**车辆的深度卷积神经网络模型识别和面部识别**

**摘要**

近年来，车辆识别已经成为智能交通监控和管理中的重要应用。本文提出了一个超过九层的深度卷积神经网络。本研究使用从多方收集的车辆数据集，使用深度学习框架Caffe验证所提出的算法。由于使用深度卷积神经网络，本文提出的模型与传统的需要车辆定位并且准确率低的机器学习车辆识别相比具有更好的性能。

**关键字：**车辆识别，人脸识别，九层网络，深度学习

# 引言

随着信息技术的发展，人工智能在人们的日常生活中变得越来越重要。因此，开发人工智能的重要性日益凸显。但是，传统机器学习算法存在识别效率低，识别效率精度低等缺陷。深度学习问世后，识别精度和效率都有了显著提高。越来越多研究人员涉足深度学习。实际上，几年前已经有人提出了神经网络。由于深神经网络需要大量数据这一条件未能满足，因此深度学习一直不温不火。幸运的是，当前的GPU与高度优化的2D卷积实现，足以支持深度卷积神经网络的训练。

最近几年提出了一些CNN模型，它们具有各自的差异和创新之处。例如Alexnet [1]，它具有五个卷积层和三个完全连接层。另外，Resnet [2]是一个热点最近。许多研究人员还将它们用于面部识别[3]，检测[4]和视频跟踪[5]。据我们所知，深度学习算法被广泛用于人脸识别。

网络模型都需要工具来实施，因此人们提出了一些深度学习框架。他们支持

不同的编程语言接口，大大加快了模型的运算速度。一些框架，如Caffe，Torch，Theano等，也可以提高模型准确度。

在本文中，我们使用Caffe框架来确定基于八层网络和九层网络的不同性能。九层网络的Top-1准确率准确度达到92.25％， 而少于8层的网络准确度低于80％。此外，相同的9层网络 使用相同的参数应用到人脸数据集，其中准确率仅仅达到80.5％

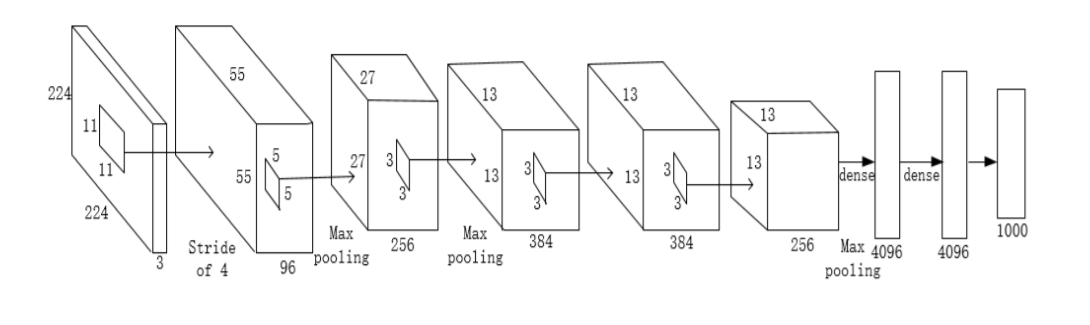
本文的其余部分安排如下：第二部分介绍了当前研究现状，第三部分介绍了九层神经网络和车辆识别处理。部分实验结果将通过第四部分中的深度学习框架Caffe进行报告，最后，在第五部分对本文进行总结。

# 研究现状

深度学习之所以流行，是因为它能够独立的从数据中学习有用的特征。此外，它使用大规模数据集来研究更复杂的概念，并使用大规模并行计算来完成优化。因此，加州大学伯克利分校的贾阳清博士研发了一种清晰有效的深度学习框架Caffe，方便研究人员实施深度学习算法。

Caffe有很多优点。首先，模型和优化是文本形式而不是代码形式，并且Caffe还提供了模型、优化设置和训练权重的定义。其次，Caffe与cuDNN结合使用。因此，Caffe可以运行最佳模型并支持大量数据[6]。 此外，Caffe易于扩展新任务和设置。开发人员可以使用一种简单的语言（例如Google protobuffer）来定义新的网络结构，并通过使用CPU或GPU来实现。最重要的是，研究人员可以发布他们的模型，并分享自己的研究结果。

