

基于改进 PSPNet 网络的古代壁画分割方法

曹建芳^{1,2}, 田晓东¹, 贾一鸣¹, 闫敏敏¹, 马尚¹

(1.太原科技大学 计算机科学与技术学院,太原 030024;2.忻州师范学院 计算机系,山西 忻州 034000)

摘要:针对传统方法在古代壁画图像分割过程中出现的目标边界模糊、图像分割效率低等问题,提出一种基于 PSPNet 网络的多分类壁画图像分割模型(PSP-M).模型首先融合轻量级神经网络 MobileNetV2,降低硬件条件对于模型训练的限制.其次通过全局金字塔模块,将不同级别的特征图拼接起来,避免了表征不同子区域之间关系的语境信息的丢失.最后利用金字塔场景解析网络嵌入壁画背景特征,减少特征损失的同时提高特征提取效率.实验结果表明,PSP-M 模型较传统的图像分割模型在训练精确度上平均提升 2%,峰值信噪比(PSNR)较实验对比模型平均提高 1~2 dB,结构相似指标(SSIM)指标较实验对比模型平均提高 0.1~0.2,实验验证了 PSP-M 模型在壁画分割方面的可行性.

关键词:壁画图像分割;金字塔池模块;深度可分离卷积;轻量级神经网络

中图分类号:TP391.41

文献标志码:A

随着计算机软硬件的不断发展,社会信息化程度逐步加深,各文物保护单位引入先进数字化技术,对传统技术进行迭代更新,提高相关工作者工作效率,减少工作量,提升了传统文化的价值和影响力.古代壁画作为一种承载传统文化的载体,在中国文化体系中有举足轻重的地位.然而,由于年代久远,以敦煌壁画和辽阳汉墓壁画为代表的中国传统壁画受到人为和自然环境的侵害,多数壁画面临内容缺损、颜色缺失等一系列问题.壁画图像数字保护的第一步是对图像进行理解,受中国传统文化的影响,古代壁画色彩鲜明,内容丰富,如何进行图像分析是壁画图像保护工作中的难题之一.

图像分割是图像理解的重要方法之一.该方法利用图像区域边界灰度的不连续性,对所划分的不同区域进行标定,以此达到分析图像的目的.当前,古代壁画分割还停留在使用传统分割方法阶段,较少涉及深度学习领域,所使用方法多用于处理灰度图像,不适用于色彩丰富的古代壁画.常见方法有以下几种,方法一是模糊 C 均值(Fuzzy C-mean,FCM)^[1-2],该算法应用范围广,已形成成熟的理论体系,但在古代壁画分割方向上,FCM 算法没有考虑空间信息,对噪声和灰度不均匀敏感,且算法会受样本不平衡的影响,分割样本与目标样本存在差异性.方法二是 K-means^[3-4],算法缺点较为明显,其中 K 值的选取需要人为给予,受主观因素影响,另外在聚类结束之前,需要经过不断迭代,而最终得到的结果只是局部最优,全局分割效果较差.方法三是 Graph Cuts^[5-6]及其改进算法 Grab Cut^[7],两种算法分别采用一次性最小化和迭代最小化使得目标和背景建模的灰度直方图和高斯混合模型(Gaussian Mixed Model,GMM)^[8]的参数更优,从而使其达到良好的分割效果,但算法分割效果会受图像复杂度和指定像素点的影响,效果较差.

深度学习领域利用神经网络系统组合图像的低层特征形成抽象的高层特征,以此表示图像元素的属性类别,发现数据的分布式特征表示,提高图像分割或分类预测的简易程度.图像分割领域常见的分割模型有全卷积神经网络(Fully Convolutional Networks,FCN)^[9],基于 FCN 改进的 SegNet(Segment Networks)^[10]网络,引入空洞卷积的金字塔场景解析网络(Pyramid Scene Parsing Network,PSPNet)^[11]和

收稿日期:2021-02-09;修回日期:2021-05-07.

基金项目:教育部人文社会科学研究规划基金(21YJAZH002);山西省高等学校人文社会科学重点研究基地项目(20190130).

作者简介(通信作者):曹建芳(1976-),女,山西忻州人,太原科技大学教授,博士,CCF 高级会员(25770S),主要研究方向为数字图像理解、大数据技术,E-mail:kcxjdj122@126.com.

DeepLab^[12-13]系列网络.根据古代壁画复杂的构图方式,利用深度学习网络的强大学习能力,本文将卷积神经网络 MobileNetV2^[14]融入 PSPNet 模型中的空间金字塔池模块,引入 Dice 损失系数^[15],形成一种新的适用于古代壁画图像分割的轻量级金字塔场景解析网络(Pyramid Scene Parsing MobileNetV2 Network, PSP-M)模型.PSP-M 网络中的金字塔池结构将不同特征平滑的连接在一个全连接层上,提取图像不同区域的信息,降低样本不平衡对壁画分割结果的影响,有效解决了 FCM 算法的弊端;模型利用卷积神经网络对壁画图像特征进行提取,消除了 K-means 算法和 Graph Cut 算法中人为因素对实验结果的影响,且 PSPNet 网络具有层次全局优先级,包含不同子区域的不同尺度信息,在分割效果上优于 K-means 和 Grab Cut.

实验从模型运行时间、模型分割精确度等多个角度证明 PSP-M 模型在古代壁画图像分割方向上的可行性和有效性,分割图像边缘信息保存相对完整.

1 相关理论

1.1 传统 PSPNet 网络模型

语义分割的目的是将图像中的每一个像素指定一个类别标签,从而提供对图像表达意义的完整理解,而像素元素位置、类别和形状的分置归类有利于计算机对图形对象的理解.常见的图像分析框架一般是基于 FCN,该类框架的弊端之一是在不同图像元素具有相似特征的时候会发生像素错误归类的现象.导致这种问题的主要原因是基于 FCN 的网络模型没有合适的策略来利用全局场景类别线索.为了获得全局图像级别特征,PSPNet 网络使用了空间金字塔池技术,利用空洞卷积,收集空间统计数据为图像全局提供信息解析,并将其命名为全局金字塔模块(Pyramid Pooling Module, PPM)^[11].此模块将图像局部信息和全局信息相结合,使图像分割结果更加准确,PPM 模型的原理如图 1 所示.

