基于超分辨率和深度神经网络的车型识别

雷 倩 郝存明 张伟平
(河北省科学院应用数学研究所 石家庄 050081)

摘 要 车型识别在视频监控系统中起着关键作用,文中利用深度神经网络和超分辨率来实现交通监控中的车型识别。利用深度卷积神经网络 CaffeNet,并采用先进的深度学习框架 CAFFE 和具有强大计算能力的 GPU 来完成对车辆的车型识别。在图像预处理阶段,采用一种基于深度学习和稀疏表示的图像超分辨率(SR)重构算法,来增强图像的细节信息。其中首先基于深度学习模型自编码器,提出一种改进模型非负稀疏去噪自编码器(Nonnegative Sparse Denoising Auto-Encoders,NSDAE)来实现字典的联合学习,然后基于稀疏表示实现车辆图像的超分辨率重构。经实验验证,在加入超分辨率处理之后,车型识别效果在精确度上得到了明显的提升。

关键词 超分辨率,深度学习,去噪自编码器,车型识别,深度卷积神经网络
中图法分类号 TP391.4 **文献标识码** A

Vehicle Recognition Based on Super-resolution and Deep Neural Networks

LEI Qian HAO Cun-ming ZHANG Wei-ping
(Institute of Applied Mathematics, Hebei Academy of Sciences, Shijiazhuang 050081, China)

Abstract Vehicle recognition plays a key role in traffic video surveillance system. In this paper, deep neural network and super-resolution were uses to realize vehicle recognition in traffic surveillance. It uses deep convolution neural network CaffeNet to complete vehicle recognition with advanced deep learning framework CAFFE and computationally powerful GPU. In the image preprocessing stage, an image super-resolution reconstruction algorithm based on deep learning and sparse representation was used to enhance the detail information of the image. First, based on auto-encoders, an improved model of nonnegative sparse denoising auto-encoders (NSDAE) was proposed to realize the dictionary joint learning. Then, the sparse representation was used to realize super-resolution reconstruction of vehicle image. Experimental results show that the accuracy of vehicle recognition is improved obviously after adding the super resolution processing.

Keywords Super-resolution, Deep learning, Denoising auto-encoders, Vehicle recognition, Deep convolutional neural networks

1 引言

视频监控系统是交通监管的一种重要手段,交通视频图像中的车型识别[1]是其中的关键技术。由于车辆外观复杂多样,且其受到背景、光照、视角等因素的影响,在实际应用中一直难以找到稳定、有效的视觉特征以进行准确识别。传统的识别算法主要是人工设计特征,常用的特征有 HOG[2],SIFT[3-4],LBP[5-6]等。基于人工选取特征可能会丢失一些内在的重要信息,而且运算量大,仅能在粗粒度(大、中、小)类别上对车型进行识别。随着车辆车型的不断增加,仅仅在粗粒度类别上进行处理已经无法满足大数据常态下高准确率、高效率的分析需求。深度学习[7-8]是最近新兴起的机器学习的一个分支,其模型能从海量的训练样本中学习更为有效的特征表示,典型的深度学习模型有深度置信网络(DBN)、深度卷积神经网络(DCNN)、自编码器(AE)等。本文将 DCNN 相关技术引入到车型识别的处理中,可以对车型进行细粒度识别或者预测,再依托实时处理系统可及时发现嫌疑车辆,精确实现交通预警。

图像的超分辨率重构[9]就是借用软件技术,结合一幅或多幅关于同一场景的低空间分辨率图像,重构关于该场景的高空间分辨率图像的过程。基于稀疏表示的图像超分辨率算法发展得很快,Yang 等[10]从压缩感知的角度实现图像超分辨率重构。基于稀疏表示的重构方法的关键在于字典的学习,传统的字典学习大部分没有考虑到高低分辨率图像块之间稀疏系数的一致性。因此,如何从庞大的训练集中获取通用性更好的字典是最近研究的难点。

