



文章编号:1001-9081(2018)02-0405-05

DOI:10.11772/j.issn.1001-9081.2017081966

基于空洞卷积的快速背景自动更换

张 浩, 窦奇伟, 莲桂凯, 姚绍文, 周 维*

(云南大学 软件学院, 昆明 650091)

(*通信作者电子邮箱 zwei@ynu.edu.cn)

摘要:针对背景更换过程复杂性较高导致传统方法效率低下并且精确度难以提高的问题,提出一种基于空洞卷积的快速图像背景更换方法——FABRNet。首先,采用VGG(Visual Geometry Group network)模型中前三部分网络结构对输入图片进行卷积和池化操作;其次,多组空洞卷积并联组合使得网络拥有足够大和足够细的感受野,并且加上残差网络结构来保证卷积过程中信息位置分布的准确性;最后,通过双线性插值算法将图片缩放到原图尺寸输出。在实验部分,与三种经典方法KNN(K-Nearest Neighbor) matting, Portrait matting 和 Deep matting 进行了对比,结果表明,FABRNet能够有效地完成背景自动更换的操作,并且在速度方面有一定的优势。

关键词:深度学习;空洞卷积;残差网络;背景更换;双线性插值

中图分类号: TP391.41 **文献标志码:**A

Fast image background automatic replacement based on dilated convolution

ZHANG Hao, DOU Qiwei, LUAN Guikai, YAO Shaowen, ZHOU Wei*

(School of Software, Yunnan University, Kunming Yunnan 650091, China)

Abstract: Because of complexity of background replacement, the traditional method is inefficient and the accuracy is difficult to improve. To solve these problems, a fast image background replacement method based on dilated convolution, called FABRNet, was proposed. First of all, the first three parts of VGG (Visual Geometry Group network) model were used for convolution and pooling operations of input images. Secondly, the combination of multiple sets of dilated convolutions were embedded into convolution neural network to make the network have a large and fine enough receptive field; meanwhile, the residual network structure was used to ensure the accuracy of the information distribution in the convolution process. Finally, the image was scaled to the original size and output by bilinear interpolation algorithm. Compared with three classical methods such as KNN (K-Nearest Neighbors) matting, Portrait matting and Deep matting, the experimental results show that FABRNet can effectively complete the background automatic replacement, and has advantages in running speed.

Key words: deep learning; dilated convolution; residual network; background replacement; bilinear interpolation

0 引言

图像背景更换是指通过图像语义分割算法提取图像的目标物体,然后应用图像融合技术将目标对象与背景图像很好地融合。随着图像处理软件日趋流行以及在图像处理和影视制作等方面的应用,背景更换的功能显得尤为重要。传统的处理方式基于图像分割操作来完成^[1],往往需要复杂的人工操作,不能够自动地完成图像背景更换^[2]。因此,一些自动化的方法被提出。Qian 等^[3]提出一种通过检测视频中静态物体的方法更换视频中的背景图像;Levin 等^[4]提出了一种封闭形式的自然图像消光解决方案,根据前景和背景颜色的局部平滑度的不同计算出图像的 alpha 通道,该方法也能实现背景更换操作。

卷积神经网络(Convolutional Neural Network, CNN)是一种多层感知机(Multi-Layer Perceptron, MLP)的优化网络结构,近年来,已经在图像处理和图像识别方面广泛应用,由此使得图像背景更换方面的研究也取得了很大的进展。Shen