此外，不同的CNN网络具有不同的应用场景。 Alexnet是一个八层神经网络，结构如图1所示。前两个卷积层的内核大小分别为11×11和5×5，其中有池化层。接下来两层是每个内核大小为3×3且没有池化层的卷积层（即3th 4th）。不过第五卷积层使用3×3内核大小并且具有池化层。此外，前两个全连接层使用dropout，以防止过拟合。



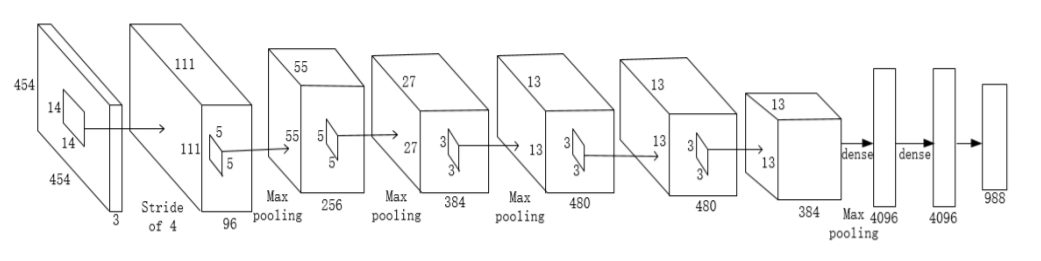
**图1 Alexnet在Caffe框架上的结构**

# 神经网络介绍

在本节中，图2总结了九层网络的体系结构。它包含六个卷积层。 和三个完全连接的层。接下来是车辆识别模型和过程描述。

本文提出的九层网络模型包括两个部分。一部分如图2所示，有卷积层，池化层和全连接层。卷积层负责特征提取和去除噪音。池化层负责减少参数和提高操作速度。 最后，全连接层负责分类和回归。

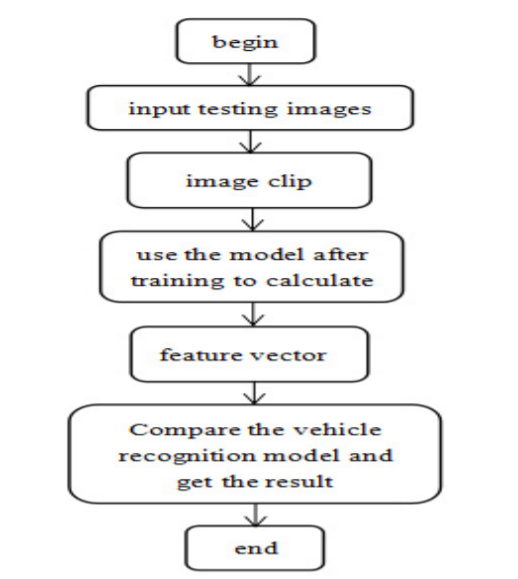
另一部分也有一些层。众所周知，激活功能对于深度学习非常重要，它使得深度学习不再是仅仅支持线性组合，而是可以近似任意函数。它赋予隐藏层意义。但是，如何选择激活函数是一个难题。在这里，提出的模型使用ReLU函数，ReLU有一些优点。首先，使用ReLU激活函数可以节省很多计算。其次，它可以避免梯度的消失[7]。再次，ReLU可能使网络稀疏并减少参数，减轻了过拟合的问题。 LRN层模仿生物神经建立局部神经元活动竞争，使得响应随着输入值的增大而增大，提高了模型泛化能力[8]。此外，过拟合是深度学习的主要问题，所以dropout是必要的，即将每个隐藏神经元的输出设置为零，每个神经元被dropout的概率为0.5 [9]。我们在图2的前三个全连接的层中使用dropout。、



