

TensorFlow 在图像识别系统中的应用

邢艳芳 段红秀 何光威

(中国传媒大学南广学院 江苏 南京 211172)

摘要: 人工智能将是未来发展的方向,深度学习则是人工智能领域的一个重要分支。随着深度学习在国内外的快速发展,以及深度可分离卷积神经网络模型的提出,极大地推动了深度学习在图像识别、文字处理和语音识别等领域的广泛应用。基于 Google 人工智能系统 TensorFlow 的深度学习开发平台,可以快速搭建出深度可分离卷积神经网络。文中采用 MobileNet 模型,Ubuntu16.04 开源 Linux 操作系统, CUDA9.0 运算平台, cuDNN7.0.5 并行架构,设计 Python 爬虫程序构建数据集,运用 TensorBoard 对模型进行可视化。通过对此类模型进行重建和训练,保存训练完成的模型,对图像实现了较高准确度的识别。完成在不同迭代次数下模型性能的测试和分析,通过与 Inception_v3 模型的准确度与迭代周期进行对比,表明 MobileNet 在移动端、嵌入式以及网络规模大小和内存限制时具有较好的推广应用价值。

关键词: 图像识别; TensorFlow; 深度可分离卷积神经网络; MobileNet

中图分类号: TP319

文献标识码: A

文章编号: 1673-629X(2019)05-0192-05

doi: 10.3969/j.issn.1673-629X.2019.05.040

Application of TensorFlow in Image Recognition System

XING Yan-fang, DUAN Hong-xiu, HE Guang-wei

(Nanguang College of Communication University of China, Nanjing 211172, China)

Abstract: Artificial intelligence will be the major development direction in the future, while deep learning is an important branch in the field of artificial intelligence. With the rapid development of deep learning at home and abroad, as well as the proposition of deep separable convolution neural network model, the deep learning has been widely applied in the field of image recognition, word processing and speech recognition etc.. Based on the deep learning development platform of Google artificial intelligence system TensorFlow, the deep separable convolution neural network can be built quickly. In this paper, the MobileNet model, Linux open source operation system Ubuntu16.04, calculation platform CUDA9.0 and parallel structure cuDNN7.0.5 are adopted to design the image recognition system and data sets are built by Python crawler program. Using TensorBoard realizes the visualization of the model. By the reconstruction and training to the model, the image recognition with high accuracy can be gained after saving the model which has been trained. The test and analysis of the model performance are finished for different iteration numbers. By comparing with the accuracy and iterative cycles of the Inception_v3 model, it shows that the MobileNet model has better promotion and application value in mobile terminal, embedded terminal and limited network scale and memory.

Key words: image identification; TensorFlow; deep separable convolution neural network; MobileNet

0 引言

众所周知,人工智能将是未来发展的方向,深度学习则是人工智能领域的一个重要分支,在图像和语音识别等领域取得了重要突破。随着深度学习的发展,相关神经网络模型也不断推陈出新。TensorFlow 是目前主流的深度学习框架,文中以该框架为基础,采用 Google MobileNet 模型,Ubuntu 16.04 开源 Linux 操作系统, CUDA 9.0 运算平台, cuDNN 7.0.5 并行架构,

选取从网络中爬取的图片作为数据集,进行训练和识别。

1 TensorFlow 框架

TensorFlow^[1-5]是 Google 公司推出的开源人工智能系统,具有灵活性、高效性以及良好的可扩展性、可移植性等特征,可应用于从智能手机到大型计算集群的多种计算环境。

收稿日期: 2018-06-12

修回日期: 2018-10-18

网络出版时间: 2018-12-21

基金项目: 江苏省高等学校自然科学研究面上资助项目(17KJB510054);教育部2017年第二批“产学研协同育人”项目(201702072056)

作者简介: 邢艳芳(1981-),女,硕士,讲师,研究方向为深度学习、嵌入式系统。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20181221.1525.040.html>

TensorFlow 的主要流程是:

(1) 产生对应的数据集。数据集可从网上下载公开数据,也可以通过爬虫自动化测试或者手工寻找等方式来制作。

(2) 对数据进行转换和归一化。将导入的数据转换成指定格式,使维度和类型符合 TensorFlow 直接处理的要求。

(3) 划分数据集。将数据集划分成训练集、测试集以及验证集。

(4) 设置相关超参数。学习过程中需要众多常量参数,包括迭代次数(iterations)、学习率(learning rate)、少批量(mini batch) 等。

