

# 基于卷积神经网络的图像识别研究综述\*

黄志强 李家

(重庆交通大学机电与车辆工程学院)

**摘要:**卷积神经网络技术作为无人驾驶以及可移动机器人等领域研究的热点,具有广阔的应用前景和科研视觉算法价值,目前传统计算机视觉算法已经被深度神经网络算法替代。将目前常用的计算机视觉算法分为传统算法和神经网络算法2类,按照计算机视觉算法的不同分类,对各类算法在图像识别领域中取得的最新研究成果及应用进行了系统的阐述,并对各种情况下进行图像识别时存在的问题提出了使用融合算法等建议,可为研究者提供一定的参考和借鉴。

**关键词:**传统算法;神经网络算法;计算机视觉算法;图像识别

## Review of Research on Image Recognition Based on Convolutional Neural Network\*

**Abstract:** Convolution neural network technology, as a hot research topic in the fields of unmanned and mobile robots, has broad application prospects and scientific research value of vision algorithms. At present, traditional computer vision algorithms have been replaced by deep neural network algorithms. The commonly used computer vision algorithms are divided into traditional algorithms and neural network algorithms. According to different classification of computer vision algorithms, the latest research results and applications of various algorithms in the field of image recognition are systematically expounded, in addition, suggestions such as using fusion algorithm are put forward for the problems existing in image recognition under various conditions, which can provide some reference for researchers.

**Key words:** Traditional algorithms; Neural network algorithms; Computer vision algorithms; Image recognition

20世纪40年代神经元的概念首次被引入计算机领域,开启了神经网络的大门。在2005年之后,神经网络因为当时解决了梯度弥散问题而得到了广泛应用、出现大量已标注的数据和GPU的快速发展。神经网络算法在图像和字符识别领域起着重要作用,如欺诈检测、人脸识别、医学病理的判断等等,还可以对股票价格、经济走势等广义上的模型做预测,其克服了传统模型不能预测复杂的非线性模型的缺点,因此越来越多的学者投身于该领域。文章从理论和算法的可行性等方面详细介绍了各种算法的优点和缺点,并针对各类算法在图像识别中存在的问题提出一些改进建议,为今后更进一步地深入研究提供了参考。

### 1 图像识别分类

图像识别是为了将不同类别的图像划分到不同的类别中,引入卷积神经网络是为了达到最小的分类误

差,获得最高的识别率。图像识别问题总体<sup>[1]</sup>上可以分为跨物种语义级分类、子类细粒度图像分类以及实例级图像分类。

跨物种语义级分类是为了识别不同种类的对象,此类算法的特点是存在较小的类内方差,较大的类间方差;子类细粒度图像分类是在同一大类实现其子类的细分,此类检测对分类器的要求更高,实例级图像分类在实现前2种图像分类的基础上,同时还需要识别不同的个体。目前广泛使用的是基于实例级的图像分类技术,在满足精度要求的同时,也能达到较好的速度要求。

### 2 图像识别算法

#### 2.1 传统检测算法

传统的图像识别算法包括微分算子边缘检测算法、Canny边缘检测算法、角点检测算法等。文章将对后

\* 基金项目:重庆市轨道交通车辆系统集成与控制重庆市重点实验室项目(CSTC2015yft-zdsys30001)

2 种常用的算法作详细说明。

1) Canny 边缘检测算法一般包含 4 个步骤<sup>[2]</sup> 滤波、梯度幅值和梯度方向计算、非极大值抑制计算、边缘检测与连接。首先通过高斯滤波函数去除图像的噪声,并对图像进行平滑处理,接着通过一阶有限差分法分别对滤波后的图像水平和垂直方向的像素点进行偏导求解,再使用非极大值抑制算法将局部最大值之外的正负梯度值设置为 0,最后通过不相同的 2 个阈值对候选边缘图像中的像素进行处理,保留两阈值范围内的像素,最终检测出物体。

传统的 Canny 边缘检测算法降噪能力较差,针对这一情况,文献[3]提出同时使用 4 个具有各向异性的 5 阶差分模板检测多个方向上的像素点,不仅能够检测上、下、左、右 4 个领域的灰度加权值,同时还能够检测对角线方向的值。为了提高 Canny 算法的自适应能力,文献[4]采用自适应中值滤波和形态学闭运算来防止多方向梯度幅值计算时边缘信息被弱化,同时利用目标与背景的最佳分离点是最优梯度下最大的类间方差与最小的类内方差这一概念,来计算 Canny 算法中的上下阈值,以此来提高其自适应能力。