2 图像超分辨率重构模型

2.1 非负稀疏去噪自编码器

对于基于稀疏表示的图像超分辨率重构,首先是需要进行字典的联合学习。首先是训练样本的获取,相当于对高分辨率图像进行降质平滑的过程,而由低分辨率图像向高分辨率进行恢复则相当于去噪、去平滑的过程。因此,引入去噪自编码器能够从输入的分布中学习得到更稳定和更具鲁棒性的特征,从而有效抑制噪声。其次,由于图像本身具有非负的特

本文受河北省科技计划基金项目(17395602D)资助。

雷 倩(1989-),女,硕士,CCF 会员,主要研究方向为人工智能,E-mail:943221957@qq.com;郝存明(1981-),男,硕士,副研究员,主要研究方向为计算机视觉;张伟平(1989-),女,硕士,主要研究方向为图像处理。

性，因此将非负性与去噪自编码器相结合。最后，去噪自编码器中隐含层关于输入的非线性映射即为关于输入部分的非线性特征降维。为了使自编码器模型在隐含层单元数较多时能够发现样本数据集的内部结构，需要在学习过程中控制隐含层单元节点的稀疏性[11]。根据以上 3 点，实现基于非负稀疏去噪自编码器的联合字典学习。

自编码器[12]可看作是使得目标输出等于输入的一种神经网络模型。去噪自编码器[13]是对传统自编码器的一种改进：对网络输入进行了加噪。去噪自编码器的训练过程与传统的自编码器非常相似，首先将初始输入 x 通过某种随机映射进行污染降质得到 $\bar{x}$，然后降质过的 $\bar{x}$ 通过自编码器的编码函数映射到隐含层：$y=f_\theta(\bar{x})$，最后在解码端得到重构结果：$z=g_{\theta'}(y)$。图 1 给出了去噪自编码器的结构模型。

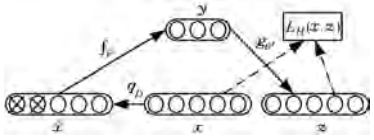

图 1　去噪自编码器的结构模型

对于输入样本的非负性，采用数据归一化的方法将样本数据映射到[0，1]。为了强调网络权重的非负性，引入一种非对称衰减机制。记原始干净的输入样本为 $x^c\in\mathcal{R}^D$，经噪声污染后的样本为 $x\in\mathcal{R}^D$，$f(\cdot)$是参数化的非线性激活函数：

$$f_i(g_i|a_i,b_i)=\frac{1}{1+e^{(-a_ig_i-b_i)}}\in[0,1]$$

其中，a_i 为斜率，b_i 为偏置量；网络以非负训练、样本集重构误差为目标函数：$E=\frac{1}{K}\sum_{i=1}^{K}\|x_i^c-\hat{x}_i\|^2$，其中 K 为训练样本数目，x_i^c 为第 i 个原始干净的非负样本。

对于重构误差函数，通过增加正则项 $\lambda\|W\|^2$ 来引入权重衰减机制。基于梯度下降法进行网络参数更新，可得到关于网络权重的衰减项 $d(w_{ij})=-\lambda w_{ij}$，网络权重的修正规则为：

$$\Delta w_{ij}=\eta(x_i-\hat{x}_i)h_j+d(\tilde{w}_{ij}) \tag{1}$$

其中，w_{ij} 是隐含层神经元 j 到输出层神经元 i 的连接权重，h_j 是神经元 j 的激活值，η 为自适应学习率。式(1)用于强调权重的非负性：

$$d(\tilde{w}_{ij})=\begin{cases}-\alpha\tilde{w}_{ij}, & \text{if } \tilde{w}_{ij}<0\\ -\beta\tilde{w}_{ij}, & \text{else}\end{cases} \tag{2}$$