(5) 选取变量和占位符。采用占位符(placeholder) 得到数据,调整变量(variables)、权重

(weights) 和偏置(biases) 构建更优化的网络,并寻求损失函数(目标函数) 的最优解。

(6) 构建模型。根据参数和初始化内容构建模型,通过操作(op)、变量和占位符来建立计算图。

(7) 定义损失函数。根据损失函数(loss function) 的决策结果,得到预测值与真实值之间的偏差。使用梯度下降算法(gradient descent) 和似然函数,避免出现梯度消失和梯度下降。

(8) 初始化和训练模型。采用交叉熵(cross entropy)、验证和训练的准确度来评估损失函数。

TensorFlow 的相关操作有:

(1) 创建张量。如表1所示,TensorFlow 主要数据结构为张量(tensor) 基本计算操作包括 add()、sub()、mul() 和 div()。

表1 常规张量

张量	函数
零张量	zero = tf.zeros([...])
单位张量	one = tf.ones([...])
常量张量	const = tf.constant([...])
随机张量中的均匀分布	tf.random_uniform()
随机张量中的指定边界正态分布	tf.truncated_normal([row, column], mean, stddev)

(2) 封装张量。var = tf.Variable(...),省略号可为自定义张量(tensor)。

(3) 通过会话(session) 的 feed_ 参数获取数据,让得出的 h 为 x 的实体(identity),实体的返回值为传入

的数据本身。

(4) 实现激励函数操作。表2展示了相关激活函数。

表2 相关激活函数

激活函数	函数	特性	相关操作
ReLU	$\max(0, x)$	连续不平滑	tf.nn.relu([])
ReLU6	$\min(\max(0, x), 6)$	运行速度快,解决梯度消失	tf.nn.relu6([])
Sigmoid	$1/(1+\exp(-x))$	反向传播项趋于0,连续平滑	tf.nn.sigmoid([])

(5) 卷积函数:用于滑动窗口、步长和卷积核。

(6) 剪裁函数:用于剪裁图像中多余的维度。

(7) 损失函数:比较预测值和目标值的差值。

(8) 实现优化器。通过二分类算法等方式实现梯度下降,指定合适的学习率,确保机器学习算法收敛。

(9) 更新变量以及最小化损失函数反向传播误差,使用较小的学习率(步长)和在一定范围内尽可能多的迭代网络会让训练易于收敛。

(10) 实现线性回归。

(11) 实现逻辑回归。

在深度学习过程中,如何监控算法是一个难点。文中采用 TensorFlow 中的可视化工具 TensorBoard,可以对算法计算过程中的各个流程图、偏置、权重以及准确度、交叉熵的参数通过图形进行可视化展示。在整体识别程序的启动后,在命令行输入以下命令: tensor-

board -logdir=retrain_logs。在终端打开一个本地端口 http://zwx-Inspiration-5437:6006,一弹出便为展示页。TensorBoard 配置过程如下:

(1) 初始化计算图,创建 write 将 TensorBoard summary 写入对应文件。

(2) 确保 writer 信息写入相应的 summary 文件夹。

(3) 设置模型参数,让模型生成所需的数据集。

(4) 将数据集分割为测试集、训练集以及验证集。

(5) 建立占位符、变量以及模型相关运算操作、损失值和优化操作。

(6) 创建 TensorBoard 的直方图汇总和标量值汇总。

2 深度学习应用

文中采用深度可分离卷积神经网络作为深度学习

的模型,深度可分离卷积(depth-wise separable convolution)相对传统卷积方式,减少了参数数量,提升了计算速度,降低了过拟合的风险^[6-8]。在进行深度可分离卷积的神经网络设计中,在有通道分离的情况下,接上一个深度卷积,即可实现空间卷积^[9-12]。

在高计算复杂度下,深度卷积神经网络相比人的视觉精度更高,将卷积层分解,便于提高效率。传统卷积层,三维卷积被认为是同时在通道空间卷积和线性通道投影,从而得到高冗余计算。将它们分离后,所需层只包括单个通道卷积和线性通道投影。此外,还提出了在输入通道和输出通道之间的拓扑连接框架,以便提高该卷积层的效率。复杂度远远低于标准卷积层,在确保准确性的前提下减少了 3-4 倍计算量。深度可分离卷积神经网络支持移动端的使用和开发,更适用于移动设备便携式开发设计。

