

文章编号: 2096-1618(2017)05-0503-05

# 基于 He-Net 的卷积神经网络算法的图像分类研究

王强, 李孝杰, 陈俊

(成都信息工程大学计算机学院, 四川 成都 610225)

**摘要:**基于 Caffe 深度学习框架和首层卷积层取反操作,提出了 He-Net 网络模型解决图像分类问题。该模型主要由 3 个卷积层和最大池化层及 3 个完全连接层组成,首层采用卷积层取反以增加有效特征信息的传递,卷积层采用较小的卷积核提取更多的纹理特征,最后采用 Softmax 分类器实现图像分类。为了使训练更快速,采用更加高效的 GPU 运算实现卷积操作。针对相同数据集,同经典的网络模型 CaffeNet、AlexNet 实验比较,He-Net 网络模型具有更高的分类正确率。

**关键词:**人工智能;深度学习;Caffe 框架;卷积神经网络;图像分类;激活函数

**中图分类号:**TP301.6

**文献标志码:**A

**doi:**10.16836/j.cnki.jcuit.2017.05.007

## 0 引言

1985 年, Hinton 等<sup>[1]</sup>在《Nature》上发表的一篇用于神经网络的反向传播 (back-propagation, BP) 算法, 推动了神经网络的飞速发展。然而神经网络面临着过拟合、网络参数高达百万甚至千万原因导致训练难度大, 训练周期过长, 导致许多学者放弃了对神经网络的研究。深度学习<sup>[2]</sup>是神经网络的机器学习的一个分支<sup>[3]</sup>。2006 年, 由机器学习领域的 Hinton 等<sup>[4]</sup>在《Science》发表文章, 首次提出深度学习的思想, 大大地解决了以上问题, 促进了神经网络的发展。目前, 经典的深度学习的网络模型共有 3 类, 卷积神经网络<sup>[4]</sup> (convolutional neural networks, CNN) 模型、受限莫尔兹曼<sup>[5]</sup> (restricted boltzmann machines) 模型及自动编码器 (Auto-encoder) 模型<sup>[6]</sup>。其中, CNN 模型应用范围较广, 该网络模型主要采用稀疏连接和权值共享, 网络结构更加类似生物神经网络结构, 有效降低网络模型复杂度, 减少权值的个数。2015 年 Yann LeCun 等<sup>[7]</sup>, 详细陈述了深度学习在语音识别<sup>[8]</sup>、图像分类<sup>[4]</sup>、目标检测<sup>[9]</sup>等领域取得的重大成果。

随着 GPU 计算平台被直接用于科学计算, 极大的提高数据的计算速度, 为深度学习提供了硬件支持。目前, GPU 应用于科学计算主要以 NVIDIA (英伟达) 公司的 CUDA 框架作为开发的底层系统。基于该架构开发许多深度学习框架, 如 Caffe、Theano、Torch、MXNet 等。其中, Caffe 框架可以很好地实现卷积神经网络模型, 是一个基于 C++/CUDA 架构, 完全开源而高效的深度学习框架, 支持命令行 Python 和 MATLAB

接口, 可实现 CPU 和 GPU 无缝切换。Caffe 给出了模型定义, 可以使用 Caffe 提供的各层类型定义卷积神经网络模型, 最优化设置及预训练的权重。另外, 采用 GPU 处理图像的速度较快, 如基于 Caffe 框架的 AlexNet<sup>[2]</sup> 模型, 在 NVIDIA Tesla K40 上处理每张 256256 图片只需要 1.17 ms。

## 1 卷积神经网络

CNN 是一种监督学习的深度结构下的机器学习模型<sup>[2]</sup>, 具有极强的适应性。该网络采用权值共享结构, 类似于生物的神经网络结构, 善于挖掘数据局部特征, 组成全局训练的特征用于图像分类。标准卷积神经网络结构如图 1 所示。卷积神经网络是通过一系列隐含层映射, 得到长度为  $C$  的概率向量  $\mathbf{Y} \in \mathbf{R}^{C \times 1}$ , 其中  $C$  是类别个数, 向量  $\mathbf{Y}$  中的每个概率值对应一个类, 采用 One-hot 编码, 通过最大值方法判定样本的类别。

