



文献引用格式:熊帅,何大治,管云峰,等. ATSC3.0 台到端及单频网系统[J]. 电视技术 2017 41(11/12):214-219.

XIONG S, HE D Z, GUAN Y F, et al. ATSC3.0 studio to transmitter link and single frequency network system[J].

Video engineering 2017 41(11/12):214-219.

中图分类号:TP391.4

文献标志码:A

DOI:10.16280/j.videoe.2017.h11.041

改进的深度卷积网络在交通标志识别中的应用

杨远飞,曾上游,甘晓楠,冯燕燕,周悦

(广西师范大学电子工程学院,广西 桂林 541004)

摘要: [目的]交通标志的采集大多是在自然环境下进行的,因为背景干扰、视角倾斜、局部遮挡等不利条件下导致采集的图像质量不高;传统的卷积神经网络(convolution neural network, CNN)只有一条通道提取特征,在面对质量不高的图像时,会出现特征提取不充分的情况导致分类准确率不高。[方法]提出了一种改进的网络模型,该模型集合了多尺度输入、并行交叉以及恒等映射的特点。[结果]能够保证特征提取的充分性与多样性并使网络性能不会随深度加深而退化。[结论]在实验阶段用该网络对德国交通标志数据库(GTSRB)进行测试,取得了97.6%的准确率。

关键词: 交通标志; 卷积神经网络; 特征提取

Application of improved depth convolution network in traffic sign recognition

YANG Yuanfei, ZENG Shangyou, GAN Xiaonan, FENG Yanyan, ZHOU Yue

(College of Electronic Engineering, Guangxi Normal University, Guangxi Guilin 541004, China)

Abstract: Most of the traffic signs are collected in the natural environment, the image quality is not high because of the adverse conditions such as background interference, angle tilt and local shelter. Because the traditional convolution neural network(CNN) has only one channel to extract features, when dealing with images of low quality, it may appear feature extraction is not sufficient, which leads to low classification accuracy. Due to the mentioned reasons, we present an improved network model, which contains the characteristics of multi-scale input, parallel crossover and identity mapping. It can ensure the sufficiency and diversity of feature extraction, and the network performance will not be degraded with depth. In the test stage, the German traffic sign database(GTSRB) was tested by this network, and an accuracy of 97.6% is achieved.

Key words: traffic sign; convolution neural network; feature extraction

现今社会交通非常发达,道路交通占据主导地位,交通标志^[1]是道路交通的重要组成部分之一。随着智能汽车以及智能驾驶的发展,对交通标志的识别受到了各界关注。交通标志的识别主要包括两个方面:检测以及识别。交通标志的检测主要包括定位、提取以及预处理;交通标志的识别主要包括特征提取以及分类。

如今图像分类与识别大多采用人工智能技术,主要是通过机器学习算法^[2-3]如支持向量机^[4]来对特征向量进行分类。一种方法是传统分类方法:通过对图像预处理、提取特征、对特征编码,生成特征向量,再由支持向量机等分类器对特征分类。这种

方法的关键在于特征选取,特征的好坏直接影响分类结果。人们已经认识到越抽象的特征,表达能力越强,分类效果越好^[5]。由于人类对于特征认识的局限性,导致人工定义的特征不够抽象,因此这种方法识别精度有限;另一种方法是深度学习方法^[6-7],如自编码器、限制玻尔兹曼机和卷积神经网络模型^[8-9]等。这种方法的好处是不需要人工定义特征,直接把图像中的像素值作为模型输入,避免了传统方法中特征提取的复杂过程。模型通过构建多个隐含层来模拟人脑学习过程,逐层对图像信息进行特征提取从而找到对图像更好的表征方法,在很大程度上能够提高分类或者识别的准确率。

基金项目:国家自然科学基金项目(11465004)

卷积神经网络在图像处理领域有很多优势^[4-7]。但是由于传统卷积神经网络只有一条通道提取特征,导致特征提取不完全;而且滤波器的尺寸也会影响网络的识别率^[12]。在对图像分类尤其是处理交通标志等质量不好的图像时识别率不高,不能达到人们对精度的要求。本文通过对卷积神经网络的原理和结构分析,提出了一种改进的网络结构,该结构是把多尺度输入、并行交叉以及恒等映射网络结合起来,在对交通标志进行分类识别时可能更有优势。