等^[1]基于深度网络全卷积网络(Fully Convolutional Network, FCN)和 matting Laplacian matrix 的方式实现了深度神经网络自动抠图功能,并获得了较好的效果;Zhu 等^[5]提出了一种基于深度学习的方法对人脸头像进行抠图实现背景更换;Xu 等^[6]采用 encoder-decoder 的方式进行图像分割中的细节优化,获得了较好的效果。卷积神经网络虽然基于局部感受视野和计算权值共享的理念,能够大幅度降低网络的计算复杂度,但是随着应用需求的复杂性增加,卷积神经网络结构复杂性越来越高,计算量同样可观。尤其是在图像分割领域中,首先需要经过计算复杂的向下卷积和池化(pooling)的过程来获得全局视野^[7],然后再通过多次反卷积操作获得原始图像尺寸的图像。为了提高图像背景更换网络模型的效率,Yu 等^[8]和 Chen 等^[9]在研究中提出了空洞卷积(Dilated Convolution),能够通过卷积核的尺寸在不增加参数并且网络在不采用向下卷积的情况下获得更大的观察能力,该方法已经在图像语义分割研究方面取得了很好的效果^[10-12]。

结合上述方法,本文采用空洞卷积和残差网络结合的方

收稿日期:2017-08-11;修回日期:2017-09-09。 基金项目:国家自然科学基金资助项目(61762089,61363021,61640306)。

作者简介:张浩(1992—),男,云南玉溪人,硕士研究生,主要研究方向:深度学习; 窦奇伟(1994—),男,内蒙古包头人,硕士研究生,主要研究方向:深度学习; 莲桂凯(1993—),男,山东烟台人,硕士研究生,主要研究方向:深度学习; 姚绍文(1966—),男,云南昆明人,教授,博士,主要研究方向:工作流、Petri 网; 周维(1974—),男,云南昆明人,教授,博士,主要研究方向:分布式处理、生物信息学。



式,提出一种端到端图像背景更换方法 FABRNet (Fast Automatic Background Replacement neural Network)。本文的工作主要包括:1)提出一种端到端的自动图像背景更换方法,采用空洞卷积,能在不增加参数数量的同时有效地控制卷积核的视野;2)在卷积网络中采用双线性插值,降低模型的计算复杂度,同时减少图像变换过程中的失真对模型准确率的影响,优化卷积效果。

1 相关工作

背景更换是一个比较复杂的工作,传统的图像背景多数基于图像分割处理完成。Yang 等^[13]最早提出对两个图像的相应像素之间的光强差进行比较,分别在场景的前景和背景区域中具有不同强度的两个红外图像(IR 图像),以形成区分图像的前景和背景区域的 alpha 通道;然后将 alpha 通道应用于场景的可见光图像,并且使用预选背景来替换原始背景。Penta 等^[14]提出一种基于图像分割技术的自动搜索合适背景、更换图片背景的方法,将搜索得到的最佳背景图片与前景图片合成更换好的图片。Swanson 等^[15]通过获取图片中明暗差异的方式来获取图片的前景图像和图像的 alpha 通道,从而通过 alpha 通道将图像中的背景去除,替换上其他所需的背景。Chen 等^[16]通过 K 最近邻(K-Nearest Neighbors, KNN)算法进行图像语义分割,得到图像的前景和背景区域,通过该方法替换背景区域也可以达到背景更换的效果。

CNN 最早由 Lecun 等^[17]提出并在手写数字识别应用中取得了突破性的进展,之后被广泛应用于图像识别^[18]、语音检测^[19]、生物信息学^[20-21]、文档分析等多个领域。因为 CNN 在图像处理中可以直接作用于图像中像素值,能够提取更广泛、更深层次和更有区别的特征信息^[22]。通过卷积神经网络提取图像特征可以避免传统方法中特征提取不充分以及特征提取过程中信息丢失的问题。目前已有学者将深度学习应用于图像语义分割和图像目标对象提取当中。Shelhamer 等^[7]率先提出一种“完全卷积”网络模型,经过像素级的训练能够得到一个端到端的图像语义分割模型,可以适应任何尺寸大小的图片输入,该方法在后续研究中得到了广泛的应用。Chen 等^[9]结合空洞卷积网络处理图像语义分割任务,通过空洞卷积可以在不增加计算量和参数个数的同时控制卷积核的视野;此外还提出了空间金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)网络模型,该模型可以划分不同尺寸的图片对象。Xu 等^[6]提出方法分为两个部分:第一部分采用的是自动编码和解码的深度卷积网络得到一个灰度图,第二部分网络是一个小型网络结构对第一部分得到的结果进行微调得到更精细的结果。Shen 等^[1]提出了一种用于人像的自动抠图方法。这种方法不需要用户交互,不仅考虑图像语义预测,还考虑像素级图像优化。