隐含层主要由以下 4 个部分组成:

(1) 卷积层: 使用若干个卷积核 (滤波器) 在图像矩阵上滑动, 对应位置做卷积运算, 最终形成新的图像矩阵, 滑动完成后即完成了对原始图像的卷积变换。

(2) 激活函数<sup>[10]</sup>: 使用一个激活函数对每个卷积结果进行矫正运算。常见的几种激活函数有 Sigmoid, Tanh, ReLU。

(3) 池采样层: 对图像进行子采样可以减少来自上层隐含层计算复杂度, 同时保持图像旋转不变性。包含 3 种池化方法: 最大值池化、均值池化、随机池化。

(4) 局部采样标准化层 (local response normalization, LRN): 对已经池采样的图像进行对比度归一化操作, 使其输出的特征信息更加平稳。

收稿日期: 2017-05-06

基金项目: 国家自然科学基金青年基金资助项目 (61602066); 四川省教育厅资助项目 (17ZA0063)

隐含层中的卷积层和池采样层是实现卷积神经网络特征提取功能的关键,最后一层输出层是一个分类器,可以采用 Softmax 回归<sup>[11]</sup>。该网络结构采用梯度下降算法<sup>[12]</sup>,根据最小化目标函数对网络中的权重参数逐层反向调整,不断的迭代训练提高网络精度。标

准的卷积神经网络存在提取特征不足等缺陷,制约着图像的分类效果,提出的 He-Net 卷积神经网络结构采用首层卷积层取反增加有效特征的传递,并且采用较小的卷积核提取更多的纹理特征,有效解决标准卷积神经网络存在的缺陷,提高图像的分类效果。

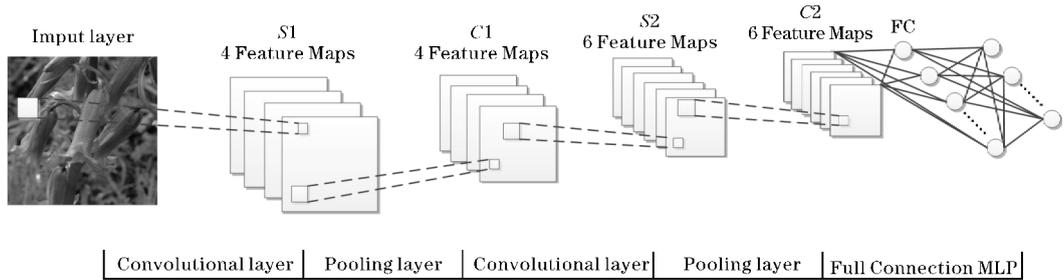


图1 标准卷积神经网络结构(C为卷积层,S为特征映射层即池采样层,FC为全连接层)

