

基于卷积神经网络的前景提取

沈锐 李丹

四川大学锦城学院 四川 成都 610065

【摘要】如今抠图技术日渐崛起,可分为自动的和手动的,手动的虽然细节处理得很好,但是效率不高,所以自动抠图应运而生,越来越多的算法也应用于此。自动抠图也称自动前景提取,前景提取的难点就在于对前景图像的边缘像素进行判断,特别是细枝末节的地方,比如手指头发等,是将其分为前景还是背景,主要还是看算法对其进行的运算和判断,所以本文主要是引用其他文章的卷积神经网络模型算法,实现图片或视频的自动前景提取的精细化,对其输入进行了改进,不仅要输入原图像(或视频)还要输入一张纯色的背景图(或视频)。

【关键词】前景提取; 图像; 算法

引言

传统的前景提取是利用 Trimap 对图像进行像素归类,将图像分为三类:分别是前景、未知区域和背景。其基本公式为: $C = \alpha F + (1 - \alpha) B$, 其中 I 为原图, C 为未知区域, F 为前景像素, B 为背景像素, α 为透明度, α 的范围为 $[0,1]$, $\alpha = 0$ 确定背景部分, $\alpha = 1$ 确定前景部分^[1]。通过公式可见,通过输入原图 I , 求取三个未知数,是不太明智的,尽管可以将前景和背景的大部分像素都划分出来,但是,未知区域中的前后景划分一直都是一个问题。

因此,这里在获得了 α 和 F 基础上,额外输入一张不含主题的背景图片 B , 通过 CS 模块 (Context Switching Block) 对输入的图片进行特征处理,然后利用残差网络和原图 I 的解码特征提取,得到了细节精细化的 α 和 F , 再将得到的 F , α 和输入的 B 代入传统公式,得到了合成图像 I' , 最后利用模型分别对真实的图像 I 和合成图像 I' 进行前后景切割,得到了两个 α 和 F , 利用损失函数调整两者的差距,对 α 值和 F 值进行调整,将最终获得的 α 、 F 和替换的背景代入公式得到想要的图形^[2]。

若是视频,则会先将视频的每一帧都提取出来,然后再进行上述的一系列操作,替换背景以后,将分开的每一帧图片连在一起,输出最终的图像。

2 研究对象的相关研究

2.1 数据

2.1.1 Adobe 数据集和 Real 数据集

Adobe 数据集中的照片如上述方法所言,都是通过模型先将前景和背景进行了切割,再与背景图 B 合成的

照片^[2]。通过此数据集训练的模型更便于前景提取,都是不一定很准确,所以这里也要用到 Real 数据集。

Real 数据集是直接利用真实图像训练的模型,此模型的优点是:得到的前景相对前者更接近原始图像;缺点在于细节处理的不到位,达不到模型的目的,相对于传统模型,没有体现出优势。

两者皆有优势,所以利用损失函数对其 α 和 F 值进行了调整,弥补了相互的缺点,其公式为: $(F, \alpha) = G(X, \theta)$, 其中 X 为图像 I , 前景 F , 背景 B 和黑白图片 M 的集合。其中 M 是由四张原图的黑白照片集成,目的是为了减少颜色对人物动作识别的影响, θ 是神经网络的权重参数。

2.1.2 数据筛选

因为此方法现在还比较新颖,网上暂时还没有可以利用的完整的数据集,所以本文采用了两种途径采取数据。

一是通过 Photoshop 对花瓣网上找到的图像进行处理,将图片中的人物抹去,得到一张纯背景的图片(其中人物空缺的部分,会自动用背景填补),如图 3 所示,这样也就有了原图 I 和背景图 B , 但是,这样做的缺点在于:在前期解码优化 α 的时候会受到一些影响,对边缘部分和细节部分的处理可能就不能达到最好的效果。

二是自己拍摄一张照片和一张背景图,如图 4 所示。这样的数据集,优点在于:优化 α 的时候,不会存在上面的情况,并且效果也好。但是在切割视频的时候就会存在隐患:一是拍摄视频的时候,可能会出现人物和背景的距离太近、光线太暗或者阴影较多等情况,在切割前后景时就会出现将背景物体判断为前景的情况。二是背景视频和原视频的时间长短不同,可能会出现加速

