

引用格式:张宜春,徐鹏举,向翔.基于边缘预测的传统文化图案语义分割算法[J].中国传媒大学学报(自然科学版),2022,29(04):19-25+56.

文章编号:1673-4793(2022)04-0019-08

# 基于边缘预测的传统文化图案语义分割算法

张宜春<sup>1\*</sup>,徐鹏举<sup>2</sup>,向翔<sup>2</sup>

(1. 中国艺术科技研究所,北京 100007;2. 北京邮电大学人工智能学院,北京 100876)

**摘要:**针对传统文化图案分割模型存在边缘贴合精度低的问题,本文从模型预测和数据标签两个角度对分割模型进行优化。首先,提出了一种基于边缘预测的迭代上采样策略,在生成预测图的阶段,将预先训练的点分类器与网络浅层特征进行融合,进一步对模糊边缘的像素点进行进一步分类,从而得到具有更高边缘质量的预测图。其次,针对像素标注存在标签模糊或错误问题,提出了一种基于标签松弛的混合损失函数,并与交叉熵损失函数相结合,支撑分割模型的训练过程。最后,在传统文化图案数据集上,仿真验证了本文所提算法的有效性,同时也证明了算法具有较强容错机制,可较好地提升分割质量。

**关键词:**传统文化图案;语义分割;边缘预测;标签松弛

中图分类号:TP391

文献标识码:A

## Boundary-prediction based traditional cultural pattern semantic segmentation

ZHANG Yichun<sup>1\*</sup>, XU Pengju<sup>2</sup>, XIANG Xiang<sup>2</sup>

(1. China Art Science and Technology Institute, Beijing 100007, China; 2. Artificial Intelligence Institute, Beijing University of Post and Telecommunication, Beijing 100876, China)

**Abstract:** Aiming at the problem of low edge fitting accuracy in traditional cultural pattern segmentation models, this paper optimizes the segmentation model from the perspectives of model prediction and data labeling. Firstly, propose an iterative upsampling strategy based on edge prediction. In the stage of generating the prediction map, the pre-trained point classifier is fused with the shallow features of the network, and the pixels of the blurred edge are further classified, so as to obtain a better prediction map with high edge quality. Secondly, for the problem of label ambiguity or error in pixel labeling, we propose a hybrid loss function based on label relaxation, which is combined with the cross-entropy loss function to support the training process of the segmentation model. Finally, on the traditional cultural pattern data set, we simulate to verify the effectiveness of the algorithm proposed in this paper, and also proves that the algorithm has a strong fault-tolerant mechanism, and can better improve the segmentation quality.

**Keywords:** traditional cultural pattern; semantic segmentation; boundary prediction; label relaxation

### 1 引言

中华民族的历史源远流长,在五千多年的华夏文明长河里,孕育了大量光辉璀璨的文化瑰宝。书画、服饰、瓷器、民俗、节日等文化元素承载了中华民族长

久以来丰富的想象力和创造力,值得在当代被更好地传承和发展。图案这种文化形式,一直以来因其内容的丰富而占据着不可替代的地位。以书画、服饰为代表的传统文化图案中包含着大量错综复杂的纹理细

基金项目:揭榜挂帅重点研发课题(课题编号:2021YFF0901701)

