

基于注意力机制和深度卷积神经网络的材质识别方法

许华杰^{1,2} 杨洋¹ 李桂兰³

1 广西大学计算机与电子信息学院 南宁 530004

2 广西多媒体通信与网络技术重点实验室 南宁 530004

3 广西壮族自治区产品质量检验研究院 南宁 530007

摘要 材质识别旨在识别自然材质图像中的主要对象及其所属材料类别。针对材质图像数据集通常数据量少、人工标注局部纹理区域困难所导致的材质识别准确率低的问题,提出了一种基于注意力机制和深度卷积神经网络的材质识别方法,该方法的核心是材质识别深度卷积神经网络(MaterialNet)。MaterialNet利用深度残差网络对图像进行特征提取,采用所提出的级联空洞空间金字塔池化的方式引入注意力机制,使网络可以通过端到端训练自适应地关注包含纹理特征的关键区域,从而有效识别材质的局部纹理特征。在FMD材质数据集上进行实验,结果表明,MaterialNet的总体识别准确率可达到82.3%,比当前主流的B-CNN和CNN+FV材质识别方法分别提高了7.2%和4.5%,对多种材质的识别准确率较高且具有参数量少、计算量小等优点。

关键词: 注意力机制;深度卷积神经网络;空洞卷积;空间金字塔池化

中图法分类号 TP391

Material Recognition Method Based on Attention Mechanism and Deep Convolutional Neural Network

XU Hua-jie^{1,2}, YANG Yang¹ and LI Gui-lan³

1 College of Computer and Electronic Information, Guangxi University, Nanning 530004, China

2 Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning 530004, China

3 Guangxi Institute of Product Quality Inspection, Nanning 530007, China

Abstract The purpose of material recognition is to identify the main objects and their material categories in natural material images. Aiming at the problem of low recognition accuracy caused by the lack of data in material image data sets and the difficulty of manually labeling local texture regions, a material recognition method based on attention mechanism and deep convolutional neural network is proposed. The core of the method is material recognition deep convolutional neural network (MaterialNet). MaterialNet uses the deep residual network to extract the features of the image, and introduces the attention mechanism by the proposed cascaded atrous spatial pyramid pooling method, so that the network can adaptively focus on the key areas containing texture features through end-to-end training, so as to effectively identify the local texture features of materials. Based on the FMD material datasets, the experimental results show that the overall identification accuracy of MaterialNet is 82.3%, which is 7.2% and 4.5% higher than the current mainstream B-CNN and CNN+FV material identification methods, respectively. The recognition accuracy of MaterialNet is high for a variety of materials, and it has the advantages of less parameters and less calculation.

Keywords Attention mechanism, Deep convolutional neural network, Atrous convolution, Spatial pyramid pooling

1 引言

材质识别是纹理图像分类研究领域的典型问题^[1]。材质识别的目标是从图像中提取主要材质对象的纹理特征,根据纹理特征判断对象所属材料类别,如木制、玻璃、塑料、金属和纤维等。材质识别在机器人、产品搜索以及室内设计的图像编辑等诸多应用中发挥了重要作用^[2]。材质识别问题的难点

在于:首先,图像拍摄的视角、光照条件以及物体表面的非刚性形变等因素可能会给物体的纹理外观带来较大的视觉影响;其次,不同材质实例的外观形态具有不同程度的相似性,也称为类间相似性,在某些相似性极高的情况下,即使人类肉眼观察图像也难以正确判断其材质^[3];此外,由于材料图像的复杂性与多样性,提取主要对象特征时易受到背景环境和周围物体的干扰,这些困难使材质识别任务极具挑战性。

到稿日期:2020-08-11 返修日期:2020-11-18

基金项目:广西壮族自治区科技计划项目(2017AB15008);崇左市科技计划项目(FB2018001)

This work was supported by the Science and Technology Plan Project of Guangxi Zhuang Autonomous Region(2017AB15008) and Science and Technology Plan Project of Chongzuo(FB2018001).

