测试精度为 98%.考虑到我们用了各种各样不同的图像,我觉得这个结果是非常令人惊喜的.

3)用我们的计算图编写脚本来进行图像分类.

我们使用了 Inception 模型原始计算图并用我自己的图片集重新训练.接着我们用 Bazel 构建我们自己的图像分类器.

```
$ bazel build tensorflow/examples/
label_image:label_image
```

注:需要运行 ./configure 对 tensorflow 进行配置.另外还需将相应的 Inception 模型文件拷贝到 label\_image/data 目录下,对 main.cc 文件的 main()函数相应的 pb 和 txt 文件要改名,以便和 data 目录下的文件一致.

4)对新图像进行分类来测试脚本.

对新的一组图像进行分类器测试.为了简便起见,我们将测试图像放在 My\_file 文件夹中.

```
$ bazel - bin/tensorflow/examples/
label_image/label_image \
--graph = My_file/retrained_
graph.pb \
--labels = My_file/retrained_
labels.txt \
--output_layer=final_result \
--image = My_file/test.jpg #
This is the test image
```

首先尝试分类一张我们立即就能识别出为 cuckoo(布谷鸟)的图片:



图 2 cuckoo(布谷鸟)

TensorFlow 也识别出来是 cuckoo(布谷鸟)! 而且识别率达 97.133%. 识别结果如下:

```
2017-03-09 09:58:36.728351: I
tensorflow/examples/label_image/main.
cc:206] cuckoo (2): 0.97133
```

```
2017-03-09 09:58:36.728383: I
tensorflow/examples/label_image/main.
cc:206] sparrow (4): 0.0141217
```

```
2017-03-09 09:58:36.728390: I
tensorflow/examples/label_image/main.
cc:206] owl (0): 0.0130435
```

```
2017-03-09 09:58:36.728396: I
tensorflow/examples/label_image/main.
cc:206] frog (3): 0.000879179
```

```
2017-03-09 09:58:36.728402: I
tensorflow/examples/label_image/main.
cc:206] lion (1): 0.000625494
```

接着我们也用了其它 5 种动物的图片,分类器也能以相当高的精确度识别出来.这说明我们的分类器是可信的.

### 3 结束语

文中我们用 Tensorflow 的预训练模型,对自己的图像建立分类器,效果很不错.预训练模型包装了 Tensorflow 的底层代码,使得使用者的难度降低,容易推广.当然 Tensorflow 应用范围很多,不局限于图像分类,还包括语音识别、序列生成等等,这有待于我们大家去研究使用它.

### [参考文献]

- [1]孙志军,薛磊,许阳明,王正.深度学习研究综述[J].计算机应用研究[J].2012,(8).
- [2]深度学习那么火,它究竟能做些什么?深度学习世界[EB/OL].www.dlworld.cn[引用日期 2016-05-3].

【编校:胡军福】

## A Fast & Accurate Image Classifier Based on the Pre-training Models of TensorFlow

CAO Da-you, XU-Shuai

(Department of Computer Science, Hanjiang Normal University, Shiyan 442000, China)

**Abstract:** Image classification is an approach for processing images to distinguish the different types of target according to individual characteristics reflected in images. For the sake of replacing the human visual interpretation, it takes advantage of the computer version to carry on the quantitative analysis to images and classifies each pixel or regions in an image or the image into one of several categories. TensorFlow is a new machine-learning system developed by Google based on DistBelief. In TensorFlow, in order to analyze and process images, one needs to transfer the complex data structure to an artificial intelligence neural network. In this paper, we present a low-cost, fast and accurate image classifier on the basis of the pretraining models of TensorFlow. Experiment results show the classifier attains accuracy 98%.

**Key words:** image classification; deep learning; artificial intelligence; conventional neural networks