其中，$\tilde{w}_{ij}=w_{ij}+\Delta w_{ij}$ 是根据式(2)进行误差修正后的新权重，α 和 β 是用来控制权重的符号。

为了实现编码的稀疏性，引入了由 Triesch[14] 于 2005 年提出的内在可塑性(Intrinsic Plasticity，IP)机制。IP 是受生物学学习规则的启发，实现对输入的稀疏编码。假定神经元节点的净输入为 x，满足分布 $f_x(x)$，相应神经元节点的输出为 $h=f(x)=\frac{1}{\exp(-(ax+b))}$。由于 f 为非线性激活函数，非负且严格单调，则 h 的分布为 $f_h(h)=f_x(x)/(\frac{\partial h}{\partial x})$。

IP 的核心思想就是对非线性激活函数中的参数 a 和 b 进行调整，借助散度距离的最小化，使得神经元的输出分布 $f_h(h)$和预期的指数分布 $f_{\exp}(h)=\frac{1}{\mu}e^{-\frac{h}{\mu}}$ 尽量逼近，从而最优化神经元的信息传输。其中，μ 是整个学习过程中输出分布的平均激活水平。参数 a 和 b 的更新规则为：

$$\begin{cases}a=a+\Delta a, & \Delta a=\eta_{IP}\left(\frac{1}{a}+x-\left(2+\frac{1}{\mu}\right)xh+\frac{1}{\mu}xh^2\right)\\ b=b+\Delta b, & \Delta b=\eta_{IP}\left(1-\left(2+\frac{1}{\mu}\right)h+\frac{1}{\mu}h^2\right)\end{cases} \tag{3}$$

其中，η_{IP}是内在可塑性学习率。

2.2　字典联合学习

NSDAE 模型具有很好的重构能力，根据以下结构来构建基于 NSDAE 的联合字典学习[15]网络模型：

1)隐含层节点数对应字典的原子个数；

2)模型的输入层分为两部分，对低分辨率图像经过上采样后提取的图像子块，以及对相应的低分辨率图像子块提取的特征；

3)模型的目标输出为原始高分辨率图像子块和相应的低分辨率图像子块提取的特征。

假设 $D_h\in K\times m$ 为输入层第一部分与隐含层节点的连接权重矩阵；$D_l\in K\times n$ 为输入层第二部分与隐含层节点的连接权重矩阵，则网络权重矩阵 $W=(D_h,D_l)^{\mathrm{T}}$ 即可看作是学习到的字典。字典学习网络结构如图 2 所示。

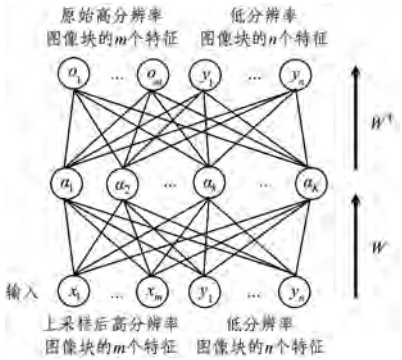

图 2　基于 NSDAE 进行联合字典学习的网络结构图

下面给出基于上述网络模型进行联合字典学习的基本步骤。

1)输入：训练样本集$\{(x_t,y_t)^{\mathrm{T}}_{t=1}\}$，目标输出 $u^p=(o,y)^{\mathrm{T}}$。

2)初始化：学习率 η_0 和 η_{IP}，规整化因子 η_g，衰减因子 α 和 β，激活函数平均输出值 μ；以非负小随机数初始化 W，a 和 b。

3)网络训练：可通过最小化训练样本集的平均重构误差来实现。

$$\begin{aligned}[W,a,b]&=\arg\min_{W,a,b}\frac{1}{T}\sum_{i=1}^{T}L(u^{p(i)},u^{o(i)})\\&=\arg\min_{W,a,b}\frac{1}{T}\sum_{i=1}^{T}\|u^{p(i)}-u^{o(i)}\|_2^2\end{aligned}$$

