

文章编号: 2095-2163(2021)03-0066-04

中图分类号: TP391.41

文献标志码: A

基于密集卷积神经网络的全卷积池化算法

宋佳霏, 宋欣霞, 杨贺群, 黄若琳

(东华大学 计算机科学与技术学院, 上海 201620)

摘要: 目前用于图像识别的大多数卷积神经网络 (CNN) 都使用相同的原理构建, 即: 卷积层、池化层、全连接层。文中使用密集卷积神经网络重新评估了用于图像识别的所有组件, 并对池化层不存在的必要性提出了质疑。经过实验, 分析发现池化层可以由步幅增加的卷积层代替, 却不会降低图像识别的准确率。研究中则在 DenseNets 上训练提出的由卷积层替代池化层的方法, 组成新的卷积神经网络体系结构, 并在多个图像分类的数据集 (CIFAR-10, SVHN) 上产生了先进的性能。本文提出了基于密集卷积神经网络 (DenseNets) 的全卷积池化算法, 提高了图像分类的准确率。最后, 在多个经典数据集上进行比较, 实验结果验证了全卷积池化算法的高效性。

关键词: 全卷积; 池化算法; DenseNets

Fully Convolutional Pooling algorithm based on Dense Convolutional Neural Networks

SONG Jiafei, SONG Xinxia, YANG Hequn, HUANG Ruolin

(School of Computer Science and Technology, Donghua University, Shanghai 201620, China)

【Abstract】 Most convolutional neural networks (CNN) currently used for image recognition are constructed using the same principles: convolutional layer, pooling layer, and fully connected layer. The paper re-evaluates all the components used for image recognition using dense convolutional neural networks and questions the need for the pooling layer to not exist. After experiments, it is found that the pooling layer can be replaced by a convolutional layer with an increased stride, but it will not reduce the accuracy of image recognition. The research also trains the proposed convolutional layer instead of the pooling layer on DenseNets to form a new convolutional neural network architecture, and produces advanced performance on multiple image classification data sets (CIFAR-10, SVHN). This paper proposes a fully convolutional pooling algorithm based on dense convolutional neural networks (DenseNets), which improves the accuracy of image classification. Finally, compared on multiple classic data sets, the experimental results verify the efficiency of the fully convolutional pooling algorithm.

【Key words】 fully convolutional; pooling algorithm; DenseNets

0 引言

目前,卷积神经网络 (CNNs) 在计算机视觉领域占据了主导地位,因其在许多应用中都具有出色的性能,例如图像分类、物体检测和姿态估计。从最初问世以来,对网络体系结构的探索一直是卷积神经网络模型研究的一部分。为了获得更高的图像分类精度, CNN 的网络架构变得越来越深,越来越复杂^[1-5]。在神经网络结构的文献中, Fahlman 等人^[6]已经研究了类似于密集网络结构的级联结构。研究中是通过逐层方式训练完全连接的多层感知器。如今,用于进行图像分类的卷积神经网络,组成结构几乎是相同的,包括卷积层、池化层以及全连接层。通

常在卷积层、池化层、全连接层之间都会使用激活函数。此后在网络的训练期间进行层数的丢弃,从而实现网络参数的优化。在 2014 年 ImageNet 挑战赛中,顶级参赛作品^[7] (Simonyan 和 Zisserman 提出的)通过在池化层之间引入多个卷积或者 Szegedy 等人^[1]提出的通过在每一个网络层中构建多个模块的混合卷积池化层。但是分析后却发现所有新增的扩展层在不同的网络体系结构中都有自己的参数和训练过程,因此在进行图像分类的任务中,需要不断实验能实现最先进的分类效果所实际需要哪些组件。根据以往的经验,可以先从最简单的卷积层开始,即整个网络的池化层完全被卷积层替代,可以使用步幅长度为 2,以此来降低图片的维度。由于

作者简介: 宋佳霏 (1996-), 女, 硕士研究生, 主要研究方向: 图像处理、推荐系统; 宋欣霞 (1990-), 女, 硕士研究生, 主要研究方向: 医学信息处理与分析; 杨贺群 (1996-), 男, 硕士研究生, 主要研究方向: 图像处理、深度学习; 黄若琳 (1997-), 女, 硕士研究生, 主要研究方向: 图像处理、联邦学习。