本文基于上述研究提出一种端到端的自动背景更换网络系统:通过 CNN 提取图像特征,实现背景自动更换;为了保证 CNN 效率,引入空洞卷积操作以取代 encoder-decoder 操作;此外,网络中还加入了双线性插值操作来有效降低上采样过程带来的图像失真率。

2 自动背景更换方法

2.1 数据集的准备

为了验证 FABRNet 的可行性,首先收集了 Shen 等^[1]提

供的 2000 张包含原始图片和 alpha 通道的公共数据集,为了扩大训练数据集,使网络得到更大的泛化能力,采用 COCO 数据集^[25]作为背景图像,对每个训练样本更换 120 张不同的背景图片,保留其输入图片和背景图片作为训练样本,从而得到 204×10^3 个训练数据,采用其中的 1700 个数据对象作为训练数据,另外 300 个数据对象作为测试数据。

此外为了提高 FABRNet 对图像细节信息的获取能力,还收集了 alphamatting.com^[23]提供的公共数据集,该数据集中提供了较为精确的 alpha 通道,但由于此公共数据集数据样本仅包含 27 个样本数据,仍采用上述的方法对数据集进行扩张,实验中使用该数据集对网络模型进行微调。

2.2 网络模型

2.2.1 空洞卷积

背景更换图像特征提取需要充分考虑局部信息和全局信息,该网络中需要一个对全局特征和细节特征都敏感的模型,因此需要设计一种新颖的 CNN 结构模型来进行图像背景更换。经典的 CNN 中通过 pooling 操作可以使卷积核获得较大的感受视野,但是对于图像背景更换需要输出的是与原图像尺寸相同的图像,所以 pooling 后得到的较小的尺寸需要通过反卷积操作扩大图像,但是如果向下 pooling 操作过多容易导致反卷积过程中信息损失量过大,而空洞卷积能够控制卷积核的 rate 的大小,从而得到不同大小的卷积视野,如图 1 所示。

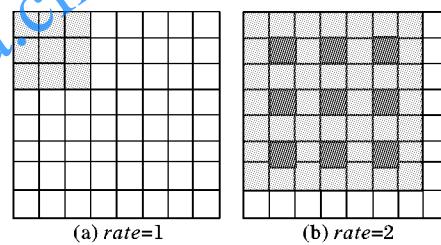


图 1 空洞卷积示意图
Fig. 1 Dilated convolution illustration

图 1(a) 中显示的是经典的 3×3 的卷积核心作用的视野效果,每次覆盖 3×3 的视野;图 1(b) 中对应的是 3×3 的 $rate = 2$ 空洞卷积,实际上卷积核的大小仍然是 3×3 ,但是卷积核的计算视野增大到了 7×7 ,而实际参数只有 3×3 。感受野大小可以表示为:

$$v = ((k_{size} + 1) \times (r_{rate} - 1) + k_{size})^2$$

此外,很容易看出, $rate = 1$ 和 $rate = 2$ 的卷积同时作用的效果和传统卷积核 7×7 单独作用的效果相同,FABRNet 通过四种不同 $rate$ 的卷积核心作用于同一输入能够感受不同视野的特征。