4)输出：高低分辨率字典 $W=(D_h,D_l)^{\mathrm{T}}$，激活函数的斜率 a 和偏置量 b。

完成字典学习后，图像超分辨率重构可按如下 3 步来实现：

1)基于非负稀疏去噪自编码器的联合字典学习。

2)基于稀疏表示的高分辨率图像子块的重构。对于低分辨率图像子块 y，先根据稀疏表示模型求解得到关于 y 的稀疏表示向量 q：$\min_q\|y-D_lq\|+\lambda\|q\|_1$，其中，$\|q\|_1$ 是 q 的 1-范数。则相应的高分辨率图像子块 x 为 $x=D_hq$。

3)基于重构子块拼贴的高分辨率初始图像的生成。基于步骤 2)利用低分辨率各重叠子块完成相应高分辨率图像子块的超分辨率重构后,需将各个重构子块按照对应位置拼接,得到初始高分辨率重构图像 X_h。

3 基于深度神经网络的车型识别

本文采用了先进的深度学习框架 CAFFE[16] 和具有强大计算能力的 GPU。DCNN 车辆特征提取的网络结构示意图如图 3所示,所有有关深度卷积网络的模型都是基于基本的卷积神经网络模型延伸而来。网络每层都是卷积、池化以及全连接层的一种或者组合。本文经过反复实验采用具有 13 层的深度神经网络 CaffeNet,其包括 4 个卷积层,3 个池化层,3 个全连接层。最后以一个全连接层的输出作为 Softmax 的输入进行分类。每个卷积层和全连接层都采用 Relu 激活函数,以缩短收敛时间;全连接层采用了 Dropout 技巧来避免过拟合。

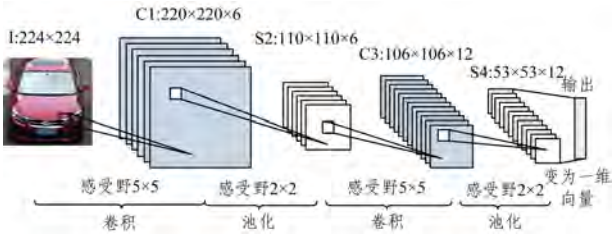

图 3 DCNN 车辆特征提取的网络结构示意图

本文采用原始图像作为网络的输入,经过深度神经网络提取特征,避免了图像分割以及人工提取特征等步骤。由于每张图像的大小不一,为了满足深度卷积神经网络的输入要求,将图像大小归一化,即将所有的车型图片都归一化为 256×256 的大小。在输入网络模型时将图片随机裁剪为 224×224 大小的图像,并且将所有图片的每个像素都减去所有训练集图片的平均值。

4 实验

4.1 数据集以及参数设置

本文以某市公路监控视频中的车辆图像作为识别对象,所有原始数据均采集于各个路口实际监控过程中的真实视频。采集到的图像包括各种尺度、光照和角度的车辆。对采集到的原始交通车辆图像通过目标检测算法进行定位和提取,并通过人工筛选剔除掉非车辆样本。本文对车辆样本图像进行了人工标注,其中将某些车标相同但车型比较少的类别归为一类,或者按车型归为几类。最后得到 240 类不同车型的车辆样本图像,共 46620 张。其中训练集 22634 张,验证集 10000 张,测试集 13986 张,标注好的部分车型图片如图 4 所示。

图 4 部分车型图片

在图像预处理步骤,本文将超分辨率重构技术用于图像预处理,使得模型对于尺度较小和比较模糊的图像都能达到较高的识别率。

联合字典学习中,所需训练样本集均是基于网络搜索得到的内容丰富的各种图像,包括风景、人物、自然图像等。实验中,样本集大小为 200000,统一设定采样因子 $s=2$。采用 4 个滤波器对低分辨率插值图像进行特征提取:

$$f_1=[-1,0,1],f_2=[-1,0,1]'$$

$$f_3=[-1,0,-2,0,1],f_4=[-1,0,-2,0,1]'$$

4.2 实验结果及分析

在使用 CaffeNet 网络提取特征前,将数据集中的所有彩色图像均转换为灰度图像。经过反复实验,设置训练迭代次数为 300000;初始学习率为 0.01,且在 100000 次迭代后以速率 0.1 倍的方式递减;动量学习率为 0.9。图 5 展示了对车辆图像的超分辨率重构结果。图 6 为随着迭代次数的变化模型正确率和损失函数的曲线图,其中正确率高达95.2%,对应的损失函数最终基本收敛至 0。图 7 展示了深度卷积网络最后一层提取的特征,可知特征明显地反映了车辆的具体信息,提取的特征具有鲁棒性。表 1 对比了加入 SR 前后的实验结果,从表中可以看出,加入超分辨率后的识别效果得到了明显提升。表 2 列出了车型识别中常用特征 HOG、PCA+SIFT 的识别效果与本文识别效果。从表 2 可以看出,在特征识别率方面,本文构建的特征提取模型的效果是最好的。

图 5 超分辨率重构结果(对图像进行了 2 倍尺度的放大)

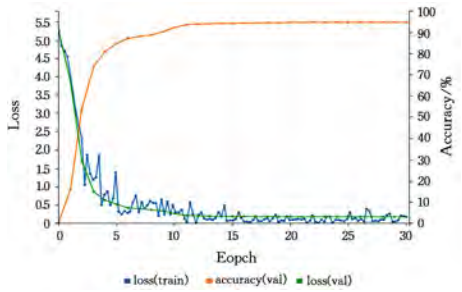

图 6 随迭代次数变化,模型的准确率和损失函数值

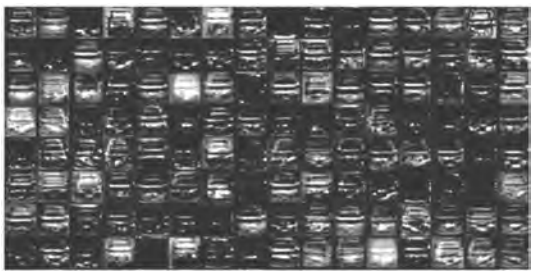

图 7 深度卷积网络最后一层提取的特征

表 1 识别准确率

(单位:%)

车型	CaffeNet	CaffeNet+SR
奥迪_A6L_07	94.1	95.2
本田_CR-V_07	94.4	95.1
别克_君越_12	94.0	95.2

表 2　特征识别的准确率比较

（单位：%）

车型	CaffeNet＋SR	PCA＋SIFT	HOG
奥迪_A6L_07	95.2	85.4	87.7
本田_CR-V_07	95.1	87.6	90.9
别克_君越_12	95.2	86.5	87.4

结束语　本文基于深度学习自编码器模型和稀疏表示实现了一种图像超分辨率重构算法。该算法将深度卷积神经网络与车型识别问题相结合，提出了基于卷积网络的车型识别框架，并将超分辨率引入车型识别的图像预处理部分，增加了车辆特征的鲁棒性。通过在市内公路监控数据上的对比实验显示，本文提出的车型识别框架和算法取得了良好的效果；但对于无法从外观信息分类的车辆图像，本文提出的方法还不够完善，有待后续研究。另外，车型识别数据库中共有 240 类车型，由于每年车型都有新增，因此需要不断更新车型数据库。

参 考 文 献

[1] BUCH N，VELASTIN S A，ORWELL J. A review of computer vision techniques for the analysis of urban traffic[J]. IEEE Transactions on Intelligent Transportation Systems，2011，12(3)：920-939.

[2] DALAL N，TRIGGS B，SCHMID C. Human detection using oriented histograms of flow and appearance[C]// European Conference on Computer Vision. 2006：428-441.

[3] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. Kluwer Academic Publishers，2004，60(2)：91-110.

[4] MA X，GRIMSON W. Edge-based rich representation for vehicle classification[C]// Tenth IEEE International Conference on Computer Vision. 2005：1185-1192.

[5] OJALA T，PIETIK，INEN M，et al. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[C]// IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002：971-987.

[6] ZHANG L，LI S Z，YUAN X，et al. Real-time object classification in video surveillance based on appearance learning[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2007：1-8.

[7] 孙志军，薛磊，许阳明，等. 深度学习研究综述[J]. 计算机研究应用，2012，29(8)：2805-2810.

