

# 基于深度神经网络的汽车车型识别

王茜,张海仙

(四川大学计算机学院,四川 成都 610065)

## 摘要:

研究基于深度学习的车型自动识别技术,运用神经网络对在各个角度下拍摄的具有复杂背景的汽车图像进行网络训练,从而达到车辆车型的自动识别的目的。采用先进的深度学习框架 Caffe 和具有强大计算能力的 GPU,使用深度神经网络 VGG16 和 AlexNet,分别对汽车图像进行网络训练与测试,并通过与传统的分类算法,K 最近邻进行对比研究。实验显示,VGG16 网络模型准确率高达 97.58%,在汽车车型识别问题上具有很大优势。

## 关键词:

深度学习;车型识别;卷积神经网络

## 基金项目:

四川省科技计划项目(No.2014GZ0005-5)

## 0 引言

深度学习的概念起源于人工神经网络,是指具有多层结构的神经网络。深度学习主要从仿生学的角度模仿神经系统的层次结构,低层次表示细节,高层次表示抽象的数据结构特征,通过逐层抽象,高度挖掘数据的本质信息,从而达到学习的目的。

在过去,多层神经网络面临着训练困难的问题,一方面是数据集限制,在不足的数据上进行训练会导致过拟合;另一方面是网络复杂性的影响。常用的反向传播算法(BP)也存在陷入局部最优、梯度弥散等问题。这无疑阻碍了神经网络的进一步研究。2006年,Hinton提出的贪婪的逐层训练算法为深度学习取得了突破性的进展<sup>[1]</sup>。贪婪的逐层训练算法解决了深度学习在训练上的难点,并且多层次的神经网络具有优异的学习能力,得到的特征具有更好的表达能力,从而可以更好地进行可视化和分类。

目前,典型的深度学习的网络模型有三类,卷积神经网络(Convolutional Neural Network)模型、受限莫尔兹曼(Restricted Boltzmann Machines)模型以及堆栈式自动编码器(Stacked Auto-encoder Network)模型。人工神经网络通过层与层完全连接的方式搭建起来,引入了庞大的参数空间,尤其是在输入尺寸比较大的情况下,网络的完全连接给计算机带来了存储和计算的沉重负担。而自然图像具有一个重要的特性:统计不变性(稳定性)。同一图像的不同区域具有相同的统计特性,这意味着我们在图像的部分区域学到的特征可以应用到整个图像中。卷积神经网络通过局部连接的方式,共享权值,进而有效地解决完全连接的问题,这也使卷积神经网络在图像处理方面具有独特的优越性。

传统的车型识别技术包括车辆检测分割、特征提取与选择、模式识别等处理。这类技术面临着诸多难点:如何在复杂背景下分割出完整的目标车辆区域是

车型识别的前提和基础,图1从各个角度下拍摄的汽车图片,从上到下依次为 SUV、皮卡车、面包车以及小轿车而目标车辆分割的质量直接关系到最后车型分类的结果;如何在汽车的众多特征中选择具有代表性的特征,并将其转化成有效的参数也格外重要;在得到特征参数后,如何正确选择和设计分类器也直接影响着最后识别的准确率。在本文中,我们的神经网络采用原始图像作为网络的输入,原始数据通过由卷积层、完全连接层以及 Softmax 层组成的深度神经网络进行处理,省去了图像分割、手工提取特征等步骤。

表 1 VGG16 网络配置

	感受野	步长	填充	输出大小
Input:224 x 224 RGB 图像				
Conv	3x3	1	1	64
Conv	3x3	1	1	64
Max-Pooling				
Conv	3x3	1	1	128
Conv	3x3	1	1	128
Max-Pooling				
Conv	3x3	1	1	256
Conv	3x3	1	1	256
Conv	3x3	1	1	256
Max-Pooling				
Conv	3x3	1	1	512
Conv	3x3	1	1	512
Conv	3x3	1	1	512
Max-Pooling				
Conv	3x3	1	1	512
Conv	3x3	1	1	512
Conv	3x3	1	1	512
Max-Pooling				
FC	-	-	-	4096
FC	-	-	-	4096
FC	-	-	-	4
Softmax				

表 2 AlexNet 网络配置

	感受野	步长	填充	输出大小
Input:224 x 224 RGB 图像				
Conv	11x11	4	0	96
Max-Pooling				
Conv	5x5	1	2	256
Max-Pooling				
Conv	3x3	1	1	384
Conv	3x3	1	1	384
Conv	3x3	1	1	256
Max-Pooling				
FC	-	-	-	4096
FC	-	-	-	4096
FC	-	-	-	4
Softmax				

Conv 表示卷积层,FC 表示完全连接层,Pooling 表示池化层,Softmax 表示分类层

1 数据集

我们数据集包含运动型多功能车(SUV)、皮卡车、面包车以及小轿车四类,训练集和测试集分别包含 998 张和 248 张图像。该数据集是从各个角度下拍摄的汽车图像,图像大小不统一,背景复杂,目标车辆在整个图像中所占比例相差较大,这些因素大大增加了车型识别的难度。