3 系统构建

较之 CPU, GPU 更适合一次性进行大量重复性工作。在深度学习卷积神经网络中,更多的是加、乘法的计算,使用 GPU 进行并发计算可以提高运算效率。因此深度学习可以非常恰当地用 GPU 进行加速。CUDA 是 NVIDIA 公司为了支持 GPU 而设计的并行计算架构,其 ISA 指令集架构和硬件计算引擎让 CUDA 更适合并行计算。文中在 Ubuntu 16.04 系统下,通过配置 GPU、CUDA 9.0 以及 cuDNN 7.0.5 的相关参数,来搭建相关平台和并行计算架构^[13]。

图像识别相关参数如下:

(1) prepare_file_system() 函数,准备 summary 需要的输出文件夹,summary 信息可以通过 TensorBoard 来看。

(2) model_info = create_model_info(FLAGS.architecture) 根据 architecture 参数,返回模型的元信息,其中 architecture 模型采用的是 MobileNetV1。

图像识别相关算法:

(1) 梯度下降优化算法:对权值进行优化,在曲线中寻找梯度方向下降最快的位置,来寻找最低点,以便让损失值最小。梯度下降算法有起始点、搭建平面的形状和学习率(步长)三个影响因素。

(2) 初始化:为了进行有效训练,让各层神经元方差不变,采用权值初始化来保证其方差不变性。

(3) 偏差/方差(bias/variance):对网络优化以后,如果网络有高偏差/高反差状态,需要进一步对参数进行调整。

(4) 正则化(regularization):为了解决高方差和拟合问题,避免神经元将非一般特征当作一般规律学习,导致验证集准确度较低。通过数据翻转、比例调整、亮

度调整、剪裁大小的方法进行数据扩充,以获取更多的训练样本,便于训练和正则化惩罚,减少过拟合和学习非一般规律的问题。

从数据输入、MobileNet 模型再到最后的输出层,各层神经网络的实现如图 1 所示。

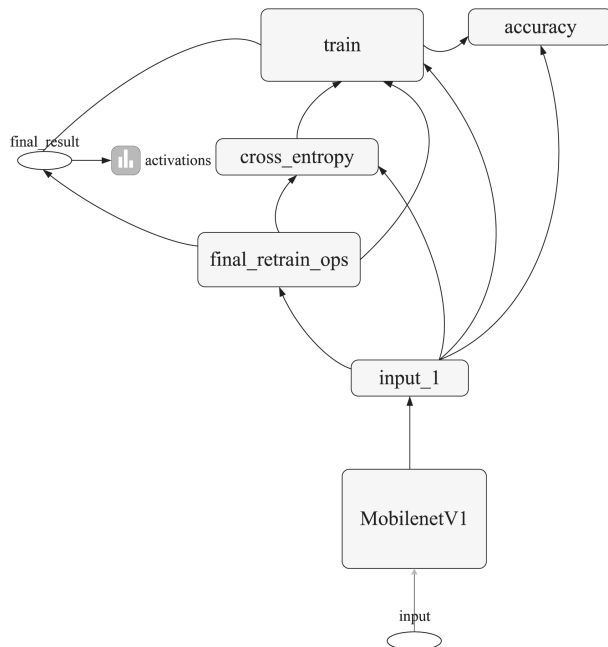


图 1 各层网络图解

4 性能测试与分析

数据集分为两部分获取:第一部分与暴力相关的包括争斗、爆炸、枪械、血腥四类,通过网络搜索下载的方式,分别存放在 fight、explosion、gun、blood 目录下。第二部分与色情(porn)相关的通过 Python 实现网络爬取,存放在 porn 目录下。数据集的 80% 作为训练样本,剩下各 10% 作为测试和验证样本。得出识别结果后,通过交叉熵来判断识别的准确度,交叉熵越低,证明通过网络迭代进行训练的效果越好^[14-15]。通过计算平均值、标准差,得出相关测试集、训练集以及验证集的准确度。如图 2 所示,1 000 次网络迭代的结果表明训练集的准确度更好,验证集的准确度一般。

