

基于Inception V3的图像状态分类技术

王旖旎

引用本文:

王旖旎. 基于Inception V3的图像状态分类技术[J]. 液晶与显示, 2020, 35(4): 389–394.

WANG-Yi ni. Image classification technology based on inception V3[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(4): 389–394.

在线阅读 View online: <https://doi.org/10.3788/YJYXS20203504.0389>

您可能感兴趣的其他文章

Articles you may be interested in

基于深度卷积生成对抗网络的图像识别算法

Image recognition algorithms based on deep convolution generative adversarial network

液晶与显示. 2020, 35(4): 383–388 <https://doi.org/10.3788/YJYXS20203504.0383>

基于深度学习的乳液泵缺陷检测算法

Defect detection algorithm of lotion pump based on deep learning

液晶与显示. 2019, 34(1): 81–89 <https://doi.org/10.3788/YJYXS20193401.0081>

角度空间三元组损失微调的人脸识别

Face recognition of triple loss fine-tuning in angular space

液晶与显示. 2019, 34(1): 110–117 <https://doi.org/10.3788/YJYXS20193401.0110>

基于深度卷积神经网络的多特征融合的手势识别

Multi-feature fusion gesture recognition based on deep convolutional neural network

液晶与显示. 2019, 34(4): 417–422 <https://doi.org/10.3788/YJYXS20193404.0417>

基于多重分形谱的木材高光谱图像纹理分类算法

Texture classification algorithm of wood hyper-spectral image based on multi-fractal spectra

液晶与显示. 2019, 34(12): 1182–1190 <https://doi.org/10.3788/YJYXS20193412.1182>

文章编号:1007-2780(2020)04-0389-06

基于 Inception V3 的图像状态分类技术

王旖旎*

(重庆商务职业学院, 重庆 401331)

摘要: 为了实现对物体状态的分类识别, 本文在 GoogLeNet 的 Inception V3 模块基础上进行了优化, 使用 Tanh 作为激活函数并结合 RMSprop, SGD 优化器提升了模型的准确率。首先采用三次卷积插值, GAN 对图像集进行预处理, 再利用 Inception 对图像进行训练, 最后结合 RMSprop 和 SGD 优化器对模型进行优化。用本文提出的模型在 20 个烹饪对象的图像上进行实验, 结果表明, 本文优化的 Inception V3 模型能够以 71.5% 的准确度对这些图像的状态进行分类, 与对比算法相比, 在分类准确度、训练损失上都有明显提升, 可以满足图像分类的可靠性、稳定性等要求。

关键词: 特征提取; 图像识别; 卷积神经网络; 模型训练

中图分类号: TP391.4 **文献标识码:** A **doi:** 10.3788/YJYXS20203504.0389

Image classification technology based on inception V3

WANG-Yi ni *

(Chongqing Business Vocational College, Chongqing 401331, China)

Abstract: In order to realize the classification and recognition of object state, the Inception V3 module on the basis of GoogLeNet is optimized Using Tanh as the activation function and RMSprop, the SGD optimizer improves the accuracy of the model. Firstly, using cubic convolution interpolation, GAN preprocesses the image set. Then, the Inception is used to train the image. Finally, the SGD optimizer is combined with RMSprop to optimize the model. Experiments on the images of 20 cooking objects using the model proposed in this paper show that the optimized Inception V3 model can classify the state of these images with 71.5% accuracy. Compared with the comparison algorithm, the classification accuracy and training loss are obviously improved, which can meet the requirements of reliability and stability of image classification.

Key words: feature extraction; image recognition; convolutional neural network; model training

1 引言

图像分类是计算机视觉各种应用的基础。从

人脸表情识别到人体姿势识别, 图像分类随处可见。在过去几年中, 随着机器学习和深度学习的进展, 图像分类邻域已经发生了深刻变化, 陆续涌现了很多有效的算法, 例如支持向量机、卷积神经

收稿日期: 2019-09-26; 修订日期: 2019-11-04.