通讯作者: 宋欣霞 Email: 154317441@qq.com

收稿日期: 2020-11-28

在实现降低维度的过程中,是通过步幅为 2 的卷积层而不是池化层实现的,有研究学者 Estrach 等人^[8]发现这也适用于神经网络的可逆性问题研究。

1 密集卷积神经网络模型

1.1 DenseNets 模型结构原理

DenseNets 采用的是密集连接机制,即互连接所有的层。每一层都会与前面所有层在 channel 维度上连接在一起,该层作为下一层的输入,实现了特征重用的功能。DenseNets 主要包括 4 个创新点,分别是:减少了梯度消失的问题;加强了图像特征的传递;更有效地利用了图像;一定程度上减少了参数数量。DenseNets 网络连接方式如图 1 所示。

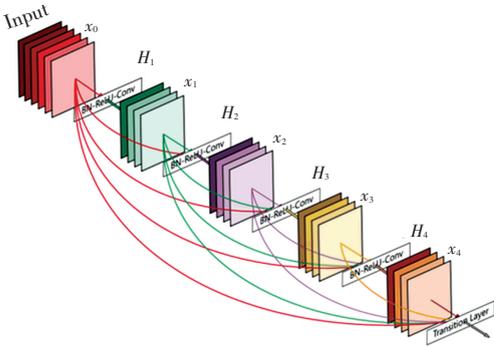


图 1 DenseNets 网络结构图

Fig. 1 DenseNets structure

在 DenseNets 中会连接前面所有层作为输入: DenseNets 网络中, $[X_0, X_1, \dots, X_{l-1}]$ 表示将 0~l-1 层的输出图像特征图做通道的合并,就像 Inception 那样。例如由图 1 可看到,第 X_3 层,是由 X_0, X_1, X_2 的输出作为输入。就是在图 1 中看到的黄色特征图的输入是由红色、绿色、紫色特征图的输出组成。即前面所有层的输出是这一层的输入。也可以认为每一层都直接连接输入层和损失函数,所以可以减轻梯度消失现象。由此可知,在 DenseNets 中,特征传递方式是直接将前面所有层的特征 concat 后传到下一层, $H_l(\cdot)$ 代表非线性转化函数,是一个组合操作,采用的是 BN+ReLU+3 * 3Conv 结构。如式(1)所示:

$$X_l = H_l([X_0, X_1, \dots, X_{l-1}]) \quad (1)$$

由于 DenseNets 采用的密集连接方式,所以需要图像特征图大小保持一致。所以在 DenseNets 网络结构中,使用 Dense Block 块和 Transition (Convolution、Pooling) 结构。Dense Block 块是包含很多层的模块,每一个层需要图像特征图大小相同,层与层之间采用密集连接的方式。这样就可以实现

在每个 Dense Block 块内部的图像特征图大小统一,在进行 concatenation 操作时就不会出现特征图大小不一致的问题。Transition 是位于 2 个不同 Dense Blocks 块之间,通过 Pooling 操作使图像的特征图大小降低。

1.2 池化模型理论

1.2.1 最大值池化算法

最经典的池化方法之一是最大值池化。该方法是将原始的输入数据,进行采样分块实现区域选择,选取每个区域位置的最大值。最大值池化的缺点是无法保留图片的更多背景信息。最大值池化算法原理可写为如下形式:

$$S_{kij}^{\max} = \max_{m, n \in R_{ij}} (F_{kmn}) + b. \quad (2)$$

1.2.2 平均值池化算法

平均值池化也是先经过采样分块后,再进行的是计算池化域中所有元素的平均值输入到下一层。平均值池化并不会只考虑最大的特征,而是考虑全局的元素值。所以,平均值池化后将不会保留特征图的最大特征信息。平均值池化算法公式可写为:

$$S_{kij}^{\text{ave}} = \frac{1}{C} \left(\sum_{m=1}^c \sum_{n=1}^c F_{kmn} \right) + b. \quad (3)$$

但需要指出的是,如果特征图中有很多零元素,则特征图的特性将大大降低。

2 全卷积池化算法

本文在实验中用步长为 2 的标准卷积层替换了 DenseNets 网络中存在的池化层。为了验证此种方案替换的可行性,这里涉及的用于定义 CNNs 中卷积层和池化层操作的数学公式为:

$$s_{i,j,u} (f) = \left(\sum_{h=-[k/2]}^{[k/2]} \sum_{w=-[k/2]}^{[k/2]} |f_g(h,w,i,j,u)|^p \right)^{1/p}. \quad (4)$$

其中, f 表示卷积神经网络中某一层产生的特征图像。该特征图可以将其描述尺寸为 $W \times H \times N$ 的 3 维数组,而 W 表示宽度, H 表示高度, N 表示通道数; k 表示池化域的大小, $k/2$ 表示池化域一半的长度; 步幅长度为 r ; $g(h,w,i,j,u) = (r \cdot i + h, r \cdot j + w, u)$ 是从 s 的位置映射到 f 的步长映射关系; p 是阶数范式。如果 $r > k$, 则池化区域不会重叠;但是,当前的卷积神经网络架构通常包括 $k = 3$ 和 $r = 2$ 的池化重叠区域。