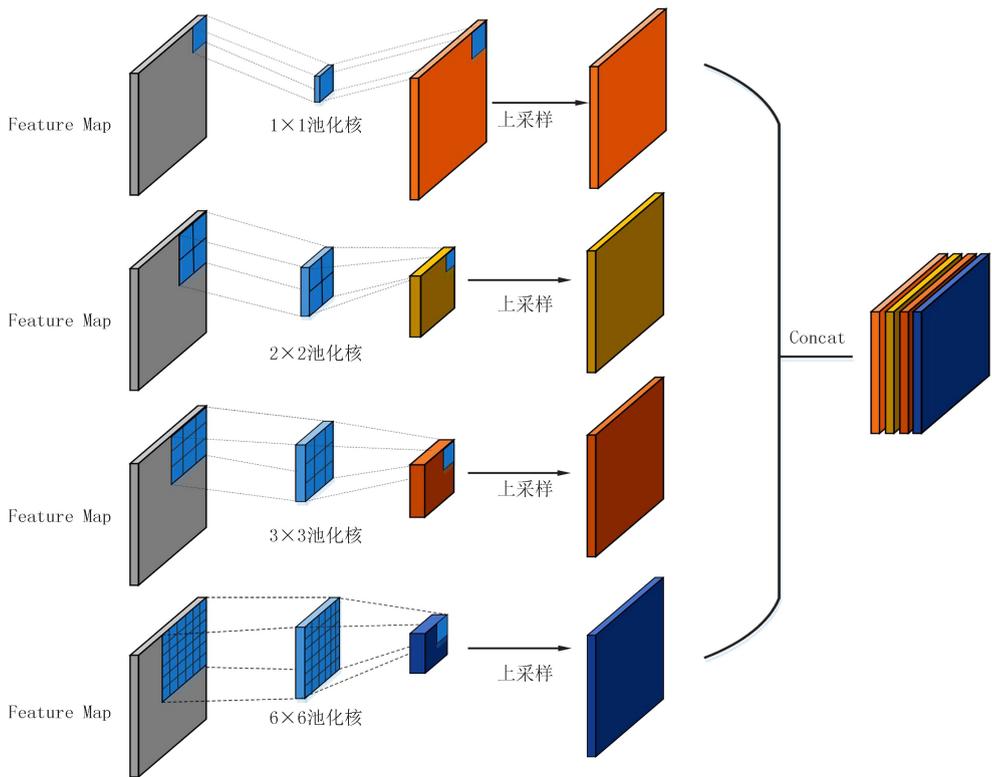


图1 全局金字塔模块原理图

Fig.1 Schematic diagram of pyramid pooling module

图 1 表示的是全局金字塔模块,是 PSPNet 网络的主要部分,该模块融合 4 种不同尺度的特征,是一种典型的四层级模块.通过 1×1 池化核的部分为粗略层级,使用全局池化生成单个 bin 输出;另外几部分将特征图划分成若干个不同的子区域,各个层池化核的大小不一,减小了模型的计算量并增大感受野.PSPNet 网

络金字塔池化模块中的不同层级输出不同尺度的特征图,由于不同层级维数不同,为保持全局特征的权重,每个金字塔层级后添加 1×1 卷积核,若某个层级维数为 n 时,通过卷积核,可将语境特征的维数降到原始特征的 $1/n$;之后通过双线性插值方法对低维特征进行上采样,恢复原始特征图尺度大小,便于与原图像进行比对;最后,PSPNet 网络模型将不同层级的特征图相融合,拼接出输入图像的全局特征,尽可能地减小小图像失真的可能性。

PSPNet 网络模型通过全局金字塔模块,将不同级别的特征图拼接起来,消除了卷积神经网络对图像判别时对输入图像固定尺寸的约束,避免了表征不同子区域之间关系的语境信息的丢失。

1.2 MobileNetV2

MobileNetV2 是较为经典的轻量级移动终端神经网络之一,是由 Google 提出的一种将网络模型性能和训练精度经过适当比例调整之后得出的计算模型.目的是解决由于传统深度学习模型参数量众多而制约模型在移动端部署的问题,是解决硬件条件对于模型训练限制的重要方法.网络的核心部分是可分离卷积操作 (Depthwise separable convolution)^[16],不同于标准卷积方式,深度可分离卷积将标准卷积操作改为两层卷积操作,其中的纵向卷积 (Depthwise convolution) 部分通过对每个输入通道执行利用单个卷积核进行滤波来实现轻量级滤波,而点卷积 (Pointwise convolution) 部分负责通过计算输入通道间的线性组合来构建新的特征.其中 MobileNetV2 在使用 3×3 的可分离卷积层的情况下,计算量相较于标准卷积少了 8~9 倍,精度损失较小。

MobileNetV2 网络引入两个新的概念,倒转残差 (Inverted Residuals) 和线性瓶颈 (Linear Bottlenecks).倒转残差模块区别于正常残差模块通道数先降维后升维的方式,采用通道数先升维后降维,降低了由于通道数少而导致特征提取能力弱的风险,防止图像信息丢失.而线性瓶颈的提出是为了提取更多的特征信息,对通道数较少模块使用线性激活.MobileNetV2 网络结构如表 1 所示。

表 1 MobileNetV2 网络结构表

Tab. 1 Network structure of MobileNetV2

输入尺寸	操作	扩张系数	输出通道数	卷积层重复数	步幅	输入尺寸	操作	扩张系数	输出通道数	卷积层重复数	步幅
$224^2 \times 3$	conv2d	—	32	1	2	$14^2 \times 96$	bottleneck	6	160	3	2
$112^2 \times 32$	bottleneck	1	16	1	1	$7^2 \times 160$	bottleneck	6	320	1	1
$112^2 \times 16$	bottleneck	6	24	2	2	$7^2 \times 320$	conv2d 1×1	—	1 280	1	1
$56^2 \times 24$	bottleneck	6	32	3	2	$7^2 \times 1\ 280$	Avgpool 7×7	—	—	1	—
$28^2 \times 32$	bottleneck	6	64	4	2	$1 \times 1 \times 1\ 280$	conv2d 1×1	—	k	—	—
$14^2 \times 64$	bottleneck	6	96	3	1						