1 卷积神经网络

1.1 卷积神经网络的结构

典型的卷积神经网络大多由输入层、卷积层、池化层(下采样层)、全连接层以及输出层组成。网络的前几层大多由卷积层以及池化层交替组成,靠近输出的后几层主要由全连接层组成。如图1所示。

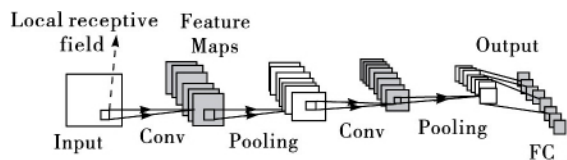


图1 卷积神经网络结构

卷积神经网络把局部感受野、权值共享以及下采样结合起来,在一定程度上具有平移、

尺度不变性。同时,权值共享方式使得网络参数大大减小,从而降低了网络复杂度,使网络更容易训练^[13]。

1.1.1 卷积层

卷积层通过可以训练的卷积核(滤波器)对上层图像进行卷积操作,再把卷积结果通过激活函数则得到该层的输出特征图,如公式1所示:

$$x_i^j = f(x_{i-1}^j * k_i^j + b_i^j) \quad (1)$$

其中 k_i^j 是卷积核的权值, b_i^j 是该层的偏置项,“*”代表卷积操作, $f(\cdot)$ 是激活函数。

1.1.2 池化层

池化层通常在卷积层后,它通过池化操作把卷积层输出特征图中相邻特征简化为一个特征,从而降低了特征图的维度,一定程度上也减少了冗余性。把池化操作后的结果通过激活函数即可得到池化层

的输出特征图,如公式2所示:

$$x_i^j = f(\beta_i^j \text{down}(x_{i-1}^j) + b_i^j) \quad (2)$$

其中 β_i^j 是池化层的权重系数, $\text{down}(\cdot)$ 表示下采样函数,常用的有两种方式,最大池化和平均池化。 b_i^j 是池化层的偏置项, $f(\cdot)$ 表示激活函数。

1.1.3 全连接层

全连接层是把前面的二维特征图整合成一维特征向量,再通过分类器对特征向量分类。全连接层的输出是对输入特征图求加权和再通过激活函数得到的,如公式3所示:

$$x_i = f(\omega_i x_{i-1} + b_i) \quad (3)$$

其中 ω_i 是全连接层的权重系数, b_i 是偏置项, $f(\cdot)$ 表示激活函数。最后一层的全连接层神经元个数即为输入类别。

1.2 卷积神经网络工作原理

卷积神经网络从本质上来说是有监督学习算法,通过前向计算得到网络的实际输出,然后再求出实际输出与希望输出的误差对网络的参数进行更新,网络训练成功后就可以得到输入与输出之间的非线性关系。网络的平均误差代价函数可以通过公式4得到:

$$E = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^C (u_k^n - x_k^n) \quad (4)$$

其中 N 为样本个数, C 为输入类别。更新权重用的算法类似于梯度下降法^[14],即按极小化误差的方式来反向调整网络的权值,可由公式5得到:

$$\omega(i+1) = \omega(i) + \alpha \frac{\partial E}{\partial \omega(i)} \quad (5)$$