基于 Caffe 框架<sup>[13]</sup>的 CNN 的具体实现过程如图 2 所示。

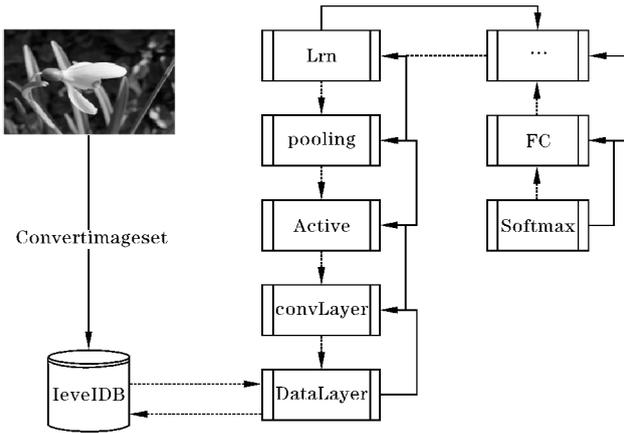


图2 Caffe 架构的 CNN 实现过程

## 2 He-Net 模型设计

基于 Caffe 框架,提出了一个 He-Net 卷积神经网络模型,其网络结构如图 3 所示。该网络使用 3 个卷积层(C1~C3),对首层卷积层取反并同原卷积层一起传递到下一层,增加了有效特征信息的传递。每一个卷积层后作用一个 ReLU 激活函数。同时每一个卷积层后面加上一个最大池化层(S1~S3),使用 3 个全连接层,最后一个全连接层的输出作为 Softmax 的输入,最终生成图像的分类结果。为避免在全连接层上出现过拟合的现象,该网络采用了一种正则化方法 Dropout。网络参数设置为第一层卷积核数目为 96,每经过一次最大池化层,卷积核数目依次递增,偏置初始化为 0.1。下面将从该模型的目标函数、首层卷积层、重叠采样层、激活函数选取几个方面介绍提出的高准确率图像分类模型。

### 2.1 标函数

模型的目标函数是 Softmax 函数,该函数是属于监督学习算法 Softmax 回归,使其同 He-Net 模型结合完成图像分类。在 Softmax 回归中,面对多分类问题,类标  $y$  可以取  $k$  个不同的值。设训练集为  $\{(x^{(1)}, y^{(2)}), \dots, (x^{(i)}, y^{(i)})\}$ ,其中,  $x^i$  代表第  $i$  个样本,  $y^{(i)}$  代表类别,  $y^{(i)} \in \{1, 2, \dots, k\}$ 。

Softmax 回归算法的代价函数如公式(1)所示。

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)}=j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] \quad (1)$$

其中  $1\{\cdot\}$  为指示函数,其规则为  $1\{\cdot\}$  值为真的表达式  $= 1$ ,  $1\{\cdot\}$  值为假的表达式  $= 0$ 。代价函数中对标记的  $k$  个可能值进行累加,将第  $i$  个样本  $x^{(i)}$  分类为类别  $j$  的概率如(2)式所示。

$$p(y^{(i)}=j | x^{(i)}; \theta) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}} \quad (2)$$

对于  $J(\theta)$  的最小化问题,使用迭代优化算法,比如:梯度下降算法,经过求导,得到梯度如(3)式所示。

$$\Delta_{\theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^m [x^{(i)} (1\{y^{(i)}=j\} - p(y^{(i)}=j | x^{(i)}; \theta))] \quad (3)$$

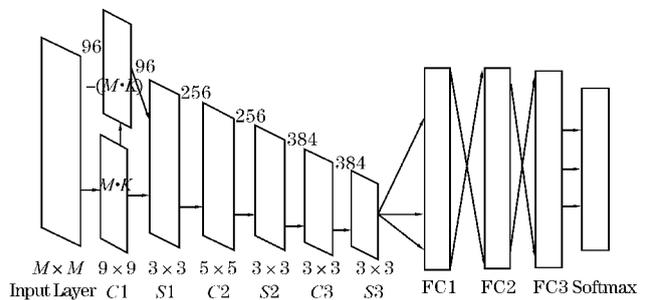


图3 He-Net 模型结构

(C表示卷积层,S表示最大池化层,FC表示全连接层)

## 2.2 首层卷积层

由于首层卷积层距离原始图像最近,且后续的层结构都取决于第一个卷积层的输出,传递的特征信息将直接影响最终的图像分类效果。为保证更多特征信息传递至下一层有利于图像的分类,模型采用将首层卷积层取反的方法。另外考虑使用较小的卷积核能够提取到细节的变化,模型采用较小的卷积核。各层的卷积核大小如图 3 所示。