如表 1 所示,MobileNetV2 包含初始的 32 个全卷积层,后接 19 个残差瓶颈层,在训练过程中使用 3×3 的卷积核.表中每行描述一个或多个相同的层的序列,重复 n 次,除第一层外,整个网络中使用常数扩展率,该决策取决于网络规模的大小,扩展率的调整有利于神经网络学习性能的提升.MobileNetV2 网络允许通过从不完全具体化的中间张量来显著减少推理过程中所需的内存占用问题,应用于壁画分割上可以减少多数嵌入式硬件设计中对主存访问的需求^[17]。

2 PSP-M 壁画分割模型

2.1 以 MobileNetV2 为基层网络的 PSP-M 模型

传统 PSPNet 模型利用残差神经网络 (Residual Networks, ResNet)^[18] 作为基层网络,原模型采用 ResNet 和空洞卷积策略提取图像特征图,通过采用金字塔场景解析网络将计算机难以解析的场景信息嵌入预测框架中,从而完成对指定图像区域的标定,达到良好的语义分割效果.在基层网络 ResNet 下,模型 PSPNet 训练性能良好,但网络深度的增加会带来额外的优化困难问题,增加分割模型的复杂度,制约模型在移动端的部署.为解决这样的问题,将轻量级神经网络 MobileNetV2 融入 PSPNet 模型中,大幅度降低网络参数量,

增加计算机硬件的适配性,改进后的模型如图 2 所示.

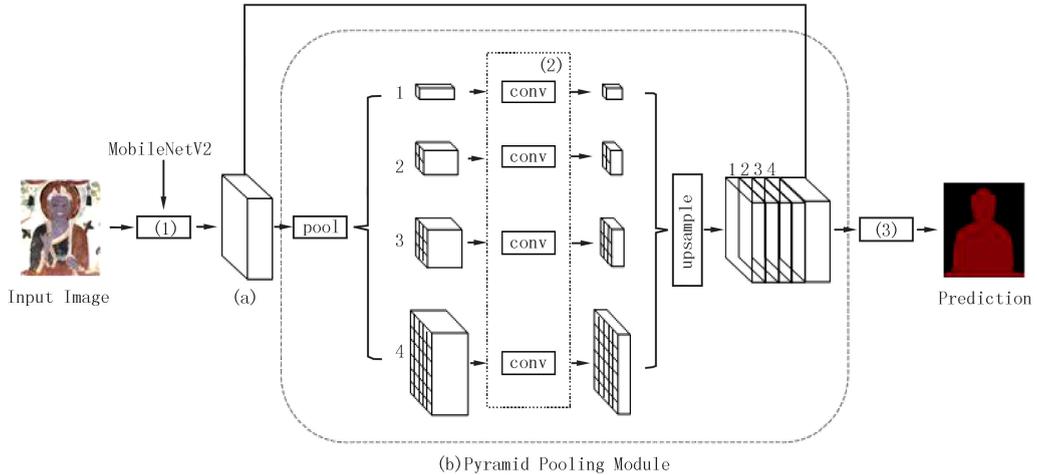


图2 PSP-M壁画分割模型

Fig.2 PSP-M mural segmentation model

如图 2 所示,模型第 1 个改进点是将标号为(1)的图像特征提取器残差卷积神经网络 ResNet 改为轻量级卷积神经网络来提取输入图像的特征图,利用深度可分离卷积网络对壁画的特征像素进行抽取,由一个卷积核负责一个通道的卷积,对输入层的每个通道进行独立的卷积运算,其中特征图的通道数量与输入层的通道数量相同,在此之后利用点卷积将之前处理过的特征在深度方向上进行加权组合,对通道进行转化,生成新的特征图(a)的同时减少神经网络的计算量.

模型的第 2 个改进点是改变传统模型的卷积网络在低维空间中使用 ReLU 函数激活的方式,因为仅当输入流形位于输入空间的低维子空间中时,ReLU 才拥有保存输入流形完整信息的能力,在 TANG 等人^[13]的实验中表明,线性层可以防止非线性函数对于图像信息的破坏,所以 PSP-M 模型在图像通道数较少时采用线性变换来代替原本的 ReLU 激活,减少图像特征的损失.

与传统方法相比,PSP-M 模型第 3 个改进点是改变传统分割方法中先降维、卷积、升维的三段式特征提取方式,融合倒转残差模块,采用先升维、卷积、降维的方法,使用直连(shortcut)结构,提高多层网络之间的梯度传播能力,与纵向卷积相匹配,将特征提取转移到高维进行,此做法的好处是卷积核的尺度远小于输出通道数,可以降低卷积层的时间复杂度和空间复杂度,设计对内存的友好,大幅度提升模型分割效率.

最后,模型还在细节方面进行优化,比如对最大池化(Maxpool)和平均池化(Avgpool)进行取舍,鉴于古代壁画分割模型的对象偏向于纹理轮廓特征,模型的池化方式选用最大池化,过滤图像无关特征信息影响,使得壁画分割效果更加鲜明.另一方面标号为(2)的部分同样引入深度可分离卷积网络,采用 shortcut 方式,跨越两到三层网络层,借鉴残差网络模型,解决深度模型中由于梯度发散而导致的特征提取误差增大问题,在原有模型的基础上进一步降低误差,从而在整体上提升特征分割精度.提取通过金字塔全局模块进行多尺度融合后的特征,与(a)进一步融合,得到结果后再次通过结构(3)的卷积模块,减少通道数量,降低模型训练复杂度,生成最终的预测图.

PSP-M 模型结合轻量级神经网络,提高了图像分割的效率,保证图像分割准确性的同时将模型计算所需的参数成倍数减少,降低预训练过程中对硬件的条件要求,减少神经网络学习代价,摆脱大中型设备依赖,提高与轻型设备的匹配度,在壁画分割上效果良好.

2.2 模型描述

PSP-M 壁画分割模型的工作流程如图 3 所示.

其中图 3(a)表示 PSP-M 模型的工作流程,图 3(b)表示每一步骤下的图像可视化结果.

模型工作可分为以下几步.

