

文章编号:1671-6833(2018)02-0011-07

基于深度卷积神经网络的羽绒图像识别

杨文柱, 刘 晴, 王思乐, 崔振超, 张宁雨

(河北大学 网络空间安全与计算机学院,河北 保定 071002)

摘 要: 由于图像中羽绒形态及其多样性,传统的图像识别方法难以正确识别羽绒分拣图像中的羽绒类型,其识别精度也难以达到实际生产的要求.为解决上述问题,构造了一种用于羽绒类型识别的深度卷积神经网络,并对其权值初始化方法进行了改进.首先利用视觉显著性模型提取羽绒图像的显著部分,然后将图像的显著部分输入到稀疏自动编码器中进行训练,得到一组符合数据集统计特性的卷积核集合.最后采用 Inception 及其变种模块实现深度卷积神经网络的构造,通过增加网络深度来提高网络的识别精度.试验结果表明,用所构造的深度卷积神经网络对羽绒图像识别的精度较传统卷积神经网络的提高了 2.7%,且改进的权值初始化方法使网络的收敛速度提高了 25.5%.

关键词: 深度卷积神经网络; 权值初始化; 稀疏自编码; 视觉显著性; 图像识别

中图分类号: TP183; TP391.41 文献标志码: doi:10.13705/j.issn.1671-6833.2018.02.015

0 引言

正确识别监控图像中的羽绒类型,是基于机器视觉^[1-2]的羽绒分拣技术的关键.在羽绒监控图像中,同一类型的羽绒形态千差万别,不同类型的羽绒形态也有相似,这使得传统的图像识别技术难以正确识别监控图像中的羽绒类型,其识别精度也难以达到实际生产的要求(正确率不小于 90%).深度卷积神经网络^[3](deep convolutional neural networks, DCNN)是一种基于特征学习的图像识别方法,其泛化能力较传统的图像识别方法有了明显提高,因此在最近几年,基于 DCNN 的图像识别方法取得了举世瞩目的成绩.基于特征学习的图像识别方法不需要事先指定应该提取的特征,而是通过迭代学习的方式寻找最适合分类的特征.将 DCNN 应用于图像识别任务时,不仅可以提高识别精度,还可以避免人工提取特征造成的人力和时间浪费,同时满足在线检测的需要.

近年来识别效果好的 DCNN 结构有 Alex-Net^[4]、VGGNet^[5]、GoogleNet^[6]、ResNet^[7]等. DC-

NN 在图像识别分类^[8]、目标检测^[9]、目标跟踪^[10]、文本识别^[11]、语音识别^[12]等方面都取得了很好的成绩.这些网络中权值初始化方法大部分采用随机初始化,如小随机数初始化^[4]、Xavier 初始化^[13]、MSRA 初始化^[14]等. DCNN 的训练方式是有监督训练,所以训练时需要大量有类标的数据来进行权值的更新调整.但当图像数据集较小时容易造成网络的前几层无法得到充分训练等问题,针对该问题的主要解决方案是对网络的卷积核进行无监督预训练,以得到尽可能符合数据集统计特性的卷积核集合.文献[15]通过训练一个稀疏自动编码器,对网络第一层卷积核权值进行初始化,在原图随机切取小块作为稀疏自动编码器的输入,训练得到的权值就是对随机切取小块的一种稀疏表达;同时,小块的选取极为重要,若识别目标在原图中占比较小,随机切取的小块取到背景的可能性较大,这样训练得到的卷积核对网络收敛贡献甚微.

针对以上问题,设计实现了一种用于识别羽绒图像的 DCNN,并利用视觉显著性模型和无监督预训练的方法对其权值进行预训练,以提高网

络收敛速度和对羽绒图像的识别精度. 首先利用视觉显著性模型从原始图像中提取其显著部分, 将显著部分截取指定规格的小块输入至稀疏自动编码器中进行无监督预训练, 得到 DCNN 第一层的卷积核, 将训练好的卷积核集合输入到针对羽绒图像识别的 DCNN 中; 网络结构采用了 Inception 模块及其变种模块^[6,16] 来提高网络的效率, 并通过增加网络的宽度和深度来提高网络识别精度; 最后利用 softmax 分类器实现对羽绒类型的识别.

