

基于迁移学习与自适应特征融合的建筑物识别

王泽泓, 刘厚泉

(中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

摘要:特定建筑物识别可应用于旅游景点地点的查询。针对现有特定建筑物识别方法特征提取困难、识别率低等问题,提出一种基于卷积神经网络的特定建筑物识别方法。针对数据规模小的限制,提出利用预训练的 VGG-16 网络进行迁移学习,以改善网络效果。为了充分利用 VGG-16 网络中各层提取出的特征,提出自适应特征融合的方法。该方法针对网络中不同层提取出的特征图的层次不同、尺度不同的特点,给每个特征图设置可学习的权重,进而融合在一起进行预测。通过从网络上爬取的 6 273 张 12 类旅游景点建筑物图片作为数据集,对提出的方法进行验证。使用 VGG-16 网络训练,最终准确率为 97.86%,处理速度为 296 fps。利用自适应特征融合改进 VGG-16 网络,改进后最终准确率为 98.93%,处理速度为 289 fps,比改进前准确率提高 1.07%。

关键词:建筑物识别;卷积神经网络;迁移学习;自适应;特征融合

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2019)12-0040-04

doi:10.3969/j.issn.1673-629X.2019.12.007

Building Recognition Based on Transfer Learning and Adaptive Feature Fusion

WANG Ze-hong, LIU Hou-quan

(School of Computer Science, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: Specific building recognition can be applied to queries about tourist attraction locations. Aiming at the difficulty of feature extraction and low recognition rate of existing specific building recognition methods, a specific building recognition method based on convolutional neural network is proposed. In view of the limitation of small data size, we propose to use the pre-trained VGG-16 network for transfer learning to improve network performance. In order to make full use of the features extracted by each layer in the VGG-16 network, we propose an adaptive feature fusion method in which each feature map is given a learnable weight and integrated to predict according to the characteristics of different layers and scales of feature maps extracted from different layers in the network. The proposed method is verified by taking 6 273 pictures of tourist attractions buildings crawled from the internet as data sets. With VGG-16 network training, the final accuracy rate is 97.86%, and the processing speed is 296 fps. The VGG-16 network is improved by adaptive feature fusion. The improved final accuracy is 98.93% which is improved by 1.07%, and the processing speed is 289 fps.

Key words: building recognition; convolutional neural network; transfer learning; adaptive; feature fusion

0 引言

特定建筑物识别具有很好的应用前景。利用建筑物识别技术,可以从图像中理解地点。面对陌生未知的建筑物,通过建筑物识别技术,游客可以识别到图像中的地点,因而建筑物识别技术可应用于旅游景点的地点识别。

近年来,研究人员在建筑物识别和物体识别方面做了许多研究。传统的建筑物识别方法通常是对大量建筑物图像手工设计特征或设计特征检测算子,并设

计相应算法进行识别。如 David Lowe 提出的尺度不变特征变换(scale-invariant feature transform, SIFT)^[1]用于在图像中检测出关键点,描述图像的局部特征,进而设计识别算法。现在有研究人员尝试利用 SIFT 结合边缘特征与关键点特征进行建筑物识别^[2]。Dalal 等提出方向梯度直方图(histogram of oriented gradient, HOG)^[3]用于计算和统计图像局部区域的梯度方向直方图来构成特征,之后结合 SVM^[4]广泛应用于图像识别。现在有研究人员尝试使用 HOG 特征进

收稿日期:2019-01-05

修回日期:2019-05-07

网络出版时间:2019-06-27

基金项目:国家级大学生创新训练计划(201710290053)

作者简介:王泽泓(1995-),男,CCF 会员(A8118G),研究方向为计算机科学与技术;刘厚泉,博士,教授,研究方向为虚拟现实及大数据分析。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190627.1105.054.html>