2) 角点检测算法<sup>[5]</sup>是通过一个固定像素窗口在图像中进行任意方向的滑动,比较滑动前后窗口中的像素灰度值,如果存在较大的变化,则可判断出该像素内存在角点。角点检测算法分为 3 类<sup>[6]</sup>:基于二值图像的角点检测、基于灰度图像的角点检测和基于边缘轮廓的角点检测。

传统的 Harris 角点检测算法精度较低,抗噪性差,文献[7]将 Sobel 算法和 Harris 算法结合,首先使用 Sobel 算法进行角点初选,将非极大值抑制算法中的矩形模板用圆周模板替代,以此来提高检测精度,最后使用临近点剔除法提高算法的抗噪性。文献[8]通过比较阶梯边缘、L 型拐角、Y 或 T 型拐角、X 型拐角和星型拐角的强度变化特性后,提出利用多尺度各向异性高斯方向导数滤波器,从输入图像中提取灰度变化的新方法,该方法能够连续地提取图像中的边缘点和角点特征。

## 2.2 神经网络算法

通过模拟人脑运转建立的神经网络算法,是一种非线性动力学系统,通过大量输入神经元构成的网络系统,能够实现各种传统算法所不能达到的目标。许多曾经严重依赖于手工提取特征的机器学习任务(如目标检测、机器翻译和语音识别),如今都被各种端到端的神经网络学习算法彻底改变了。神经网络学习算法<sup>[9]</sup>主要分为卷积神经网络、注意力神经网络、自编码

神经网络、生成网络和时空网络 5 大类。

卷积神经网络是所有其他复杂网络的基础,但是由于其结构简单,不能在复杂环境中实现较高精度的分类任务,为此研究者提出了一些比较典型的 CNN 框架<sup>[10-12]</sup> LeNet、AlexNet、GooleNet、VGGNet、ResNet。一般情况下研究者会将其融合使用,使其效果更佳。文章将重点介绍卷积、注意力、生成对抗神经网络和金字塔神经网络。

### 2.2.1 注意力神经网络

注意力神经网络包含编码器和解码器 2 部分<sup>[13]</sup> 编码器为采用加权组合的方式对输入的参数进行编码,得到一组向量,解码器采用 RNN 结构对该组向量进行解码,引入注意力机制的神经网络更容易分析图像信息集中的那一部分,从而会忽略一些细节信息,因此文献[14]通过 ResNet 提出利用注意力机制实现特征的软阈值化,网络中输出的阈值是各个特征通道的绝对值的平均值与一组 0 和 1 之间的系数的乘积,这样能够确保所有的阈值都为正数,而且不会使所有的输出都为 0,使得该网络能够适用于噪声含量不同的样本中,避免了人工设置阈值所需要的专业知识不足的问题。此外,针对低分辨率图像在较深的网络中会丢失大量信息的问题,文献[15]提出了通道注意力机制,通过对特征通道之间的相互依赖性建模来自适应地重新缩放每个通道的特征,设计出一个 RIR(Residual In Residual)架构,其中 RG(Residual Group)作为基本模块,LSC(Long Skip Connection)则用来进行粗略的残差学习,在每个 RG 内部则叠加数个简单的残差块和 SSC(Short Skip Connection)。LSC、SSC 和残差块内部的短连接可以允许丰富的低频信息直接通过恒等映射向后传播,这可以保证信息的流动,加速网络的训练。

### 2.2.2 生成对抗神经网络

生成对抗神经网络由生成器和判别器组成<sup>[16]</sup> 生成器首先接收一个随机噪声信号,产生相应的虚拟样例,判别器则将虚拟样例与真实样例作比较,并给真实样例尽可能大的概率,给虚拟样例尽可能小的概率,反复迭代后,直至判别器无法区分样例的真实情况。但是该方法在训练时不稳定,不能生成离散的数据,由此文献[17]提出了一种端到端多任务生成对抗网络(MT-GAN),其生成器是一个超分辨率网络,它可以将小的模糊图像采样成精细的图像,并恢复详细的信息,以便进行更精确的检测。鉴别器是一个多任务网络,它用真/假分数、对象类别分数和边界盒回归偏移量来描述每个输入图像块。同时将鉴别器中的分类和回归损失反



向传播到生成器中,使生成器恢复更多的细节以便于检测。文献[18]提出了一种卷积编/解码器框架来提取图像轮廓,并由生成性对抗网络来支持,以提高轮廓质量。传统的图像到图像模型只考虑预测值与地面真实值之间的损失,而忽略了结果的数据分布与地面真实值之间的相似性。基于这一观察结果,文章提出的生成对抗神经网络旨在提高检测的准确率。所得到的方法包含 2 个模型,即编码器-解码器模型和鉴别器网络,编码器-解码器模型使用来自 VGG16 预训练模型的微调的二进制交叉熵损失来更新权重,而鉴别器网络采用地面事实和预测轮廓作为判别的输入。