1 网络权值初始化及其改进

1.1 常用的权值初始化方法

DCNN 的权值初始化主要是对卷积层和输出层的卷积核参数和偏置进行初始化. 权值初始化关系到 DCNN 的训练时长, 优秀的初始权值会使网络以较快的速度达到权值最优, 从而提高训练速度. 偏置项通常初始化为一个较小的常数或 0, 卷积核参数初始化通常采用随机初始化方法. 目前常用的随机初始化方法有小随机数初始化、Xavier 初始化和 MSRA 初始化等. 小随机数初始化是将参数初始化为小的随机数, 打破对称性, 使得权值服从均值为 0、标准差为 0.01 的高斯分布, 但当随机数取值太小时, 就会有明显的弊端——反向传播过程中梯度很小, 对较深的网络更是会造成梯度弥散的问题, 同时参数收敛速度也会降低. 针对小随机数初始化方法存在的问题, Xavier 等人提出权值服从均值为 0、方差为 $1/n$ 的均匀分布(n 为输入神经元的数量), 可以提高网络训练的收敛速度. 为了使其更适用于 ReLU 激活函数, MSRA 初始化对 Xavier 初始化进行了改进, 其权值服从均值为 0、方差为 $2/n$ 的高斯分布.

利用无监督预训练对 DCNN 的卷积核权值进行初始化也是常用的权值初始化方法之一. 通过构造一个稀疏自动编码器(sparse autoencoder, SAE)对网络权值参数进行预训练, 使得输入层和输出层尽可能保持一致; 在 SAE 训练过程中参数得以更新, 形成符合数据特性的初始值. 自动编码器是一种用于尽可能复现输入信号的神经网络^[17]. 稀疏自动编码器则是在自动编码器的基础上加入了稀疏限制, 对隐含层进行了约束, 使其变得稀疏.

自动编码器的网络结构如图 1 所示. L_1 、 L_2 、 L_3 层分别代表网络输入层、隐含层和输出层. 若

稀疏自编码输入层维度为 6 维, 隐含层节点数量设置为 3, 这样迫使隐含层节点学习得到输入数据的压缩表示方法, 即用 3 维数据重构出 6 维数据.

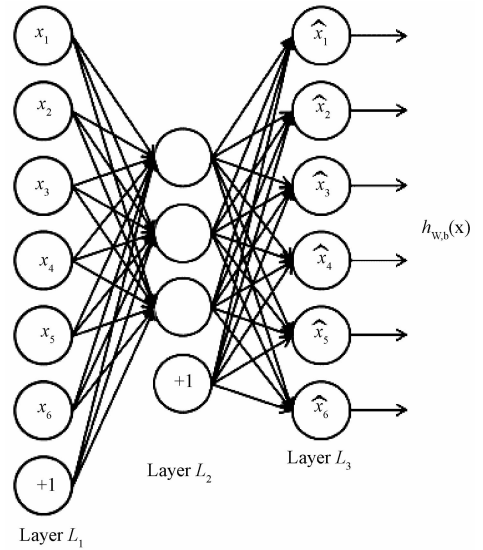


图 1 自动编码器的网络结构

Fig.1 Network structure of autoencoder

自动编码器损失函数如下:

$$\begin{aligned} J(W, b) &= \left[\frac{1}{m} \sum_{i=1}^m J(W, b; x^{(i)}, y^{(i)}) \right] + \\ &\quad \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \\ &= \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \| h_{W,b}(x^{(i)}) - y^{(i)} \|^2 \right) \right] + \\ &\quad \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2, \end{aligned} \quad (1)$$

其中:第 1 项为均方差项;第 2 项是权重衰减项, 用于减小权重以防止过拟合.