迭代 100 次的结果会发现交叉熵逐步减少,最后保持在 0.15~0.2 左右,交叉熵越小,说明预测值越接近真实值。在学习率 0.01 更换为 0.001 以后,发现验证集平均准确度从 87% 上升到 89%。在进行 1 000 次迭代实验以后,训练结果的验证集准确度在 88% 左右,学习率为 0.001,可见数据集较小(总共 1 400)对网络迭代有一定影响。在进行训练集、测试集少批量网络迭代以后,从原来少批量 100 变为 200 时,准确度有小幅提高,说明 200 也为网络迭代所能承受的批量,而当少批量数为 300 时,测试集准确度却降低了,可见批量过大时不仅使得计算速度变慢,而且网络模型承受

力不足。

由图3数据可见,枪械类的识别准确率为0.976,

血腥类为0.010,争斗类为0.010,色情类为0.002,爆炸类为0.000 1,得出该图是属于枪械类的敏感图片。

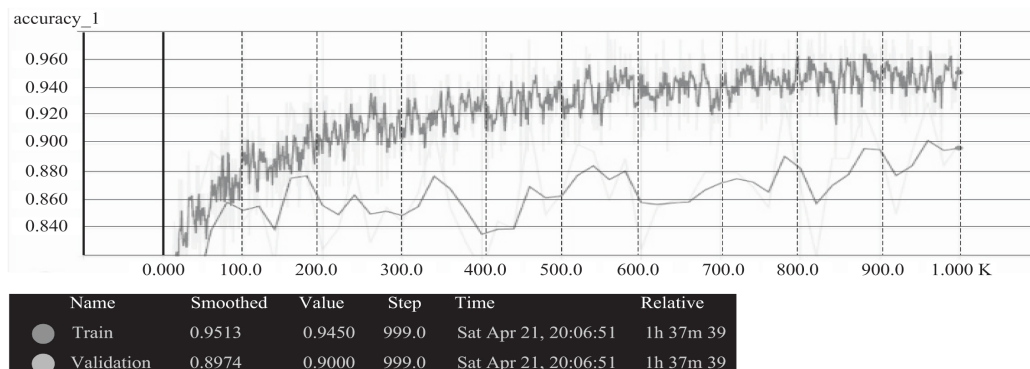


图2 验证集、训练集准确度图解

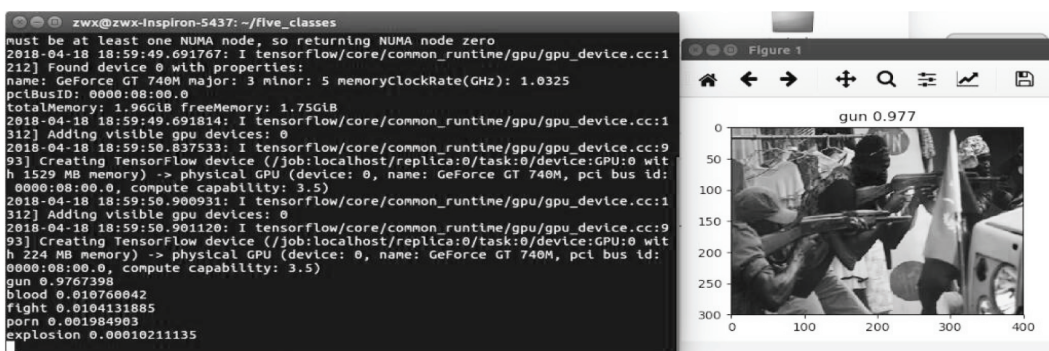


图3 Matplotlib 展示分类枪械类识别结果图解

由图4数据可见,争斗类的识别准确率为0.992,血腥类为0.002,枪械类为0.005,色情类为0.000 4,爆炸类为0.000 2,得出该图是属于争斗类的敏感图片。

由图5数据可见,爆炸类的识别准确率为0.954,争斗类为0.000 1,枪械类为0.001,色情类为0.002,血腥类为0.04,得出该图是属于爆炸类的敏感图片。



图4 Matplotlib 展示分类争斗类识别结果图解



图5 Matplotlib 展示分类爆炸类识别结果图解

测试结果显示,五类图片相应的识别准确度还是很高的,通过模型搭建再到各类权重、偏置更新,再到算法优化,让识别结果完整。通过文中设计的 100 次网络迭代,MobileNet 模型测试准确度为 86%,更换为 Inception_v3 模型的测试准确度为 87.2%,但是 MobileNet 模型网络迭代周期仅为 5 分钟,远低于 Inception_v3 模型的迭代周期(30 分钟)。因此,MobileNet 在移动端、嵌入端以及网络规模大小和内存限制时具有较好的应用价值。