其中 α 是学习率。网络就是通过前向运算与误差反向传播不断交替来调整权重等参数,从而来极小化误差,使网络达到收敛要求为止。

2 基于改进卷积神经网络的交通标志识别

2.1 并行交叉的网络模型

并行交叉网络主要有两种形式,一种是基于人眼视觉原理,网络结构如图2所示。

人眼观察物体时,通过两条视觉神经进行信息传递,通过漆状体和视交叉后形成交叉混合信息以供大脑分析。图二给出的网络结构完全是根据人眼

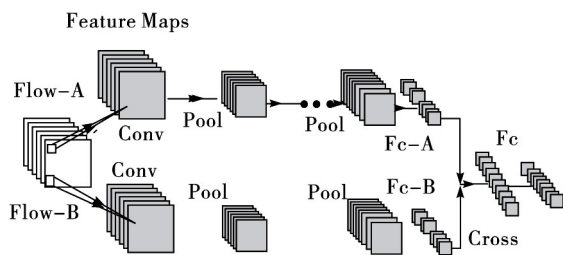


图2 并行交叉模型

视觉原理设计,通道 A 和 B 对输入图像提取特征然后进行特征传递,在全连接层把 A 和 B 的特征交叉混合到一起,最后用分类器进行分类。但是这种方法相较于传统卷积神经网络会增加网络的复杂度,使网络训练变的相对困难。

另一种是网络设计时使用并行模块,把多个并行模块叠加起来形成并行交叉模型。并行模块如图 3 所示。

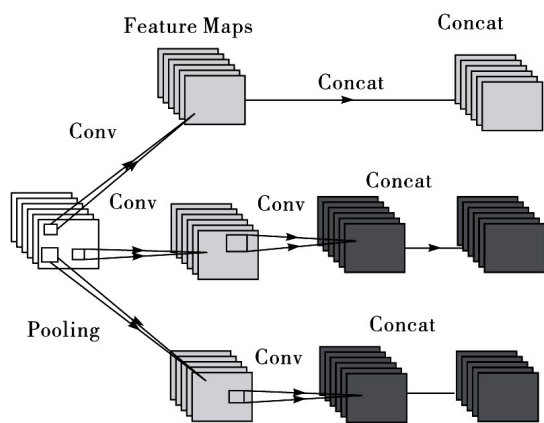


图3 并行模块

这种并行模块也是结合了人眼视觉原理,用多个卷积层通路同时对上层特征图提取特征,也可以用池化层来增加特征多样性,再把提取的特征混合在一起继续作为下一层的输入特征图。在卷积通路中,可以采用 1×1 的卷积核来有效减少参数数量^[15],降低网络复杂度,使网络训练更加容易。

并行交叉模型主要是通过不同通道提取不同特征来确保特征提取的充分性,同时增加特征的多样性和鲁棒性,以此来提高网络的分类精度。

2.2 恒等映射的网络模型

恒等映射的网络模型是基于 ResNet^[16] 网络设计的。ResNet 网络是卷积神经网络在 2015 年的新变革,它提出了“残差”的概念,网络通过对“残差”

学习,在深度和精度上比传统卷积神经网络更加强大。残差模块如图 4 所示。

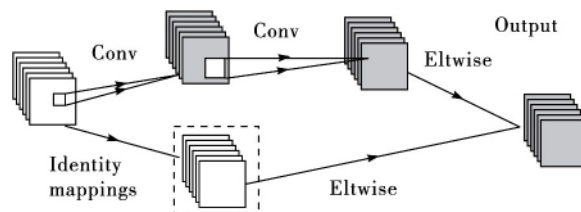


图4 恒等映射模块

从图四可以看出,残差模块其实是把上层特征图和卷积层提取的特征进行组合,把组合之后的特征作为下一层的输入。组合的方式有两种,一种是 Concat,一种是 Eltwise, Eltwise 共有三种形式,相加、点乘和取最大值。不同点在于,假如特征图个数为 n ,Concat 之后就有 $2n$ 个特征图,而 Eltwise 之后还是 n 个特征图。残差网络就是通过这种方式对每一层做了参考,形成了残差函数,这种残差函数更容易优化,不仅能够增强网络的鲁棒性还能提高网络分类精度。

2.3 多尺度输入的网络模型

尽管传统 CNN 对尺度和平移变换具有较强的鲁棒性,但是文献^[17]已经提出:输入图像尺寸的变化依然会导致网络最终识别结果不同。多尺度网络模型主要是针对输入图像尺寸来说的,可以把输入图像的不同尺寸分别作为网络输入,使用多条网络通道来传递图像特征以保证特征的多样性。多尺度网络模型如图 5 所示。