## 2.3 重叠采样

标准的卷积神经网络的 Max-Pooling (最大池化)层不会采用重叠采样,即假设步长大小为  $S$ ,池化层核大小为  $K$ ,那么  $S=K$ 。在 AlexNet 中采用重叠采样,使用  $K=3, S=2$ ,这样可以减少过拟合。模型网络结构也采用重叠采样,发现减少层数的同时,可以保证空间的不变性。

## 2.4 激活函数选取

模型选择 ReLU 作为激活函数,其函数表达式为  $f(x) = \max(0, x)$ ,当输入特征值小于 0 时,输出为 0,当输入特征值大于 0 时,输出等于输入。ReLU 主要具有如下优点:

(1) ReLU 得到的随机梯度下降收敛速度快。

(2) ReLU 只需要一个阈值就可以得到激活值,计算复杂度低。

## 3 实验结果与分析

### 3.1 实验设置

所有对比实验均在 Caffe 框架下完成,配置环境使用 ubuntu14.04 操作系统,2 个 NVIDIA Tesla K40 GPU 的高性能工作站。并且对比模型 He-Net、AlexNet 和 CaffeNet,3 种网络结构具体设置如表 1(0~6 层)和表 2(7~13 层)所示,表中的卷积层或最大池化层(Max-Pooling 层)的参数按照“Conv<卷积核大小/池化层大小>-<步长>-<填充边界>-<数量>”定义,如模型 AlexNet 中第一个卷积层的参数为“Conv11-4-0-96”,表示卷积核大小为  $11 \times 11$ ,步长为 4,填充边界为 0,卷积核个数为 96。每一个卷积层后面作用一个激活函数 ReLU。

3 种网络模型的配置文件中 base\_lr(初始学习率)均设为 0.001, max\_iter(最大迭代次数)均设为 500 次,采用 GPU 进行训练,每 50 次迭代进行一次验证,并输出识别率。

表 1 3 种网络结构设置(0~6 层)

模型	0 输入层	1 卷积层	2 Max-Pooling	3 卷积层	4 Max-Pooling	5 卷积层	6 Max-Pooling
He-Net	256×256	Conv 9-4-0-96	MP 3-2-0-96	Conv 5-0-2-256	MP 3-2-0-256	Conv 3-0-1-384	MP 3-2-0-384
CaffeNet	256×256	Conv 11-4-0-96	MP 3-2-0-96	Conv 5-0-2-256	MP 3-2-0-256	Conv 3-0-1-384	MP 3-2-0-384
AlexNet	256×256	Conv 11-4-0-96	MP 3-2-0-96	Conv 5-0-2-256	MP 3-2-0-256	Conv 3-0-1-384	无

表 2 3 种网络结构设置(7~13 层)

模型	7 卷积层	8 卷积层	9 Max-Pooling	10FC	11FC	12FC	13 分类器
He-Net	无	无	无	2048	2048	5	Softmax
CaffeNet	Conv 3-1-1-384	Conv 3-1-1-256	MP 3-2-0-256	4096	4096	5	Softmax
AlexNet	Conv 3-1-1-384	Conv 3-1-1-256	MP 3-2-0-256	4096	4096	5	Softmax

### 3.2 实验一

实验使用的真实物体图像 RO 数据集具有大巴车、恐龙、大象、鲜花和马的 5 个类,每个类 200 张图片,采用交叉验证法<sup>[14]</sup>,随机选取 250 张作为训练,50 张作为验证。原始图像大小均为  $256 \times 256$ ,图像均为 jpg 格式,该数据集部分图像如图 4 所示。