### 2.2.3 金字塔网络

金字塔神经网络的出现大大提高了小物体的识别率,其将原始图片缩放为不同的大小<sup>[19]</sup>,分别对这些图片提取特征进行预测,由于每次特征提取和预测都是独立进行的,这样会造成不同层级之间信息不能共享,增加网络的计算时间。为此,文献[20]使用特征转化和特征融合 2 个模块,将图像金字塔得到的浅层空间信息和细节信息融合到主干网络中,让网络能够提取足够的浅层图像信息和来自深层语义的信息,从而可以减少特征未对准和细节丢失的影响。所得到的特征也既包含了丰富的语义信息,又补充了浅层的空间细节信息,从而提升网络对小目标检测的性能。为了克服使用多尺度图片金字塔计算量大的问题,文献[21]设计了自上而下、自下而上、融合-分割 3 种模式,分别检测小型、大型、中型物体,最后使用混合特征金字塔网络将其特征融合使用,以继承 3 种模式的优点。

### 2.2.4 自编码神经网络

自编码网络由负责降维的编码器和负责升维的解码器组成<sup>[22]</sup>,其目的是为了在输出层重新构建输入数据,让输出数据与输入数据相同,这就可以看作是一种无监督学习方法。自编码神经网络最简单的方法可以通过 2 个恒等函数表达式来完成,但是这样会造成隐藏层无法被表达,多尺度信息丢失。针对这些问题,文献[23]提出分两阶段提取图像的深度多尺度信息,先通过堆叠稀疏自动编码器(SSAE)提取图像的多尺度信息,再通过一维平均合并策略来降低特征维数,这样可使神经网络在隐藏层神经元较多的情况下依然能够提取样本的深层特征和结构。而文献[24]利用稀疏自动编码器降低图像的特征维度,并重新构建图像中的感性区域,再从该区域进行特征提取,该方法可减少无用特征所占用的计算空间,提高最终的识别率。

## 3 结论

文章主要分析了传统神经网络和卷积神经网络的优缺点,对其原理以及当前的运用现状进行了研究。当今图像识别中已广泛应用卷积神经网络技术,未来图像识别应用的环境将更加复杂,单独使用某一类卷积神经网络技术时会存在一定的局限性,为了提高特殊环境下的识别率,如在目标颜色与背景颜色一致,多个目标重叠在一起,识别的目标遮挡严重等极端环境下,可将不同的卷积神经网络技术相结合,充分利用各自的优点,同时可以针对不同物体的特殊结构使用相应的神经网络,进一步提高物体的识别率。

### 参考文献

- [1] 张强,李嘉峰,卓力.车辆识别技术综述[J].北京工业大学学报,2018,44(3):382-392.
- [2] 商景辉.基于 Canny 算子的边缘检测算法研究[J].信息与电脑(理论版),2020,32(2):39-41.
- [3] 王益艳.基于多方向的各向异性边缘检测算法[J].计算机与数字工程,2020,48(1):167-169.
- [4] ZHANG Xiao, CHEN Fuen. Lane Line Edge Detection Based on Improved Adaptive Canny Algorithm[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1549(2): 022131.
- [5] MOKHTAR M K, MOHAMED F, SUNAR M S, et al. Image Features Detection and Tracking for Image Based Target Augmented Reality Application[C]//2019 IEEE Conference on Graphics and Media (GAME). Erfurt, Germany: IEEE, 2019: 38-43.
- [6] 朱思聪,周德龙.角点检测技术综述[J].计算机系统应用,2020,29(1):22-28.
- [7] 董立红,彭业勋,符立梅.基于 Sobel 边缘检测的圆周 Harris 角点检测算法[J].西安科技大学学报,2019,39(2):374-380.
- [8] ZHANG Weichuan, SUN Changming. Corner detection using multi-directional structure tensor with multiple scales[J]. International Journal of Computer Vision, 2020, 128(2): 438-459.
- [9] WU Zonghan, PAN Shirui, CHEN Fengwen, et al. A comprehensive survey on graph neural networks[J/OL]. IEEE Transactions on Neural Networks and Learning Systems, 2020. (2020-03-24) [2020-07-30]. <https://pubmed.ncbi.nlm.nih.gov/32217482/>. DOI:10.1109/TNNLS.2020.2978386.
- [10] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J/OL]. (2017-04-17) [2020-07-30]. <https://arxiv.org/pdf/1704.04861.pdf>.
- [11] ZHANG Xiangyu, ZHOU Xinyu, LIN Mengxiao, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City, American: IEEE, 2018: 6848-6856.
- [12] XIE Saining, GIRSHICK R, DOLLAR P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu, American: IEEE, 2017: 1492-1500.
- [13] CHAUDHARI S, POLATKAN G, RAMANATH R, et al. An attentive sur-