行建筑物识别^[5]。但是,以上方法都需要人工设计特征或者人工设计特征提取算子。而如何提取特征以及充分利用提取出的特征,是建筑物图像识别中最关键的问题。

为了解决上述问题,文中引入卷积神经网络(convolutional neural network, CNN)^[6]。CNN 广泛应用于各种视觉识别任务。Maxime Oquab 等提出的迁移学习^[7-9]可以有效解决难以获得高质量的建筑物的大型数据集的问题,因为浅层神经网络提取出的特征是比较通用的特征。Karen Simonyan 等提出对卷积网络的特征图进行融合^[10-12],以充分利用更多的特征,得到更好的效果。因此,文中基于迁移学习对预训练的 VGG-16 网络^[10]进行改进,设计能够自适应特征融合的卷积神经网络应用于建筑物图像识别任务中。

1 自适应特征融合的建筑识别方法

1.1 方法概述

文中将卷积神经网络应用于建筑物图像识别,同时利用迁移学习,利用在 ImageNet 数据集上预训练的 VGG-16 网络的权重,并对网络结构进行改进,使其能够自适应地进行特征融合,充分利用各个层提取出的特征,以提高 VGG-16 网络的性能。

1.2 数据来源与预处理

文中的建筑物图像是利用网络爬虫在网上获取的,这些图像共 6 273 张,共 12 个类别(目前数据集及文中算法已在 <https://github.com/zehong1995/structureRecognition> 开源)。由于从网络上爬取的图片尺寸不一,所以首先使用三次样条插值方法对原图进行降采样,得到相同尺寸 124 × 124 的图片。建筑物图像预处理见图 1。

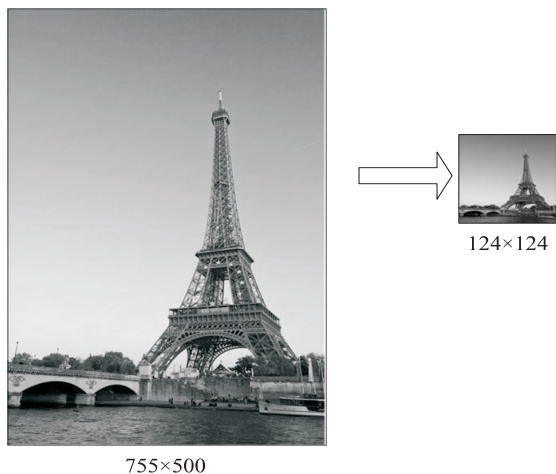


图 1 建筑物图像预处理

1.3 自适应特征融合的网络结构

由于卷积神经网络中浅层可以学习到比较通用的低级特征,深层可以学到特定的高级特征,且网络层数

越多,整个网络在样本空间中就可以拟合更复杂的函数。所以,为了更充分地利用各层提取出的特征,提出自适应特征融合方法(见图 2),此网络由 VGG-16 改进而来。VGG-16 网络是 2014 年由牛津大学的视觉几何组(visual geometry group)提出,该网络共有五个卷积块,每个卷积块后都是最大池化层,最后三层为全连接层。文中只使用 VGG-16 的前五个卷积块,之后将第二至第五个卷积块中池化层前的特征图取出,分别进行自适应的特征映射,得到四个映射的特征向量,再与第五个卷积块中最大池化层提取出的特征图相融合。在对不同层次特征进行自适应特征映射部分,由于每层的特征图的尺度不同,所以文中先使用 1 × 1 卷积^[13]提取特征,之后使用全连接层使得每个特征图都能映射为相同长度的特征向量,再对每个特征向量添加一个可学习的权重 γ_i ,再使用求和方法将 5 个相同尺度的特征向量进行融合,最后使用三层全连接层对提取出的特征向量进行分类。