或者慢放的情况。

除了以上两种数据集，还采用了除了人物主题以外的图像，但是结果出乎意料，很难提取其他的前景物体。因为模型在最开始对图像中的主题粗略提取的时候就失败，得到的 α 图一片漆黑，完全不能区分其他主题和背景的区别，若是强行输入模型，就会报错。

2.2 参数选择

在实践中，由于使用者拍摄图片的工具可能不一样，可能是用手机，可能是用三脚架固定相机，两者拍出来的照片或者视频有一定的差别，此时就需要对参数进行调整。

用手机可能就会出现，拍摄时手轻微的颤抖，导致背景图片和原图出现细微的差别，从而影响未知区域的判断，此时，就需要对参数进行调整，选择 real-hand-held 模式，如果适用固定相机拍摄，可以选择 Adobe 数据集上的训练模型 syn-comp-adobe，也可以选择实际数据训练模型 real-fixed-cam，但是因后者模式效果较差，所以不推荐使用后者训练模型。

本文拍摄时使用的是手机，但是在实际操作中是固定在了一个地方进行拍摄的，所以这里使用的参数是 syn-comp-adobe。

2.3 算法详述

2.3.1 卷积神经网络

由于这里是对图像进行识别和特征提取，若是使用传统的神经网络不仅很难实现，而且还会产生大量的冗余。因为传统神经网络的每一层输入只能来源于前一层的输出，但是相对于卷积神经网络，每一层都会进行降采样，类似于将图片压缩一样，通过滤波器将图片的尺寸变小，目的是避免过拟合现象。

而卷积神经网络是由卷积层和池化层构成的特征提取器，其中的每一个神经元只和图像的一部分连接，同一特征图的神经元可共享权值，共享权值的优点在于减少神经网络各层之间的连接，同时又降低了过拟合的风险，且减少了冗余^[9]。

2.3.2 卷积层

卷积层的主要作用是对输入图像进行处理，图区图像的主要特征。

卷积层通过滑动窗口（神经元）读取载入的数字图像，但是又不可能一次性直接读取数据的所有特征，所以这里就采用一个神经元对应一个特征的方式，读取图像中的每一个特征，简而言之，若是将所有的神经元的特征集合起来就相当于是一张原图。读取操作具体如图 1^[9] 所示：

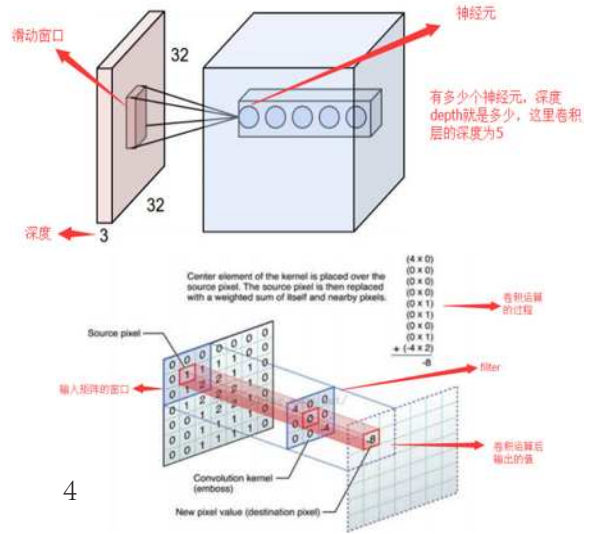


图 1^[9] 神经元提取图像特征

一个神经元像一个固定大小的正方形窗口，从左向右，从上向下读取数据图像的值。但是由于输入的数据图像不可能都一样大小，所以在读取数据的时候可能会遇到，多出一列或几列的情况，这时，卷积神经网络会自动的扩大数据矩阵。并且，这里需要注意的是，一般的彩色图像都是有 RGB 三个通道，所以图像的深度一般是三层，但是，在抠图的时候，会用到 α 通道，所以这里的深度是 4。