步骤 1 输入图像.

步骤 2 利用 MobileNetV2 网络中的纵向卷积和点卷积配合提取输入图像的特征信息,形成特征图.

步骤 3 将特征图进行最大池化,利用 4 层级金字塔模块获取语境信息,4 个层级的池化核大小分别对应整体特征图像、1/2 特征图像和小部分图像特征,这些特征融合可以形成图像全局特征.

步骤 4 通过双线性插值直接对低维特征图进行上采样,使全局金字塔模块每层的特征图都恢复原始特征图尺寸.

步骤 5 不同层级的特征图拼接为最终的金字塔池化全局特征.

步骤 6 通过一个卷积层后生成最终预测图,分割流程结束.

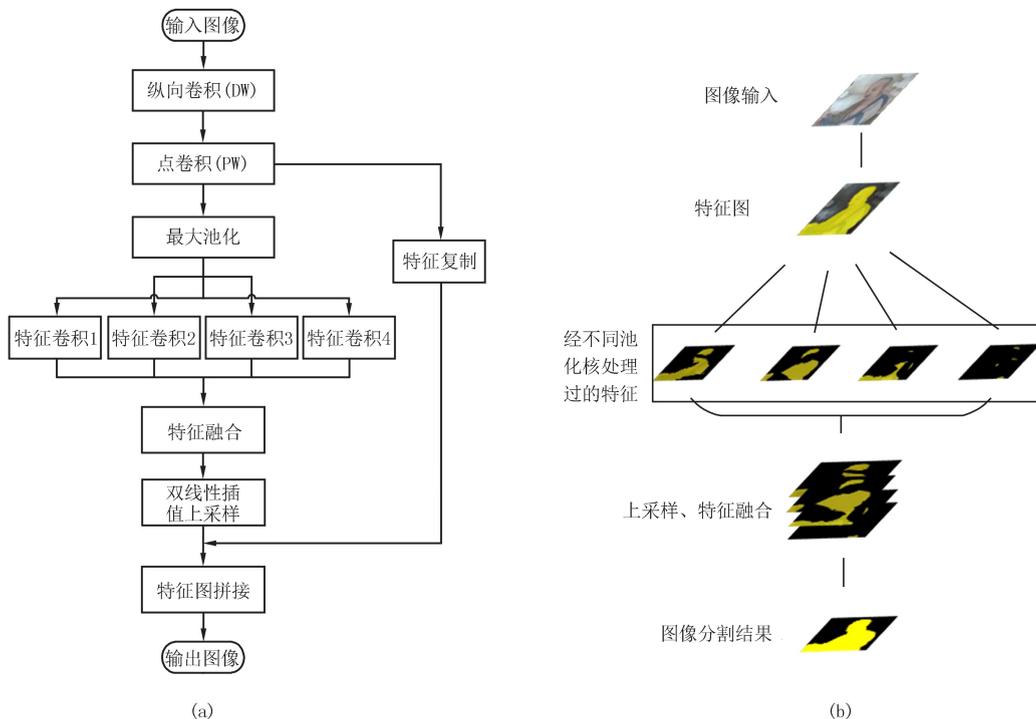


图3 PSP-M模型流程图

Fig.3 Flow chart of PSP-M model

3 实验结果

3.1 实验环境与实验数据来源

实验环境基于 Window10 操作系统,PC 端处理器型号为 Inter Core i7-9750H,显卡版本为 NVIDIA GeForce 1660Ti,实验平台为 JetBrains PyCharm Community Edition 2019,语言为 python,利用 TensorFlow 深度学习框架,结合 Keras 库来训练和测试本文模型,使用计算机视觉和机器学习软件库 Opencv 和标注软件 Lableme 来处理数据集.

实验数据集分为训练数据集和测试数据集,包含 6 种不同类型标签,共计 500 张图片,来自于《中国敦煌壁画全集》和五台山壁画图像扫描图,利用 OpenCV 提供的 resize 函数将不同类型、不同大小的图片改为像素为 224×224 分辨率的图片,所得结果整合为原始数据集.深度学习领域网络训练模型对应的数据集包含的图像数量成千上万,为解决图像分割过程中由于数据集小而出出现的过拟合问题,利用数据增强(Date Augmentation)和迁移学习(Transfer Learning)的方式对数据集进行处理.首先使用 Scikit-image 执行 Rotation 等数据增强指令,利用这种方式,数据集图片增加到 2 000 张,训练集和测试集的比例为 9 : 1,数据集具体信息如表 2 所示.

表 2 数据集相关信息

Tab. 2 Dataset related information

标签集	background	animal	build	cloud	disciple	fo	total
图片数	2 000	422	380	400	382	416	2 000

表 2 标签分别对应数据集中的背景、动物、屋舍、祥云、信徒、佛像 6 类。PSP-M 模型使用单通道标注图作为数据集,在进行数据增强之后,使用图像标注软件对每张图片的主要前景进行逐点标注,利用扫描图和标注图生成的单通道灰度图与原始数据集相结合,共同构成实验数据集。模型训练阶段使用迁移学习方法提取公共数据集 PASCAL Consortium2012 参数,对预训练权重和参数进行优化。优化的方式是将神经网络最后前几层进行冻结,利用自己的 softmax 单元替换原模型的对应单元,在数据集图像较少的前提下,提高分割模型的性能。

模型使用的损失函数由两部分构成,第一部分为常见的交叉熵损失函数(Cross Entropy Loss),当 PSP-M 模型利用 Softmax 函数对像素点进行分类的时候使用;另一部分是 Dice 损失函数,Dice 系数用于计算两个样本的相似度,计算公式如(1)式所示:

$$s = \frac{2 | X \cap Y |}{| X | + | Y |} \quad (1)$$

(1)式的原理是将预测结果和真实结果作交,之后乘 2 再除以预测结果和真实结果绝对值的和,为了体现损失函数的收敛性,将 Dice 损失函数取 1 减去 Dice 系数的值,模型 Dice 损失函数变化如图 4 所示。