通信作者:许华杰(hjxu2009@163.com)

随着深度学习(deep learning)的发展,深度卷积神经网络(Deep Convolutional Neural Network, DCNN)凭借其强大的特征学习能力在各种视觉任务中取得了重要突破^[4-5]。在大规模图像数据集(如 ImageNet)中,训练好的 DCNN 模型可以作为通用的特征提取器,在解决材质识别问题上, CNN 特征相对于传统手工纹理特征具有更好的效果^[3,6-7]。但在 ImageNet 数据集上,训练得到的 CNN 模型更多的是关注对象整体形状,比较适合物体识别等问题,对于需要关注图像的局部纹理特征和几何不变性的材质识别任务的作用有限,相关模型的识别性能仍有很大的提升空间^[6]。研究显示,目前材质识别领域尚不存在一个类似于 ImageNet 的大规模材质图像数据集,现有的材质图像数据集大多仅包含类标签,很少给出材质纹理区域标注,这制约了深度学习模型对材质纹理的学习与分析,导致材质识别准确率较低^[1]。

针对上述问题,本文基于注意力机制提出了一种材质识别深度卷积网络 MaterialNet,在深度残差网络(ResNet-18)的基础上进行改进,通过引入注意力机制使网络能够自适应地关注局部纹理特征关键区域;提出了级联空洞空间金字塔池化(C-ASPP)作为注意力机制的引入方式,提高了网络的细节纹理特征提取能力。实验结果表明,C-ASPP 的采用使 MaterialNet 具备了在仅给予类标签的情况下学习到特征关键区域定位的能力;与当前主流材质识别方法相比,其识别准确率更高,且具有参数量少、计算量小等优点,即使在算力和存储空间受限的设备上也适用。

2 相关工作

2.1 材质识别方法

MIT 的研究者指出:与常见的物体识别和场景识别问题不同,计算机自动材质识别问题通常用纹理分类方法来解决^[8]。目前,DCNN 在纹理分类领域中的应用研究比较具有代表性的成果如下:牛津大学视觉几何研究小组 Cimpoi 等提出的 DCNN 特征与传统特征汇聚方法 Fisher Vector(FV)相结合的纹理特征提取方法^[6];Bu 等提出的基于深度卷积网络的局部性和稀疏性约束的纹理分类方法^[9];Lin 等提出的双线性卷积神经网络(B-CNN)^[10]等。目前在材质识别上效果较好的是 B-CNN 方法,其可基于材质数据集直接进行端到端训练。该方法的主要思想是使用在 ImageNet 上预训练的两个 DCNN 模型对图像进行特征提取,将两个模型输出的特征矢量的外积运算结果作为图像特征表示^[10]。该方法既能检测材质对象及关键区域,又能分别对这些区域提取局部特征,最后结合关键区域的局部纹理特征和图像全局特征信息进行分类。以文献^[6,10]中的方法为代表的材质识别方法尽管有的取得了较好的识别效果,但这些方法往往将图像表示成一个上万维甚至更高的特征矢量,具有高存储需求和高计算复杂度,在资源受限的设备上难以训练和实现。

2.2 注意力机制

注意力机制的原理是模仿人类视觉的大脑信号处理机制,即通过快速扫描全局图像,聚焦重点关注的目标区域,加大投入到该区域的注意力,以获取更多目标区域的细节信

息^[11-12]。用于计算机视觉领域的注意力机制则主要有空间注意力机制和通道注意力机制两大类。空间注意力机制将注意力放在空间的某些区域上,并通过时间以及视角的移动对信息进行融合,建立对场景的完整表示;通道注意力机制则更关注特征的通道信息,其中比较具有代表性的工作是 2016 年自动驾驶公司 Momenta 提出的 SE-Net^[13]。

注意力机制是一种弱监督的关键区域定位策略,具备学习目标定位和识别的能力,即使在只给予类标签的情况下也能较准确地从图像中获取关键位置信息^[14]。因此,本文引入注意力机制来解决材质识别问题。

3 材质识别深度卷积神经网络

本文提出的基于注意力机制的材质识别深度卷积网络(MaterialNet)包含两个分支:特征提取-分类分支和注意力分支。特征提取-分类分支利用深度卷积神经网络对材质图像进行特征提取,根据全局特征向量得到分类结果;注意力分支通过分析特征提取-分类分支学习到的特征来自适应地关注包含纹理特征的关键区域信息。MaterialNet 的网络结构如图 1 所示。