其实,池化层可以看作是按特征图大小进行卷积,其中将激活函数替换为 p 范式。因此,在这里有一个疑问,为什么要在网络层中引入这种特殊的层。

也就是说池化层为什么可以帮助 CNNs 更好地提取特征。本次研究给出 3 种解释： p 范式表示的 CNNs 更加具有不变性；通过池化操作后的空间维数会减少，在较高层中网络有可能会输入较大的图像；池化操作的特征提取功能可以使网络优化变得更加容易。

研究中先假设只有通过池化执行的降维操作对 CNNs 有着至关重要的作用。现在就可以用步长大于 1 的普通卷积层替换池化层。比如说对于 $k=3$ 且 $r=2$ 的池化层，可将其替换为具有相应步幅和卷积核大小的卷积层，输出通道数等于输入通道数。其实删除池化层的研究已然有了一些工作成果，LeCun 等人^[9]提出将池化层称为子采样层，这就指出后续研究可以使用不同的操作进行子采样。

3 实验分析

3.1 实验环境

在本次研究中，所有的实验都是在内存大小为 32 G RAM，操作系统是 Ubuntu16.04 LTS 的计算机上运行。使用开源的机器视觉库 OpenCV，深度学习框架为 TensorFlow。可公开获得的 cuda-convnet^[10] 软件包用于使用单个 NVIDIA GeForce GTX 1080TI GPU 进行实验，另外 CPU 版本为 Intel (R) Core(TM) i7-8700 CPU@3.20 GHz 3.19 GHz。以上算法均采用 Python 语言实现。

3.2 数据集

本文所使用的数据 CIFAR-10 数据集包括 10 个类别，是由 32×32 像素的自然彩色图像组成的集合。研究中遵循 CIFAR-10 的通用实验协议，即选择 50 000 张图像进行训练，选择 10 000 张图像进行测试。实验中网络的参数是通过 10 000 张在验证集上的错误率来选择的。过程中仍然遵循标准 Nesterov Adam 优化器，其 *momentum* 为 0.9，*mini-batch* (最小批量) 为 128。

街景门牌号码 (SVHN) 数据集，包含了 32×32 彩色数字图像。训练集中有 73 257 张图像，测试集中有 26 032 张图像，验证集有 531 131 张图像。仿真时是按照 3:1 的比例划分训练集和测试集，同时使用 *momentum* 为 0.9 的动量和 0.000 1 的权重衰减来训练网络。最小批量大小设置为 128，初始学习率设置为 0.001。同样，训练 100 个 *epochs*，每个 *epoch* 有 20 次迭代。

3.3 实验结果及分析

为了验证本文提出的全卷积替代池化的算法对

图像分类的效果，文中是以 DenseNets 网络结构为基础，比较了 3 组使用不同池化方法的实验。与最大值池化和平均值池化进行了比较，以此证明网络性能的提高。实验在 2 个基准图像分类数据集上评估了性能，包括 CIFAR-10 和 SVHN 数据集。

在 CIFAR-10 数据集上的实验结果见表 1，在 DenseNets 中针对 all-conv pooling 操作后 100 个 *epochs* 后图像分类在训练集和测试集上的准确率的变化趋势见图 2。从表 1 和图 2 中可以看出，DenseNets 中所提出的 all-conv pooling 在测试集上的准确性分别为 90.48%。所提出的 all-conv pooling 在测试集上的准确率明显高于 DenseNets 中平均值池化和最大值池化的准确率。这也证明了所提出的 all-conv pooling 优于 DenseNets 中现有的最先进的最大值池化和平均值池化的结果。另外，还可以在图 2 中看到，随着 DenseNets 中的 3 种池化方法在 CIFAR-10 数据集上训练的 *epochs* 增加，测试集上的准确率和训练集上的准确率的总体趋势是越来越好。研究中提出的 all-conv pooling 是具有最高的图像分类准确率。这表明所提出的 all-conv pooling 优于 DenseNets 中的其他 2 种经典池化方法，即：平均值池化和最大值池化，并且可以解决过拟合问题。