Dice 损失函数不会受到图片大小的影响,训练过程中倾向于对图像前景区域的挖掘,消除由于样本不均衡对分割结果造成的影响。但是,在使用 Dice 损失函数的时候,正样本为小目标时会产生严重的震荡,当小目标有部分像素的预测发生错误,其 loss 值会有大幅度变动,所以模型中结合了交叉熵损失函数,交叉熵损失函数变化如图 5。

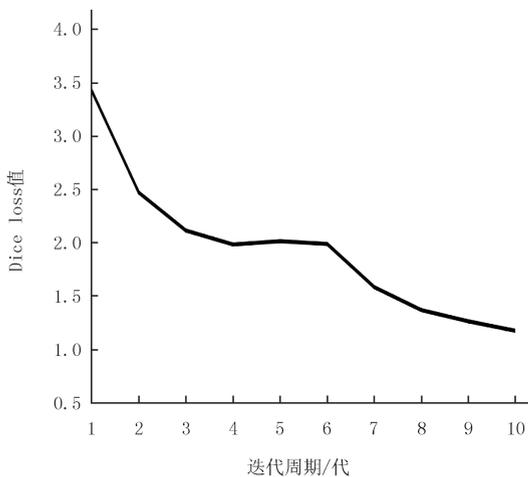


图4 Dice损失值变化图

Fig. 4 Dice loss value change chart

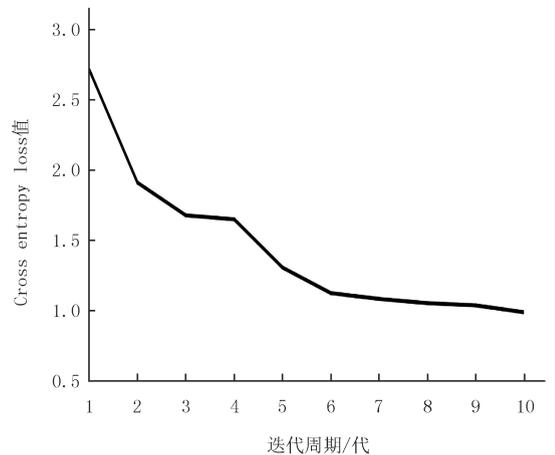


图5 交叉熵损失值变化图

Fig. 5 Change chart of cross entropy loss value

训练过程中,每 10 epoch 为一代,设置 batch_size 大小为 8,一世代提取 250 次 batch,更新 250 次参数,学习率为 $1e-5$,利用回调函数对训练集和测试集损失值进行监督,该值 3 次未下降,则降低学习率,如果损失值超过 3 次未下降,则表示模型训练流程结束,模型分割精度变化如图 6 所示。

如图 6 所示,分割模型在前 3 代的时候,精确度提升快,5、6 代时精确度呈现上下波动的态势,而 8 代之后,精确度重新趋于稳定,模型于第 10 代时训练终止,学习率达到最优。

3.2 对比实验数据分析

为验证 PSP-M 模型的优良特性,特从模型分割用时、精确度、分割效果等 3 个方面来对模型进行评价。首先,在自制数据集的基础上,选取 SegNet 图像分割模型、PSPNet 图像分割模型、DeeplabV3+ 图像分

割模型、文献[17]提出的多分类轻量级网络分割模型 (Multi-Class DeeplabV3 + MobileNetV2, MC-DM),实验平台为 JetBrains PyCharm Community Edition 2019,语言为 python,利用 TensorFlow 深度学习框架,结合 Keras 库来训练和测试上述模型;选取传统分割模型 FCM,Grab Cut 在 MATLAB R2018 平台进行测试,所有模型的实验耗时如表 3 所示。

表 3 中,传统模型 FCM 耗时最长,Grab Cut 模型耗时最短,但是 Grab Cut 分割边界模糊,图像背景和前景混乱,相对于其他模型,分割效果

较差,若想要达到良好效果,需要人为标注大量目标点,时间远超其他模型。DeeplabV3+模型是图像分割领域较为出色的模型之一,使用 Encoder-Decoder^[19]结构,将图像特征信息进行多尺度融合,减少图像空间信息丢失,由于模型相对复杂,预测结果耗时较长。MC-DM 是基于 DeeplabV3+的改进模型,将轻量级神经网络嵌入原模型,降低了硬件条件的局限性,提高模型工作效率,但该模型在分割效率上与 PSP-M 模型存在一定差距。在对比实验中,PSPNet 选用残差神经网络 ResNet50 作为基础网络,PSP-M 模型选用 MobileNetV2 作为基础网络层,PSP-M 模型在分割效率上要优于 PSPNet 模型。

表 3 分割模型实验耗时

Tab. 3 Time consuming of segmentation model experiment

模型	SegNet	PSPNet	DeeplabV3+	MC-DM	FCM	Grab Cut	PSP-M
时间/s	33.4	28.4	33.9	29.4	35.7	3.6	17.4

在训练精确度方面,由于传统模型 FCM 和聚类算法 K-means 对图片的色彩敏感度低,常用于灰度图像分割,Grab Cut 算法分割图像的精确度会随用户提供图像内容的差异而改变,存在人为因素的影响,所以本文选取 SegNet,PSPNet,DeeplabV3+,MC-DM 与 PSP-M 模型的参数量作比,对比结果如表 4 所示。

表 4 模型训练精确度对比

Tab. 4 Comparison of model training accuracy

模型	SegNet	PSPNet	DeeplabV3+	MC-DM	PSP-M
精确度	0.816 1	0.828 1	0.852 8	0.849 5	0.843 4

如表 4 所示,SegNet 模型采用最大池化的方式计算出池化的索引,通过这种方式计算相应的编码器的非线性上采样,节省了上采样学习的过程。但是在壁画分割领域,由于壁画构图复杂,SegNet 模型不能充分利用图像各像素点之间的关系,缺乏图像上下文推理能力。而 PSPNet 网络利用具有全局优先级,且包含不同子区域之间的不同尺度信息的全局金字塔模块,融合 4 种不同金字塔尺度特征,解决了复杂场景中图像的理解问题,在精确度上比 SegNet 网络提高 1%。DeeplabV3+模型利用 Xception^[11]网络为底层网络,结合空间金字塔模块 ASPP(Atrous Spatial Pyramid Pooling, ASPP)^[20],恢复图像的空间信息,优化了图像的分割边界。MC-DM 模型将轻量级神经网络与 DeeplabV3+相结合,在保证精确度的前提下,提高模型的分割效率,两者训练精确度相近。而 PSP-M 通过改变模型底层网络,针对分割过程中存在的样本不均衡问题,设计不同的损失函数,优化了模型特征提取模块,节省分割时间,在精确度方面,与 MC-DM 模型相近,略逊于 DeeplabV3+模型,比 PSPNet 网络提升约 2%,比 SegNet 网络提升约 3%。