在计算过程中, 上述损失函数常导致网络收敛很慢, 计算复杂度过高. 稀疏自编码在自动编码器的基础上加入 L1 正则化限制, 使得大部分神经元处于抑制状态, 少数处于激活状态, 使隐含层节点的输出均值尽可能为 0. 稀疏自动编码器的损失函数表示为:

$$J_{sparse}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} KL(\rho \| \hat{\rho}_j), \quad (2)$$

其中, KL (Kullback-Leibler)距离表达如下

$$KL(\rho \| \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j}, \quad (3)$$

式中: ρ 为稀疏性参数, 通常取接近 0 的较小的数; $\hat{\rho}$ 为隐含层节点输出均值, 表达式为

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m [a_j^{(2)}(x^{(i)})]. \quad (4)$$

加入稀疏限制使得 $\hat{\rho}_j$ 尽可能逼近 ρ ,从而确保多数神经元处于抑制状态,以增加网络的稀疏性. $a_j^{(2)}(X)$ 为输入向量 X 对隐含层单元 j 的激活值.

1.2 基于视觉显著性和无监督预训练的权值初始化方法

首先利用视觉显著性算法来提取图像的显著部分,然后随机截取显著部分图像输入至稀疏自动编码器中进行无监督预训练. 采用谱残差算法

(spectral residual approach, SRA) 来获取图像的显著部分,其主要原理是在原图中将背景剔除,即可得到图像的前景,也就是图像的显著部分^[18],如图 2 所示. 若 CNN 的第一层卷积操作需要 8 个 5×5 大小的卷积核,那么首先在显著图上随机截取 N 个 5×5 的小块作为稀疏自编码的输入层,隐含单元个数设置为 8,经过稀疏自编码的预训练得到的 W 大小为 25×8 ,再将 W 变为 $8 \times 5 \times 5$ 即可得到卷积神经网络第一层卷积核集合 W .

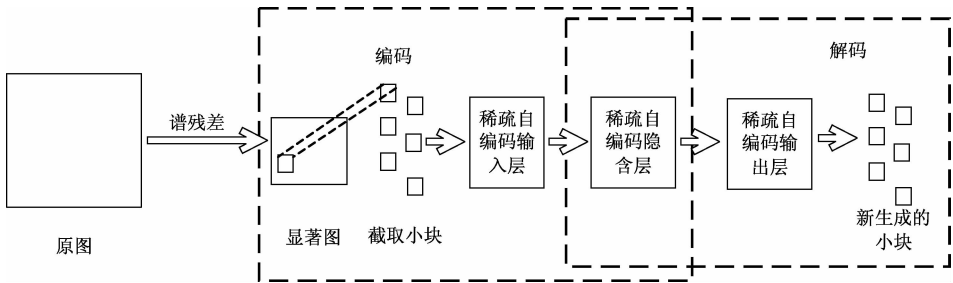


图 2 基于视觉显著性和 SAE 预训练框架

Fig. 2 Training framework based on visual significance and SAE

2 基于 DCNN 的羽绒图像识别网络结构

深度卷积神经网络的网络结构对最终的识别精度有着重要的影响. 本文数据集为在生产线上实际采集的羽绒图像经过剪裁后的子图像,子图像大小均为 100×100 ,比 MNIST 和 CIFAR 数据集中图像的分辨率要大很多. 为提高网络效率和识别精度,采用 Inception 及其变种模块来构建基本网络结构.

Inception 模块的主要特点是在加大了网络深度和宽度的同时不增加计算量,还提高了计算资源的利用率. Inception 模块结构如图 3 (a) 所示,由 1×1 、 3×3 、 5×5 的卷积操作和 3×3 的池化操作组成,通过设定 1×1 卷积核的数量,实现通道数的降维或升维,同时对不同通

道的特征进行融合. Inception 模块的优势主要体现在两个方面:①使用小卷积核在减少网络训练参数的同时降低了网络计算复杂度;②使用不同大小的卷积核对同一特征图进行特征提取. 随后研究者又提出将较大的卷积核分解成 2 个小卷积核,如图 3 (b) 所示,用 2 个 3×3 的卷积核代替 5×5 的卷积核,节约了计算时间,为增加卷积核数目提供便利条件. 图 3 (c) 为 Inception 模块的另一个变形,经过卷积和池化操作后的特征图大小是前一层的 $1/2$,该方法使用了 2 个并行化的模块,不仅减少了网络的计算量,还有效地避免了池化所造成的信息损失.