作者简介(\*为通讯作者):张宜春(1978-),男,副研究员,主要研究领域为文化与科技融合。Email:zhangyichun@vip.sina.com

节和语义信息,如图1中依次展示的是蓝色缎绣彩云金龙夹朝袍<sup>[1]</sup>和恽寿平的花鸟画<sup>[2]</sup>。近年来,国家文化大数据体系建设作为宣传文化界的新基建,是我国重点发展方向,中华文化素材库作为国家文化大数据体系的供给端,其相关技术研究对传承与弘扬传统文化有很大的帮助。在以上背景下,需要设计实现传统文化图案分割系统,利用特定的分割算法将图像中的文化元素完整地提取出来,经过一定的预处理和信息标注,构建中华文化素材库。图像分割是计算机视觉领域的重点研究方向,伴随着计算机软硬件的发展,图像分割技术不断取得突破性成果,解决了实际生活中的种种难题。传统的图像分割方法通常关注目标的边缘、纹理、颜色等低层视觉特征,辅以点、线、框等人工交互手段作为指导信息来对一张图像上的所有像素进行前景和背景的分类,从而获得提取的目标。该类方法在简单图像上的分割已经进行了较为成熟的应用,但仍然存在缺失语义信息、计算效率低下等缺点。随着深度学习的不断发展,分割技术也分别在语义分割<sup>[3]</sup>、实例分割<sup>[4]</sup>、全景分割<sup>[5]</sup>等全新子领域上取得了突破,且每个子领域都有大量研究工作不断推进该技术的发展。其中,语义分割依赖大量人工标注的数据集进行训练,可以对一张数字图像上的所有像素进行语义标注,目前已经在自动驾驶<sup>[6]</sup>、医学图像分析<sup>[7]</sup>等场景下有了较为成熟的应用。

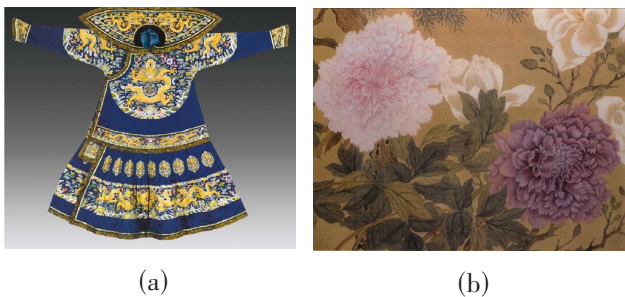


图1 (a)蓝色缎绣彩云金龙夹朝袍(b)恽寿平《花卉》

传统文化图案通常具有颜色多样和纹理繁杂的特点,针对该类图案的分割任务具有以下两方面的需求:(1)能够提取出完整的文化元素。(2)能够对提取出的文化元素标注相应的语义信息,以支撑素材库的建设和下游文化元素的分析应用。传统分割算法对图像低层特征的复杂度十分敏感,且大多只能处理单目标分割任务,因而难以高效地提取图案中的元素。在语义分割方面,不同于自动驾驶等领域,文化图案对分割结果的边缘质量有着更高的要求。对于边缘

模糊的分割子图,其质量无法达到素材库的要求,同时也给下游的分析应用带来困难。因此,我们需要改进现有算法,提升预测边缘的质量,从而满足传统文化图案分割的需求。本文首先构建了传统文化图案语义分割数据集,其内容载体为服饰、丝巾、国画等,这些图案主要来源于书籍扫描和相机采集。为了提高分割结果的边缘质量,本文针对现有模型中预测模块的上采样流程做出改进,提出一种基于边缘预测的多次迭代上采样预测模块,在恢复预测图分辨率的同时选取可能的边缘像素点进行分类,从而提升预测结果的边缘细节。同时,本文在构建数据集时发现,传统文化图像数据集的制作过程中存在着不可避免的边缘像素标注错误问题。模型针对该问题提出了一种基于标签松弛的损失函数,结合原有的交叉熵损失函数共同作用于模型的监督训练,有效降低了该类问题对模型训练误差的影响,提升了模型的性能和鲁棒性。最终,通过对比实验验证了以上改进的有效性,提出的模型满足传统文化图像分割需求,能够高效地提取图像中的文化元素。

## 2 相关工作

基于深度学习的分割方法按照结果的不同可分为语义分割,实例分割,全景分割等,按照分割目标的不同可分为图像分割,视频分割<sup>[8]</sup>,三维点云分割<sup>[9]</sup>等,其中图像的语义分割是最为关键的研究领域之一。与传统分割方法不同,语义分割在划分图片不同区域的同时能够对区域附加语义信息,实现对图片像素级别的语义信息标注。深度学习之前的语义分割通常基于条件随机场,在像素或者超像素上计算一元势能量和二元势能量,从而分割图像。Jonathan等人<sup>[3]</sup>在2014年提出了使用FCN替换传统CNN中末尾的全连接层,以接受任意尺寸大小的输入,其对输入的RGB图像多次进行卷积和池化操作,得到小分辨率的特征图,并通过反卷积等方法恢复分辨率,得到与原始图像同样大小的预测图。该流程也成为了语义分割的标准范式。Olaf等人<sup>[10]</sup>提出的U-net网络通过残差连接将下采样阶段的图像低层特征拼接至上采样层,并以此恢复图像的细节信息。Liang-Chieh等人<sup>[11-14]</sup>提出的DeeplabV3+网络框架通过增加编解码模块和空洞卷积金字塔池化模块提取不同尺度的图像特征,再将特征信息拼接以获得更多的上下文信息。Barret等人<sup>[15]</sup>提出了自学习的策略,通过中间模