通过对比实验,3 种网络模型的识别率随着迭代次数的变化曲线如图 5 所示。

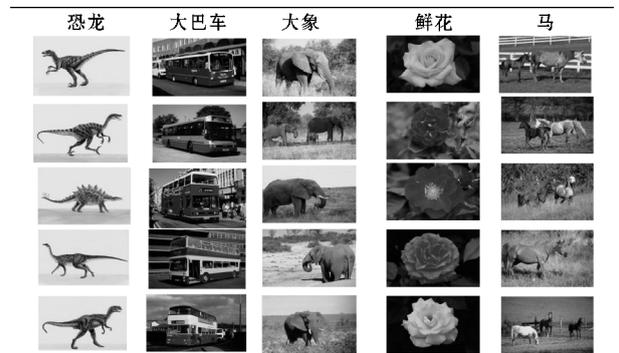


图 4 部分 RO 数据集

由图5可知;当迭代50次时,CaffeNet的识别率为70%,AlexNet的识别率为22%,而He-Net网络的识别率已为80%;当迭代次数为100次时,CaffeNet的识别率为90%,AlexNet的识别率为77%,而He-Net网络的识别率已为93%。

提出的He-Net网络模型的最高识别率为93%,CaffeNet网络模型的最高识别率为90%,而AlexNet网络模型的最高识别率为84%。另外,He-Net网络模型在迭代50次时开始收敛,均优于CaffeNet的100次和AlexNet的150次。

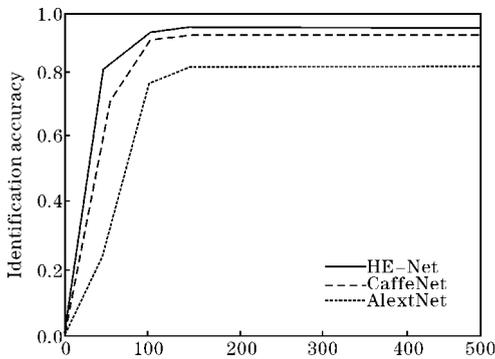


图5 实验一:3种网络模型随着迭代次数变化的识别率

### 3.3 实验二

实验采用牛津大学图像识别小组 Maria-Elena Nilsback 和 Andrew Zisserman 收集的花卉种类识别 Flower 数据集<sup>[15]</sup>。该数据集图像相似度高,具有17类,每类80张图片,共1360张图片,其中每类选70张作为训练,10张作为测试。该数据集部分图像如图6所示。



图6 部分 Flower 数据集

通过对比实验,3种网络模型的识别率随着迭代次数的变化曲线如图7所示:

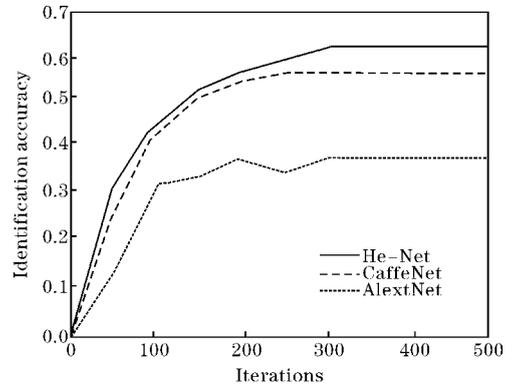


图7 实验二:3种网络模型随着迭代次数变化的识别率

由图7可知;当迭代50次时,CaffeNet的识别率为24%,AlexNet的识别率为15%,而He-Net网络的识别率已为43%;当迭代次数为100次时,CaffeNet的识别率为43%,AlexNet的识别率为33%,而He-Net网络的识别率已为63%。

提出的He-Net网络模型,对于相似度很大的花卉图像分类依旧不错,分类正确率高达63%,明显高于CaffeNet网络模型的54%和AlexNet网络模型的39%。另外He-Net网络模型在迭代150次左右便开始收敛,均优于CaffeNet模型的200次和AlexNet的300次。

### 3.3 实验三

实验为了验证He-Net网络模型的第一层卷积核大小,对于精准度和收敛时迭代次数的影响,改变第一个卷积核的大小,分别记录了最高正确率和收敛时迭代次数,结果如表3所示。

表3 第一层卷积核大小的影响

数据集	第一层卷积核大小	正确率/%	收敛时迭代次数
RO	11×11	90	140
	10×10	90	140
	9×9	93	80
	8×8	91	180
	5×5	89	220
Flower	11×11	58	150
	10×10	58	100
	9×9	63	100
	8×8	62	140
	5×5	60	140