那么，神经元又是怎么通过窗口读取的数据只提取图像的某一特征的呢？

神经元会将获得的四个通道的数据矩阵和当前层对应的权重，分别进行内积计算之和，再加上偏置项，就得到了最后的特征值。窗口滑动一次就会得到一个新的特征值，若是输入为一个 $7 \times 7 \times 4$ 的数据图像，窗口滑动步数为 2 时，最后可能会形成一个 $3 \times 3 \times 3$ 的特征矩阵（滤波器）。

正是因为神经网络的这个特点，所以本文在进行人物抠图的时候就会利用滤波器对输入的原图、背景图、 α 图像和黑白图集，分别进行特征提取，然后将原图中得到的特征和其他图像提取出的特征进行对比和选择，以此来优化 α 图形的边缘部分。

2.3.3 残差网络

残差网络解决了由于网络层数太深，而造成的梯度消失的问题，以及网络深度的增加导致降级和更高的训练误差^[4]。其本质是通过走捷径，通过将输入直接加到后面的卷积层中通过激活函数创建模型，简化了神经网络。

本文使用到的残差网络是将一个七层残差网络作为两个三层的残差网络的输入，然后分别得到 α 和 F。所有的 ResBLK 都以 $256 \times 128 \times 128$ 的空间分辨率工作。

每个残差网络都是先经过三层的卷积网络，然后对像素矩阵进行批量归一化（BN），再使用 ReLU 激活函数，最后再次进行卷积和 BN。

2.3.4 ReLU 激活函数

激活函数是神经网络中必不可少的一部分，其本质是一种非线性函数，功能是通过非线性分布数据建立有效的模型，它会影响到整个模型的收敛效果^[6]。传统的神经网络常用的是 Sigmoid 函数，但是它没有办法解决梯度消失的问题，所以，本文采用 ReLU 函数，不仅解决了此问题，还加速了直接监督训练。其公式如图 2：从公式可以看出激活函数的输出是没有负值的，其输出范围为 0 到无穷。ReLU 的缺点会出现“神经元坏死”现象，当参数更新以后，ReLU 的神经元就不会再有激活的功能，导致梯度永远是 0 的现象。

$$f(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$

图 2 ReLU 函数公式

2.4 实践应用

本文的前景提取应用广泛，不仅可以满足图像的需要，还可以满足视频的需求。在图像方面，可以实现寸照换底，艺术照换背景等，在视频方面可以实现不用绿屏，也可以进行特效处理。如下图是实际应用上述数据集的一些结果。



图 3 对合成图像进行抠图

如图 3 所示，可以清晰地看见，一开始的前景提取并没有达到想要的效果，地上的影子、人物的边缘以及帽子轮廓都可以明显地看出锯齿的形状，所以这里将得到的图形作为输入，再次放入模型，细化其前景，可以明显地看出人物的边缘有明显的改善，帽子的轮廓也清晰可见。缺点在于，老奶奶的头发后面的阴影和背景颜色都较深，所以在判断前景时会出现以上情况。



图 4 对真实拍摄图像进行抠图

如图 4 所示，由于拍照时处于背光的状态，所以刚开始对图像进行切割的时候会明显的收到背后阴影的影响，误将后面的图像认为是前景的一部分，但是在后面根据模型中背景图像和原图的多次对照，还是可以识别出前景和背景的区别，并对人物的边缘进行细化。对头发部分的处理也很到位，但是仍然有一部分的瑕疵就在于衣袖部分的处理，还是受到了影响。

由图 3 和图 4 可以看出，虽然两张图像的来源不同，但是对人物的边缘部分都进行了细化和哑光处理，使前景的形象更加完整，并且尽管图像的光线并不明亮，但还是可以进行前景提取。唯一的缺点在于，很容易受到颜色的影响，若是颜色相近则很容易误判为相同的部分。