型对未标注数据进行预测,从而扩充训练集来解决标注成本过大的问题。Xiangtai等人<sup>[16]</sup>通过对解耦分割目标的边缘和主体进行多任务学习,提升分割结果的内部一致性和边缘精细程度。Alexander等人<sup>[17]</sup>提出的基于图形渲染的分割方法,将模型的计算集中于物体的边缘,并混合不同粒度的特征用于像素点分类,提升了分割物体的边缘质量。Xia等人<sup>[18]</sup>提出的期望最大化注意力网络是先在CNN输出的预测图上通过期望最大化算法计算出一组低秩的基,并将其用来重建特征图,而后计算注意力,这种做法在降低了复杂度的同时,也提升了模型性能。

近两年,除了全卷积神经网络外,基于Transformer网络的分割算法也有大量工作进展。Ashish等人<sup>[19]</sup>在自然语言处理领域提出了Transformer网络结构,能够在一次计算中获得任意长度的上下文信息,克服了循环神经网络效率过低和卷积神经网络上下文距离有限的缺点。Alexey等人<sup>[20]</sup>将图片切分成256个图像块,加上一个可学习的分类头作为Transformer的输入,通过分类头对图像进行分类,成功将Transformer网络应用在计算机视觉领域。Sixiao等人<sup>[21]</sup>提出的SETR网络利用Transformer作为骨干网络,提取图像特征作为后续语义分割头的输入,首次将Transformer网络应用到语义分割任务。Enze等人<sup>[22]</sup>提出的Segformer对Transformer进行了多处改进以适应图像分割任务,在编码器部分利用卷积而不是位置编码来学习位置信息,在切分图像块时保留部分重叠确保局部连续性。为了增加模型的计算效率,其采用了多层感知机作为解码器,有效提升了分割的性能。

交互式的图像分割方法也因深度学习技术的引入而有了许多创新工作。该类分割方法需要人工交互作为额外输入,但在交互一定次数后准确率通常比自动分割高。当前,交互式的图像分割技术在现实场景中被广泛使用。Ning等人<sup>[23]</sup>提出将用户的前景点标记和背景点标记作为额外的通道与RGB通道共同作为全卷积神经网络的输入。Won-Dong等人<sup>[24]</sup>提出一种用于交互分割的反向传播优化结构,通过优化目标函数的反向传播来调整用户交互距离图,有效减少了交互次数。Zheng等人<sup>[25]</sup>提出了初始交互点注意力模块,由于用户初始交互点通常位于目标物体中心,具有位置指导能力,因而在此基础之上计算注意力图。该工作同时提出了一种结构完整性策略,通过设定阈值的方法判断像素点是否被错分,提升了模型的

性能及容错能力。

语义分割的主要应用领域如自动驾驶,医学图像分析,地质检测等对于分割结果的边缘准确度要求并不严格,这也就导致了主流的语义分割算法无法直接应用于传统文化图案分割任务。常用的语义分割网络,例如上文提到的FCN、EMANet等,在下采样阶段计算得到特征图后,通常采用双插值或转置卷积的方法来使其恢复到原始图像的大小,得到最后的预测图。双线性插值是数字图像处理领域经典的上采样方法,是对单一方向线性插值的扩展,其在原始图像上选取距离最近的四个点,并在x轴和y轴两个方向上各进行一次插值,得到最终结果。转置卷积方法是将普通卷积核的输入输出的大小进行颠倒,即从多个像素点输入对应一个像素点输出变为一个像素点输入对应多个像素点输出,通过这样运算方法对特征图进行上采样。