为了保持网络输入大小一致性,我们将原始图像调整为统一的 256x256x3 尺寸。然后,我们计算出所有图像 RGB 三个通道的均值,对输入数据进行 0 均值标准化处理。在网络训练测试时,选取 224x224x3 的样本作为输入。

2 网络结构

本文采用文献[2]中提出的深度神经网络 VGG16。VGG16 网络在 ILVRC 2014(ImageNeLarge-Scale Visu-



图 1

al Recognition Challenge 2014) 挑战赛中获得第一名。该网络主要包含 5 个堆栈式的卷积神经网络 ConvNet, 3 个完全连接层以及 1 个 Softmax 层, 是一个“网络中的网络”架构。每个 ConvNet 由多个卷积层组成, 后面紧跟着 Max-Pooling 层。在卷积和池化之后, 经过 3 层的完全连接, 最后一个完全连接层的输出作为 Softmax 层的输入, 生成车型分类的结果。该网络增加了非线性的 ReLU 层, 卷积层和完全连接层的输出都经过 ReLU 的处理, 这可以大大缩短了网络训练的时间。此外, 该网络还使用一种正则化方法 Dropout, 避免在完全连接层上出现过拟合(Overfitting)的现象。

表 3 VGG16, AlexNet 以及 KNN 分类结果

模型/算法	准确率
VGG16	97.6%
AlexNet	93.0%
KNN	50.4%

而另外一个网络 Alexnet<sup>[3]</sup>, 该网络相比于 VGG 简单, 该网络主要由 5 个卷积层, 3 个完全连接层和 Softmax 层组成, 部分卷积层后紧跟着 Max-Pooling 层。该网络同样应用了非线性的 ReLU 层, 重叠池化(Overlapping Pooling) 的方式在某种程度上降低了过拟合的几率。

### 3 实验

我们的深度神经网络 VGG16 和 AlexNet 搭建在

Caffe 框架, 运行在 GeForce GTX TITAN X GPU 工作站, 运算效率很高。单一的网络训练只需要 2 个小时, 测试一张图片仅需要大约 0.2 秒。除了使用网络进行训练和测试外, 我们还采用了经典的分类算法 KNN<sup>[4]</sup>对车型进行分类, 表 2 为我们的实验结果。从表中可以看出, VGG16 网络对汽车车型分类有最好的表现, 准确率高达 97.6%, 其次为 AlexNet 网络, 为 93.0%, 而传统的 KNN 算法对于背景复杂的图片, 分类能力最弱, 仅有 50.4%。图 2 为 VGG16 网络错误分类的个别图像, 从图中可以看出, 面包车是完全正确分类的。错误分类的三个 SUV 车型中, 第一个是因为车型结构太类似皮卡车, 第二个是由于车的颜色由两部分组成, 而红色部分刚好与皮卡车车型雷同, 第三个的背景混入了其他车, 这对分类结果也造成了一定影响。对于图像中仅包含车头的车型识别, 其本身也存在一定难度, 不同车型可能从正面角度观看的差异性不大, 人也可能做出错误判断。

### 4 结语

本文主要采用深度学习方法, 结合先进的深度学习框架 Caffe 和强大计算能力的 GPU, 对四类汽车进行车型识别。实验结果显示, VGG16 网络的准确率最高, 而传统的经典分类算法准确率只有 VGG16 的一半左右。由此可见, 深度神经网络具有强大的学习能力, 在图像分类问题上具有强大的优势。在今后的工作中, 我们会尝试调整深度神经网络的结构, 将其应用到更多



图 2 VGG16 预测错误样本

类型的图像分类问题中去。

参考文献:

- [1]Hinton G E, Osindero S, Teh Y W. A Fast Learning Algorithm For Deep Belief Nets[J]. Neural Computation, 2006, 18(7):1527–1554.
- [2]Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25:2012.
- [3]Simonyan K, Zisserman A, Simonyan K, et al. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Eprint Arxiv, 2014.
- [4]K.P Soman, Shyam Diwakar, V.Ajay. Insight into Data Mining Theory and Practice. China Machine Press.

作者简介:

王茜(1990–),女,四川眉山人,在读硕士研究生,研究方向为机器智能

通讯作者:张海仙(1980–),女,河南邓州人,博士研究生,副教授、研究方向为机器智能,Email:zhanghaixian@scu.edu.cn

收稿日期:2015–11–12

修稿日期:2015–12–10

## The Depth of Vehicle Recognition Based on Neural Network

WANG Qian,ZHANG Hai-xian

(College of Computer Science, Sichuan University,Sichuan 610065)

Abstract:

Studies the recognition of vehicle types based on deep learning methods. Deep neural network is trained to classify automobile images which are shot from different angles with complex background. Uses the cutting-edge Deep Learning architecture, Caffe and a powerful computational platform, GPU. VGG16 network and AlexNet network are trained and tested for this task respectively. Moreover, applies the classical algorithm, K-Nearest Neighbor for comparison. The result suggests that VGG16 network outperform other methods by a big margin with the accuracy of 97.58%.

Keywords:

Deep Learning; Vehicle Recognition; Convolutional Neural Network