基于深度学习的纹理图像语义识别

王军敏,张诗梦
(平顶山学院 信息工程学院,河南 平顶山 467036)

摘　　要:传统的纹理图像识别方法主要对底层的视觉特征进行检测,从而导致对高层语义特征的识别能力不足. 为了解决上述问题,提出一种基于深度学习的纹理图像语义特征识别方法. 该方法首先把在 ImageNet 数据集上预先训练的模型作为骨干网络构建模型,然后对输入的纹理图像进行数据增强和预处理,并输入模型进行训练和验证,最后利用训练获得的最佳模型对测试图像进行语义识别. 在语义描述纹理库(Describable Texture Dataset, DTD)上进行仿真实验,结果表明该方法在纹理图像的语义识别方面具有良好的性能.

关　键　词:纹理图像识别;深度学习;语义识别;特征提取
中图分类号:TP391　　**文献标识码:**A　　**文章编号:**1673 - 1670(2022)02 - 0040 - 04

0　引言

纹理图像识别技术具有广泛的应用领域和重要的应用价值,例如可应用于生物特征识别[1]、遥感图像分类[2]、工业产品的瑕疵检测[3]等. 传统的纹理图像识别方法主要是对纹理图像的底层视觉特征进行提取和识别,例如:灰度共生矩阵(Gray Level Co-occurrence Matrix, GLCM)算法[4]主要检测不同像素对出现的统计特征,局部二值模式(Local Binary Pattern, LBP)算法[5]和完备的局部二值计数(Completed Local Binary Count, CLBC)算法[6]主要检测局部邻域像素与中心像素的灰度关联信息. 语义特征是图像表达出的接近人类认知和理解的概念,是一种更高级和抽象的特征,对提高图像的高级表达和理解具有重要作用. 由于传统方法对纹理图像语义特征的提取和识别能力不足,从而导致这些方法在纹理图像语义场景下的识别性能较差. 考虑到深度学习模型能够模拟人脑分层提取特征的机制,能够提取从底层到高层、从具体到抽象的图像特征,因此提出一种基于深度学习的纹理图像语义识别方法,以便更好地提取和识别纹理图像的语义特征.

1　本文方法

1.1　图像数据集的准备

DTD 纹理库[7]是对现实场景中的纹理样本进行拍摄获取的彩色纹理图像数据集,一共包含 47 个类别,每类均有 120 个样本图像. 这些样本图像分别用可描述的语义属性作为类别标记,例如条带状(banded)、辫子状(braided)、网格状(grid)、Z 字形(zigzagged)、螺旋状(spiralled)等. 图 1 展示了 DTD 纹理库中的一部分样本图像及其对应的语义属性标签.

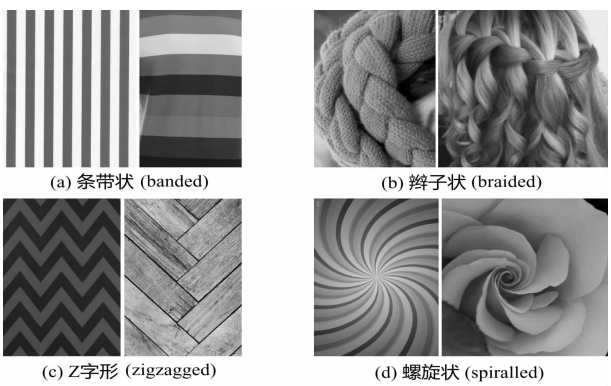

图 1　DTD 纹理库的部分样本图像及其语义属性标签

从图 1 可以看出,DTD 纹理库中的样本图像含有丰富的纹理特征,即便是同类别的样本图像也呈现出不同的底层视觉特征,因此不方便利用底层视觉特征进行统一标记,采用语义属性则能很好地描述这些样本图像的类别,并且符合人类的认知和描述习惯.

DTD 纹理库的训练集、验证集和测试集的划分如下:由于 DTD 纹理库的每类含有 120 个样本,

收稿日期:2021 - 09 - 30
基金项目:河南省科技厅科技攻关项目(202102210331);平顶山学院博士科研启动基金项目(PXY-BSQD-202004)
作者简介:王军敏(1982—　),男,河南省平顶山市人,工学博士,平顶山学院信息工程学院讲师,主要从事图像识别、机器学习研究.

所以从每类选取 60 个样本用于训练,20 个样本用于验证,剩下的 40 个样本用于测试,从而构建本文的训练集、验证集和测试集.