在上述两种上采样方法中,前者更加高效,后者更加灵活,但不论哪种上采样方法,其输出的特征图都存在一定程度上的边缘过度平滑和细节丢失问题。U-net通过多次上采样和添加残差链接的方式,来拼接上下采样过程中相同大小特征图丢失的边缘细节。DeepLab采用编解码结构,在编码阶段进行下采样操作,在解码阶段进行上采样操作,并通过残差连接恢复细节。SegFix通过生成边缘图和方向图的方法来用目标内部像素标签指导边缘像素标签的确定<sup>[26]</sup>,从而获得详细的边缘信息。以上方法都取得了不错的效果,但是仍然不能达到传统文化图案分割的要求。

### 3 算法流程

基于对现有语义分割算法的分析,本文提出新的预测模块来更好地得到细节完整的预测图像。记 $f$ 为卷积层输出的特征图,通过标准卷积层输出的粗糙预测图为 $p_0$ ,对其进行多次上采样迭代。每次分为以下四个步骤:(1)对前一次低分辨率预测图 $p_i$ 进行双线性插值上采样,结果记为 $p'_i$ 。(2)在 $p'_i$ 上采取一定的策略选取 $N$ 个不确定点。(3)对被选取的点进行特征构建,特征需包含网络浅层信息和语义信息。(4)用一个预先训练好的点分类器对被选取的点进行分类并更新 $p'_i$ ,得到精细预测图 $p_{i+1}$ 。预测过程中不断重复以上流程,直到得到与原始输入图像大小一致的 $K$ 类标签预测图 $p_{final}$ ,整体网络结构如图2所示。

## 4 基于边缘预测的上采样流程

上采样流程包含双线性插值、边缘点选取、特征构建、分类与更新四个部分。在迭代过程中,通过选取可能的边缘点并对其分类,不断完善预测图的边缘细节,确保最终输出的预测图符合传统文化语义分割的需求。选点个数 $N$ 的大小在训练阶段和推理阶段是不同的:在训练阶段为初始特征图中点的数量,在推理阶段为超参数 $c$ 和初始特征图中点数量的最小值。改进的上采样流程在每次迭代过程中只对高不确定性的点进行分类,这样不仅能够得到精细的边缘细节,同时也具有很高的计算效率。

### 4.1 选点策略

出于对计算效率的考量,选点策略在推理阶段和训练阶段并不相同。

在推理阶段的一次上采样迭代过程中,双线性插值后的预测图 $p_i$ 需要选取合适的 $N$ 个点进行预测分类。由于本章算法改进的目的是获得目标边缘的细节,所以被选点需要尽量处于物体边缘,因而模型需要对 $p_i$ 上的点的不确定性进行量化。具体的量化方

法为:将量子中每个像素点的 $K$ 类概率进行排序,选取最高概率和次高概率计算绝对差值 $\Delta P$ , $\Delta P$ 与该点的不确定性成反比。通过以上方法可以量化 $p_i$ 上所有像素点的不确定性,并选取不确定性最高的 $N$ 个点进行分类并更新。不断迭代计算,直到预测图与原图大小一致。

在训练阶段,迭代的过程在反向传播时计算效率较低,为了提升模型的效率,需要采用一种新的选点策略,来模拟推理过程中可能被选择的点的分布。模型提出一种混合选点策略,与推理阶段的选点策略相近,结合使用超参数 $\alpha$ 和 $\beta$ ,具体过程可分为以下三个步骤:(1)过采样:在特征图像素点的均匀分布上随机采样 $\alpha N$ 个点作为候选点,其中 $\alpha > 1$ 。(2)重要点采样:用与推理阶段相同的方法对这 $\alpha N$ 个点的不确定性进行量化,即计算类别概率中最高项和次高项的绝对插值 $\Delta P$ ,并选取不确定性最高的 $\beta N$ 个点,其中 $\beta < 1$ 。(3)剩余点采样:剩余的 $(1-\beta)N$ 个点从特征图的均匀分布上随机采样。通过上述的选点策略可以用较高的计算效率来选取合适的点进行训练,模拟推理阶段边缘点的分布。