用 $f_{i,j,k}^n, 1 \leq n \leq 4$ 分别表示第二至第五个卷积块池化层前取出的特征图。首先用 1 × 1 卷积将特征图进行降维,得到 $p_{i,j,k}^n$,再用 W^n 分别对 $p_{i,j,k}^n$ 进行自适应的特征映射,得到特征向量 θ^n ,再用可学习的权重 γ^n 对特征向量进行加权求和,完成特征融合。

$$p_{i,j,k}^n = \sum_j \sum_k \sum_t \text{filter}_{i,j,k}^n \times f_{i,j,k}^n \quad (1)$$

$$\theta^n = \max(W^n p^n + \text{bias}, 0) \quad (2)$$

$$f_v = \theta^n + \sum_{n=1}^4 \gamma^n \theta^n \quad (3)$$

基于上述结构,对提取出的中间特征先进行 1 × 1 卷积,增强模型的非线性性质,之后对特征进行特征映射,将不同尺寸的特征向量映射成为相同尺寸,再使用激活函数 Relu^[14]进行非线性变换。由于不同层次提取的特征的重要性不同,又难以人工判断,所以最后需要进行一个加权求和,其中前四个映射好的特征需要进行一个可学习的加权求和,第五个特征权重设置为固定值 1。在这里,将第五个卷积块输出的特征的权重设置为常量 1,而将其他的特征的权重设置为可学习的,有如下几方面的考虑:

(1) 因为 VGG-16 网络能够较好地提取特征,所以文中不对第五个卷积块输出的特征图设置可学习的权重。

(2) 对于前面层提取的特征,设置可学习的权重,一方面是为了使模型能够自己学习到最佳的权重,即能够自适应地进行特征融合,另一方面是避免在前面层提取的特征对最终效果有负面影响的情况下,能够使得该分支很容易地学到恒等映射,即对于式 3,希望有:

$$fv = \theta^n$$

(4)

那么,只需要模型学习到

$$\gamma^n = 0$$

(5)

这样就不会对模型效果有任何负面效果。

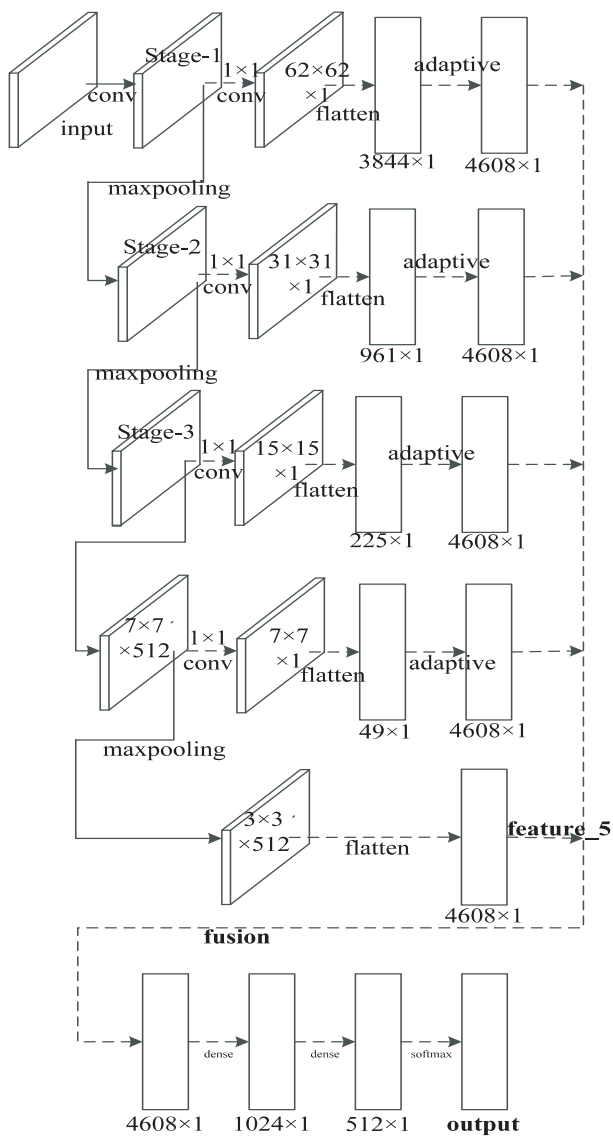


图 2 自适应特征融合网络的详细结构

2 实验过程与结果

2.1 训练过程

文中训练了两个网络,用以对比提出的自适应特征融合网络的有效性。一是对在 ImageNet 数据集上预训练的 VGG-16 网络进行训练,二是对自适应特征融合网络进行训练。将数据集分为两部分,5 051 张作为训练集,1 222 张作为测试集。