表 1 不同池化方法在 CIFAR-10 上测试的准确率

Tab. 1 The accuracy of different pooling methods tested on CIFAR-10

池化方法	DenseNets
average pooling	89.77
max pooling	89.87
all-conv pooling	90.48

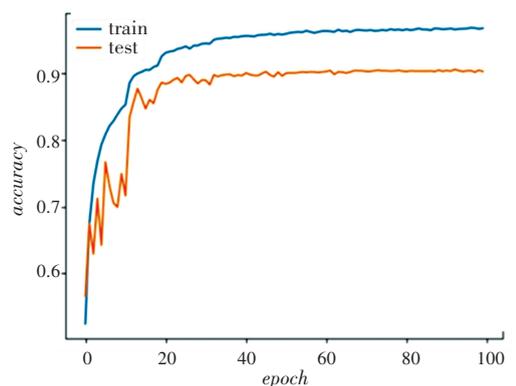


图 2 模型准确率的变化趋势对比

Fig. 2 Comparison of the change trend of model accuracy

在 SVHN 数据集上的实验结果见表 2，在实验过程中同样训练 100 个 *epochs*。图 3 展示了在测试

集上的准确率和训练集上的准确率。由图3可以看出,在测试集上 DenseNets 中 all-conv pooling 的准确率为 94.70%。具有 all-conv pooling 的 DenseNets 超过了传统的平均值池化和最大值池化,获得当前的最佳结果。另外,由图3中也可以看到,在 DenseNets 中,对于3种池化方法,随着 SVHN 数据集上训练轮数的增加,测试集上的准确率和训练集上的准确率的总体趋势会越来越好。all-conv pooling 具有最高的测试准确率。研究所提出的 all-conv pooling 通过比较得知,证明了所提出的 all-conv pooling 的优越性。

表2 不同池化方法在 SVHN 上测试的准确率

Tab. 2 The accuracy of different pooling methods tested on SVHN

池化方法	DenseNets
average pooling	94.21
max pooling	94.27
all-conv pooling	94.70

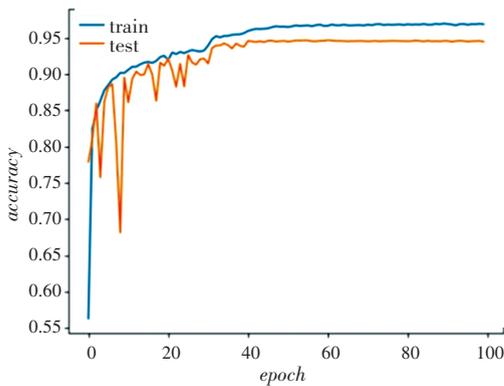


图3 测试集和训练集上的准确率对比

Fig. 3 Comparison of accuracy on test set and training set

4 结束语

针对卷积神经网络池化层中经典的最大值池化和平均值池化无法更详细地提取图像特征图的有效信息。本文对 DenseNets 网络结构的池化层进行了优化,提出了新的卷积层替代池化层的方法从而学

习到原图像特征图中更多的有用特征。本文利用 CIFAR-10 和 SVHN 数据集对 all-conv pooling 方法进行测试,2组数据集均不做数据增强处理,并根据同一评价指标对比分析了相应的图像结果,实验结果表明,all-conv pooling 方法能提高图像分类的准确度。

参考文献

- [1] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MASSACHUSETTS: IEEE, 2015:1-9.
- [2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.
- [3] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016:770-778.
- [4] XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated residual transformations for deep neural networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, Hawaii: IEEE, 2017:5987-5995.
- [5] HOWARD A G, ZHU Menglong, CHEN Bo, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. arXiv preprint arXiv:1704.04861, 2017.
- [6] FAHLMAN S E, LEBIERE C. The cascade-correlation learning architecture [M]// TOURETZKY D S. Advances in Neural Information Processing Systems2(NIPS). Los Altos, CA: Morgan Kaufmann, 1990:524-532.
- [7] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.
- [8] JOAN B, ARTHUR S, LECUN Y. Signal recovery from pooling representations [C]//1st International Conference on Machine Learning (ICML 2014). Beijing, China: International Machine Learning Society (IMLS), 2014:1585-1598.
- [9] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [10] KRIZHEVSKY A. cuda-convnet [EB/OL]. <http://code.google.com/p/cuda-convnet/>.

(上接第65页)

- [37] KIPF A, KIPF T, RADKE B, et al. Learned cardinalities: Estimating correlated joins with deep learning [C]// 9th Biennial Conference on Innovative Data Systems Research (CIDR 2019). Asilomar, CA, USA: www.cidrdb.org, 2019:1-7.
- [38] HASAN S, THIRUMURUGANATHAN S, AUGUSTINE J, et al. Deep learning models for selectivity estimation of multi-attribute queries [M]//MAIER D, POTTINGER R, DOAN A, et al.

Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020. Portland, OR, USA: ACM, 2020: 1035-1050.

- [39] YANG Zongheng, LIANG E, KAMSETTY A, et al. Deep unsupervised cardinality estimation [J]. Proceedings of the VLDB Endowment, 2019, 13(3): 279-292.