2.2.2 FABRNet 结构

结合图像背景更换的特定需求,本文提出一种新颖的网络结构,如图 2 所示。网络主要分为三个主要部分:第一部分采用了 VGG(Visual Geometry Group network)模型中的部分结构,该结构已经在计算机视觉中取得了显著的成果,这部分中包含三个阶段,第一阶段和第二阶段中分别包含两个卷积层和一个 maxpooling 层,第三阶段包含三个卷积层和一个向下卷积层;第二部分采用五种不同大小的卷积核和不同的 $rate$ 的空洞卷积对网络前一层输出的结果进行卷积操作,为了保持图像中信息位置分布,该部分还增加了残差网络结构;最后一部分为了得到原图像尺寸,结构中采用 1×1 的卷积对得到的结果进行微调之后输出 3 通道的图像,为了减小图像缩放



中的误差,网络中额外增加了双线性插值操作得到与原图大小相同的尺寸。

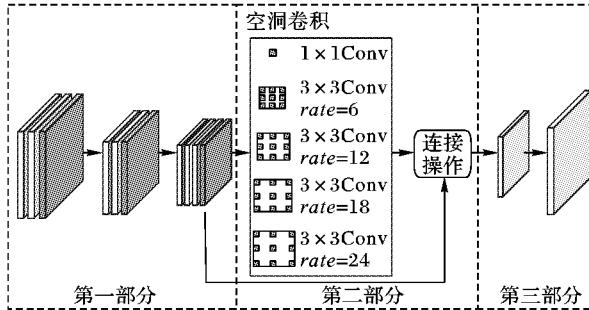


图2 FABRNet的网络结构
Fig. 2 Architecture of FABRNet

2.2.3 双线性插值

双线性插值是图像缩放中比较常用的算法,模型中使用该方法将卷积过程中缩小的图片放大到原图像尺寸,通过计算预测得到的图片与真实图片之间的损失,通过反向传播的方法得到最优模型。具体计算过程如下:

假设图片中存在点 $p(x, y)$ 需要通过插值方式得到,但是我们已经知道图片中存在确定的点 $q_1(x_1, y_1), q_2(x_2, y_2), q_3(x_3, y_3), q_4(x_4, y_4)$, 满足函数 $y = f(x)$ 。如图3所示,为了得到点 p 的像素值,首先需要在横坐标方向进行插值得到点 t_1, t_2 的坐标值,插值的方法为:

$$f_{t_1} = \frac{x_2 - x}{x_2 - x_1} f_{q_1} + \frac{x_2 - x}{x_2 - x_1} f_{q_2}$$

$$f_{t_2} = \frac{x_2 - x}{x_2 - x_1} f_{q_3} + \frac{x_2 - x}{x_2 - x_1} f_{q_4}$$

同理得到 y 方向的插值:

$$f_p = \frac{y_2 - y}{y_2 - y_1} t_1 + \frac{y_2 - y}{y_2 - y_1} t_2$$

通过上述三个公式即可得到 $p(x, y)$:

$$f(x, y) \approx \frac{f_{q_1}(x_2 - x)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} + \frac{f_{q_2}(x - x_1)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} + \frac{f_{q_3}(x_2 - x)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} + \frac{f_{q_4}(x - x_1)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)}$$

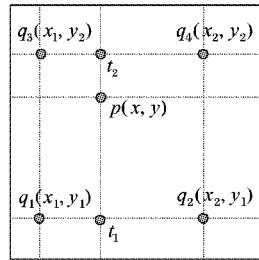


图3 双线性插值示意图
Fig. 3 Bilinear interpolation illustration

如图2所示,FABRNet为双线性插值操作层,目的是将图片放大到原始图像尺寸,通过该方法:一方面可以取代复杂的反卷积操作,降低模型的计算复杂度;另一方面可以减少图像变换过程中的失真对模型准确率的影响。