由于所采用数据集的规模较小,为了避免过拟合并提高模型效果,两个网络的训练都采用相同的数据增强方法(见图 3)^[15],包括旋转、水平翻转、垂直翻转、宽度放缩、高度放缩。在测试时为了贴近真实场景,不采用数据增强。

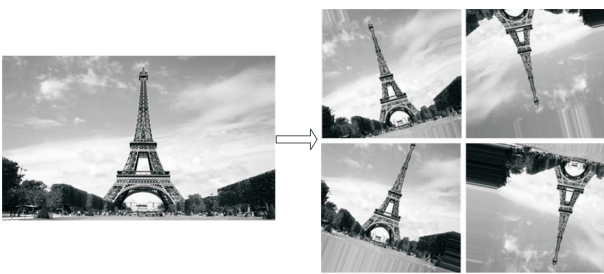


图 3 数据增强方法

在训练过程中,两个网络都采用了多类交叉熵损失函数,VGG-16 网络采用 Adam^[16] 优化算法,学习率为 0.000 01,经过 250 个 epoch 的迭代。自适应特征融合网络采用 Adam 优化算法,学习率为 0.000 01,首先冻结 VGG-16 的五个卷积块的权重,训练迭代 50 个 epoch,再解冻卷积块的权重,对整个网络训练迭代 200 个 epoch,共训练迭代 250 个 epoch。

2.2 实验平台

实验是在一台具有 Intel Core i7-4790K CPU(4.0 GHz×8)、32 GB RAM、一个 4 GB 的 Nvidia Geforce GTx970 图形处理器、64 位 Ununtu 18.04 的计算机上进行的。运行环境为 Python3.6 和 Keras 2.2.4(后端为 Tensorflow 1.12.0)。

2.3 评价指标

通过识别准确率、召回率和处理速度对模型进行评估。召回率是指对某一个类别,该类别中能够被正确识别的比例。处理速度是指每秒钟模型能够预测多少张图片。

2.4 实验结果对比

经过上述相同 epoch 轮次的训练,最终 VGG-16 网络的识别准确率为 97.86%,自适应特征融合网络的识别准确率为 98.93%,比 VGG-16 网络识别效果提高了 1.07%,对比如表 1 所示。图 4 中分析了两个网络的召回率对比,图 5 中详细分析了两个网络的结构差别。

