

基于非特定类别图像前景主体分割的深度学习算法研究

陈祥龙, 李海军

引用本文

陈祥龙, 李海军. 基于非特定类别图像前景主体分割的深度学习算法研究[J]. 计算机科学, 2024, 51(11A): 231000071-9.

CHEN Xianglong, LI Haijun. [Study on Deep Learning Algorithm for Foreground Subject Segmentation of Non-specific Category Images](#) [J]. Computer Science, 2024, 51(11A): 231000071-9.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[拟态防御中基于ANP-BP的执行体异构性量化方法](#)

ANP-BP Based Executive Heterogeneity Quantification Method in Mimicry Defense
计算机科学, 2024, 51(11A): 231000005-6. <https://doi.org/10.11896/jsjx.231000005>

[基于图神经网络的银行交易欺诈检测方法](#)

Bank Transaction Fraud Detection Method Based on Graph Neural Network
计算机科学, 2024, 51(11A): 240200024-8. <https://doi.org/10.11896/jsjx.240200024>

[基于CNN结合BiGRU的恶意流量分类算法研究](#)

Study on Malicious Traffic Classification Algorithm Based on CNN Combined with BiGRU
计算机科学, 2024, 51(11A): 231100106-9. <https://doi.org/10.11896/jsjx.231100106>

[MB-ATMK:融合属性权重和时序元知识的多行为序列推荐模型](#)

MB-ATMK:Multi-behavior Sequential Recommendation Integrating Attribute Weights and Temporal Meta-knowledge
计算机科学, 2024, 51(11A): 231100047-9. <https://doi.org/10.11896/jsjx.231100047>

[基于参数化量子线路的量子神经网络数据分类](#)

Parameterized Quantum Circuits Based Quantum Neural Networks for Data Classification
计算机科学, 2024, 51(11A): 231200112-7. <https://doi.org/10.11896/jsjx.231200112>

基于非特定类别图像前景主体分割的深度学习算法研究

陈祥龙 李海军

三亚学院信息与智能工程学院 海南 三亚 572022

三亚学院陈国良院士团队创新中心 海南 三亚 572022

(1482103902@qq.com)

摘要 通过在 Mobile-Unet 网络的基础上加入 SENet 通道注意力机制来改进图像前景主体分割算法。改进后的算法引入深度可分离卷积来减小模型参数量,同时利用跳跃连接和多尺度特征融合来提高模型的分割精度。在训练过程中,采用了带空洞卷积的空间金字塔池化模块来增加感受野,提高模型对于大尺度物体的识别能力。实验结果表明,改进后的算法在 PASCAL VOC2012 数据集上达到了 96% 的 MIOU(Modular Input/Output Unit)分割精度,准确率达到 0.971,优于现有的多种图像分割算法,例如 FCN 全卷积神经网络算法。在速度方面,模型对于每张图片的处理时间为 1.7~2.5 s,改进后的算法相对于传统的全卷积神经网络具有更快的推理速度,适合于在移动设备上实现实时图像分割。通过对比实验,比较了改进前和改进后的 Mobile-Unet 模型以及 FCN 模型对于明亮条件下和昏暗条件下图像前景主体分割的效果,并得出了改进后的 Mobile-Unet 模型具有最好效果的结论。最终进行算法的部署,设计了 GUI 可视化操作界面,并生成 .exe 可执行文件。

关键词: 主体分割;神经网络;感受野;参数量;分割精度

中图分类号 TP183

Study on Deep Learning Algorithm for Foreground Subject Segmentation of Non-specific Category Images

CHEN Xianglong and LI Haijun

School of Information and Intelligent Engineering, University of Sanya, Sanya, Hainan 572022, China

Academician Guoliang Chen Team Innovation Center, University of Sanya, Sanya, Hainan 572022, China

Abstract By incorporating SENet channel attention mechanism on the basis of Mobile Unet network, the image foreground subject segmentation algorithm is improved. The algorithm introduces deep separable convolution to reduce the number of model parameters, while utilizing skip connections and multi-scale feature fusion to improve the segmentation accuracy of the model. During the training process, a spatial pyramid pooling module with hollow convolution is used to increase the receptive field and improve the model's recognition ability for large-scale objects. Experimental results show that the improved algorithm achieves 96% MIOU(Modular Input/Output Unit) segmentation accuracy on the PASCAL VOC2012 dataset, with an accuracy rate of 0.971, which is superior to various existing image segmentation algorithms, such as the FCN fully convolutional neural network algorithm. In terms of speed, the processing time of the model for each image is between 1.7 s and 2.5 s. The improved algorithm has a faster inference speed compared to traditional fully convolutional neural networks, making it suitable for real-time image segmentation on mobile devices. Through comparative experiments, the effectiveness of the Mobile Unet models before and after the improvement, as well as the FCN model, in foreground subject segmentation of images under bright and dim conditions is compared, and the conclusion is drawn that the improved Mobile Unet model has the best performance. Finally, the algorithm is deployed, a GUI visualization operation interface is designed, and an .exe executable file is generated.