2.3 损失函数

模型采用的损失函数为预测生成图片与真实图片像素值的欧几里得距离,并且加上常数 ε 以保证损失函数可微,具体表示形式如下:

$$L = \sum_{i,j}^m \sqrt{(\alpha_p^{ij} - \alpha_g^{ij})^2 + \varepsilon}$$

其中: α_p^{ij} 表示通过神经网络得到的预测图像像素值; α_g^{ij} 表示真实图像的像素值; ε 是非常小的值,在实验中被设置为 10^{-6} 。可以很容易计算出损失函数的倒数为:

$$\frac{\partial L}{\partial \alpha_p^{ij}} = \sum_{i,j}^m \frac{\alpha_p^{ij} - \alpha_g^{ij}}{\sqrt{(\alpha_p^{ij} - \alpha_g^{ij})^2 + \varepsilon}}$$

该损失函数可以计算出预测图像与标准图像之间的像素差值,计算产生的误差通过反向传播的方法回传到网络中的每个节点,可以帮助网络学习到图像中主体特征、背景特征以及前景与背景之间的边缘特征等。

3 实验结果与分析

3.1 实验环境

实验中训练数据集中有 204×10^3 个训练样本,网络输入是大小为 224×224 像素的目标图片和背景图片,当模型在该数据集下训练趋于稳定后再采用 alphamatting.com^[23] 提供的数据集对模型进行微调,使得模型能够学到更多细节特征,尤其是边缘细节特征。模型的训练和测试基于 tensorflow^[24] 框架实现,具体结构如图2所示。首先采用 VGG-16^[26] 中训练好的参数对网络中第一部分进行初始化,其余部分采用随机的初始化方式,训练过程中学习率的范围设置为 $[10^{-6}, 10^{-3}]$,调整不同的学习率使网络达到最优效果。

模型的训练在 NVIDIA TESLA T40M 显卡上完成,训练过程大概需要 4 个 epochs,训练时间大概是 7 小时;测试过程中还对比了模型在 GPU 和 CPU (intel i7 3.40 GHz) 两种设备上的表现能力。

3.2 评价指标

实验中通过公共数据测试集评测模型的有效性,并且与其他传统的机器学习 (KNN matting^[16])、深度学习的方法 (Portrait matting^[1] 和 Deep matting^[6]) 进行了比较分析。首先采用两个通用衡量指标对实验结果进行评测,分别是 Rhemann 等^[23] 提出的 $G(\alpha^p, \alpha^{gt})$ 和 $C(\alpha^p, \alpha^{gt})$,它们已经被广泛应用于比较图片之间的差异。其中: $G(\alpha^p, \alpha^{gt})$ 表示预测图片与真实图片之间的梯度误差, $C(\alpha^p, \alpha^{gt})$ 表示预测图片与真实图片的像素之间的均方差。计算公式如下:

$$G(\alpha^p, \alpha^{gt}) = \frac{1}{k} \sum_i \| \nabla \alpha_i^p - \nabla \alpha_i^{gt} \|$$

$$C(\alpha^p, \alpha^{gt}) = \frac{1}{k} \sum_i (\alpha_i^p - \alpha_i^{gt})^2$$

其中: α^p 表示的是预测图像, α^{gt} 是通过手工更换背景得到的图像, k 是图像的像素点数, ∇ 表示计算图像的梯度。

3.3 实验预测分析

实验中首先比较了 FABRNet 与 KNN matting^[16]、Portrait matting 和 Deep matting 在不同数量测试数据下的平均梯度差和像素内容差(为了能减少误差,实验中分别测试 5 次取平均值)。从表1中可以看出,FABRNet 在不同数量的数据集的预测结果比其他三种方法都好。以 1000 个测试样本为例,FABRNet 模型的 $G(\alpha^p, \alpha^{gt})$ 为 3.82×10^{-2} ,比其他三种方法中表现最好的还低 0.35×10^{-2} ;同样 FABRNet 模型的 $C(\alpha^p, \alpha^{gt})$ 为 2.94×10^{-1} ,比其他三种方法中表现最好的还低 0.09×10^{-1} 。另外在测试样本量为 200 和 500 时也有同样的表现趋势,最主要的原因有两方面:首先,深度卷积网络能识别复杂的图片特征,充分提取背景更换需要的特征信息,并且空洞卷积有效兼顾了网络中全局特征和细节特征的提取;其