基金项目: 2019 年重庆市教育委员会科学技术研究项目 (No. KJQN201904402); 重庆商务职业学院 2018 年度校级重点课题 (No. 2018XJKTZD06)

Supported by Science and Technology Research Project of Chongqing Education Commission, 2019 (No. KJQN201904402); Chongqing Business Vocational College 2018 Annual School Key Topic (No. 2018XJKTZD06)

* 通信联系人, E-mail: wangyini0917@126.com

网络等^[1-2]。深度学习已被证明是最有效的机器学习技术之一,其核心思路是构建多层神经元来训练网络。所有训练网络都使用数十个层,在网络中构建上千个参数,但是由于层级过深导致了较高的计算复杂度和过度拟合。对象状态分类在机器人、医疗、天文、游戏等领域的各种应用中非常有价值,例如识别物体的状态在人工智能技术中就很重要。在机器人领域,操作处于不同状态的物体需要不同的抓取动作,这需要在机器人的抓取动作中进行不同的行为。

识别物体状态问题的一种典型方法是使用形状匹配算法^[3],这类方法是从对象-对象-交互信息的三元组中获取对象状态来实现图像分类。这种方法可以连接和模拟同一帧中物体的运动和特征,训练过程由标记了的视频序列完成。这类算法通常基于贝叶斯网络,主要用对象特征、人类行为和对象反应作为参数^[4]。刘成颖等利用 PSO 优化了 LS-SVM 模型,实现了道具磨损状态的识别^[5]。杨昌其利用 BP 神经网络建立模型^[6],并对空中管制员的疲劳进行分类,用来评估空中交通管制员的疲劳状态。但是上述算法的神经网络深度较浅,优化的目标太过复杂,导致其只适合对状态分类较简单的情况进行识别,识别率不高。

2 基于 Inception V3 的图像识别算法

2.1 图像预处理

深度学习使用具有数百个隐藏层的神经网络,并且需要大量的训练数据。构建良好的神经网络模型需要仔细考虑网络的体系结构以及输入数据格式。数据收集和预处理是深度学习^[7]的重要组成部分。预处理的第一步是将数据分区为训练和测试数据集。本文对数据集中的图像执行多种预处理技术,分别是图像恢复、图像插值放大、图像增强和归一化。

数据划分:将数据划分为训练和测试数据集,在分类之前先用随机算法将数据进行随机混合,以便每个训练和验证集都没有任何特殊性。对于此项目,数据分类以 9:1 的比例完成。90% 的数据用于训练,剩余的 10% 用于测试训练结果。在划分完成之后,输入数据到图像预处理器中,对图像进行恢复和扩充等预处理,生成批量的规范化和增强数据。

图像缩放:本文在不同的图像大小和不同的 epoch 数量上训练模型。如果图像尺寸较大,则模型需要更多的 epoch 才能达到良好的准确度。为了让训练的时间复杂度更加稳定,本文对输入图像最终确定了图像尺寸 256×256 。重新缩放图像会按给定的缩放系数调整大小,由于重新缩放可能会严重影响精度,为了尽量不丢失图像的信息,本文利用 Cubic-convolution 插值算法对图像进行有理插值^[8]。

规范化图像输入:规范化可确保每个输入参数具有相似的数据分布,利用这一特性,规范化操作可以让训练时网络的收敛更快。本文通过从每个像素中减去平均值并将该结果除以标准偏差来完成数据的标准化。在本文的实验中,图像像素值在 $[0, 255]$ 范围内归一化。

数据扩充:如果在小型数据集上训练深度神经网络,可能导致过度拟合或者可能在测试数据上表现不佳,数据扩充通过增加数据集的大小防止过度拟合^[9]。由于本文主要针对食物的数据进行处理,而这类数据集本身比较小,因此利用数据扩充作为预处理技术的最后一步。本文的数据扩充主要利用生成对抗网络 (Generative Adversarial Networks, GAN) 来实现^[10],该网络可以不断学习图像特征,利用生成器不断生成伪造图片,利用判别器不断判断生成图片的真伪,最终达到纳什均衡,从而生成出逼真的图片。扩充后数据集的总大小约为 30 000 个图像。

2.2 优化的 Inception V3

本文利用 Inception V3 进行转移学习。在实际应用中,有时可能存在没有足够数据来训练网络以获得有效结果的情况。在这种情况下,可以使用转移学习。用于预训练网络的训练可能仅对一个域是足够的,并且我们拥有的数据可能在不同的特征空间中。在这些情况下,应用学习转移方式赋予网络已训练好的参数信息,将大幅提高神经网络的学习性能。