1.2　深度学习模型的构建

深度卷积神经网络(Convolutional Neural Network, CNN)是深度学习技术的一个重要分支,目前在图像和语音识别领域获得了广泛的应用. 但是,从头开始训练一个深度 CNN 模型需要大规模的训练样本,并且对硬件平台的计算能力要求很高,因此为了降低对训练样本数量和计算平台性能的要求,利用在大规模图像数据集 ImageNet 上预先训练好的深度 CNN 模型作为本文模型的骨干网络. 由于这些预训练模型的参数已经在 ImageNet 图像数据集上经过了充分的预训练,已经具备了初步的、较好的特征提取能力,所以可在预训练模型的基础上添加所需的后端模块并进行微调训练,从而方便快捷地构建针对不同应用场景的深度学习模型.

笔者构建的深度学习模型包括骨干网络和后端模块两部分,其基本结构如图 2 所示.

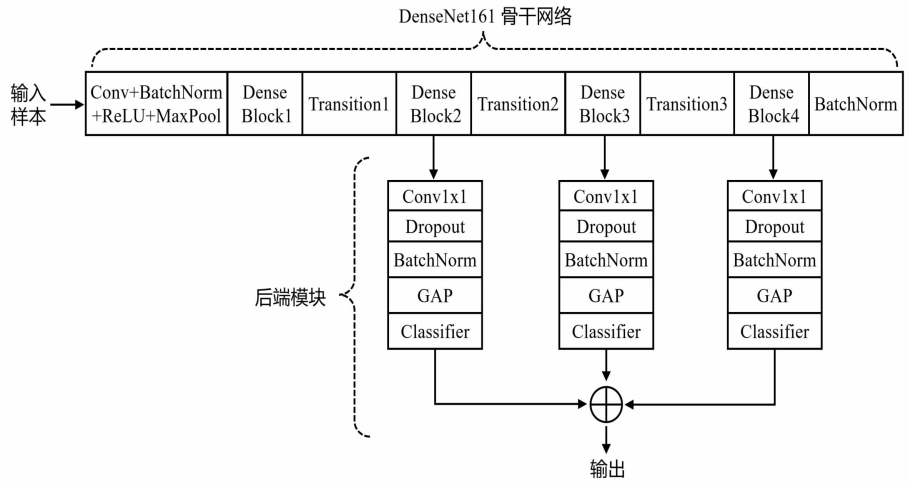

图 2　本文深度学习模型的基本结构

1)骨干网络的主要作用是作为特征提取器,即从输入纹理图像中提取不同层次的纹理特征. 笔者采用在 ImageNet 数据集上预先训练好的 DenseNet161 模型[8]作为骨干网络,DenseNet161 模型的结构如表 1 所示,主要利用 4 个 DenseBlock 模块提取特征. 因此,为了提取不同层次的语义特征,本文方法分别从 DenseNet161 模型的 DenseBlock2、DenseBlock3、DenseBlock4 的输出端提取低、中、高三个层次的语义特征,以便送入后端模块做进一步处理.

表 1　DenseNet161 模型的基本结构

模块序号	模块名称
1	Conv2d + BatchNorm2d + ReLU + MaxPool2d
2	DenseBlock1 + Transition1
3	DenseBlock2 + Transition2
4	DenseBlock3 + Transition3
5	DenseBlock4 + BatchNorm2d
6	Classifier

2)后端模块的主要作用是对骨干网络提取的特征进行降维、抑制过拟合,并对骨干网络提取的不同层次的语义特征进行特征融合,作为最终的分类输出. 其中:Conv1x1 卷积模块和全局平均池化(Global Average Pooling, GAP)层可以显著降低模型的参数规模;Dropout 层在训练阶段随机丢弃一些神经元之间的连接,相当于每次仅训练一个子网络,可有效抑制模型的过拟合;BatchNorm 层对批样本的特征进行归一化,能够加快训练速度,有效防止梯度消失和梯度爆炸,并且能够有效抑制过拟合;分类器 Classifier 使用一个全连接层将输出端的个数调整为待识别纹理图像的目标类别个数,并计算分类得分. 最后,将多个分类器的输出进行特征融合,作为模型最终的预测结果,根据预测值的大小即可判断当前输入样本所属类别的标签.

1.3　深度学习模型的训练和验证

在训练阶段,把训练样本集送入模型进行充分训练,不断优化调节模型的参数,可使深度学习模型获得更好的识别能力. 为了获得良好的训练效果,对训练阶段的超参数进行如下设置.