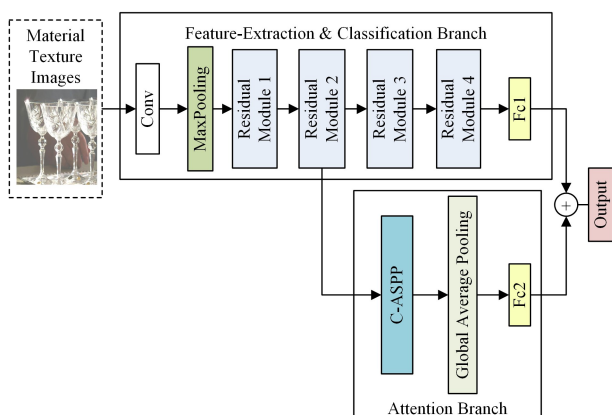


图 1 基于注意力机制的材质识别深度卷积神经网络(MaterialNet)

Fig. 1 The deep convolutional neural network for material recognition based on attention mechanism (MaterialNet)

3.1 特征提取-预测分支

特征提取-分类分支(Feature-Extraction & Classification Branch)的结构如图 1 上半部分所示,该分支使用已经在 ImageNet 数据集上预训练的 ResNet-18 模型^[15]来完成特征提取和分类。ResNet-18 中的卷积层和全连接层(带有权重的层)共计 18 层,其网络结构具体由一个卷积核大小为 7×7 的卷积层(Conv)、一个内核大小为 3×3 的池化层(最大池化 Max Pooling)、四组残差模块(residual module)和一个全连接层(Fc1)构成,每组残差模块内有两个标准残差块结构(residual block)。由于训练 MaterialNet 使用的 FMD 材质数据集含有 10 类材质,因此网络最后的全连接层包含 10 个神经元。定义这 10 个神经元的输出为标签分类向量。

3.2 级联空洞空间金字塔池化

由于材质图像的纹理特征通常有较大的类间相似性,区别往往体现在图像关键区域和细节纹理中,因此增强注意力分支的细节纹理特征学习能力对整体提升材质识别性能有十

分重要的作用。受文献[16-17]的启发,本文提出了一种级联空洞空间金字塔池化(C-ASPP)作为注意力分支的核心模块。

图2给出了C-ASPP的结构组成。从上至下看,C-ASPP分为4个分支;第一个分支包含1个 3×3 卷积层,该分支直接对输入特征进行卷积运算;第二个分支包含3个级联的 3×3 空洞卷积层,扩张率分别为2,4,2;第三个分支有3个级联的 3×3 空洞卷积层,扩张率分别为3,6,3;第四个分支有3个级联的 3×3 空洞卷积层,扩张率分别为4,8,4。

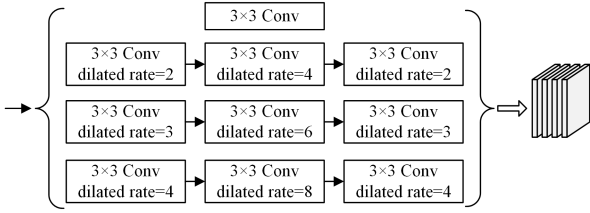


图2 级联空洞空间金字塔池化(C-ASPP)

Fig.2 Cascaded atrous spatial pyramid pooling(C-ASPP)

卷积操作中可以通过增大卷积核尺寸来增大感受野,使特征提取过程中获得的信息量更大,但同时大的卷积核也增加了网络参数数量和计算量。空洞卷积通过对 3×3 卷积设置不同的膨胀系数使其获得不同尺度的感受野,在不额外增加参数数量和计算量的情况下增大卷积核感受野,进而使网络获得不同尺度的上下文信息。对C-ASPP的4个分支设置不同的扩张率是为了获取输入特征的多尺度信息,从而提高学习能力。C-ASPP在每个分支上采取扩张率先增大后减小的策略,扩张率先逐渐增大可以使计算和采样更密集,避免因扩张率突然增大而导致局部特征信息丢失等问题;扩张率后减小可以聚合局部信息,增强细节纹理的特征提取。C-ASPP内所有卷积层均采用ReLU非线性激活函数。4个C-ASPP分支的输出特征在通道维度上进行合并,将有序的空间信息转化为相对无序的特征信息,起到了弱化空间结构的作用,从而提升特征的几何不变性。

3.3 注意力分支

注意力分支(Attention Branch)对特征提取-分类分支学习到的特征进行分析,利用注意力机制挖掘类标签和局部特征关键区域之间的相关性,通过调整通道的权重配比以达到更好的识别效果,其结构如图1下半部分所示。注意力分支由一个级联空洞空间金字塔池化(C-ASPP)模块、一个全局平均池化层(GAP)和一个全连接层(Fc2)组成,全连接层(Fc2)的输出为注意力分支的预测向量。