为直观感受各个模型的分割效果,在数据集中随机抽取 4 张不同种类壁画图像进行语义分割,以分割出单一种类壁画图像为基准,其他图像元素为背景,对分割结果进行像素级图像标注,实验模型对比如图 7

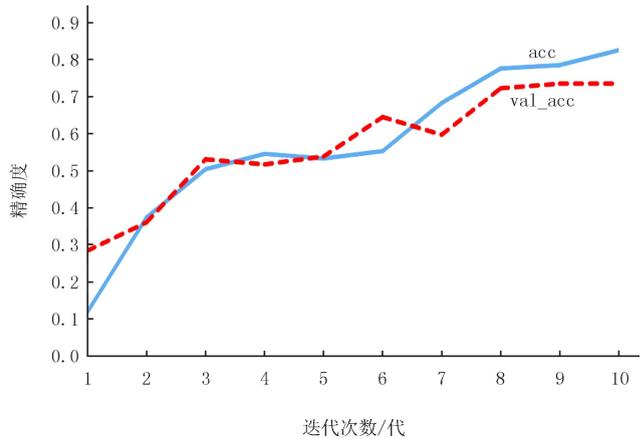


图6 训练精度变化图

Fig.6 Map of training accuracy

所示.

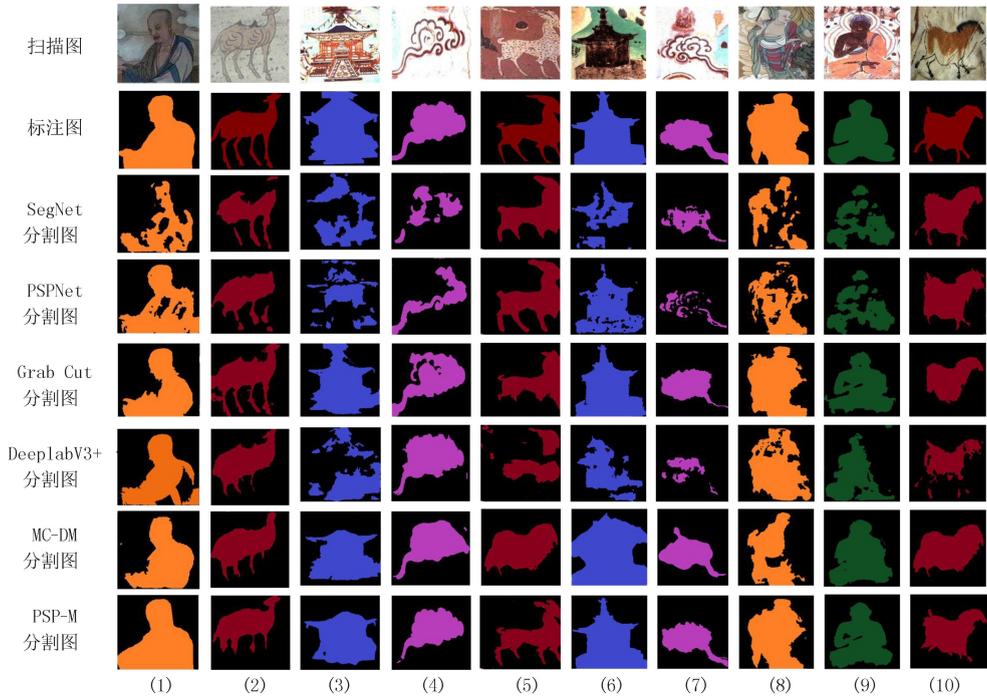


图7 实验模型对比图

Fig.7 Comparison of experimental models

图7第一行所示是壁画扫描图,第二行所示是利用图像标注软件,通过标注锚点勾勒而成的标注图.其余几行是在不同图像分割网络下的图像分割效果图.SegNet模型是较早利用Encoder-Decoder结构的模型之一,其连续下采样可以将图像特征压缩为很小的图像索引,但这会导致图像空间信息的重叠,连续上采样之后图像会出现中心信息缺失、图像分割边缘不连续等问题.PSPNet模型中结合了残差神经网络,引入残差块的概念,提高了模型性能,这也间接增加了网络宽度,使得模型计算能力下降,分割图像边缘连续性较SegNet模型有一定的提升,在单类别图像分割时,会有中心细节缺损的情况发生.Grab Cut模型的图像分割效果会随人为标注目标点的增多而变优,受人为因素的影响较大,分割效果只做参考.DeeplabV3+模型使用了空间金字塔模块和Encoder-Decoder结构,分割效果良好,但由于网络深度的增加,参数空间扩充,训练难度增大,模型容易受到过拟合问题影响,分割效果不稳定,MC-DM模型结合轻量级神经网络,减少了模型参数,网络训练时间减少,模型分割效果有了一定提升,但在图像细节处理上存在不足,容易将图像外的点标为分割点.PSP-M模型在减少模型参数的同时,解决由于样本不均衡导致的图像分割问题,优化了模型的分割效果.

实验采用人为评价、峰值信噪比(Peak Signal to Noise Ratio, PSNR)、结构相似性(Structural Similarity, SSIM)作为分割结果的评价指标.首先将分割结果制作成图,随机抽取100名学生作为考察对象,记录在学生眼中效果分割效果最好的图像,统计结果表明50%的人表示PSP-M的效果最优,其次是MC-DM模型,有18%的人支持此模型.由于DeeplabV3+模型在分割效果中存在个体差异化,导致训练精确度最高的该模型只有10%的学生支持,SegNet模型与PSPNet模型得到的支持率最低,分别为4%和6%.