次,网络模型中最后一层通过双线性插值的方法调整图片的大小,有效减少了图像变换过程中的失真对模型准确率的影响,从而减小了整个网络的误差。

为了能更直观地体现FABRNet的生成结果,图4展示了FABRNet与其他三种模型得到的背景更换后的结果图片。从图4可以看出,KNN matting^[16]和Portrait matting^[1]两种方法容易将背景区域分割为前景内容,且两种方法中得到的图像边缘信息过于明显;而Deep matting^[6]方法中出现了前景区域被分割为背景区域的情况;本文提出的FABRNet模型通过增

加空洞卷积和残差网络的方式使网络有效获取到了图片的全局视野,并且通过不同大小的空洞卷积组合实现不同大小的视野信息,有效解决了上述这两种缺陷。此外,从图中还可以看出,本文方法可以有效弥补结果图片中边界信息过于明显的缺陷。但是也可以发现,相比KNN matting^[16]和Portrait matting^[1],FABRNet模型也引入了一个新的问题:在得到的结果图片与原图之间存在微弱色差,这是因为在CNN的计算过程中存在数值计算误差,导致结果图片中像素与真实像素之间有略微的差异。

表1 不同测试样本量的结果比较

Table 1 Comparison of the results of different test samples

方法	200个测试样本		500个测试样本		1000个测试样本	
	$G(\alpha^p, \alpha^{gt}) \times 10^{-2}$	$C(\alpha^p, \alpha^{gt}) \times 10^{-1}$	$G(\alpha^p, \alpha^{gt}) \times 10^{-2}$	$C(\alpha^p, \alpha^{gt}) \times 10^{-1}$	$G(\alpha^p, \alpha^{gt}) \times 10^{-2}$	$C(\alpha^p, \alpha^{gt}) \times 10^{-1}$
KNN matting ^[16]	12.78	4.12	13.47	5.93	11.37	4.31
Portrait matting ^[1]	10.54	3.12	10.96	4.17	8.96	4.06
Deep matting ^[6]	5.51	2.16	6.01	4.59	4.17	3.01
FABRNet	4.79	2.05	5.12	4.34	3.82	2.94