将特征提取-预测分支中第二个残差序列输出的特征作为注意力分支的C-ASPP模块的输入,C-ASPP模块得到的特征通过全局平均池化层将每个通道上的二维特征图压缩为一个实数,作为该通道的权重,最后经过一层全连接层将上一层的权重信息整合并作为预测向量输出。注意力分支的计算过程如式(1)所示:

$$AB = \omega(FC(GAP(CP_1(F) \oplus CP_2(F) \oplus CP_3(F) \oplus CP_4(F)))) \quad (1)$$

其中,AB代表特征图F经过注意力分支后的输出结果, ω 代表Softmax函数,FC代表全连接层操作符,GAP代表全局平

均池化操作符, \oplus 代表通道维度上的串联操作, CP_1, CP_2, CP_3, CP_4 分别代表C-ASPP模块的4个分支。由于网络本身具备学习能力,注意力分支通过学习特征在通道维度上的权重分布来获取每个通道对于图像特征的重要程度,并据此得到预测向量;通过在训练过程中不断优化通道的权重配比,来提升有效的特征通道,并抑制无效的特征通道。

3.4 网络训练

首先将材质图像输入到特征提取-分类分支的ResNet-18模型,模型中的卷积层对图像进行特征提取,最后在全连接层Fc1得到分类向量,表示为 y_{clf} 。注意力分支得到基于通道注意力机制的分类向量,表示为 y_{att} 。对两个分支的分类向量 y_{clf} 和 y_{att} 进行求和运算,得到MaterialNet最终输出的材质识别结果。使用随机梯度下降(SGD)算法最小化损失函数来优化网络权重,所采用的交叉熵损失函数的计算式为:

$$L(y, y_{clf}, y_{att}) = - \sum_{j=1}^K y_j \log(y_{clf,j} + y_{att,j}) \quad (2)$$

其中, y 为输入图像的真实标签, j 表示K类材质中的第j类。随着训练的进行,网络在两个分支的共同作用下不断进行迭代优化,从而得到最终的MaterialNet材质识别网络模型。

4 实验及其结果分析

本文开展3方面的实验以测试所提出的级联空洞空间金字塔池化C-ASPP和材质识别深度卷积网络MaterialNet的效果和性能,其中实验1的目的是测试C-ASPP的效果,实验2分析引入注意力机制对模型的影响并通过可视化的方式探究实验结果背后的原因,实验3将MaterialNet与目前效果较好的主流方法进行对比。

4.1 实验数据和参数

实验采用的数据集为FMD材质数据集(Flicker Material Database)^[18],包含织物(fabric)、植物(foliage)、玻璃(glass)、皮革(leather)、金属(metal)、纸质(paper)、塑料(plastic)、石材(stone)、水(water)、木材(wood)10个材质类别,每类材质有100个样本,样本均来自真实场景下的材料实例。实验将数据集按7:1:2的比例划分出训练集、验证集和测试集,每组实验重复3次,取3次实验结果的平均值作为最终的实验结果。借鉴相关文献的经验,训练过程的学习率、动量、权重衰减等超参数分别设为0.01、0.9和0.001,每次训练共循环20个轮次(Epoch),batch-size设为128。实验代码在Pytorch框架下实现,实验环境如下:Windows10 64操作系统,四核Intel Core i7 CPU(2.7 GHz),NVIDIA GeForce GTX1080Ti显卡(GPU),16GB内存。

4.2 实验结果及分析

实验1 为了验证所提出的级联空洞空间金字塔池化C-ASPP的有效性,用几种典型的空洞空间金字塔池化模块替换MaterialNet的注意力分支中的C-ASPP模块,在其他条件不变的情况下训练网络,对比MaterialNet在使用C-ASPP模块和其他的空洞空间金字塔池化模块下的识别性能,将识别准确率(accuracy)作为评价指标(简称识别率),实验结果如表1所列。

根据表 1 可以看出,在注意力分支上使用 C-ASPP 比使用文献[17,19-20]提出的相关空间金字塔池化结构获得的识别率更高,说明 C-ASPP 相比上述几种空间金字塔模块更适合 MaterialNet,用在 MaterialNet 中更有利于实现对 FMD 材质数据集的有效识别,原因是 C-ASPP 的多个级联空洞卷积并行分支可以更好地学习特征的多尺度信息,将特征在通道维度上合并可以弱化 CNN 特征的空间结构信息,使得 MaterialNet 更适合处理纹理图像。