创建的模型将在对数据集进行训练时在预训练模型上添加了一些图层。整个模型如图 1 所示。

网络的第一层是卷积层。卷积层以张量阵列的形式生成输出^[11-12]。

下一层是批量标准化层 (BN 层)。批量标准化提供了一种将输入逼近 0 或对输入数组中

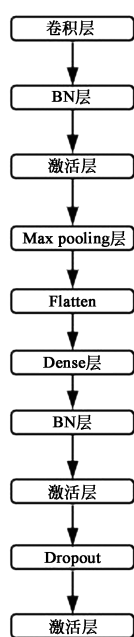


图1 本文提出的模型架构

Fig.1 Model architecture presented in this paper

所有值求平均的方法。这可以在网络层的开始和中间使用,该技术有助于更快地学习并获得更高的准确性。在深度卷积神经网络中,随着数据进一步流动,权重和参数会调整它们的值。模型使用中间数据,会导致计算量太大而增加复杂度,通过规范每个小批量数据可以避免这个问题。

网络中的下一层是一个具有整流线性单元(ReLU)作为激活功能的激活层。ReLU 由于良好的收敛性被广泛应用在深度卷积神经网络,其中 Tanh 的性能最高,能将收敛性提高 6 倍^[13]。ReLU 的数学公式见式(1)。

$$y = (e^x - e^{-x}) / (e^x + e^{-x}) \quad (1)$$

另一层是最大池化层。进行最大池化的目的是对数据进行下采样,这样可以通过减少数据的维度来避免过度拟合。最大池化的另一个好处是它通过减少参数数量来降低计算成本。在本文的模型中,将数据的维度从 8×8 降到 2 个 4×4 ,然后从 4×4 降低到 2 个 2×2 。总体来说,在网络中使用两组卷积、批量标准化、激活和最大池化层可以获得更高的准确度。

下一层进行 Flatten^[14-16]操作。Flatten 将前一层的二维输出转换为单个长连续线性向量,Flatten 是在数据流进入全连接层之前的必要操

作,全连接层只是一个人工神经网络(ANN)分类器,需要 Flatten 将输入数组转换为特征向量^[17]。

再下一层是全连接层或稠密层。全连接层是完全连接的神经网络层,其中每个输入节点连接到每个输出节点。它将 Flatten 的向量作为输入,并将 n 维度张量作为输出。使用稠密层可以解决消失梯度问题,同时加强了特征传播,促进了特征重用,并大幅减少了参数的数量。

接下来是激活层,然后是 Dropout 层用于正则化。正则化有助于通过降低过度拟合的风险,从而更好地推广模型。如果数据集的大小与需要学习的参数数量相比太小,则存在过度拟合的风险。Dropout 层随机删除网络中的某些节点及其连接。我们可以使用带有隐藏或输入图层的 Dropout。通过使用 Dropout,网络中的节点对其他节点的耦合性变得更加不敏感,这使得模型更加稳定。

网络中的最后一层是具有 Softplus 激活功能的激活层。Softplus 函数将输出收敛于 $0 \sim 1$ 之间。通常最好在最后一层网络中使用 Softplus 来解决分类问题。它用于分类概率分布。Softplus 函数的数学公式见式(2)。

$$\partial(x) = \lg(1 + e^x) \quad (2)$$

2.3 训练过程

模型的训练分两步完成。首先,仅在给定数据集上利用本文的模型进行训练。训练完成之后,通过训练顶部 4 层基础模型并保持底层的其余部分不参与训练来完成微调。实验在 2 个优化器上完成:RMSprop 和随机梯度下降(Stochastic Gradient Descent, SGD)。首先,使用优化器 RMSprop 编译模型,仅训练新层并保持基层不可训练。RMSprop 通过将学习率除以指数衰减的平方梯度平均值来完成优化^[18],本文中 RMSprop 的学习率设置为 0.002。模型的目标函数是交叉熵损失函数(Categorical crossentropy loss),该目标函数通常用在数据集中存在。通过将 epoch 保持为 20, 50, 100, 同时保持 batch 大小为 32 来进行实验。为了避免过度拟合,第一轮训练的 batch 大小(32)选得相对较小。在训练期间保存了最佳模型参数(在验证损失和准确性方面),以便可以在第二步中继续优化。在训练的第二步,通过使用给定数据集训练前 4 层基础模型来完成微调,同时冻结剩余的下面几层。在该过程中使用在训练第一步中获得的最佳模型,通过

