

基于改进SIFT特征和神经网络结合的场景识别

郭昊琛, 闫帅帅, 刘天鹤

(河北工业大学人工智能与数据科学学院, 天津 300401)

摘要: 基于深度学习的场景识别作为计算机视觉领域的重要方向, 目前仍存在部分问题, 如仅提取图像的高层语义特征而缺失了图像的底层特征, 针对这个问题, 提出基于改进SIFT特征与深度神经网络相结合的室内RGB-D图像识别方法。首先提取图像的SIFT特征, 然后利用随机森林算法根据重要度对SIFT特征进行筛选, 然后结合基于ResNet的深度神经网络, 并提出基于深度直方图与深度均值直方图的深度损失函数, 加速模型的收敛。实验结果表明, 算法可以在NYUD v2数据集上达到71.52%的识别率, 有效提升了室内场景识别的准确率。

关键词: 改进SIFT特征; 深度神经网络; 损失函数; 深度直方图

中图分类号: TP183 **文献标识码:** A

Scene Recognition Based on Improved SIFT Features and Neural Network

GUO Haochen, YAN Shuaishuai, LIU Tianhe

(School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China)

Abstract: As an important direction of computer vision, scene recognition based on deep learning still has some problems, such as only extracting the high-level semantic features and missing the bottom features of an image. To solve this problem, the paper proposes an indoor RGB-D image recognition method based on improved SIFT features and deep learning neural network. Firstly, the SIFT features of images are extracted, the SIFT features are filtered according to the importance degree by means of the Random Forest Algorithm, and then the depth loss function based on the depth histogram and the depth mean histogram is proposed to accelerate the convergence of the model by combining the ResNet-based deep neural network. The experimental results show that the algorithm can achieve 71.52% recognition rate on NYUD V2 data set, and effectively improve the accuracy of indoor scene recognition.

Keywords: improved SIFT features; deep learning neural network; loss function; depth histogram

1 引言(Introduction)

随着“人工智能”热潮的到来, 场景识别作为其中的关键技术之一, 已经成为图像处理领域的重要研究问题, 场景识别技术的进步可以极大地推动智能机器人、图像检索、视频检索等领域的发展。Lowe^[1]于1999年提出, 并于2004年加以完善了尺度不变特征变换(Scale-Invariant Feature Transform, SIFT), 该特征是用来侦测与描述影像中的局部性特征, 它在空间尺度中寻找极值点, 并提取出其位置、尺度、旋转不变量, 对于旋转、位移、缩放甚至遮挡都具有一定的稳定性。Wang等人^[2]提出了MR-CNNs(Multi-Resolution CNNs)网络, 使用了多分辨率的设计, 分别提取低分辨率的全局信息和高分辨率的细节信息, 更加全面的利

用图像信息。

2 基于传统特征和深度学习的场景识别(Scene recognition based on traditional features and deep learning)

2.1 改进SIFT特征算法

由于SIFT特征是从尺度空间检测极值点, 导致SIFT特征具有边缘效应, 即图像的边缘处容易检测到极值点, 但其中一部分极值点是对分类无效的, 去除掉这些无效的边缘点, 分类效果就会提升, 但无差别去除边缘上的特征点会丢失部分在边缘上的有效特征点, 导致损失一部分底层信息^[3]。

本文使用随机森林算法对SIFT特征进行筛选, 根据SIFT特征点与Canny边缘点的欧式距离, 赋予特征权重, 以此来

筛选SIFT特征,在保持底层信息不丢失的情况下,筛去大部分重要度低的SIFT特征,保留重要度高的SIFT特征。具体步骤为:

(1)构建尺度空间 $L(x,y,\sigma)$,如式(1)所示:

$$L(x,y,\sigma) = G(x,y,\sigma) \otimes I(x,y) \quad (1)$$

其中, $I(x,y)$ 为原图, $G(x,y,\sigma)$ 为高斯核函数, σ 称为尺度空间因子,它是高斯正态分布的标准差,反映了图像被模糊的程度,其值越大图像越模糊,对应的尺度也就越大。如式(2)所示:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

(2)构建高斯金字塔,两层之间相差比例因子 k ,若每组有 S 层, $k = 2^{\frac{1}{S}}$ 。

(3)相邻两层相减,得到DoG(差分高斯)响应图像,如式(3)所示:

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \quad (3)$$

(4)比较每个像素与周围8个像素,以及相邻的上下两层DoG响应图像9个像素,共26像素,若该像素大于或小于所有的像素,则为特征点。

(5)计算以特征点为中心、以 $3 \times 1.5\sigma$ 为半径的区域图像的幅角和幅值,每个点 $l(x,y)$ 的梯度的模 $m(x,y)$ 和方向 $\theta(x,y)$ 可通过式(4)和式(5)求得:

$$m(x,y) = \sqrt{|l(x+1,y) - l(x-1,y)|^2 + |l(x,y+1) - l(x,y-1)|^2} \quad (4)$$

$$\theta(x,y) = \arctan \frac{l(x,y+1) - l(x,y-1)}{l(x+1,y) - l(x-1,y)} \quad (5)$$