表 1 使用不同空间金字塔池化方法时的识别率

Table 1 Recognition accuracies with different spatial pyramid pooling methods

CNN	Accuracy/%
SPP ^[18]	75.6
ASPP(deeplabv2) ^[20]	77.8
ASPP(deeplabv3) ^[17]	79.6
C-ASPP	82.3

实验 2 为了考察引入注意力机制带来的性能增益,实验将 MaterialNet 模型和未加入注意力分支的 ResNet-18 网络模型(以下简称 ResNet-18)所能达到的识别率在相同的实验条件下进行对比。

首先,考察本文方法对于各种类型材质所体现出的识别效果。表 2 列出了两种模型对于 FMD 数据集所包含的 10 类材质所能实现的识别率。由表 2 可知,加入了注意力分支的 MaterialNet 对于 10 类材质的识别率均不低于 ResNet-18,且表中最后一行所示的总体识别率由后者的 72.2% 提高到了 82.3%,说明引入注意力机制能够有效地提升识别率。

表 2 两种模型在 FMD 数据集上的识别率

Table 2 Accuracies of two models on the FMD dataset

Category	Accuracy/%	
	ResNet-18	MaterialNet
Fabric	88.9	88.9
Foliage	83.3	88.9
Glass	77.8	94.4
Leather	83.3	83.3
Metal	33.3	61.1
Paper	77.8	83.3
Plastic	72.2	94.4
Stone	66.7	72.2
Water	72.2	72.2
Wood	66.7	66.7
Overall	72.2	82.3

为了进一步测试加入注意力分支的 MaterialNet 模型的性能,对其进行过拟合分析,ResNet-18 和 MaterialNet 模型训练过程的准确率变化如图 3 所示。

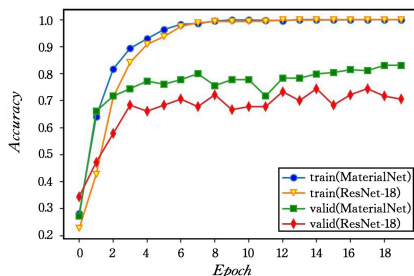


图 3 两个模型的准确率变化曲线

Fig. 3 Accuracy changes of two models

由图 3 可见,MaterialNet 模型的训练准确率在前期不断上升,约 8 个 Epoch 后趋于平稳,验证准确率整体也呈现先上升后平稳的趋势,经过 20 个 Epoch 后准确率达到 0.8 左右,未出现明显的过拟合的情况。ResNet-18 模型相比 MaterialNet 模型,验证准确率较低,在训练过程的后期有略微下降的趋势,出现了轻微的过拟合现象。导致该结果的一个原因是注意力分支的加入使 MaterialNet 模型能够更好地适用于材质数据集识别任务,减小了发生过拟合的概率;另一个原因是 MaterialNet 是在 ImageNet 数据集预训练的基础上再使用 FMD 数据集进行训练的,因此用于模型训练的整体数据量实际上并不少,而且训练只进行了 20 个 Epoch,在训练结果趋于稳定时就及时终止训练,避免了过拟合情况的发生。

其次,考察注意力机制用在 MaterialNet 中提升识别率的原因和机理。类激活映射(Class Activation Mapping, CAM)^[21]是 Zhou 等提出的 CNN 特征可视化方式,CAM 以类似热图的方式对图像中的不同区域在网络识别过程中的重要程度进行标示,通过 CAM 可以直观了解网络更关注哪些图像区域。表 3 列出了两种网络对 6 种材质图像的 CAM(限于篇幅,只列出了 10 种材质中视觉效果较为明显的 6 种),表 3 中第二列是待识别的原始材质图像,第三列是 ResNet-18 的 CAM 效果,第四列是 MaterialNet 的 CAM 效果。

表 3 两种网络模型对 6 类材质图像的 CAM 效果

Table 3 CAMs of 6 types of material images in two network models

Category	Input Image	ResNet-18	MaterialNet
Fabric			
Foliage			
Glass			
Paper			
Stone			
Wood			