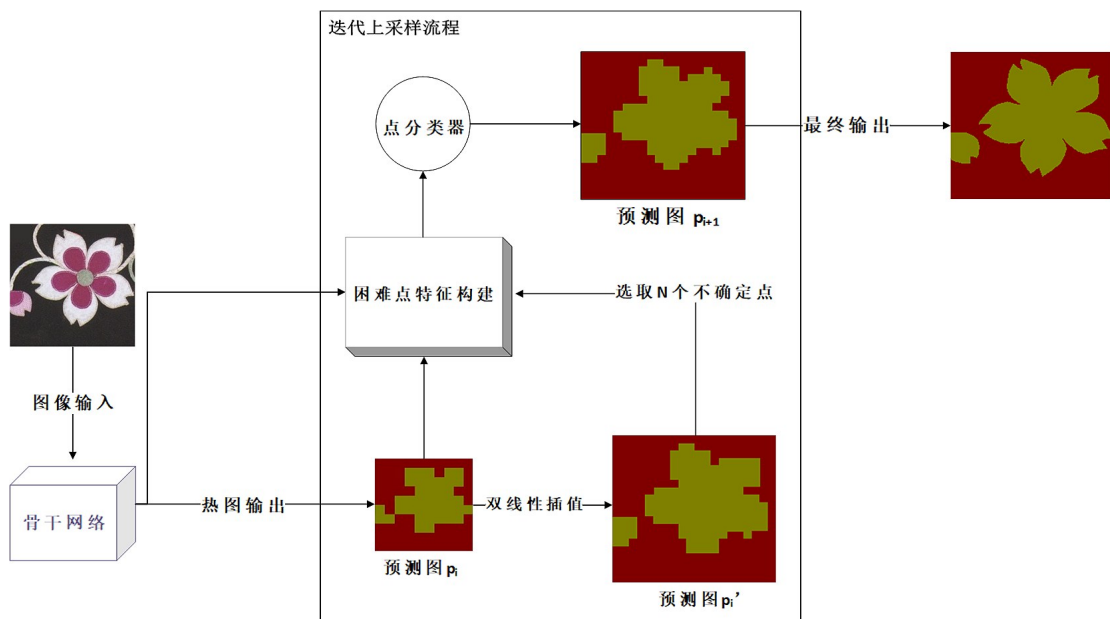


图2 本文算法流程图

### 4.2 基于点的特征提取与分类

模型需要对在预测图 $p_i$ 上被选择的点进行特征提取和标签分类。提取的点的特征需包含低层的细粒度边缘信息和高层的粗粒度语义信息,两者通过拼接操作来进行特征的构建,以此作为下游点分类任务

的输入。细粒度边缘信息通常位于网络模型的浅层,由于点是一个二维坐标,因而可以通过插值的方法在残差网络的浅层特征图 $res_2$ 处提取其特征向量,同时也可以拼接 $res_2$ 到 $res_5$ 的特征向量作为特征表达,这也是一种常见的图像低层特征细节提取方案。细粒

度的浅层特征已经可以用来恢复边缘细节,但是仍然存在缺失上下文信息和语义信息不足的问题,并且随着选择的浅层特征图变化,特征向量也会受到很大干扰,故细粒度的边缘信息需要补充拼接语义信息来完善点的特征提取。粗粒度的语义特征为被选取的点在粗糙预测图 $p_i$ 处的特征向量,其包含该点相对于不同分割类别的概率。提取出的点的混合特征是下视点分类器的输入。点分类器是一个经典的多层感知机结构。由于对于每个选取点,多层感知机都会进行标签分类,因而可以用标准的分割任务损失函数去训练参数。