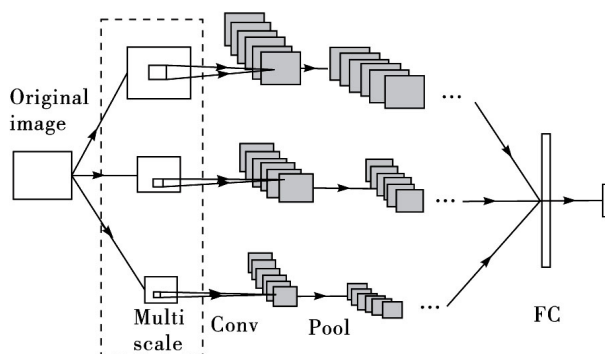


图5 多尺度网络模型

可以看出,输入图像的不同尺寸共同作为网络输入以供卷积层提取特征,每个尺寸为一个网络通道,特征提取与传递在各自的通道进行,在全连接层

时各个通道进行交叉,把特征进行整合,最后通过分类器输出结果。这种模型在原理上和并行交叉模型类似,不同点在于它每个通道是基于尺寸大小不同的图像,而并行交叉每个通道的输入尺寸相同。这种方法不仅可以保证特征提取的充分性,也在一定程度上增强了网络的鲁棒性,从而提高网络分类精度。

2.4 改进的网络结构

通过以上分析,可以对网络结构进行如下改进:用多尺度输入的网络模型,每个尺度对应的通道采用并行模块和恒等映射相结合的网络结构,各个尺度在全连接层进行交叉。交叉时各个尺度下权值的更新通过式 6 得到:

$$\begin{cases} \frac{\partial E}{\partial \omega_{l-1}^1} = \frac{\partial E}{\partial g_l} \frac{\partial g_l}{\partial g_{l-1}^1} \frac{\partial g_{l-1}^1}{\partial \omega_{l-1}^1} \\ \frac{\partial E}{\partial \omega_{l-1}^n} = \frac{\partial E}{\partial g_l} \frac{\partial g_l}{\partial g_{l-1}^n} \frac{\partial g_{l-1}^n}{\partial \omega_{l-1}^n} \end{cases} \quad (6)$$

其中 l 代表层数, g^1 和 ω^1 是尺度 1 中的变换函数和权值矩阵。 g^n 和 ω^n 是尺度 n 中的变换函数和权值矩阵。这样,误差依次在各个尺度下进行反向传播。

3 实验

3.1 实验设置

实验所用的数据集为 GTSRB 德国交通标志数据库,该数据库包括了 43 类的交通标志,共 51839 张图片。包括 39209 张训练集和 12630 张测试集。所有的图像都是在真实环境下采集的,其中包括了大量的低分辨率、不同光照强度、局部遮挡、运动模糊等不利条件下的图像。如图 6 所示。



图 6 不利条件下的部分交通标志

实验主要包括两个内容:1. 传统卷积神经网络、并行交叉模型、恒等映射模型在 GTSRB 数据集上的识别率。2. 不同尺度输入以及多尺度输入的网络在 GTSRB 数据集上的识别率,网络的整体结构如图 7 所示。

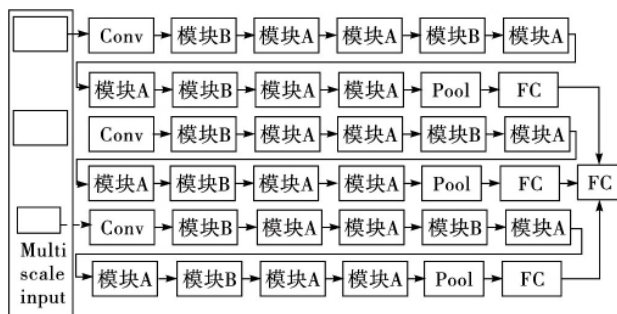


图 7 网络整体结构

可以看出每个尺度对应的通道完全相同,主要是由模块 A 和模块 B 组成。模块 A 是并行交叉和恒等映射相结合的网络结构,模块 B 是为了改变特征图维度。模块 A 的结构如图 8 所示,各层参数设置如表 1 所示。