针对羽绒图像识别构造的 DCNN 结构如图 4 所示. 输入层图像大小为 100×100 ;C1 层是步长

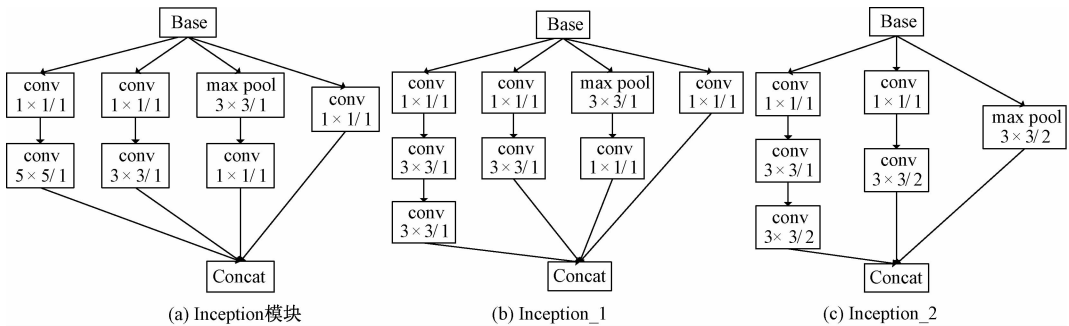


图 3 Inception 模块及其变种
Fig. 3 Inception module and its variants

为 2 的卷积层,其卷积核大小为 3×3 ,本层产生 32 个大小为 50×50 的特征图;C2 层为步长为 1 的卷积层,卷积核大小为 3×3 ,本层产生 64 个大小为 48×48 的特征图;S1 为池化层,采样窗口大小为 3×3 ,窗口滑动步长设置为 2,本层产生 64 个大小为 24×24 的特征图. Inception_1 层具体结构如图 3(b) 所示. 采用 padding 方式进行卷积运算,得到的特征图大小与前一特征图大小相同,即得到 164 个大小为 24×24 的特征图. Inception_2 层具体结构如图 3(c) 所示,得到 114 个大小为 12×12 的特征图;C3 层是步长为 1 的卷积层,卷积核大小为 3×3 ,本层产生 64 个大小为 10×10 的特征图;FC1 层为全连接层,将 C3 层的

64 个特征图连接成一个特征向量;FC2 层由 8 个神经元构成特征向量,对应 8 种类别输出,FC2 全连接层和输出层构成一个 softmax 分类器.

3 试验设计与结果分析

试验环境采用河北大学信息技术中心的超算平台,使用其中一个独立的计算节点,该计算节点配置 64 个 GPU,可大幅提升训练速度. DCNN 程序采用基于 Python 的 Tensorflow 深度学习框架,而基于视觉显著性和无监督预训练的过程采用 MatlabR2016a 实现. 数据集为经过剪裁的 2 300 张羽绒图像,图像像素大小均为 100×100 ,共 8 类(5 个单独类和 3 个混合类)如图 5 所示.