实验表明,正确率大小以和收敛时迭代次数受第一层卷积核大小影响较大,对于He-Net网络结构模型,最优的第一层卷积核大小应该选择9×9。

## 4 结束语

提出了一个基于 Caffe 框架的卷积神经网络模型 He-Net,模型结构采用将首层卷积核取反增加有效特征信息的传递,并且采用重叠采样的最大池化等方法,通过同主流分类模型 CaffeNet、AlexNet 的对比试验,模型具有更高的图像识别率。且验证了该模型中第一层卷积核最优大小为  $9 \times 9$ ,精准度最高,迭代次数最少时收敛。在今后的工作中,将调整模型的参数,将它运用到更多实际分类问题中。

致谢:感谢成都信息工程大学科研基金(KYTZ201608)对本文的资助

## 参考文献:

- [1] Hinton G E, Rumelhart D E, Williams R J. Learning internal representations by back-propagating errors[J]. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1985, 1.
- [2] 韩小虎,徐鹏,韩森森.深度学习理论综述[J]. *计算机时代*, 2016, (6): 107-110.
- [3] Haykin S, Network N. A comprehensive foundation [J]. *Neural Networks*, 2004, 2.
- [4] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]. *Advances in neural information processing systems*. 2012; 1097-1105.
- [5] Hinton G E. A practical guide to training restricted boltzmann machines [M]. *Neural networks: Tricks of the trade*. Springer Berlin Heidelberg, 2012; 599-619.
- [6] Zhang J, Shan S, Kan M, et al. Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment [C]. *European Conference on Computer Vision*. Springer International Publishing, 2014; 1-16.
- [7] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521 (7553): 436-444.
- [8] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups [J]. *IEEE Signal Processing Magazine*, 2012, 29 (6): 82-97.
- [9] Ciodaro T, Deva D, De Seixas J M, et al. Online particle detection with neural networks based on topological calorimetry information [C]. *Journal of Physics: Conference Series*. IOP Publishing, 2012, 368(1).
- [10] Schmidhuber J. Deep learning in neural networks: An overview [J]. *Neural networks*, 2015, 61: 85-117.
- [11] Hung C, Nieto J, Taylor Z, et al. Orchard fruit segmentation using multi-spectral feature learning [C]. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013: 5314-5320.
- [12] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition [J]. *Neural computation*, 1989, 1 (4): 541-551.
- [13] 陈致远.基于区域梯度统计分析与卷积神经网络的条码定位算法研究[D].上海:上海交通大学, 2015.
- [14] Yadav S, Shukla S. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification [C]. *Advanced Computing (IACC)*, IEEE 6th International Conference on. IEEE, 2016: 78-83.
- [15] Nilsback M E, Zisserman A. Automated flower classification over a large number of classes [C]. *Computer Vision, Graphics & Image Processing. ICVGIP'08. Sixth Indian Conference on. IEEE*, 2008: 722-729.

## Research on Image Classification based on HE-Net Convolutional Neural Networks

WANG Qing, LI Xiao-jie, CHEN Jun

(College of Computer Sciences, Chengdu University of Information Technology, Chengdu 610225, China)

**Abstract:** Based on the Caffe deep learning framework and the first convolution layer inversion operation, this paper proposes a He-Net network model to achieve image classification problem. The model consists of three convolution layers and the max-pooling layer, and three fully connection layers. The first convolution layer adopts the inversion operation to add the transmission of the effective feature information and uses a smaller convolution kernel to extract more texture features. Finally, the softmax classifier is employed to identify image classifications. In order to make training faster, we use a more efficient GPU to achieve convolution operations. For the same data set, compared with the classic network model CaffeNet, AlexNet experiment, He-Net network model has a higher classification accuracy.

**Keywords:** artificial intelligent; deep learning; Caffe frame; convolution neural network; image classification; activation function