表 1 两个模型的实验结果

模型	准确率/%	处理速度/fps
VGG-16	97.86	296
自适应特征融合网络	98.93	289

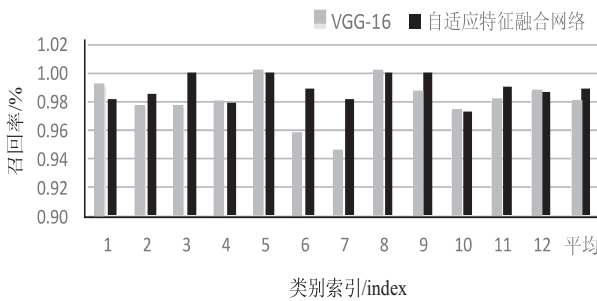


图 4 两个网络的召回率对比

根据实验结果,说明了提出的自适应特征融合网络在准确率上和召回率上都优于 VGG-16 网络,且处理速度与 VGG-16 模型大致相当。

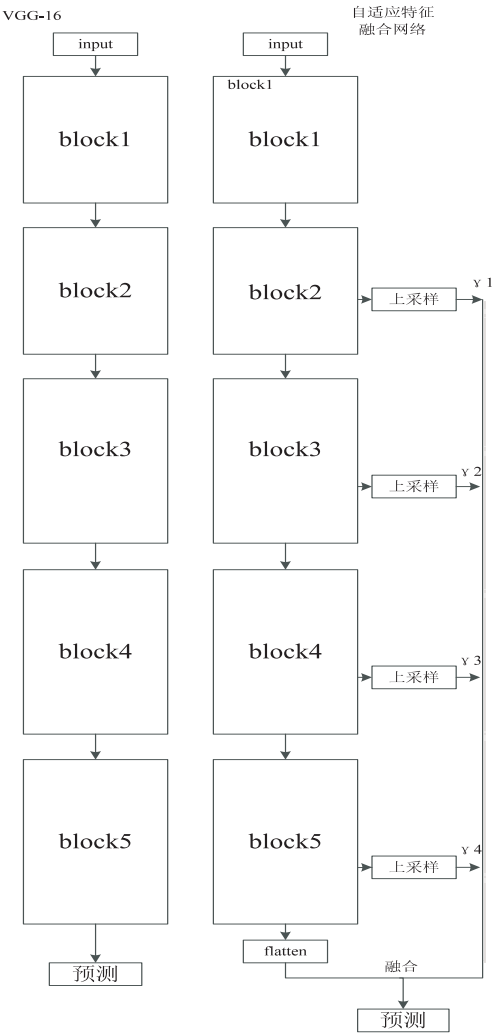


图 5 两个网络结构的详细对比

3 结束语

特定建筑物的识别在旅游景点的地点识别上具有广阔的应用前景,但是传统方法需要进行繁杂的特征设计和特征提取。对此,文中提出一种基于迁移学习的识别方法,从 VGG-16 网络迁移而来,并进行自适应的特征融合改进。实验结果表明,提出的自适应特征融合网络提高了 VGG-16 网络的性能,能很好地解决建筑物识别问题。

参考文献:

[1] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004,60(2):91-110.

[2] 王君秋,查红彬. 结合兴趣点和边缘的建筑物和物体识别方法[J]. 计算机辅助设计与图形学学报,2006,18(8):1257-1263.

[3] DALAL N,TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE computer society conference on computer vision & pattern recognition. San Diego,CA,USA: IEEE,2005:886-893.

[4] DRUCKER H,BURGES C J C,KAUFMAN L,et al. Support vector regression machines[C]//Advances in neural information processing system. Cambridge: MIT Press, 1997: 155-161.

[5] 杨松,李盛阳,邵雨阳,等. 基于改进 HOG 特征的建筑物识别方法[J]. 计算机工程与应用,2018,54(7):196-200.

[6] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Proceedings of the 25th international conference on neural information processing systems. Lake Tahoe, Nevada: Curran Associates Inc.,2012:1097-1105.

[7] OQUAB M, BOTTOU L, LAPTEV I, et al. Learning and transferring mid-level image representations using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Columbus, OH, USA: IEEE,2014:1717-1724.

[8] 李晓雨,聂秀山,崔超然,等. 基于迁移学习的图像检索算法[J]. 计算机科学,2019,46(1):73-77.

[9] 宋光慧. 基于迁移学习与深度卷积特征的图像标注方法研究[D]. 杭州:浙江大学,2017.

[10] SIMONYAN K,ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International conference on learning representation. [s. l.]: [s. n.],2015:562-575.

[11] 许庆勇,江顺亮,黄伟,等. 基于多特征融合的深度置信网络图像分类算法[J]. 计算机工程,2015,41(11):245-252.

[12] 刘晴,唐林波,赵保军,等. 基于自适应多特征融合的均值迁移红外目标跟踪[J]. 电子与信息学报,2012,34(5):1137-1141.

[13] LIN M,CHEN Q,YAN S. Network in network[C]//International conference on learning representation. [s. l.]: [s. n.],2014:786-795.

[14] NAIR V,HINTON G E. Rectified linear units improve restricted Boltzmann machines[C]//Proceedings of the 27th international conference on machine learning. Haifa, Israel: Omnipress,2010:807-814.

[15] 高友文,周本君,胡晓飞. 基于数据增强的卷积神经网络图像识别研究[J]. 计算机技术与发展,2018,28(8):62-65.

[16] KINGMA D P,BA J. Adam:a method for stochastic optimization[C]//International conference on learning representation. [s. l.]: [s. n.],2015:379-393.