PSNR是最普遍和使用最为广泛的一种图像评价指标,表示信号最大可能功率和影响它表示精度的破坏性噪声功率的比值,数值越大表示图像相似性越强,假设 $x, y \in \mathbf{R}^{m \times n}$ 两张图像,其中 x 是 y 的噪声近似,则PSNR定义为:

$$P(x, y) = 10 \lg \frac{255^2}{\frac{1}{nm} \|x - y\|_F^2}. \quad (2)$$

4 个样本的 PSNR 对比结果如表 5 所示.

表 5 模型 PSNR(dB)对比
Tab. 5 PSNR(dB) comparison

样本	1	2	3	4	5	6	7	8	9	10
SegNet	12.78	25.86	16.19	16.67	20.76	10.91	13.59	8.53	21.03	21.45
PSPNet	16.28	25.8	15.25	15.07	22.09	13.98	11.75	10.01	19.21	22.92
Grab Cut	14.33	26.54	16.53	20.75	20.64	19.26	14.31	14.94	25.80	22.80
DeeplabV3+	9.81	26.36	15.45	21.78	22.36	13.29	16.19	10.09	24.82	20.05
MC-DM	16.17	26.75	17.74	21.97	18.55	11.43	15.11	11.95	26.72	20.82
PSP-M	18.29	26.13	19.85	22.76	23.82	19.49	19.01	14.91	27.63	23.27

实验过程中,SegNet 模型和 PSPNet 模型表现相对稳定,在分割样本 2 这种轮廓分明,构造简单的图像时,6 种模型分割效果相近.而在分割样本 1、3 这种构图相对复杂的图像时,DeeplabV3+ 模型出现了两极分化现象,对于个别图像的分割效果较差,MC-DM 模型和 PSP-M 模型表现相对较好,PSP-M 模型的 PSNR 的数值又比 MC-DM 模型多出 1~2 dB,Grab Cut 模型分割结果受人为因素影响,PSNR 数值的高低只能作为参考,不具有对比性.

由于人眼的视觉对于误差的敏感度并不是绝对的,其感知结果会受到周围环境、光感等许多因素的影响而产生变化,所以会出现主观感受好而 PSNR 值反而低的现象出现,为此引入另一个评价指标——结构相似性指标 SSIM.SSIM 同样是一种用以衡量两张数字图像相似度的指标,相比于 PSNR,SSIM 在图片结构品质上的衡量更符合人眼对于图片结构质量的判断,范围为 $-1 \sim 1$,数值越大,图片结构相似性越高.结构相似性的基本原理是依据相邻像素间的关联关系,定义结构性失真的衡量方式,样本 SSIM 对比结果如表 6 所示.

表 6 模型 SSIM 对比
Tab. 6 SSIM comparison

样本	1	2	3	4	5	6	7	8	9	10
SegNet	0.645	0.813	0.584	0.745	0.725	0.613	0.806	0.588	0.746	0.890
PSPNet	0.784	0.813	0.520	0.716	0.771	0.762	0.741	0.682	0.648	0.877
Grab Cut	0.853	0.847	0.543	0.795	0.747	0.767	0.886	0.853	0.872	0.901
DeeplabV3+	0.544	0.845	0.569	0.864	0.884	0.893	0.871	0.696	0.738	0.842
MC-DM	0.817	0.861	0.717	0.873	0.831	0.648	0.763	0.767	0.900	0.869
PSP-M	0.849	0.812	0.726	0.879	0.875	0.898	0.904	0.854	0.914	0.917

表 6 中,DeeplabV3+ 模型分割后的个别图像 SSIM 值略有下降,其他模型下的 SSIM 数值与模型下的 PSNR 数值趋势保持一致,PSP-M 模型的总体表现最优.综合主客观 3 项评价指标可得,PSP-M 模型分割效果良好,图像边缘清晰,细节保存良好,适用于古代壁画分割领域.

4 结 论

中国古代壁画是中国劳动人民智慧的结晶,是中国文明的表现形式之一,每一张壁画都有特殊的历史文化背景,是当代人们了解传统文化的珍贵途径.但随着岁月侵袭,这些镌刻于墙壁上的瑰宝受到了不同程度的损坏,大量精美壁画出现颜色脱落、承载体裂纹、图像残缺等问题,对壁画信息的汲取工作造成了一定程度的影响,所以如何将壁画中表达的内容通过技术手段再次呈现是文物保护工作中的一个重点,也是一个难点问题.通过实验研究发现,FCM,K-means 等传统图像分割方法在古代壁画图像分割方面有明显的局限性,无法较好地适用于壁画图像特征的分析,在此背景下将卷积神经网络模型运用于古代壁画图像分割领域是一种新的尝试.本文将轻量级神经网络与分割效果强大的 PSPNet 网络相结合,利用全局金字塔模块融合图像不同尺度的特征,在提高分割精度的同时,降低模型对于时间的损耗;引入 Dice 损失函数,对图像区域进行

多点分析,解决样本不均衡对于分割效果的影响.文章最后通过大量对比实验,客观分析不同分割模型的优缺点,验证了新模型在壁画分割领域的可行性.但是不管使用怎样的模型,都面临着特征信息缺失的问题,各种多尺度融合网络只能尽可能还原特征信息,对于某些锐点较多的图像,特征还原能力较差,这也是在图像分割领域没有一个普适性模型的重要原因之一.

随着人工智能的不断发展,深度学习网络模型渐渐地融入人们的生活中,许多科学工作者将目光转向对神经网络的学习和研究工作.对于历史研究者来说,让古代壁画和其他古文物重新在世人面前焕发光彩是他们一生的追求.科学没有捷径,图像技术的更新迭代让文物的复苏变成了可能.