通过表 3 中 CAM 的对比可以看出,相比 ResNet-18,MaterialNet 能更准确地捕捉到材质识别所需的纹理信息、对于识别决策有重大影响的关键区域(CAM 中的热点区域),这就解释了 MaterialNet 比 ResNet-18 能实现更高材质识别率的原因。

实验 3 将 MaterialNet 与文献[6]和文献[10]中提出的方法进行对比,所选取的对比方法都是目前主流的、效果较好的基于 CNN 的材质识别方法。分别从识别率(accuracy)、参数量(params)和计算量(FLOPs)这 3 个方面评价几种

材质识别网络模型的性能表现。FLOPs 反映了实际运算中加法和乘法的次数,Params 反映了模型的大小,两者主要针对卷积层和全连接层而言。对于卷积层,FLOPs 和 Params 之间有如下关系:

$$FLOPs = Params \times H \times W \quad (3)$$

其中, H 代表输出特征图的高, W 代表输出特征图的宽;而对于全连接层,FLOPs 数目等于该层的参数量 Params。实验结果如表 4 所列。

表 4 不同模型在 FMD 数据集上的测试结果

Table 4 Comparative experiment results with different models on the FMD dataset

Model	Accuracy/%	Params	FLOPs
VGG-D+FV ^[6]	75.1	—	—
B-CNN(VGG-M) ^[10]	74.1	1.1×10^9	7.51×10^9
B-CNN(VGG-D) ^[10]	77.8	1.11×10^9	15.38×10^9
B-CNN(ResNet-18) ^[10]	74.2	2.5×10^7	1.82×10^9
MaterialNet	82.3	1.3×10^7	2.43×10^9

表 4 中的第 1 行对应于文献[6]提出的 CNN(VGG-D)+FV 方法实现的模型,该方法利用 Fisher 向量对 CNN 特征进行全局纹理图像特征表示,以增强特征的几何不变性。由于文献[6]中并未给出方法参数量和计算量方面的相关信息,因此实验中不考虑其所对应的这两个量。表 4 中第 2—4 行分别对应于采用了 VGG-M, VGG-D 和 ResNet-18 网络的 B-CNN 方法^[10]实现的模型,该方法以端到端的方式训练两个相同的 CNN 网络,通过相互协调作用来完成物体对象或局部区域的检测与特征提取。对于 CNN,一般来说,参数量主要集中在全连接层,计算量主要集中在卷积层。以 VGG 网络为基础的 B-CNN(VGG-M)和 B-CNN(VGG-D)中含有多个全连接层且神经元个数较多,而以 ResNet-18 网络为基础的 B-CNN(ResNet-18)和 MaterialNet 中全连接层数少且神经元个数较少,因此以前一种网络为基础的模型的参数量远大于后一种;其次,由于上述两种网络的卷积层数量相差不大,因此它们的计算量规模基本上处于同一数量级。这就是导致表 4 中以 VGG 为基础的网络模型的参数量和计算量之间数量级相同而以 ResNet-18 为基础的网络模型的参数量和计算量之间相差两个数量级的原因。

从识别准确率看,本文提出的 MaterialNet 取得了最高的精度,达到了 82.3%,比位居其后的 B-CNN(VGG-D)和 VGG-D+FV 两种模型分别高出 4.5%和 7.2%;MaterialNet 的参数量为 1.3×10^7 ,是几种对比模型中最低的;MaterialNet 的计算量为 2.43×10^9 FLOPs,在几种对比模型中相对较低。实验结果说明,MaterialNet 可以有效提升材质识别率,同时其具有较低的参数量,所需的计算量也较少。导致这一结果的原因有两个方面:首先,MaterialNet 根据特征提取-分类分支和注意力分支两个分支的输出结果和相关损失函数计算公式,通过端到端训练优化网络权重,根据输入图像自适应地学习图像特征;其次,注意力机制的引入有助于网络更有效地识别材质图像中包含纹理特征的关键区域,从而可以使深度适中的 ResNet-18 获得以往需要更深层网络才能达到的识别率,显著降低了所需的参数量和计算量。