对比贝叶斯分类算法：

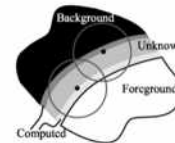


图 5 对未知区域的像素进行分类

贝叶斯算法属于概率统计中的一种分类算法。其基本原理就是基于条件 Y 的情况下计算出 X 的概率 P1，然后再通过得到的概率 P1 反过来求得在 X 发生的条件下，Y 发生的概率 P。将其用到前景提取时，先是基于 Trimap 对图像进行标记，然后通过圆形窗口对未知区域的像素逐渐采样，圆形窗口的中心是待计算的像素点^[7]，为了使得未知区域的像素点都可以得到划分，可以调节圆形的半径大小，尽量包含所有的像素，如图 6 所示。

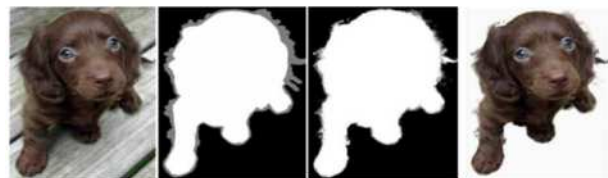


图 6 贝叶斯算法对毛发部分的处理

图 6 所示为利用贝叶斯算法实现的三分图^[8]，其优点在于，可以识别动物，并且大部分的未知区域，划分很详细，可以清晰地看见前景中的毛发。但缺点在于，狗狗的毛看起来并不是特别的自然，看起来有点粗糙，没有哑光处理。

3 结束语

本文基于原背景，直接进行自动抠图，不需要使用绿屏可以看出，在人物图像方面的前景提取，无论是 PS 处理过的人物图像还是自然真实的图像，受到前景和背景颜色的影响较大，有时很难区分其为前景还是背景，而且，前景提取的主题有限，只能进行人物提取。但是，此算法优点在于，对毛发等细节提取很有效果，并且对

前景边缘进行了哑光处理,让前景在新背景上看起来没有那么突兀,看起来很自然,并且此算法不仅适用于图像的前景提取,还可以用到视频中。

【参考文献】

- [1] 顾抑扬,杨昭,陆伟,安然.基于 Trimap 自动生成的前景提取算法 [C].中国计算机学会多媒体技术专业委员会、中国图象图形学学会多媒体专业委员会、中国计算机学会普适计算专业委员会、ACM SIGCHI 中国分会、中国自动化学会.第六届和谐人机环境联合学术会议(HHME2010)、第19届全国多媒体学术会议(NCMT2010)、第6届全国人机交互学术会议(CHCI2010)、第5届全国普适计算学术会议(PCC2010)论文集.中国计算机学会多媒体技术专业委员会、中国图象图形学学会多媒体专业委员会、中国计算机学会普适计算专业委员会、ACM SIGCHI 中国分会、中国自动化学会:中国计算机学会多媒体技术专业委员会,2010:578-583.
- [2] Soumyadip Sengupta, Vivek Jayaram, Brian Curless, Steve Seitz, and Ira Kemelmacher-Shlizerman. Background Matting: The World is Your Green Screen. University of Washington.
- [3] CharlesQAQ. 卷积神经网络(CNN)详解 [EB/OL].(2018-09-16). <https://www.cnblogs.com/duanhx/p/9655223.html>.
- [4] 刘卫铭,辛月兰,姜星宇.结合跳跃连接的残差网络图像语义分割[J].信息技术,2020,44(06):5-9.
- [5] CSDN.深度残差网络[EB/OL].(2017-10-17).https://www.sohu.com/a/198462299_115128.
- [6] 牟晋娟.深度神经网络中激活函数的研究[J].电脑编程技巧与维护,2019(12):59-61.
- [7] 李晓颖,周卫星,吴孙瑾,李丹,胡晓晖.基于贝叶斯抠图的图像背景虚化技术[J].电脑知识与技术,2017,13(28):211-214.
- [8] 刘俭.贝叶斯抠图算法研究及其在视频抠图中的应用[D].昆明:云南大学,2018.
- [9] Bobby0322.一文让你理解什么是卷积神经网络[EB/OL].(2019-02-21). <https://www.jianshu.com/p/1ea2949c0056>.