**图2 九层神经网络的Caffe框架**

如图2所示，第一卷积层具有454×454×3输入图像，96个尺寸为14×14的卷积核和一个使用最大池化函数[10]的池化层。第二个卷积层具有256个大小为5×5的卷积核，并使用最大值池化层。第三卷积层具有384个3×3大小的卷积核，还具有使用最大值函数的池化层。此外，接下来的两个卷积层（即第4个，第5个）没有任何池化层和规范化层[11]。但是第四卷积层有480个3×3大小的卷积核， 第五个卷积层有380个3×3大小的卷积核。第六个卷积层有384个3×3大小的卷积核，并使用最大值池化。此外，全连接的层每个都有4096个神经元， 我们在前两个全连接的层中使用dropout。最后，将获得988个类。

此外，要实现车辆识别，它仍然需要几个步骤，如图3所示。首先需要使用一些测试图像，并将图像裁剪为454×454。其次，训练后的模型将用来计算。然后，我们可以获得一些特征向量并与车辆识别模型进行对比。最后，将得到结果。



**图三 车辆识别流程**

# 实验结果

在本节中，我们评估两种情况。然后，我们从不同的位置收集了大量的车辆数据集，车辆数据集包括998个类别和近90000张图像，并分为72000张训练图像 和18000个验证图像。人脸数据集包括500种类别和超过5000张图像，并分为 4000个训练图像和1000个验证图像。我们评估前1名和前5名的准确率。

所有实验我们都使用相同的学习率，如表1所示。 图像被随机采样到尺寸为454×454规格。并使用批处理大小为128的SGD [12] ，梯度下降学习率始于0.02，每55000次迭代除以10，然后对模型进行多达 45×10^4次迭代。权重衰减和动量分别为0.0005和0.9。

**表1 模型参数设置**

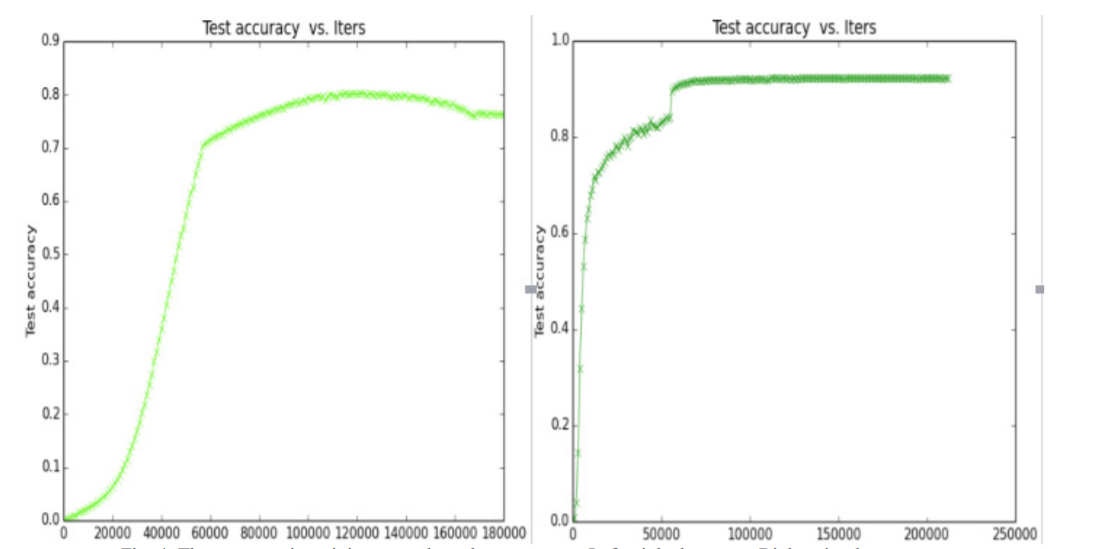
|  |  |  |
| --- | --- | --- |
| **参数** | **值** | **含义** |
| **Batch size** | **128** | **每次训练使用的图像数量** |
| base\_lr | **0.02** | **初始学习率** |
| lr\_policy | **Step** | **学习率变化趋势** |
| stepsize | **55000** | **每次更改学习率需要达到的迭代次数** |
| max\_iteration | **450000** | **迭代次数最大值** |
| momentum | **0.9** | **超参数** |
| weight\_decay | **0.0005** | **超参数** |