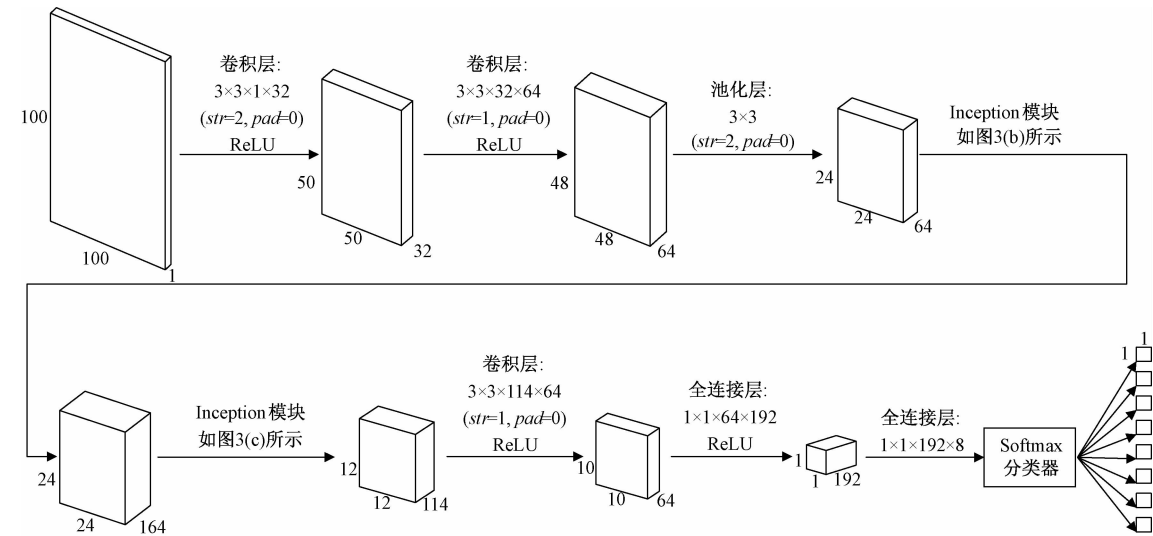


图 4 网络结构

Fig. 4 Structure

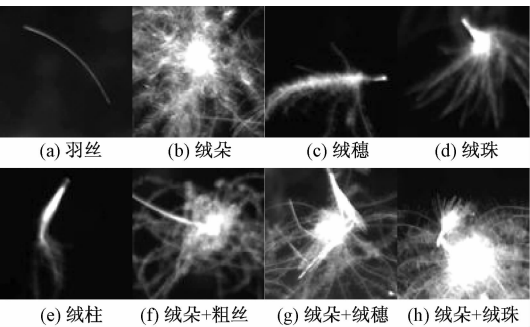


图 5 羽绒图像类别

Fig. 5 Categories of down image

3.1 试验步骤

步骤 1:首先利用谱残差模型提取原图显著部分并切割,提取若干张图像显著区域,如图 6 所示.

步骤 2:因为针对羽绒图像识别提出的框架第一层卷积层需要 32 个 3×3 的卷积核,故将上

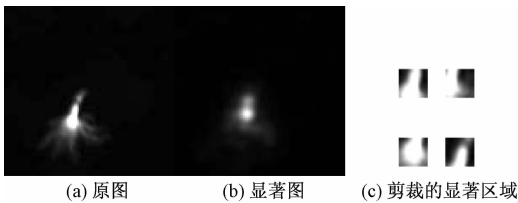


图 6 提取图像显著区域

Fig. 6 Extracting saliency area from image

一步中得到的若干显著区域随机切取 N 个 3×3 的小块,输入至稀疏自动编码器中,其中,稀疏自动编码器的隐含单元个数设置为 32,训练得到的稀疏自动编码器的权值系数大小为 6×32 ,将其格式转换为 $32 \times 3 \times 3$ 即可得到 DCNN 第一层卷积核集合 W ,图 7 为训练的卷积核集合.

步骤 3:将数据集输入至图 4 的深度卷积神经网络中训练并测试,得到识别错误率. 其中训

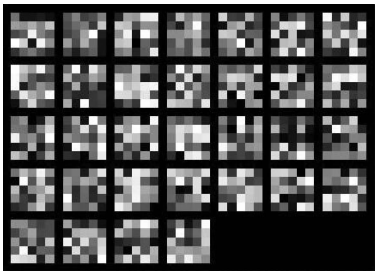


图7 卷积核集合

Fig.7 Convolutional kernels

练集包含 2 000 张羽绒图像,测试集包含 300 张羽绒图像.

3.2 试验结果与分析

本试验利用支持向量机 (support vector machine, SVM) 作为传统图像识别分类的代表进行对比试验. 将待识别的图像进行小波变换和图像分割,提取变换系数的主成分以及分割得到的目标形状特征作为特征向量,该特征表示能力强、特征维数较低,对笔者试验数据集有较好的表示能力,且计算效率较高. 但其计算错误率和训练时长相较于 LeNet-5 卷积神经网络仍然较高,如表 1 所示.

表 1 支持向量机与卷积神经网络对比实验

Tab.1 The comparison between SVM and CNN

算法	识别错误率/%	训练时长/s
SVM	39.6	79.8
LeNet-5	10.3	46.9

为了验证图 4 网络结构及笔者提出的基于视觉显著性和稀疏自编码预训练的权值初始化方法的有效性,设计了 6 种不同的试验,如表 2 所示.