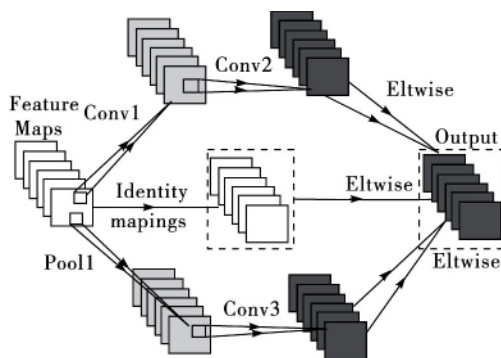


图 8 模块 A 结构

表 1 模块 A 参数设置

操作类型	参数设置
Conv1	$K=1, S=1, P=0$
Conv2	$K=3, S=1, P=1$
Pool1	$K=3, S=1, P=1, AVE$
Conv3	$K=1, S=1, P=0$
整合方式	Eltwise/SUM

模块 B 的结构和模块 A 相比仅仅少了恒等映射(可参考图 8),参数设置如表 2 所示。

表2 模块B参数设置

操作类型	参数设置
Conv1	$K=1, S=1, P=0$
Conv2	$K=3, S=2, P=1$
Pool1	$K=3, S=2, P=1, AVE$
Conv3	$K=1, S=1, P=0$
整合方式	Eltwise/SUM

网络整体参数设置如表3所示(一个尺度输入下)。在对图像进行预处理时,将所有图像缩放为 80×80 、 64×64 以及 48×48 的像素作为网络输入。网络训练时采用图像扩增技术,对每幅图像分别从左上角、左下角、右上角、右下角和中间裁剪 72×72 、 56×56 和 40×40 的像素大小,然后进行水平翻转,这样训练集扩增为原来的10倍;最后对图像进行去均值处理。测试时,则只取图像裁剪后的中间图像。除此之外,在实验中用ReLU(Rectified linear unit)类型的激活函数代替传统卷积神经网络中的sigmoid激活函数,并且卷积层后面加入了归一化层(Batch-Normalization, BN [18]),这样做的好处是使卷积层的输出具有相同分布,很大程度上减少了梯度消失、梯度爆炸或过拟合等问题。

表3 网络整体参数设置

操作类型	卷积核大小,步长,Pad,特征图个数
Conv	$K=5, S=1, P=2, N=32$
模块B	$N=32$
模块A	$N=32$
模块A	$N=32$
模块B	$N=64$
模块A	$N=64$
模块A	$N=64$
模块B	$N=128$
模块A	$N=128$
模块A	$N=128$
Pool	$K=2, S=2, P=0, N=128$
FC	$N=43$

所有的网络结构都是使用caffe框架来部署的,各模型的实验结果都是在NVIDIA-GTX 1070的GPU上完成的。

3.2 实验结果

表4列出了不同的CNN模型在GTSRB数据集

上的识别率,用的输入尺度为 64×64 。表5列出了三种不同尺度输入在GTSRB数据集上的识别率以及多尺度共同输入时网络的识别率。

表4 不同CNN模型的识别率

网络模型	准确率 %
传统CNN	89.9
并行交叉模型	94.9
恒等映射模型	95.8
并行交叉以及恒等映射	96.3

表5 不同尺度输入时网络的识别率

输入尺度	准确率 %
通道1(80×80)	95.9
通道2(64×64)	96.3
通道3(48×48)	95.6
通道1+2	97.1
通道1+3	96.7
通道2+3	97.0
通道1+2+3	97.6

3.3 实验结果分析

从表四可以看出,并行交叉模型、恒等映射模型在GTSRB数据集上的识别率分别可达到94.9%,95.8%,比传统卷积神经网络89.9%的准确率高,两者结合后识别率可达到96.3%,和传统传统卷积神经网络相比有了大幅度提高。从表五可以看出,以 80×80 、 64×64 和 48×48 大小作为网络输入可以分别达到95.9%、96.3%和95.6%的识别率;把它们共同作为网络输入时,可达到97.6%的准确率,远远超过传统卷积神经网络的识别率。我们可以得到以下结论:

(1) 并行交叉网络、恒等映射网络对整体识别率是用贡献的。

(2) 并行交叉和恒等映射结合后网络的识别率可以进一步提升。

(3) 不同尺度输入会导致最终结果不同,用多尺度共同输入时网络识别率可以进一步提升。

4 结束语

本文通过对传统卷积神经网络、多尺度模型、并行交叉模型以及恒等映射模型的分析提出改进的网



络结构,此结构结合了上述网络的优点。在对交通标志识别时,可以发现,输入图像尺寸的不同会影响网络的最终识别率,最后把三种不同尺度的图像一起作为网络的输入,提高了网络识别的准确率。但是,多尺度输入相对于单尺度来说参数更多,网络更复杂,网络更容易出现过拟合等问题。接下来,需要对网络进行进一步优化,在保证准确率的同时降低网络参数。同时,用该网络对其它数据集进行测试,增加该网络的泛化能力。

参考文献:

- [1] AN D, MEIER U, MASCI J, et al. 2012 Special Issue: Multi-column deep neural network for traffic sign classification[J]. Neural Networks the Official Journal of the International Neural Network Society, 2012, 32(1): 333-338.
- [2] HOLZINGER A. Interactive Machine Learning (iMI) [J]. Informatik-Spektrum, 2016, 39(1): 64-68.
- [3] HONG M, RAZAVIYAYN M, LUO Z Q, et al. A Unified Algorithmic Framework for Block-Structured Optimization Involving Big Data: With applications in machine learning and signal processing[J]. IEEE Signal Processing Magazine, 2016, 33(1): 57-77.
- [4] HOU S, LOU K, RAMANI K. SVM-based Semantic Clustering and Retrieval of a 3D Model Database[J]. Computer-Aided Design and Applications, 2013, 2(1): 155-164.
- [5] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: a review and new perspectives. [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 35(8): 1798-1828.
- [6] DENG L, YU D. Deep Learning: Methods and Applications [J]. Foundations & Trends in Signal Processing, 2014, 7(3): 197-387.
- [7] Schmidhuber J. Deep Learning in neural networks: An overview. [J]. Neural Networks, 2014, 61: 85.
- [8] SHIN H C, ROTH H R, GAO M, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning [J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1285.
- [9] CHEN Y H, KRISHNA T, EMER J S, et al. Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks[J]. IEEE Journal of Solid-State Circuits, 2016, 52(1).
- [10] CHEN H, QI X, YU L, et al. DCAN: Deep contour-aware networks for object instance segmentation from histology images. [J]. Medical Image Analysis, 2017, 36: 135-146.
- [11] 蔡晓东, 甘凯今, 杨超, 等. 基于多分支卷积神经网络的车辆图像比对方法[J]. 电视技术, 2016, 40(11): 116-120.
- [12] COATES A, NG A Y, LEE H. An Analysis of Single-Layer Networks in Unsupervised Feature Learning [J]. Journal of Machine Learning Research, 2011, 15: 215-223.
- [13] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012: 1097-1105.
- [14] LI L, CHEN Y, XU T, Et al. Integration of Bayesian regularization back-propagation neural network and particle swarm optimization for enhancing sub-pixel mapping of flood inundation in river basins[J]. Remote Sensing Letters, 2016, 7(7): 631-640.
- [15] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2015: 1-9.
- [16] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016: 770-778.
- [17] PLA-CIVERA J, RUIZ-FUERTES M I, GARCIA-MUNOZ L H, et al. Optimizing certification-based database recovery [C]// International Symposium on Parallel and Distributed Computing. IEEE, 2007: 30-30.
- [18] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [J]. Computer Science, 2015 (7): 631-640.



作者简介:

- 杨远飞(1992—) 硕士生,主研深度学习;
曾上游(1974—) 工学博士,主要研究方向为非线性动力学、计算神经科学;
甘晓楠(1991—) 硕士生,主研深度学习;
冯燕燕(1993—) 女,硕士生,主研深度学习;
甘晓楠(1993—) 硕士生,主研深度学习。

收稿日期:2017-04-05