[8] HINTON G E，SALAKHUTDINOV P R. Reducing the dimensionality of data with neural networks[J]. Science，2006，313(5786)：504-507.

[9] 沈焕锋，李平湘，张良培，等. 图像超分辨率重建技术与方法综述[J]. Optical Technique，2009，35(2)：196-203.

[10] YANG J C，WRIGHT J，HUANG T，et al. Image super-resolution as sparse representation of raw image patches[C]// IEEE Conference of Computer Vision and Pattern Recognition (CVPR). 2008：1-8.

[11] CHO K. Simple Sparsification Improves Sparse Denoising Autoencoders in Denoising Highly Noisy Images[C]// Proceedings of the 30th International Conference on Machine Learning (ICML-13). 2013：432-440.

[12] ALAIN D，OLIVIER S. Gated Autoencoders with Tied Input Weights[C]// International Conference on Machine Learning. 2013：154-162.

[13] VINCENT P，LAROCHELLE H，BENGIO Y，et al. Extracting and composing robust features with denoising autoencoders[C]// International Conference on Machine Learning. 2008：1096-1103.

[14] TRIESCH J. A gradient rule for the plasticity of a neuron's intrinsic excitability[C]// International Conference on Artificial Neural Networks：Biological Inspirations. Springer-Verlag，2005：65-70.

[15] YANG J C，WANG Z W，LIN Z，et al. Coupled dictionary training for image super-resolution[C]// IEEE Transactions on Image Processing. 2012：2359-2390.

[16] JIA Y，SHELHAMER E，DONAHUE J，et al. Caffe：Convolutional Architecture for Fast Feature Embedding[J]. Eprint Arxiv，2014：675-678.

（上接第 197 页）

[3] 水泳. 虚拟现实中连续碰撞检测算法研究[D]. 合肥：中国科学技术大学，2013.

[4] BACIU G，WONG S K，SUN H. RECODE：an image-based collision detection algorithm[J]. Journal of Visualization & Computer Animation，1999，10(4)：181-192.

[5] BACIU G，WONG W S K. Image-based techniques in a hybrid collision detector[J]. IEEE Transactions on Visualization and Computer Graphics，2003，9(2)：254-271.

[6] 范昭炜，万华根，高曙明. 基于图像的快速碰撞检测算法[J]. 计算机辅助设计与图形学学报，2002，14(9)：805-810.

[7] 于海军，马纯永，张涛，等. 基于图像空间的快速碰撞检测算法[J]. 计算机应用，2013，33(2)：530-533.

[8] GOVINDARAJU N K，REDON S，LIN M C，et al. CULLIDE：Interactive collision detection between complex models in large environments using graphics hardware[C]// Proceedings of the ACM SIGGRAPH/ EUROGRAPHICS Conference on Graphics Hardware. Eurographics Association，2003：25-32.

[9] KIM D，HEO J P，HUH J，et al. HPCCD：Hybrid parallel continuous collision detection using CPUs and GPUs[C]// Computer Graphics Forum. Blackwell Publishing Ltd，2009，28(7)：1791-1800.

[10] 杜鹏，唐敏，童若锋. 多核加速的并行碰撞检测[J]. 计算机辅助设计与图形学学报，2011，23(5)：833-838.

[11] 邹益胜，丁富国，周晓莉，等. 一种基于图像空间的碰撞检测算法[J]. 系统仿真学报，2011，23(5)：944-949.

[12] DU P，ZHAO J Y，PAN W B，et al. GPU Accelerated Real-Time Collision Handling in Virtual Disassembly[J]. Journal of Computer Science and Technology，2015，30(3)：511-518.

[13] DU P，LIU E S，SUZUMURA T. Parallel continuous collision detection for high-performance GPU cluster[C]// Proceedings of the 21st ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games. ACM，2017：4.

[14] 徐文鹏，王玉琨，刘永和. 计算机图形学基础(OpenGL 版)[M]. 北京：清华大学出版社，2014：213-232.