Keywords Subject segmentation, Neural network, Receptive field, Parameter quantity, Segmentation accuracy

1 引言

计算机视觉领域中,图像分割算法的研究是重要的研究方向之一。前景主体分割是一项具有挑战性的任务,在这个任务中,需要准确地将图像中的前景主体进行分割,以便进行后续的处理和分析。传统的基于规则的图像分割方法在面对复杂场景和多样化图像时效果不佳^[1]。因此,近年来越来越多的研究者开始关注基于深度学习的前景主体分割算法。

前景主体分割在各个领域中有广泛应用,如视频处理、自动驾驶和医学图像分析^[2]等。目前,图像分割的深度学习技术可以分为两种类型:全卷积神经网络(FCN)和编码-解码结构型网络(encoder-decoder)。FCN 通过将全连接层替换为卷积层的方式,能够处理任意大小的输入图像;而编码-解码结构型网络则分为编码和解码两个阶段,分别用于提取特征和还原图像。在 FCN 方面,经典方法有 U-Net, SegNet, DeepLab 等,它们通过上采样或跳连接模块保留语义和空间信息^[3]。另外的方法有 FCN + CRF, PSPNet, UNet++ 等,

通过注意力机制、空洞卷积、空间金字塔池化^[4]等提高准确率和鲁棒性。

Liang-Chieh 等提出了 DeepLab 方法,结合了深度卷积网络、空洞卷积和全连接条件随机场,取得了在语义分割任务上的优秀结果^[5]。此外,Wu 等的研究还引入了金字塔场景解析网络(PSPNet),该网络利用金字塔池化操作来获取不同尺度的上下文信息,从而提升了图像分割技术的性能^[6]。还有一些其他的网络被提出,如 RefineNet,DeepLab v3+ 等,它们采用了改进的结构和技术,能够有效地解决过拟合和干扰等问题,从而提高图像分割的准确性和效率^[7]。

传统方法的模型参数量大,模型训练和预测时间较长,不适合应用于移动端的图像前景分割。并且,对于黑暗环境下的图像前景分割的效果不佳,分割边缘锯齿化。为了解决上述问题,本文使用了 Mobile-Unet 轻量化网络,大大减少了模型的参数量,图像前景分割速度大大提升,平均每张图片处理时间为 1.7~2.5 s。同时在模型中还加入了 SENet 通道注意力机制,来解决黑暗条件下图像前景主体分割边缘

锯齿化的问题,并且还在分割前对图像进行锐化操作,以此来凸显主体特征,提升了分割的准确度,对主体的边缘切割效果更好。

2 相关基础理论

2.1 MobileNet 网络

MobileNet 模型通过层级特征提取和抽象化的方式,获取不同尺度的特征图用于多目标检测。在这个过程中,底层特征提供了更细节和精细的信息,适用于识别占比较小的目标;而高层特征提供了更抽象的信息,适用于识别占比较大的目标。然而,在卷积操作和下采样过程中,MobileNet 网络易于丢失高层特征中细小目标的信息,导致检测不成功^[8]。此外,MobileNet 的基础网络缺乏互补的特征提取过程,导致不同特征层无法充分提取特征,尤其是在处理复杂的多目标检测任务时,对小目标的检测能力较弱^[9]。

2.2 U-Net 网络

U-Net 网络是全卷积神经网络的一种变形,由特征提取网络和上采样网络组成,其网络结构如图 1 所示。

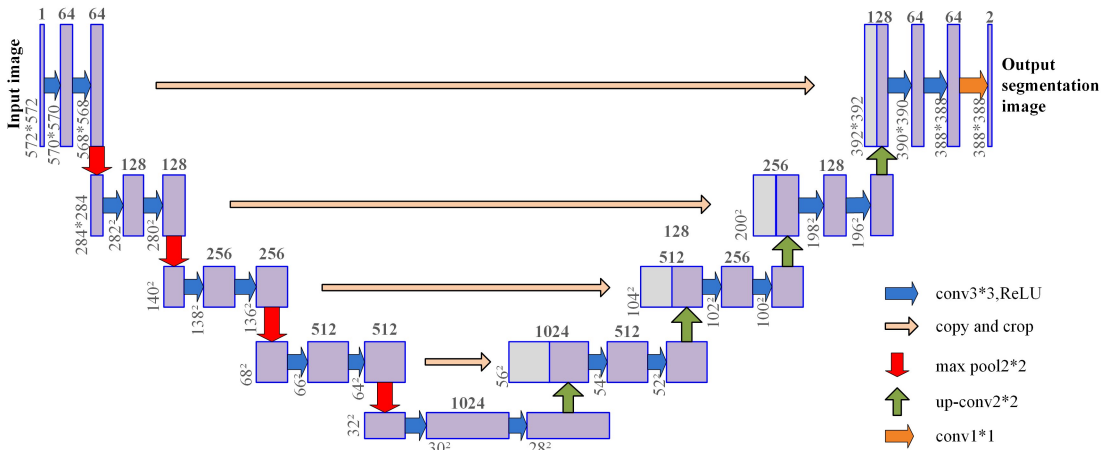


图 1 U-Net 网络结构图^[1]