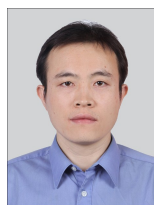
结束语 针对材质图像数据集通常数据量少、人工标注

局部纹理区域困难所导致的材质识别准确率低的问题,提出了一种基于注意力机制和深度卷积神经网络的材质识别方法。该方法以基于注意力机制的材质识别深度卷积神经网络 MaterialNet 为核心,利用深度残差网络对材质图像进行特征提取;提出级联空洞空间金字塔池化结构 C-ASPP,通过学习特征的多尺度信息提升其几何不变性,并以该结构为核心在网络中引入注意力机制,实现弱监督的材质纹理关键区域定位。在 FMD 材质数据集上的实验结果表明,所提方法达到的材质识别率为 82.3%,高于目前主流的材质识别方法,同时该方法具有参数量少、计算量小等优点。在实际应用中,本文方法在红木等木材识别方面的应用也获得了较好的效果,相关结果将在后续成果中加以展示。提高方法的鲁棒性以适应复杂场景下的多标签材质识别是下一步的研究方向。

参考文献

- [1] LIU L, ZHAO L J, GUO C Y. Texture Classification: State-of-the-art Methods and Prospects[J]. Acta Automatica Sinica, 2018, 44(4): 584-60.
- [2] BELL S, UPCHURCH P, SNAVELY N, et al. Material Recognition in the Wild with the Materials in Context Database[C]// The 2015 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Boston, MA, USA, 2015(1): 3479-3487.
- [3] CIMPOI M, MAJI S, KOKKINOS I, et al. Describing textures in the wild[C]// The 2014 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Columbus, OH, USA: IEEE, 2014: 3606-3613.
- [4] DENG R, LIN J C, YANG H Z. Building Identification Based on Deep Learning[J]. Journal of Chongqing Technology and Business University(Natural Science Edition), 2019, 36(4): 17-22.
- [5] YANG W G, HUAI Y T. Flower Image Enhancement and Classification Based on Deep Convolution Generative Adversarial Network[J]. Computer Science, 2020, 47(6): 176-179.
- [6] CIMPOI M, MAJI S, VEDALDI A. Deep filter banks for texture recognition and segmentation[C]// The 2015 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Boston, Massachusetts, USA: IEEE, 2015: 3828-3836.
- [7] LIU L, CHEN J, PIEGUTH P, et al. From BoW to CNN: Two decades of texture representation for texture classification[C]// Proceedings of International Journal of Computer Vision, 2019(127): 74-109.
- [8] SHARAN L, ROSENHOLTZ R, ADELSON E H. Accuracy and speed of material categorization in real-world images[J]. Journal of Vision, 2014, 14(9): 1-24.
- [9] BU X Y, WU Y W, GAO Z, et al. Deep convolutional network with locality and sparsity constrains for texture classification[J]. Pattern Recognition, 2019(91): 34-46.
- [10] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear CNN models for fine-grained visual recognition[C]// The 2015 IEEE International Conference on Computer Vision(ICCV). Santiago, Chile: IEEE, 2015: 1449-1457.
- [11] XU K, BA J, KIROS R, et al. Show, Attend and Tell: Neural

- Image Caption Generation with Visual Attention[C]//International Conference on Machine Learning (ICML). PMLR, 2015: 2048-2057.
- [12] LIU Y, JIN Z. Fine-grained Image Recognition Method Combining with Non-local and Multi-region Attention Mechanism[J]. Computer Science, 2021, 48(1): 197-203.
- [13] HU J, LI S, GANG S. Squeeze-and-Excitation Networks[C]//The 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA, 2018: 7132-7141.
- [14] BA J, MNH V, KAVUKCUOGLU K. Multiple Object Recognition with Visual Attention[EB/OL]. <https://arxiv.org/pdf/1412.7755.pdf>.
- [15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//The International Conference of Computer Vision and Pattern Recognition (CVPR). 2016:770-778.
- [16] LI X, WANG W, HU X, et al. Selective kernel networks[C]//The international Conference of Computer Vision and Pattern Recognition (CVPR). 2019:510-519.
- [17] CHEN L, PAPANDEOU G, SCHROFF F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation [EB/OL]. <https://arxiv.org/pdf/1706.05587.pdf>.
- [18] SHANRAN L, LIU C, ROSENHOLTZ R, et al. Recognizing materials using perceptually inspired features[J]. International Journal of Computer Vision, 2013, 103(3): 348-371.
- [19] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [20] CHEN L C, PAPANDEOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 40(4): 838-848.
- [21] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discrimination localization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:2921-2929.



XU Hua-jie, born in 1974, Ph.D, associate professor, is a senior member of China Computer Federation. His main research interests include artificial intelligence, acoustic signal recognition and computer vision.