训练前 2 层和前 4 层基础模型进行实验,并且训练前 4 层给出最佳结果。此步骤中使用的优化器是 SGD。添加前 4 层的 Inception 会导致学习参数数量增加,从而增加过度拟合的可能性。因此采用 SGD 优化器,其收敛速度较慢,可以降低过度拟合的情况。学习率 0.000 2,衰减 = e^{-6} 和 momentum = 0.5。第一阶段的训练消耗了 150 个 epoch,同时保存了每个 epoch 产生的所有模型。因为模型有可能随着时间的推移而停止前进,从而产生过度拟合的模型,因此选择最佳模型时本文综合考虑了损失函数和最短运行时间。

3 实验和结果分析

本文算法在 Tensorflow 平台完成编程,并在 GTX1080Ti 的 GPU、32 G 内存、Xeon-E3 CPU、windows 平台下进行实验。

本文使用的数据是 20 种肉类和植物食物(苹果、西瓜、奶酪、牛肉、三文鱼等)的图像,具有 6 种不同的物理状态(完整、切块、切片、磨碎、糊状、榨

汁)。原始训练数据集提供了大约 6 231 个图像,扩充后达到 30 000 个图像。图片实例如图 2 所示。



图 2 测试数据集样例
Fig.2 Data set of sample test

在训练期间进行了各种实验以产生最佳模型,同时利用 GAN 对图片进行扩张。在对比实验中,本文采用了基于 SVM 的图像分类器、BP 神经网络、ResNet 与本文算法进行对比。

表 1 算法结果对比
Tab.1 Comparison of algorithm results

算法	训练准确率/%	训练损失	验证准确率/%	验证损失
SVM	61.2	1.82	62	1.85
BP 神经网络	63.2	1.01	67.1	1.83
ResNet	68.3	0.098	71.2	1.21
本文算法	71.5	0.092	74.3	1.09

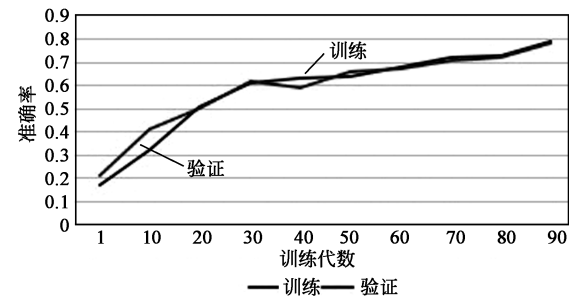


图 3 训练和验证的准确率对比
Fig.3 Comparison of accuracy of training and verification

这些实验可以添加更多层以改善分类结果。在这种情况下,使用预训练网络(GoogLeNet)对

这组图层进行微调可以获得最佳效果。使用之前的训练中选择的最佳模型进行微调,进一步改善了结果。图 3 显示了显示训练和验证准确率对比。

4 结 论

提出了一个基于 Inception V3 结构的神经网络食物状态识别算法,在识别前先对食物图像进行数据划分、图像缩放等预处理,再在原有 Inception V3 的基础上采用两阶段训练的方式提高了模型的准确度。在 20 个烹饪对象的对比实验中,选择了 SVM、BP 神经网络、ResNet 进行参照。

实验结果表明,本文算法能有效提高识别准确率,与对比实验中的次优算法相比,本文算法训练准确率提高了 3.2%。通过不同 epoch 的训练结果

和测试数据可以看出,本文算法在识别过程中具有较好的性能,可以满足图像分类的可靠性、稳定性等要求,为物体状态识别领域提供有益参考。

参 考 文 献:

- [1] 郭敬明,何昕,魏仲慧.基于在线支持向量机的 Mean Shift 彩色图像跟踪[J].液晶与显示,2014,29(1):120-128.
GUO J M, HE X, WEI Z H. New mean shift tracking for color image based on online support vector machine [J]. *Chinese Journal of Liquid Crystals and Displays*, 2014, 29(1): 120-128. (in Chinese)
- [2] 李彦冬,郝宗波,雷航.卷积神经网络研究综述[J].计算机应用,2016,36(9):2508-2515.
LIY D, HAO Z B, LEI H. Survey of convolutional neural network [J]. *Journal of Computer Applications*, 2016, 36(9): 2508-2515. (in Chinese)
- [3] 黄伟国,顾超,朱志奎.用于目标识别的 PCA-SC 形状匹配算法[J].光学 精密工程,2013,21(8):2103-2110.
HUANG W G, GU C, ZHU Z K. PCA-SC shape matching for object recognition [J]. *Optics and Precision Engineering*, 2013, 21(8): 2103-2110. (in Chinese)
- [4] 王飞,李定主.模式识别中贝叶斯决策理论的研究[J].科技情报开发与经济,2007,17(7):165-166.
WANG F, LI D Z. Research on Bayesian decision-making theory in pattern recognition [J]. *Sci-Tech Information Development & Economy*, 2007, 17(7): 165-166. (in Chinese)
- [5] 刘成颖,吴昊,王立平,等.基于 PSO 优化 LS-SVM 的刀具磨损状态识别[J].清华大学学报:自然科学版,2017,57(9):975-979.
LIU C Y, WU H, WANG L P, et al. Tool wear state recognition based on LS-SVM with the PSO algorithm [J]. *Journal of Tsinghua University: Science and Technology*, 2017, 57(9): 975-979. (in Chinese)
- [6] 杨昌其,谭娟,仇争平.基于 BP 神经网络的管制员疲劳状态识别研究[J].航空计算技术,2018,48(1):17-20,25.
YANG C Q, TAN J, QIU Z P. Controller's fatigue recognition based on BP neural network [J]. *Aeronautical Computing Technique*, 2018, 48(1): 17-20, 25. (in Chinese)
- [7] 刘建伟,刘媛,罗雄麟.深度学习研究进展[J].计算机应用研究,2014,31(7):1921-1930,1942.
LIU J W, LIU Y, LUO X L. Research and development on deep learning [J]. *Application Research of Computers*, 2014, 31(7): 1921-1930, 1942. (in Chinese)
- [8] WANG X, DING Y, LIU M Y, et al. Efficient implementation of a cubic-convolution based image scaling engine [J]. *Journal of Zhejiang University Science C*, 2011, 12(9): 743-753.
- [9] 李俭川,秦国军,温熙森,等.神经网络学习算法的过拟合问题及解决方法[J].振动、测试与诊断,2002,22(4):260-264.
LI J C, QIN G J, WEN X S, et al. Over-fitting in neural network learning algorithms and its solving strategies [J]. *Journal of Vibration Measurement & Diagnosis*, 2002, 22(4): 260-264. (in Chinese)
- [10] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [J]. *ArXiv*, 2015,1511:06434.
- [11] CHEN J C, PATEL V M, CHELLAPPA R. Unconstrained face verification using deep CNN features [C]//*Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision*. Lake Placid, NY, USA: IEEE, 2016: 1-9.
- [12] KRON G. Tensor analysis of networks [J]. *Students' Quarterly Journal*, 1996, 36(143): 171.
- [13] FAN E G. Extended Tanh-function method and its applications to nonlinear equations [J]. *Physics Letters A*, 2000, 277(4-5): 212-218.
- [14] 杨江涛,唐军,王玉波,等.用于应力测量的可调谐光栅[J].光学 精密工程,2018,26(7):1596-1603.
YANG J T, TANG J, WANG Y B, et al. Tunable grating for stress measurement [J]. *Optics and Precision Engineering*, 2018, 26(7): 1596-1603. (in Chinese)
- [15] 郁晓晖,高志山,袁群.宽光谱长工作距弱荧光信号检测显微物镜设计[J].光学 精密工程,2018,26(7):1588-1595.

- YU X H, GAO Z S, YUAN Q. Design of board spectrum and long work distance microscope objective for weak fluorescence signal detection [J]. *Optics and Precision Engineering*, 2018, 26(7): 1588-1595. (in Chinese)
- [16] 林旭东, 刘欣悦, 王帅, 等. 桌面 97 单元自适应光学系统性能测试[J]. 光学 精密工程, 2016, 24(6): 1272-1280.
- LIN X D, LIU X Y, WANG S, *et al.* Performance testing of a desk-top 97-element adaptive optical system [J]. *Optics and Precision Engineering*, 2016, 24(6): 1272-1280. (in Chinese)
- [17] CHEN A J H, WANG B C W, JENG C J H. A preliminary study of ANN implementing image filters [C]//*Proceedings of 2013 International Symposium on Next-generation Electronics*. Kaohsiung, China: IEEE, 2013: 181-184.
- [18] KESKAR N S, SOCHER R. Improving generalization performance by switching from Adam to SGD [J]. *arXiv*, 2017, 712:07628.

作者简介:



王旖旎(1981—),女,重庆人,硕士,副教授,2012 年于重庆师范大学获得硕士学位,主要从事智能计算、数据挖掘、软件技术等方面的研究。E-mail:wangyini0917@126.com