## 基于车辆数据集的八层神经网络和九层神经网络

我们首先评估八层神经网络（Alexnet）和九层神经网络，并进行比较。从表2的结果可以看出，基于车辆数据集，九层网络的准确率要高于八层网络。换句话说，就准确率而言，九层网络的性能更好。为了揭示原因，在图4中，我们比较了它们在训练过程中的准确率。

**表2.八层网络和九层网络的Top-1和Top-5准确率**

|  |  |  |
| --- | --- | --- |
|  | **Top-1准确率** | **Top-5准确率** |
| 8层神经网络 | 79.86% | 90.57% |
| 9层神经网络 | 92.25% | 97.51% |



**图4 关于两种网络训练处理的准确率。左：八层网络。右：九层网络**

如图4所示，在准确率方面，九层网络表现更好，准确率超过8层网络12％。这说明添加一层网络后，准确率有了明显提高。

## 基于车辆和人脸数据集的九层神经网络

此外，我们使用九层网络来测试人脸数据集。将图像切割成454×454尺寸。此外，实际上，人脸只占整个图像数据集的小部分，而车辆占据了整个图像数据集的大部分。表3总结了人脸数据集上的结果。

**表3.模型在人脸数据集上的准确率**

|  |  |  |
| --- | --- | --- |
|  | **Top-1准确率** | **Top-5准确率** |
| 人脸数据集（使用9层网络模型） | 80.55% | 91.22% |

显然，基于面部数据集的准确率远低于基于车辆数据集的准确率。我们估计原因是有效部分在整个图像中所占的比例不同。该问题将在未来进一步研究。

# 总结

据我们所知，越来越多的研究致力于计算机视觉识别，例如人脸识别和车辆识别。在本文中，我们提出了一个九层网络，使用从多个角度来拍摄的车辆数据集。结果表明，九层网络在车辆识别方面取得了破纪录的结果，使用该方法和深度学习框架Caffe，车辆识别的准确率达到了92.2％以上。另外，在识别精度方面，八层网络比九层网络低很多。

此外，还有许多潜在的研究值得我们不断探索和发掘。我们不断优化算法模型，以提高识别的准确性和速度。最后，我们将使用其他计算机视觉上的新模型，以供将来的识别使用。

# 参考文献

1. K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun. What is the best multi-stage architecture forobject recognition? In International Conference on Computer Vision, pages 2146–2153. IEEE, 2009.

2. J. Sánchez and F. Perronnin. High-dimensional signature compression for large-scale image classification. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 1665–1672. IEEE,2011.

3. Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep Residual Learning for Image Recognition. In Computer Vision and Pattern Recognition (CVPR), 2016

4. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks (NIPS), 2012

5. Steve Lawrence, C. Lee Giles, Ah Chung Tso, Andrew D. Back. Face Recognition: A Convolutional Neural Network Approach. IEEE Transactions on Neural Networks, 1997, 8(1):98-113

6. Y. LeCun, K. Kavukcuoglu, and C. Farabet. Convolutional networks and applications in vision. In Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on, pages 253–256. IEEE, 2010.

7. Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, et al. Handwritten digit recognition with a backpropagation network. In Advances in neural information processing systems, 1990.

8. S. Gidaris and N. Komodakis. Object detection via a multi-region & semantic segmentation-aware cnn model. In ICCV, 2015.

9. K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In ECCV, 2014.

10. Haoxiang Liy, Zhe Linz, Xiaohui Shenz, Jonathan Brandtz, Gang HuayA Convolutional Neural Network Cascade for Face Detection. In CVPR, 2015.

11. AJ SpinkˈRAJ TegelenboschˈMOS BumaˈLPJJ Noldus. The EthoVision video tracking system—A tool for behavioral phenotyping of transgenic mice. Physiology &Behavior, 2001, 73(5):731-44

12. S SaltiˈA CavallaroˈSL Di. Adaptive appearance modeling for video tracking: survey and evaluation.ǉIEEE Transactions on Image ProcessingǊ, 2012, 21(10):4334-48

13. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, vZ. Huang, A. Karpathy, A.Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. arXiv:1409.0575, 2014.

14. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Le-Cun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In ICLR, 2014.

15. J TsitsiklisˈD BertsekasˈM Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithmsǉIEEE Transactions on Automatic ControlǊ, 1984, 31(9):803-812