### 4.3 标签松弛

全监督语义分割算法依赖大量人工标注的数据集,数据集的质量直接影响模型训练的效率 and 预测的准确性。在语义分割数据集的制作过程中,人工标注环节具有一定主观性,实验表明 Ada20K 数据集制作者在部分图片标注后经过半年再次对其进行标注,两次标注结果中标签相同的像素平均仅有 82.4%<sup>[27]</sup>。出现不同标签的原因有标注错误、命名错误、标注粒度不同等等。如图 3 所示,在第 3 节构建的传统文化数据集中,边缘像素的标注更为困难,主要有以下三个原因:首先是目标边缘十分复杂导致的标注困难,由于鸟的翅膀、蝴蝶的触须等反映在图像上的细节繁多,因而常常有像素被标注错误的情况发生。其次是较粗的边缘线条导致难以确定边缘像素点标签,人工在标注粗线条上的像素点类别时依赖其主观经验,难以确定统一标准。最后是中间透明度带来的标注困难问题,由于传统纺织品和山水画中常出现物体边缘是由前景到背景的透明度过渡的技艺,因而确定其统一的边界也十分困难。基于以上原因,在传统文化数据集构建的过程中不可避免地出现大量边缘像素标签标注错误的情况。这类标注错误会导致模型的损失计算出现误差,进而影响模型的性能和泛化能力。CityScapes 数据集将标注数据分为了粗糙标注和精细标注两种<sup>[28]</sup>,但即使在精细标注的数据集中,对于人体、树木等边缘像素的标注仍然缺少大量细节,这是自动驾驶场景下对该类边缘细节要求不高导致的。传统文化图案分割任务对分割结果的边缘质量有着很高的要求,因而为了减少训练数据噪声的影响,除了需要在构建数据集时避免人为地标错以外,还需要为模型引入一定的容错机制。

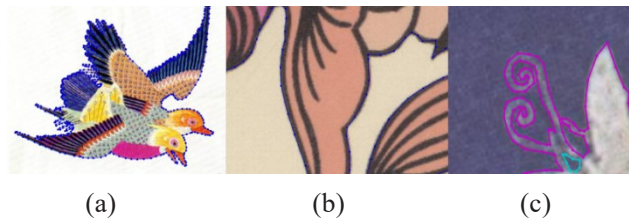


图3 边缘像素标注困难样本示例(a)复杂边缘细节(b)粗边缘线条(c)边缘中间透明度

鸟	鸟	鸟	鸟
鸟	鸟	鸟	鸟
鸟	鸟	花	鸟
花	花	花	鸟
花	花	花	鸟

鸟	鸟	鸟	鸟
鸟	鸟+花	鸟+花	鸟+花
鸟+花	鸟+花	鸟+花	鸟+花
鸟+花	鸟+花	鸟+花	鸟+花
花	花	鸟+花	鸟+花

(a)

(b)

图4 松弛前后像素的语义标签(a)标签松弛前(b)标签松弛后

#### 4.3.1 标签松弛处理

针对文化图案边缘标签存在误差的问题,提升现有算法对该类标注噪声的鲁棒性,本文提出了标签松弛方法,扩充像素点的类别标签空间。具体地,对于每个边缘像素点,即存在相邻不同标签的点,令该像素点标签为A,相邻像素点标签为B。标签松弛方法会计算像素点标签并集的概率 $P(A \cup B)$ 的最大似然估计,以此替代原本单标签的概率估计。又由于在语义分割中不同标签是互斥的,因而有以下计算公式:

$$P(A \cup B) = P(A) + P(B) \quad (1)$$

其中 $P(\cdot)$ 为每个标签类别的SoftMax概率。标签松弛计算过程中采用类似卷积的操作,对于图像的每个点计算一定范围内所有像素标签的并集,并用计算得到的标签集合替代初始标签。该卷积操作仅修改边缘像素的标签类别,对于非边缘像素,因其邻域标签与自身相同,故标签类别在计算前后并无改变。标签松弛计算实现了标签空间从单一类别到多类别集合的扩充。由于最终生成的预测图每个像素点仅对应一个标签,因而该操作仅作用于训练过程。

标签松弛操作需要确定松弛的范围,即确定松弛计算的卷积核大小。该卷积核的大小为超参数,卷积核过大会导致边缘标签混乱而无法起到监督作用,卷积核过小则无法起到提升模型鲁棒性的作用。范围的选取与图像分辨率大小以及标注的精细程度有关,本文松弛范围选取为1,即针对每个像素计算其周围

八邻域像素点标签的并集。如图4为距离1的松弛处理前后标签图示。

#### 4.3.2 基于标签松弛的融合损失函数

模型通过标签松弛操作得到一个扩大标签空间的语义图,在此语义图基础上需设计合适的损失函数,保证训练过程中一定计算效率的同时能够量化预测图与标注图之间的差异,用于反向传播参数的更新。本网络模型采用了基于标签松弛的损失函数,并与语义分割任务常用的交叉熵损失函数结合,提出一种作用于模型训练阶段的融合损失函数。