(6)将每个特征点的模 $m(x,y)$ 和方向 $\theta(x,y)$ 投影到八个方向(每 45° 取一个方向)上,取最大值为主方向,当存在80%最大值的方向时,设为辅方向,其余方向舍去,此时就得到了SIFT特征点的位置信息、尺度信息和方向信息。

(7)使用高斯滤波器平滑图像,消除噪声,然后计算图像中每个像素点的梯度强度和方向,应用非极大值(Non-Maximum Suppression, NMS)抑制,以消除边缘检测带来的杂散响应,再应用双阈值(Double-Threshold)检测来确定真实的和潜在的边缘,最后通过抑制孤立的弱边缘最终完成边缘检测,得到图像的Canny边缘算子 $C(x,y)$ 。

(8)求出每个SIFT特征点 $l(x,y)$ 与距离最近的边缘算子 $C(x,y)$ 的欧式距离 S_1 ,如式(6)所示,并求出 \bar{S}_1 。

$$S_1 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (6)$$

\bar{S}_1

(9)给每个特征赋予 $\frac{1}{\bar{S}_1}$ 的权重,并放入随机森林,并建立袋外数据(OOB),每一颗决策树都随机抽取一半的特征点,另一半放入袋外数据,并随机加入噪声,计算GINI指数,如式(7)所示:

$$GINI = \frac{\sum_{i=1}^n (c_i \times \omega_i)^2}{\sum_{i=1}^n (c_i \times \omega_i)} \quad (7)$$

其中, n 是节点数量, c 是抽取的特征数, ω 是权重。

(10)归一化每一个特征的GINI,得出重要度VIM,并按重要度进行筛选,得到改进的SIFT特征。

2.2 基于ResNet的深度神经网络

随着深度学习的不断发展,网络模型也不断增多,主要改进方向之一就是加深模型层数,基于卷积神经网络的深度神经网络模型,往往通过不断交叉叠加非线性函数和卷积网络层增强整个网络模型的表达能力,故更深的网络通常代表着更佳的性能。但是随着网络深度的快速增加,过深的网络模型不但没有提高模型精准度,反而饱受梯度消失,梯度爆炸等问题困扰,导致了更高的训练误差。针对上述问题,何凯明等人^[4]提出了基于残差网络结构的ResNet网络模型,该模型与之前提出的分类网格相比,在收敛性能、分类准确率和训练速度等方面都有了较大提升,本文主干网络采用ResNet 50网络。

目前多尺度的图像金字塔网络大多只采用网络最后一层的特征,一般高层特征的语义信息比较丰富,但是目标位置比较粗略,Lin等人^[5]提出了特征金字塔网络(Feature Pyramid Networks, FPN)做预测。算法采用一个自底向上的路线、一个自顶向下的路线和横向连接(Lateral Connection)。其中,横向连接为自底向上的特征图像经过一个 1×1 的卷积与自顶向下的特征图像经过一个两倍上采样产生的特征图像进行融合,再采用 3×3 的卷积核对每个融合结果进行卷积,目的是消除上采样的混叠效应。

对于感兴趣区域的获取,Ren等人^[6]提出了Faster-RCNN网络,采用区域生成网络(Region Proposal Networks, RPN)算法。其本质是基于滑窗的无类别检测器,对于每一个可能的物体都生成九个滑窗,然后利用非极大值抑制对这些滑窗进行筛选,得到最终的感兴趣区域。

2.3 基于深度直方图的损失函数

本文在处理深度信息时,采用了深度直方图特征,将深度信息投影到0—255的坐标轴上,根据每像素的深度值得到深度直方图,然后再将每类场景的深度直方图求出均值,得到深度均值直方图。

参考了SVM的合页损失后,本文为深度值加上权重与偏置,希望其可以无限的接近样本真实类别的深度均值,然后引进一个松弛变量 Δ ,使得当样本加权深度信息值与该样本真实类别均值深度值之差的二范数大于 Δ 时,取损失,否则不计入损失,用来惩罚与真实类别的均值深度值相差过大的样本,如式(8)所示:

$$loss1 = \frac{1}{M} \min_{\omega, b} \sum_{j=1}^N \sum_{i=1}^T \max(0, \|x_{i,j} \omega + b - \bar{X}\| - \Delta) \quad (8)$$

其中, ω 为权重, b 为偏置, M 为样本总数, N 为每类样本数, T 为类别数, Δ 为界值。

本文分类器使用的是SoftMax分类器,其损失函数如式(9)所示:

$$loss2 = -\sum_{k=1}^T y_k \log S_k \quad (9)$$

4 结论(Conclusion)