1)数据增强与归一化. 为了增加训练样本的数量,对现有的训练样本进行数据增强操作,具体包括:先对原始图像进行随机缩放并截取 224 × 224 像素的中心区域图像,进行水平方向的随机翻折;同时,由于预训练的 DenseNet161 模型是在 Im-

ageNet 数据集上训练的,要求输入图像归一化到该数据集的均值[0.485, 0.456, 0.406]和标准差[0.229, 0.224, 0.225],所以,为了满足预训练模型对输入的要求,把当前输入图像的 RGB 三个颜色通道分别归一化为均值[0.485, 0.456, 0.406],标准差为[0.229, 0.224, 0.225].

2)批尺寸. 如果一次把全部训练样本送入模型进行训练,容易产生计算设备内存不足的问题,所以采取把训练样本分批送入模型进行训练的策略,设置批尺寸为 40,即每次把 40 个训练样本送入模型进行训练,以便获得合适的梯度估计和参数调节效果.

3)冻结参数和可训练参数. 笔者构建的深度学习模型一共有 498 个命名参数,其中前 480 个命名参数属于骨干网络,后面 18 个命名参数属于后端模块. 因此,笔者冻结预训练模型的参数,只对新增的后端模块的参数进行微调训练.

4)学习率. 该参数控制模型寻找最优解的速度,如果学习率过大则可能错过最优解,如果学习率过小则会导致训练时间过长或限于局部最优解. 因此,为了获得良好的训练效果,根据实践经验,设置初始的学习率为 0.000 1.

5)丢弃概率. 由于笔者所用的 DTD 数据集不属于大规模数据集,容易导致模型训练的过拟合,所以为了有效抑制模型的过拟合现象,设置丢弃概率 $P=0.7$.

6)优化器. 选择综合性能较好的 Adam 算法优化器.

7)损失函数. 由于笔者研究多类别分类问题,所以本文模型的损失函数采用分类交叉熵函数.

8)训练轮数. 设最大训练轮数为 50 轮.

在对模型进行训练之后,为了检验模型的识别能力并选择性能最佳的模型,需要对模型进行验证,即把验证数据集送入训练后的模型,获得验证精度. 如果验证精度不够理想,需要继续对模型进行微调训练,最后保存验证精度最高的模型,将其作为最佳的训练模型.

1.4 深度学习模型的测试

在测试阶段,首先将测试集的样本进行归一化,即把样本图像的 3 个颜色通道分别归一化为均值[0.485, 0.456, 0.406],标准差为[0.229, 0.224, 0.225],然后把测试样本分批送入验证阶段保存的最佳模型,最后获得预测结果,并计算测试精度. 在测试阶段,不需要对测试的样本图像进行数据增强操作.

2 实验与结果分析

实验环境:计算机(Intel Core i5CPU@3.10 GHz, 16 GB RAM), Python 3.7, PyTorch 1.8.1.

2.1 实验步骤

本文方法的具体实验步骤如下:

Step 1 利用 DenseNet161 骨干网络和后端模块构建深度学习模型;

Step 2 对样本图像进行归一化,即把样本图像的 RGB 3 个颜色通道调整为均值[0.485, 0.456, 0.406],标准差为[0.229, 0.224, 0.225],并调整图像尺寸为 224×224 像素;

Step 3 对训练样本进行数据增强,扩充训练样本的数量,并设置模型的超参数值,包括优化器、学习率、批尺寸、损失函数等;

Step 4 设置模型的冻结参数和可训练参数,把数据增强后的训练集样本分批送入本文模型,对模型的可训练参数进行微调和优化;

Step 5 把验证集的样本分批送入模型进行验证,当模型在验证集上获得最佳验证精度时,保存此时的模型参数,即获得最佳的深度学习模型;

Step 6 利用训练和验证所得的最佳深度学习模型对测试样本进行测试,获得测试集的识别结果.

2.2 实验结果分析

图 3 展示了本文方法在训练和验证阶段的识别精度变化曲线. 从图 3 可以看出,在经过 5 个 Epoch 之后,本文方法的训练和验证精度都已超过 65%,体现出本文方法具有良好的学习能力和训练效果.

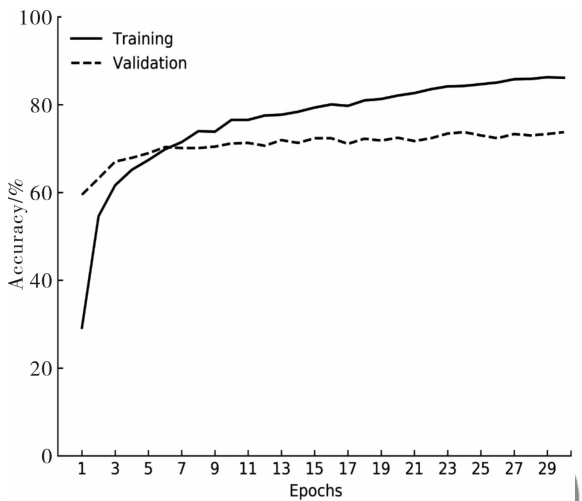

图 3 训练和验证精度曲线

为了进一步评估本文方法的性能,将本文方法与其他先进算法在 DTD 纹理库测试集上的识别精

度进行对比,结果如表 2 所示.