标签松弛用于处理数据集边缘像素点的标注误差,模型需要在松弛后的标签图上计算损失函数。令 $N$ 为 $3 \times 3$ 大小的松弛窗口内像素标签的集合, $P(\cdot)$ 为类别的softmax概率, $C$ 表示某一像素点的标签。定义基于标签松弛的边缘损失函数为如下形式:

$$\mathcal{L}(\text{boundary}) = -\log \sum_{C \in N} P(C) \quad (2)$$

当 $|C|=1$ 时,该损失函数退化为标准的one-hot标签交叉熵损失,当 $|C|$ 和标签数相等时,函数会忽略该点的损失贡献。该概率损失在修改后的标签空间上计算时,不同的类别间互相独立,且在 $P(C)=1$ 时取到最小值。边缘被错标注的像素点在一般的交叉熵损失函数下会贡献大量损失代价,不易于模型的正确收敛。而本文提出的标签松弛处理则可以提升模型对这类训练数据的容错能力。

该损失函数也可用于边缘预测上采样模块中点分类器的训练。在训练多层前馈网络时,通过对标注图像的标签松弛化处理,计算其边缘损失,能够有效提升点分类器的准确性和鲁棒性。基于标签松弛的损失函数整体上能够降低边缘像素点噪声的惩罚,但也弱化了边缘像素的监督能力,单独使用会导致预测图的边缘像素类别混乱以及物体内部存在语义标签噪声等问题。本文采用的损失函数结合了基于标签松弛的边缘损失函数和传统的交叉熵损失函数,并用超参数调节权重。其中标签松弛后的标签空间上的语义图仅用于计算边缘损失,这样能够确保模型在不丢失边缘信息监督的情况下提升对噪声的鲁棒性。该损失函数不仅运用于语义分割网络的训练优化,同时也将用于边缘预测上采样模块中点分类器的训练。利用融合损失函数训练点分类器是由于标注图像的边缘语义噪声对多层感知机的性能影响较大,因而需要融合损失函数降低噪声的干扰,提升分类模糊像素点时的准确率。

## 5 实现及结果

本节采用飞桨PaddlePaddle的2.2版本作为深度学习框架,对BP-EMA模型进行训练。实验采用的操作系统为Ubuntu16.04,使用的显卡为NVIDIA K80,显存为24G,CUDA版本为10.2。实验首先在ImageNet数据集上对ResNeXt-50进行预训练,这种在大型数据集上进行预训练并在下游任务中微调参数是目前研究领域的标准做法。在语义分割任务中,首先对输入的传统图案数据进行数据增强,采用添加噪声、图片旋转、水平翻转、随机缩放的方式来提升模型的鲁棒性,随后对图片进行随机裁剪,以此作为输入预处理。边缘预测上采样模块中,训练阶段选点策略的超参数 $\alpha$ 和 $\beta$ 分别设置为3和0.75,推理阶段的超参数 $c$ 设置为8096,即一张 $1024 \times 2048$ 大小图片在输出步长为16的特征图下点的数量,其余训练参数均为默认设置。在测试集上选择语义分割常用的MIoU和能够表现边界贴合度的Boundary-IoU两种评价指标来量化模型性能。

为了验证本文提出的BP-EMA算法在传统文化图像分割任务上的优越性,本文选取几种主流的语义分割算法,分别在传统文化数据集上训练并测试,结果如表1所示。从表中数据可知BP-EMA网络在提升了分割准确度的同时获得了更多的边缘细节。