参 考 文 献

- [1] 张瑞垚,周平.基于鲁棒加权模糊聚类的污水处理过程监测方法[J/OL].自动化学报;1-15.[2020-12-17].<https://doi.org/10.16383/j.aas.c200392>.
ZHANG R Y,ZHOU P.Monitoring method of sewage treatment process based on robust weighted fuzzy clustering[J/OL].Acta Automatica Sinica;1-15.[2020-12-17].<https://doi.org/10.16383/j.aas.c200392>.
- [2] 乔颖婧,高保禄,史瑞雪,等.融合 Tamura 纹理特征的改进 FCM 脑 MRI 图像分割算法[J].计算机科学,2021,48(8):111-117.
QIAO Y J,GAO B L,SHI R X,et al.Improved FCM brain MRI image segmentation algorithm based on tamura texture feature[J].Computer Science,2021,48(8):111-117.
- [3] PARK J H,KANG Y J.Evaluation index for sporty engine sound reflecting evaluators' tastes,developed using K-means cluster analysis [J].International Journal of Automotive Technology,2020,21(6):1379-1389.
- [4] QIN X M,LI J L,HU W,et al.Machine learning K-means clustering algorithm for interpolative separable density fitting to accelerate hybrid functional calculations with numerical atomic orbitals[J].The Journal of Physical Chemistry A,2020,124(48):10066-10074.
- [5] VEKSLER O.Efficient graph cut optimization for full CRFs with quantized edges[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(4):1005-1012.
- [6] 廖苗,刘毅志,欧阳军林,等.基于非线性增强和图割的 CT 序列肝脏肿瘤自动分割[J].计算机辅助设计与图形学学报,2019,31(6):1030-1038.
LIAO M,LIU Y Z,OUYANG J L,et alAutomatic segmentation of liver tumor in CT volumes using nonlinear enhancement and graph cuts[J].Journal of Computer-Aided Design & Computer Graphics,2019,31(6):1030-1038.
- [7] 于月娜,梁光明,刘任任.宫颈细胞核质自动分离算法研究[J].计算机与数字工程,2021,49(2):371-377.
YU Y N,LIANG G M,LIU R R.Study on automatic separation algorithm of cervical cytoplasm[J].Computer & Digital Engineering,2021,49(2):371-377.
- [8] SONG Z Y,ALI S,BOUGUILA N.Background subtraction using infinite asymmetric Gaussian mixture models with simultaneous feature selection[J].IET Image Processing,2020,14(11):2321-2332.
- [9] 杨亚男,张宏鸣,李杭昊,等.结合 FCN 和 DenseCRF 模型的无人机梯田识别方法研究[J].计算机工程与应用,2021,57(3):222-230.
YANG Y N,ZHANG H M,LI H H,et al.Research on UAV terrace recognition method based on FCN and dense CRF model[J].Computer Engineering and Applications,2021,57(3):222-230.
- [10] SUN Q W,CHEN W,CHAO J G,et al.Flsnet: fast and light segmentation network[J].Journal of Physics: Conference Series,2020,1518(1):012047.
- [11] 申瑞超,柴晓冬,李立明,等.基于前景分割的扣件状态识别算法[J].铁道标准设计,2021,65(12):28-34.
SHEN R C,CHAI X D,LI L M,et al.Fastener state detection based on foreground segmentation[J].Railway Standard Design,2021,65(12):28-34.
- [12] CHEN L C,PAPANDREOU 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,2018,40(4):834-848.
- [13] TANG W,ZOU D S,YANG S,et al.A two-stage approach for automatic liver segmentation with Faster R-CNN and DeepLab[J].Neural Computing and Applications,2020,32(11):6769-6778.
- [14] HU L D,GE Q.Automatic facial expression recognition based on MobileNetV2 in Real-time[J].Journal of Physics: Conference Series,2020,1549(2):022136.
- [15] HUANG Q,SUN J F,DING H,et al.Robust liver vessel extraction using 3D U-Net with variant dice loss function[J].Computers in Biology and Medicine,2018,101:153-162.
- [16] 高淑萍,赵清源,齐小刚,等.改进 MobileNet 的图像分类方法研究[J].智能系统学报,2021,16(1):11-20.
GAO S P,ZHAO Q Y,QI X G,et al.Research on the improved image classification method of MobileNet[J].CAAI Transactions on Intelligent Systems,2021,16(1):11-20.

- [17] 曹建芳,田晓东,贾一鸣,等.改进 DeepLabV3+模型在壁画分割中的应用[J].计算机应用,2021,41(5):1471-1476.
CAO J F, TIAN X D, JIA Y M, et al. Application of improved DeepLabV3+ model in mural segmentation[J]. Journal of Computer Applications, 2021, 41(5): 1471-1476.
- [18] XUE F, JI H B, ZHANG W B. Mutual information guided 3D ResNet for self-supervised video representation learning[J]. IET Image Processing, 2020, 14(13): 3066-3075.
- [19] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [EB/OL]. [2021-01-13]. <https://arxiv.org/abs/1802.02611v2>.
- [20] 宋廷强,李继旭,张信耶.基于深度学习的高分辨率遥感图像建筑物识别[J].计算机工程与应用,2020,56(8):26-34.
SONG T Q, LI J X, ZHANG X Y. Building recognition in high-resolution remote sensing image based on deep learning[J]. Computer Engineering and Applications, 2020, 56(8): 26-34.

Segmentation method of ancient murals based on improved PSPNet

Cao Jianfang^{1,2}, Tian Xiaodong¹, Jia Yiming¹, Yan Minmin¹, Ma Shang¹

(1. Editorial College of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China; 2. Computer Department, Xinzhou Teachers University, Xinzhou 034000, China)

Abstract: In order to solve the problems of fuzzy boundary and low efficiency of traditional methods in ancient mural image segmentation, a multi classification mural image segmentation model based on PSPNet(PSP-M) is proposed. Firstly, the model integrates the lightweight neural network mobileNetV2 to reduce the limitation of hardware conditions for model training. Secondly, through the global pyramid module, the feature maps of different levels are spliced to avoid the loss of context information representing the relationship between different sub regions. Finally, the pyramid scene analysis network is used to embed the mural background features to reduce the loss of features and improve the efficiency of feature extraction. The experimental results show that the training accuracy of PSP-M model is 2% higher than that of the traditional image segmentation model, the peak signal-to-noise ratio(PSNR) of the model is 1-2 dB higher than that of the experimental comparison model, and the structural similarity index(SSIM) of the model is 0.1-0.2 higher than that of the experimental comparison model. The experimental results verify the feasibility of PSP-M model in mural segmentation.

Keywords: mural image segmentation; Pyramid pooling module; depth separable convolution; lightweight neural network

[责任编辑 陈留院 赵晓华]