本文提出了基于改进SIFT特征与深度神经网络相结合的RGB-D图像识别算法,在NYUD v2数据集上的实验结果表明,本文算法识别率可以达到71.5%,在剔除干扰类别后,可以提升至93.47%,识别率得到有效的提升,损失函数也可以稳定快速的收敛。

本文使用了改进的SIFT特征,改善了SIFT特征的边缘特性,又将改进的SIFT特征于深度卷积网络相结合,解决了深度神经网络高层语义特征丰富但底层信息缺失的问题,实验结果表明算法具有良好的鲁棒性。

参考文献(References)

- [1] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [2] Wang L, Guo S, Huang W, et al. Knowledge guided disambiguation for large-scale scene classification with multi-resolution CNNs[J]. IEEE Transactions on Image Processing, 2017, 26(4): 2055-2068.
- [3] 张春林, 陈劲杰. 基于改进SIFT和RANSAC的物体特征提取和匹配的研究[J]. 软件工程, 2018, 21(11): 6-9.
- [4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770-778.

(上接第17页)

说明太原市PM_{2.5}污染除一次排放之外, 气态污染物的二次合成也占一定的比重。PM_{2.5}和PM₁₀回归分析得到回归模型为 $PM_{2.5} = 0.63PM_{10} - 11.76$ ($R^2 = 0.8427$), 回归方程拟合度较好; PM_{2.5}和其他气态污染物多元线性回归模型为 $PM_{2.5} = 0.24SO_2 + 20.08CO + 11.54$ ($R^2 = 0.4844$), 拟合度检验效果一般, 考虑三者之间还有其他因素的影响, PM_{2.5}的形成除了气态污染物的二次合成, 主要还有区域污染物的一次排放。

参考文献(References)

- [1] Jihua Tan, Leiming Zhang, Xueming Zhou, et al. Chemical characteristics and source apportionment of PM_{2.5} in Lanzhou, China[J]. Science of the Total Environment, 2017, 601-602: 1743-1752.
- [2] Ryou H G, B J H, Kim S. Source apportionment of PM₁₀ and PM_{2.5} air pollution, and possible impacts of study characteristics in South Korea[J]. Environmental Pollution, 2018, 240: 963-972.
- [3] Li R, Hardy R, Zhang W, et al. Chemical Characterization and Source Apportionment of PM_{2.5} in a Nonattainment Rocky Mountain Valley[J]. Journal of Environmental Quality, 2018, 47: 238-245.
- [4] 李子伊. 基于Excel和R语言的成绩统计分析——以福

- [5] Lin T-Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [6] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]. Advances in neural information processing systems, 2015: 91-99.
- [7] Gupta S, Arbelaez P, Malik J. Perceptual organization and recognition of indoor scenes from RGB-D images[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 564-571.
- [8] Song X, Herranz L, Jiang S. Depth CNNs for RGB-D scene recognition: learning from scratch better than transferring from RGB-CNNs[C]. Thirty-First AAAI Conference on Artificial Intelligence, 2017.
- [9] Herranz-Perdiguero C, Redondo-Cabrera C, López-Sastre R J. In pixels we trust: From Pixel Labeling to Object Localization and Scene Categorization[C]. IEEE/R SJ International Conference on Intelligent Robots and Systems, 2018: 355-361.

作者简介:

郭昊琛(1994-), 男, 硕士生. 研究领域: 图像处理. 本文通讯作者.

闫帅帅(1993-), 男, 硕士生. 研究领域: 深度学习.

刘天鹤(1991-), 男, 硕士生. 研究领域: 深度学习.

建省某中学期中、期末英语考试成绩为例[J]. 英语教师, 2017, 17(24): 44-51.

- [5] 周芸韬. 基于R语言的大数据处理平台的设计与实现[J]. 现代电子技术, 2017, 40(02): 53-56.
- [6] 武茗馨. 基于R语言的金融大数据审计应用研究[D]. 南京审计大学, 2017.
- [7] 李雄英. 基于R语言的统计教学应用初探[J]. 高教学刊, 2017(01): 50-51.
- [8] 董健卫, 陈艳美, 孟盼, 等. 回归分析与基于MIV的RBF神经网络在PM_{2.5}的相关因素分析中的应用[J]. 数学的实践与认识, 2017, 47(10): 127-136.
- [9] 张红, 董小刚, 李群. PM_{2.5}浓度影响因素的主成分回归分析[J]. 长春工业大学学报, 2017, 38(02): 105-110.
- [10] 杜续, 冯景瑜, 吕少卿, 等. 基于随机森林回归分析的PM_{2.5}浓度预测模型[J]. 电信科学, 2017, 33(07): 66-75.

作者简介:

解蕾(1982-), 女, 硕士, 讲师. 研究领域: 软件开发, 数据管理.

狄光智(1977-), 女, 博士, 副教授. 研究领域: 软件开发, GIS二次开发.