表 2 识别错误率对比

Tab.2 The comparison of identify error rate

算法	数据集迭代次数					%
	5	10	15	20	25	
LeNet-5	10.3	7.3	7.0	6.3	6.0	
LeNet-5 + SAE	8.7	7.0	6.0	5.7	5.3	
LeNet-5 + SAE + SR	8.0	6.0	5.7	4.7	4.3	
图 4 结构	9.3	7.0	6.0	5.7	5.0	
图 4 结构 + SAE	8.3	6.7	5.3	4.7	4.3	
图 4 结构 + SAE + SR	7.6	5.7	5.0	4.0	3.3	

由表 2 可以看出,利用稀疏自编码预训练算法对 LeNet-5 网络初始权值进行改进,较未加入 LeNet-5 的网络降低了对羽绒图像的识别错误率,

说明稀疏自编码预训练对卷积神经网络训练有一定的贡献. 对比加入稀疏自编码预训练算法的 LeNet-5 网络与利用视觉显著性和稀疏自编码算法对 LeNet-5 网络初始权值进行改进,可以看出加入了视觉显著性和稀疏自编码预训练算法的 LeNet-5 网络进一步降低了识别错误率,并随着数据集迭代次数的增加而降低.

由表 2 还可以看出,通过对比图 4 网络结构和经典 LeNet-5 网络对羽绒图像的识别训练,图 4 网络结构对羽绒图像的识别效果更好,对比经过稀疏自编码预训练的 LeNet-5 网络,在数据集迭代 25 次后,图 4 网络结构对羽绒图像识别有较低的错误率,说明笔者提出网络结构对羽绒图像识别的有效性;同时,在图 4 网络结构中加入稀疏自编码预训练算法,识别效果也有了一定的提高;在此基础上加入谱残差算法和稀疏自编码预训练算法对网络权值进行初始化,在数据集迭代 25 次后,该网络对羽绒图像的识别正确率达到了 96.7% .

图 8 为 LeNet-5 网络与笔者提出的网络结构针对羽绒图像训练的时间对比曲线. 因羽绒图像易混,若图像分辨率小将更难区分,所以本数据集分辨率大小均为 100 × 100. 由图 8 可以看出,在网络训练中,笔者提出的网络结构较 LeNet-5 训练时间更短,收敛速度更快,且与本机 (Linux Ubuntu-16) 训练速度相比,在超算平台上的训练速度更快.

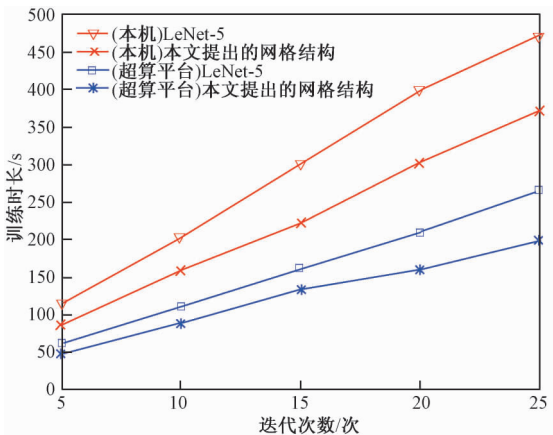


图 8 训练时长

Fig.8 Training time

4 结论

针对羽绒图像识别构造了一个深度卷积神经网络,并对权值初始化方法进行了改进,主要贡献如下:①利用视觉显著性模型提取原图显著区域,

并利用显著区域进行无监督预训练,训练得到的权值系数即为 DCNN 卷积核的初始权值.利用这种初始化方法,在样本数较小的情况下,DCNN 也能以较快速度达到权值最优,且识别精度有所提升;②网络结构主要采用 Inception 模块,适当加大了网络深度和宽度,提高了网络效率和识别精度.本试验的数据集是经过裁剪的子图像,而实际生产中的图像分辨率极高且图像内容更复杂,如何实现高分辨复杂图像的识别,是下一步将要进行的工作.