(下转第 31 页)



图 15 后桥系统多通道试验

## 8 结论

文章基于架构带宽进行后悬架下控制臂的设计优化,得出以下结论。

1) 架构项目的底盘结构件开发,需在设计之初就考虑架构带宽里各参数对零件载荷和耐久性能的影响,以最恶劣的参数组合作为设计计算的依据。

2) 找到了调试件参数对后下控制臂载荷和耐久的影响规律,弹簧刚度越大、稳定杆直径越大,则后下控制臂受载越恶劣。

3) 基于架构带宽完成了后下控制臂的设计优化,并开发零件级和系统级的台架验证方法对实际零件进行验证,零件通过台架验证并通过整车路试。

4) 提出了基于架构项目的后悬架下控制臂设计思路和方法,可拓展应用于其他底盘结构件。

### 参考文献

- [1] 陈虹. 上汽自主品牌产品开发的架构策略[J]. 汽车工程, 2010, 32(6): 461-465.
- [2] 张志飞, 陈仁, 徐中明, 等. 面向多目标的汽车悬架控制臂拓扑优化研究[J]. 机械工程学报, 2017, 53(4): 114-121.
- [3] 王霄峰. 金属疲劳寿命预计——局部应力-应变法[M]. 北京: 清华大学出版社, 2007: 145-157.
- [4] 何为军. CATIA 环境中的设计 CAE 及应用[M]. 北京: 机械工业出版社, 2012: 5-27.
- [5] 祝小元, 方宗德, 申闪闪, 等. 汽车悬架控制臂的多目标拓扑优化[J]. 汽车工程, 2011, 21(2): 138-141.
- [6] 何小静, 上官文斌. 汽车悬架下控制臂的有限元分析[J]. 噪声与振动控制, 2012, 32(3): 125-128.
- [7] 武文超, 方毅博, 宁倩, 等. 基于雨流法的汽车底盘件疲劳耐久试验方法研究[J]. 汽车技术, 2013(12): 46-49.

(收稿日期 2020-08-11)

(上接第 13 页)

- vey of attention models[J/OL]. (2019-03-05)[2020-07-31]. <https://arxiv.org/pdf/1904.02874.pdf>.
- [14] ZHAO Minghang, ZHONG Shisheng, FU Xuyun, et al. Deep residual shrinkage networks for fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2019, 16(7): 4681-4690.
  - [15] ZHANG Yulun, LI Kunpeng, LI Kai, et al. Image super-resolution using very deep residual channel attention networks[C]/Proceedings of the European Conference on Computer Vision. Munich, Germany: ECCV, 2018: 286-301.
  - [16] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]/Advances in neural information processing systems. Montreal, Canada: NIPS, 2014: 2672-2680.
  - [17] ZHANG Yongqing, BAI Yancheng, DING Mingli, et al. Multi-task Generative Adversarial Network for Detecting Small Objects in the Wild[J]. International Journal of Computer Vision, 2020, 128: 1-19.
  - [18] YANG H, LI Y, YAN X, et al. ContourGAN: Image contour detection with generative adversarial network[J]. Knowledge-Based Systems, 2019, 164: 21-28.
  - [19] CHEN Kai, CAO Yuhang, LOY C C, et al. Feature Pyramid Grids[J/OL]. (2020-03-07)[2020-07-31]. <https://arxiv.org/abs/2004.03580>.
  - [20] LIU Ziming, GAO Guangyu, SUN Lin, et al. IPG-Net: Image Pyramid Guidance Network for Small Object Detection[C]/Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, American: IEEE, 2020: 1026-1027.
  - [21] LIANG Tingting, WANG Yongtao, ZHAO Qijie, et al. MFPN: A Novel Mixture Feature Pyramid Network of Multiple Architectures for Object Detection[J/OL]. (2019-12-23)[2020-07-31]. [https://blog.csdn.net/weixin\\_42096202/article/details/103672371](https://blog.csdn.net/weixin_42096202/article/details/103672371).
  - [22] 肖思宇, 吴丁娟. 基于神经网络与模糊理论的模糊自编码器[J]. 信息技术, 2020, 44(4): 5-9.
  - [23] ZHANG Lu, JIAO Licheng, MA Wenping, et al. PolSAR image classification based on multi-scale stacked sparse autoencoder[J]. Neurocomputing, 2019, 351: 167-179.
  - [24] DAS R, WALIA E. Partition selection with sparse autoencoders for content based image classification[J]. Neural Computing and Applications, 2019, 31(3): 675-690.

(收稿日期 2020-08-10)