实验的可视化结果如图5所示,观察可见标签松弛后的边缘预测迭代上采样可以有效提升预测结果的边缘质量。

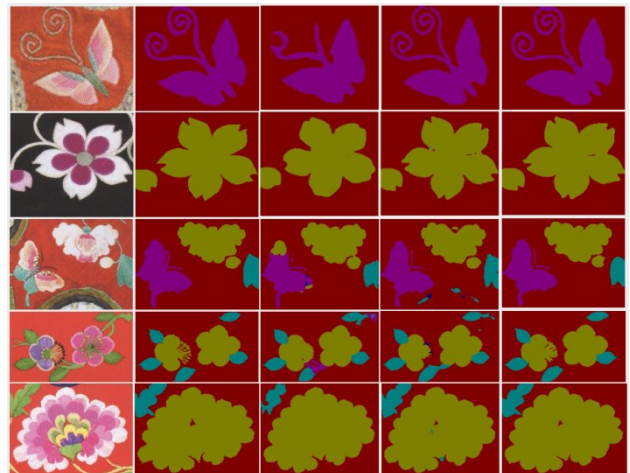


图5 预测结果可视化。

图5第一列为原图,第二列为原真语义图,第三列为采用原始上采样的结果图,第四列为采用边缘预测迭代上采样的结果图,第五列为标签松弛后采用迭代

边缘预测上采样的结果图。

表1 实验对比结果

方法	骨干网	MIoU%	Boundary-iou%
FCN	ResNet-50	68.21	52.07
PSPNet	ResNet-50	74.78	57.59
DeepLabV3+	ResNet-50	77.03	60.22
DNLNet	ResNet-50	80.42	65.53
DecoupledSegNet	ResNet-50	81.56	67.23
BP-EMA	ResNet-50	82.03	69.62

## 6 总结及未来工作

本文提出了一种基于边缘预测的迭代上采样模块,能够为潜在的边缘点构建特征并精确分类,恢复预测图的边缘细节。同时针对标注数据集时不可避免的语义噪声,提出了一种基于标签松弛的融合损失函数作为容错机制。通过以上对模型的改进能够有效获取精细的分割边缘,满足传统文化图像语义分割的需求。本文进行大量对比试验,验证了算法模型的有效性,同时对实验结果进行可视化处理,直观地展现模型对边缘分割的改进。本文的算法模型在训练之后采取了模型蒸馏和模型剪枝操作,提升了预测速度,但在对大批量图像进行语义分割时仍需要一定的处理时间。因此,后续可以在算法的运行效率上尝试突破,提升系统处理用户需求的效率。

### 参考文献(References):

- [1] 李萍.天朝衣冠——清代龙袍述论[J].荣宝斋,2012(03):30-43
- [2] 郑艳.恽寿平、蒋廷锡与清初宫廷花鸟画[J].中国国家博物馆馆刊,2011,101(12):93-104.
- [3] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE Press, 2015: 3431-3440.
- [4] He K, Gkioxari G, P Dollár, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2): 386-397.
- [5] Kirillov A, He K, Girshick R, et al. Panoptic Segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach: IEEE Press, 2019: 9396-9405.
- [6] 王中宇,倪显扬,尚振东.利用卷积神经网络的自动驾驶场景语义分割[J].光学精密工程,2019,27(11):2429-2438.
- [7] Milletari F, Navab N, Ahmadi S. V-Net: fully convolutional neural networks for volumetric medical image segmentation [C]//2016 Fourth International Conference on 3D Vision, IEEE, 2016: 565-571.
- [8] Gers F A, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM[J]. Neural Computation, 2000, 12(10): 2451-2471.
- [9] Brostow G J, Shotton J, Fauqueur J, et al. Segmentation and recognition using structure from motion point clouds [C]//Proceeding of 10th European Conference on Computer Vision, Marseille, 2008: 44-57.
- [10] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, Cham, 2015: 234-241.
- [11] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[DB/OL]. arXiv:1412.7062.
- [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, 2017, 40(4): 834-848.
- [13] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE Press, 2017: 252-263.
- [14] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 801-818.
- [15] Zoph B, Ghiasi G, Lin T Y, et al. Rethinking pre-training and self-training [J]. Advances in Neural Information Processing Systems, 2020, 33(1): 3833-3845
- [16] Li X, Li X, Zhang L, et al. Improving semantic segmentation via decoupled body and edge supervision [C]//European Conference on Computer Vision, Springer, Cham, 2020: 435-452.
- [17] Kirillov A, Wu Y, He K, et al. Pointrend: Image segmentation as rendering [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2020: 9799-9808.
- [18] Li X, Zhong Z, Wu J, et al. Expectation-maximization at