参考文献:

- [1] 陈继华,李勇,田增国,等.基于机器视觉的机械式表盘自动读表技术的实现[J].郑州大学学报(工学版),2015,36(3):101-105.
- [2] 张震,刘博,李龙.一种多特征提取及融合的虹膜识别方法[J].郑州大学学报(工学版),2017,38(1):63-67.
- [3] LECUN Y, BOSER B, DENKER J S, et al. Back-propagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4):541-551.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems (NIPS). California: MIT Press 2012: 1097-1105.
- [5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International Conference on Learning Representations (ICLR). San Diego: arXiv:1409.1556v6 [cs. CV], 2015.
- [6] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE computer society, 2015:1-9.
- [7] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE computer society, 2016:770-778.
- [8] IOANNOU Y, ROBERTSON D, SHOTTON J, et al. Training convolutional neural networks with low-rank filters for efficient image classification[J]. Journal of bacteriology, 2016, 167(3):774-783.
- [9] QU L, HE S, ZHANG J, et al. RGBD salient object detection via deep fusion[J]. IEEE transactions on image processing, 2017, 26(5):2274-2285.
- [10] WANG L, OUYANG W, WANG X, et al. STCT: sequentially training convolutional networks for visual tracking[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE computer society, 2016:1373-1381.
- [11] WU Y C, YIN F, LIU C L. Improving handwritten Chinese text recognition using neural network language models and convolutional neural network shape models[J]. Pattern recognition, 2016, 65(C):251-264.
- [12] MITRA V, FRANCO H. Time-frequency convolutional networks for robust speech recognition[C]// IEEE Automatic Speech Recognition and Understanding (ASRU). Scottsdale: IEEE computer society, 2015: 317-323.
- [13] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[J]. Journal of machine learning research, 2010, (9):249-256.
- [14] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification[C]// IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE computer society, 2015:1026-1034.
- [15] 王冠皓,徐军.基于多级金字塔卷积神经网络的快速特征表示方法[J].计算机应用研究,2015,32(8):2492-2495.
- [16] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the Inception architecture for computer vision[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE computer society, 2016:2818-2826.
- [17] ZENG K, YU J, WANG R, et al. Coupled deep autoencoder for single image super-resolution[J]. IEEE transactions on cybernetics, 2016, 47(1):27-37.
- [18] HOU X, ZHANG L. Saliency detection: a spectral residual approach[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Minneapolis: IEEE computer society, 2007:1-8.

Down Image Recognition Based on Deep Convolution Neural Networks

YANG Wenzhu, LIU Qing, WANG Sile, CUI Zhenchao, ZHANG Ningyu

(School of Cyber Security and Computer, Hebei University, Baoding 071002, China)

Abstract: Because of the scale and the various shapes of down in the image, it was difficult for traditional image recognition method to correctly recognize the type of down image and got the required recognition accuracy, even for the Traditional Convolutional Neural Network (TCNN). To solve the above problems, a Deep Convolutional Neural Networks (DCNN) for down image recognition was constructed, and a new weight initialization method was proposed. Firstly, the salient regions of down images were cut from the images using the visual saliency model. Then, these salient regions were used to train a sparse autoencoder and get a collection of convolutional filters, which accord with the statistical characteristics of dataset. At last, a DCNN with Inception module and its variants was constructed. To enhance the recognition accuracy, the depth of the network was deepened. The experiment results indicated that the constructed DCNN increased the recognition accuracy by 2.7% compared to TCNN, when recognizing the down in the images. The convergence rate of the proposed DCNN with the new weight initialization method was improved by 25.5% compared to TCNN.

Key words: deep convolutional neural networks; weights initialization; sparse autoencoder; visual saliency; image recognition

(上接第 5 页)

Improved TLD Object Tracking Algorithm Based on FAST Feature Points

MAO Xiaobo, ZHOU Xiaodong, LIU Yanhong

(School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: TLD was a new long-term object tracking algorithm in recent years. It is different from the traditional tracking algorithm in that it combined the traditional detection algorithm and the tracking algorithm, and then the learning module was used to study the characteristics of the object. In this paper, an improved TLD object tracking algorithm based on FAST feature points was proposed to ensure that the selected feature points could be tracked correctly and reliably to improve the accuracy of the tracker. At the same time, for the tracking process, the learning module template cumulative effects is obvious, reduced the real-time performance. Using a dynamic template management mechanism, when the number of templates reached the threshold, by comparing the similarity between the template and the current target, the specific template was deleted and the number of templates was kept constant. Experiments results showed that the improved algorithm had higher tracking precision and real-time performance.

Key words: TLD; long-term; object tracking; feature point