5 结束语

文中采用的 TensorFlow 人工智能框架,便于图像识别等任务的实现和开发,测试结果验证了该框架的高效性、便捷性以及可行性。随着深度学习在工业以及学术上的巨大发展,值得被进一步的学习、研究和探讨。

参考文献:

- [1] 王一超,韦建文.基于高性能计算平台的 TensorFlow 应用探索与实践[J].实验室研究与探索,2017,36(12):125-128.
- [2] 章敏敏,徐和平,王晓洁,等.谷歌 TensorFlow 机器学习框架及应用[J].微型机与应用,2017,36(10):58-60.
- [3] 刘帆,刘鹏远,李兵,等.TensorFlow 平台下的视频目标跟踪深度学习模型设计[J].激光与光电子学进展,2017,54(9):277-285.
- [4] 靳涛,张永爱.TensorFlow 平台下基于深度学习的数字识别[J].信息技术与网络安全,2018,37(4):74-78.
- [5] 樊雷.一种基于 TensorFlow 的 DCGAN 模型实现[J].电脑知识与技术,2017,13(36):219-221.
- [6] 孙志远,鲁成祥,史忠植,等.深度学习研究与进展[J].计算机科学,2016,43(2):1-8.
- [7] 吴超,邵曦.基于深度学习的指静脉识别研究[J].计算机技术与发展,2018,28(2):200-204.
- [8] 袁文翠,孔雪.基于 TensorFlow 深度学习框架的卷积神经网络研究[J].微型电脑应用,2018,34(2):29-32.
- [9] 曲之琳,胡晓飞.基于改进激活函数的卷积神经网络研究[J].计算机技术与发展,2017,27(12):77-80.
- [10] 徐胤,袁浩巍,李智.基于卷积神经网络和 TensorFlow 的手写数字识别研究[J].上海电气技术,2018,11(1):31-34.
- [11] 杨远飞,曾上游,甘晓楠,等.改进的深度卷积网络在交通标志识别中的应用[J].电视技术,2017,41(11):214-219.
- [12] WANG Min, LIU Baoyuan, FOROOSH H. Factorized convolutional neural networks[C]//2017 IEEE international conference on computer vision workshop. Venice, Italy: IEEE, 2017: 545-553.
- [13] EXL L. A GPU accelerated and error-controlled solver for the unbounded Poisson equation in three dimensions[J]. Computer Physics Communications, 2017, 221: 352-357.
- [14] 卢宏涛,张秦川.深度卷积神经网络在计算机视觉中的应用研究综述[J].数据采集与处理,2016,31(1):1-17.
- [15] 王振,高茂庭.基于卷积神经网络的图像识别算法设计与实现[J].现代计算机,2015(7):61-66.

《计算机技术与发展》投稿要求

1. 文稿具有创新性,内容科学可信,结构严谨,图表清晰,文稿必须使用 word 排版,应在 4 页以上(A4 纸,Word03 默认页面,5 号宋体,包括图表在内),文稿包括:题名(20 字以内),作者姓名,作者单位(包含地区名和邮编),摘要(须达到 300 汉字以上,包括:目的、方法、结果、结论,英译文同步、准确),关键词(4~8 个),中图分类号,正文,参考文献(总数应在 15 篇以上,外文文献不少于 5 篇),与中文对照的英文(包括:题名、作者姓名、作者单位、摘要、关键词)。提供所有作者的简介:姓名、性别、出生年、学位(职称)、研究方向。

2. 文中的图表要符合国家规范,一般不应超过 6 个,图形尽量用 visio 处理,版面要清晰、紧凑、美观,图中文字采用六号宋体或 Times New Roman,打印清晰,不用彩色,坐标图中要注计量单位、符号,表格一律采用三线表(不画竖线)。

3. 文中所有物理量和单位应符合国家标准或国际标准,外文字母、单位、符号的大小写、正斜体、上下角标及易混字母应标注清楚,标量用白斜体字母(可以加下标)表示,矢量、矩阵和张量用黑斜体字母(可以加下标)表示。数学公式必须用数学编辑器处理。

4. 文稿章节编号采用三级标题:一级如 1, 2, 3, ...; 二级如 1.1, 1.2, 1.3, ...; 三级如 1.1.1, 1.1.2, 1.1.3, ...。

5. 参考文献尽量选用近几年计算机专业期刊论文,未公开发表的资料请勿引用。参考文献序号按文中引用顺序排列。