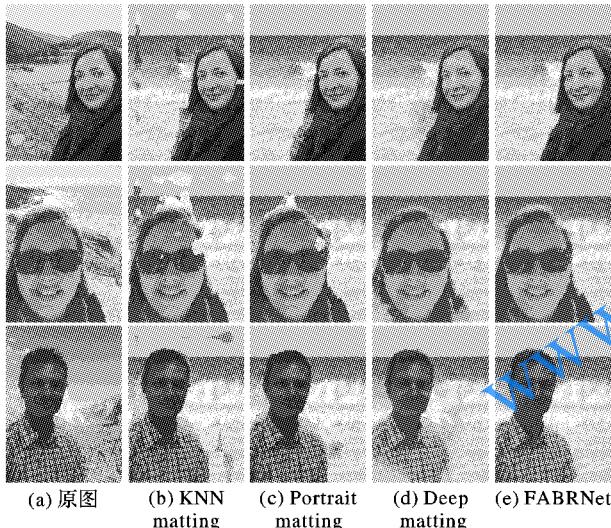


图4 不同方法背景更换结果比较

Fig. 4 Background replacement result comparison of different methods

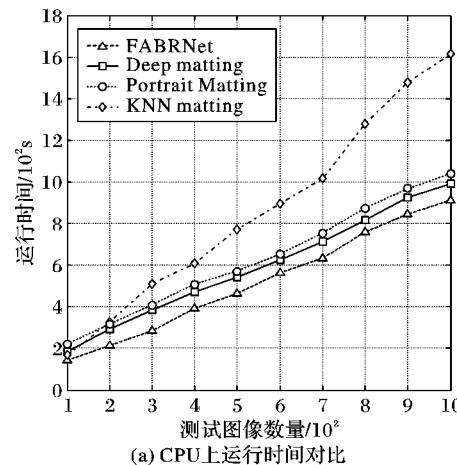
3.4 FABRNet 性能分析

实验中还测试了不同模型分别在CPU和GPU上预测100~1000张图片背景更换过程的时间消耗,结果如图5所示。图5(a)为在CPU(intel i7 3.40 GHz)上加载训练好的模型进行预测过程所消耗的时间,FABRNet相比其他方法在10个不同数量的测试数据集中消耗的时间都是最少的,而且数据量越大,与其他三种方法差异就越明显,而传统的KNN方法随着数据量的增大其效率会有所下降;图5(b)是在GPU上运行所消耗的时间对比,可以看出这四种方法的运行时间消耗趋势基本相似。FABRNet的计算时间消耗无论是在GPU还是CPU上均有明显优势,这是因为其在模型中采用了空洞卷积和双线性插值的方式,使得网络计算量相比其他几种方法有所降低。

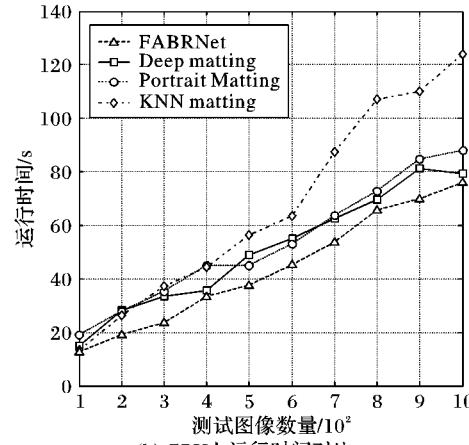
4 结语

本文提出的FABRNet方法是一种高效的图像背景替换方法,基于具有优良图像处理能力的CNN模型构建,将传统的卷积操作替换成空洞卷积操作,并引入双线性插值操作。

该方法和传统CNN相比,空洞卷积能够在不增加参数数量的同时,扩大卷积操作过程中的视野感受区域,而双线性插值操作的引入,也使得处理后的图片更为接近真实图片。实验结果表明,FABRNet方法在效果上和运算速度上均优于对比模型。除此之外,FABRNet方法也有一些待优化之处,如模型参数选取的合理性、图像缩放阶段产生的误差以及生成图像和原图像之间的色差等,这也将是我们下一步深入研究的内容。



(a) CPU上运行时间对比



(b) GPU上运行时间对比

图5 不同方法之间时间消耗比较

Fig. 5 Time consumption comparison of different models



参考文献:

- [1] SHEN X, TAO X, GAO H, et al. Deep automatic portrait matting [C]// ECCV 2016: Proceedings of the 2016 European Conference on Computer Vision. Cham: Springer, 2016: 92–107.
- [2] LI B. Image background replacement method: US, US6912313 [P]. 2005-06-28.
- [3] QIAN R J, SEZAN M I. Video background replacement without a blue screen [C]// ICIP 99: Proceedings of the 1999 International Conference on Image Processing. Piscataway, NJ: IEEE, 1999, 4: 143–146.
- [4] LEVIN A, LISCHINSKI D, WEISS Y. A closed-form solution to natural image matting [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(2): 228–242.
- [5] ZHU B, CHEN Y, WANG J, et al. Fast deep matting for portrait animation on mobile phone [C]// MM '17: Proceedings of the 2017 ACM on Multimedia Conference. New York: ACM, 2017: 297–305.
- [6] XU N, PRICE B, COHEN S, et al. Deep image matting [J/OL]. arXiv: 1703.03872, (2017-04-11) [2017-05-16]. <https://arxiv.org/abs/1703.03872>.
- [7] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640–651.
- [8] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [J/OL]. arXiv: 1511.07122, (2016-04-30) [2017-05-16]. <https://arxiv.org/abs/1511.07122>.
- [9] CHEN L-C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J/OL]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017 [2017-05-26]. <http://ieeexplore.ieee.org/document/7913730/>.
- [10] CHEN L-C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs [J/OL]. arXiv: 1412.7062, (2016-06-07) [2017-04-08]. <https://arxiv.org/abs/1412.7062>.
- [11] CHEN L-C, YANG Y, WANG J, et al. Attention to scale: scale-aware semantic image segmentation [C]// CVPR 2016: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2016: 3640–3649.
- [12] MOESKOPS P, VETA M, LAFARGE M W, et al. Adversarial training and dilated convolutions for brain MRI segmentation [C]// DLMIA 2017, ML-CDS 2017: Proceedings of the 2017 Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer, 2017: 56–64.
- [13] YANG Y, SOINI R F. Background replacement for an image: US, US5574511 [P]. 1996-11-12.
- [14] PENTA S K. Background replacement [C]// SIGGRAPH 2008: Proceedings of the 2008 International Conference on Computer Graphics and Interactive Techniques. New York: ACM, 2008: Article No. 59.
- [15] SWANSON R L, ADOLPHI E J, SURMA M J, et al. Method and apparatus for background replacement in still photographs: US, US7834894 [P]. 2010-11-16.
- [16] CHEN Q, LI D, TANG C-K. KNN matting [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(9): 2175–2188.
- [17] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.
- [18] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// CVPR 2014: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 580–587.
- [19] SWIETOJANSKI P, CHOSHAK A, RENALS S. Convolutional neural networks for distant speech recognition [J]. IEEE Signal Processing Letters, 2014, 21(9): 1120–1124.
- [20] SINGH R, LANCHANTIN J, ROBINS C, et al. DeepChrome: deep-learning for predicting gene expression from histone modifications [J]. Bioinformatics, 2016, 32(17): i639–i648.
- [21] PRASOON A, PETERSEN K, IGEL C, et al. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network [C]// MICCAI 2013: Proceedings of the 2013 International Conference on Medical Image Computing and Computer-Assisted Intervention, LNCS 8150. Berlin: Springer, 2013: 246–253.
- [22] 唐智川, 张克俊, 李超, 等. 基于深度卷积神经网络的运动想象分类及其在脑控外骨骼中的应用 [J]. 计算机学报, 2017, 40(6): 1367–1378. (TANG Z C, ZHANG K J, LI C, et al. Motor imagery classification based on deep convolutional neural network and its application in exoskeleton controlled by EEG [J]. Chinese Journal of Computers, 2016, 40(6): 1367–1378.)
- [23] RHEMANN C, ROTHER C, WANG J, et al. A perceptually motivated online benchmark for image matting [C]// CVPR 2009: Proceedings of the 2009 International Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2009: 1826–1833.
- [24] ABADI M, AGARWAL A, BARHAM P, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems [J/OL]. arXiv: 1603.04467, (2016-03-16) [2017-04-06]. <https://arxiv.org/abs/1603.04467v1>.
- [25] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]// ECCV 2014: Proceedings of the 2014 European Conference on Computer Vision, LNCS 8693. Cham: Springer, 2014: 740–755.
- [26] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J/OL]. arXiv: 1409.1556, (2015-04-10) [2017-03-05]. <https://arxiv.org/abs/1409.1556>.

This work is partially supported by the National Natural Science Foundation of China (61762089, 61363021, 61640306).

ZHANG Hao, born in 1992, M. S. candidate. His research interests include deep learning.

DOU Qiwei, born in 1994, M. S. candidate. His research interests include deep learning.

LUAN Guikai, born in 1993, M. S. candidate. His research interests include deep learning.

YAO Shaowen, born in 1966, Ph. D., professor. His research interests include workflow, Petri network.

ZHOU Wei, born in 1974, Ph. D., associate professor. His research interests include distributed computing, bioinformatics.