表 2　在 DTD 纹理库测试集上的识别精度

方法	识别精度/%
GLCM[4]	11.57
LBP[5]	29.08
CLBC[6]	34.60
VGG16[9]	64.68
DenseNet161[8]	69.84
本文方法	73.36

从表 2 可以看出:1)传统的底层视觉特征提取算法(GLCM 算法、LBP 算法和 CLBC 算法)的识别精度较低,这表明图像的底层视觉特征不能准确地描述其语义特征,导致该类算法对纹理图像语义特征的识别能力较弱;2)基于深度学习的算法(VGG16 算法、DenseNet161 算法和本文算法)的识别精度显著超过了传统底层视觉特征算法,表明深度学习方法具有良好的语义特征学习和识别能力;3)本文方法在 DTD 纹理库上取得了最高的识别精度 73.36%,超过了其他算法的识别精度,表明本文方法能够更好地提取纹理图像的语义特征,在描述和鉴别纹理图像的语义特征方面具有优越的性能,因此本文方法在纹理图像语义识别的场景中具有更好的应用价值.

3　结论

为了实现纹理图像语义特征的识别,利用在 ImageNet 数据集上预先训练的 DenseNet 模型为骨干网络,在此基础上增加后端模块构建深度学习模型,并通过对后端模块的微调训练获得最佳的深度学习模型. 该方法降低了深度学习模型构建和训练的难度. 同时,在 DTD 纹理库上的实验结果表明,本文方法在纹理图像的语义特征提取和识别方面具有良好的性能,弥补了传统底层视觉算法对纹理图像语义特征识别能力不足的问题,具有重要的应用价值.

参考文献:

[1]胡敏, 滕文娣, 王晓华, 等. 融合局部纹理和形状特征的人脸表情识别[J]. 电子与信息学报, 2018(6): 1338 – 1344.

[2]邱中原,朱希安,郁建林. 结合纹理特征和边缘信息的遥感影像分类[J]. 遥感信息,2018(3):130 – 135.

[3]MEI S, YANG H, YIN Z P. An unsupervised-learning-based approach for automated defect inspection on textured surfaces[J]. IEEE Transactions on Instrumentation and Measurement, 2018, 67 (6):1266 – 1277.

[4]陈焕栩, 解浩, 张建文, 等. 基于灰度共生矩阵纹理特征的局部放电模式识别[J]. 电力系统保护与控制, 2018(5):25 – 30.

[5]王昊鹏, 李慧. 基于局部二值模式和灰度共生矩阵的籽棉杂质分类识别[J]. 农业工程学报, 2015(3): 236 – 241.

[6]ZHAO Y, HUANG D S, JIA W. Completed local binary count for rotation invariant texture classification[J]. IEEE Transactions on Image Processing, 2012, 21(10):4492 – 4497.

[7]CIMPOI M, MAJI S, VEDALDI A. Deep filter banks for texture recognition and segmentation[C]. Boston: IEEE Conference on Computer Vision and Pattern Recognition, 2015:3828 – 3836.

[8]HUANG G, LIU Z, van der MAATEN L, et al. Densely connected convolutional networks[C]. Honolulu: IEEE Conference on Computer Vision and Pattern Recognition, 2017:4700 – 4708.

[9]SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]. Toronto: International Conference of Legal Regulators, 2015.

(责任编辑:王彦江)

Semantic Recognition of Texture Image Based on Deep Learning

WANG Junmin, ZHANG Shimeng

(School of Information Engineering, Pingdingshan University, Pingdingshan, Henan 467036, China)

Abstract: Traditional texture image recognition methods mainly detect the low-level visual features, resulting in inadequate recognition of high-level semantic features. To solve the problem, this paper proposes a method for recognizing the semantic feature of texture image based on deep learning. Firstly, the model pre-trained on the ImageNet dataset as the backbone network is employed to build our model. Secondly, the data augmentation and preprocessing of the input texture images are implemented, and then the training and validation dataset are sent to our model for training and validation. Finally, the best model obtained by training is used for semantic recognition of test images. The experimental results on the semantic Describable Texture Dataset (DTD) show that the proposed method has good performance in the semantic recognition of texture images.

Key words: texture image recognition; deep learning; semantic recognition; feature extraction