Fig. 1 Diagram of U-Net network structure^[1]

图 1 中特征提取网络是一个 CNN 卷积神经网络,每个结构包括两个 3×3 卷积层、一个 ReLU 激活函数和一个 2×2 最大池化层。通过池化层的变换,特征通道数会增加一倍。另外,上采样网络采用 2×2 上采样层,用于将特征通道数减少一半,并与对应的特征提取网络的特征图进行拼接,从而生成新的特征图。随后,通过两个 3×3 卷积层加上 ReLU 激活函数的处理,将 64 维通道特征映射到 2 维图像上,从而得到最终的图像分割结果^[10]。整个网络通过不断提取和拼接特征图来处理图像,并产生准确的目标分割结果。

2.3 基于 MobileNet 的 U-Net 网络

本文使用了一种有效的图像显著性检测算法,该算法在 U-Net 网络解码器阶段加入了自底向上的图像显著性检测算法提取的显著图,通过与 MobileNet 结合,可以将高层全局信息和低层局部信息相融合,实现了更加精确的显著图像分割。具体来说,本文算法将自底向上提取的显著图与解码器阶段的特征图进行特征融合,通过生成与特征图相同大小的先验图来缩小显著图。此外,为了提取更关键的通道信息,在每个上采样层中引入了通道注意力模块。最后,采用加权结构的损失函数对神经网络进行优化学习。

该模型结合了 MobileNet 和 U-Net 的网络结构,包括编码器和解码器阶段,如图 2 所示,其中蓝色模块表示编码器阶段的下采样,紫色模块表示解码器阶段的上采样,黄色模块表示自底向上的图像显著性检测算法提取的显著图,橙色模块表示注意力机制模块。该模型通过特征融合的方式将显著图和解码器阶段的特征图连接起来。具体的连接方式是将显著图进行多尺度缩小,使其与 U-Net 解码器阶段的特征图大小相匹配,然后通过 Concat 通道连接的方式将它们融合在一起^[11]。这样做可以增加显著图的细节信息。此外,模型还在每一层特征图中添加了 4 个不断增大分辨率的先验特征图,共加入了 4 层先验特征,逐渐丰富了图像信息。网络模型参数如表 1 所列。

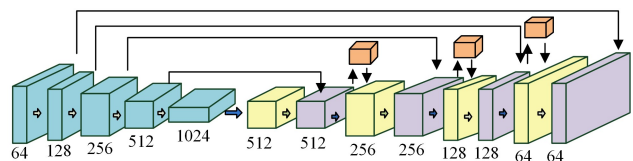


图 2 模型示意图

Fig. 2 Model diagram

表 1 网络模型参数
Table 1 Network model parameters

layer	Filters	Size	Input	Output	
dsp_1	conv_1	64	1*1/1	512*512*3	512*512*64
	Dwise_1	64	3*3/1	512*512*64	512*512*64
	max_1		2*2/2	512*512*64	256*256*64
dsp_2	Dwise_2	128	3*3/1	256*256*64	256*256*128
	Dwise_3	128	3*3/1	256*256*128	256*256*128
	max_2		2*2/2	256*256*128	128*128*128
dsp_3	Dwise_4	256	3*3/1	128*128*128	128*128*256
	Dwise_5	256	3*3/1	128*128*256	128*128*256
	Dwise_6	256	3*3/1	128*128*256	128*128*256
	max_3		2*2/2	128*128*256	64*64*256
	Dwise_7	512	3*3/1	64*64*256	64*64*512
dsp_4	Dwise_8	512	3*3/1	64*64*512	64*64*512
	Dwise_9	512	3*3/1	64*64*512	64*64*512
	max_4		2*2/2	64*64*512	32*32*512
	Dwise_10	512	3*3/1	32*32*512	32*32*512
dsp_5	Dwise_11	512	3*3/1	32*32*512	32*32*512
	Dwise_12	512	3*3/1	32*32*512	32*32*512
	max_5		2*2/2	32*32*512	16*16*512
conv_trans_1	512	3*3/2	16*16*512	32*32*512	

其中激活层中使用的激活函数是 ReLU 函数,如式(1)所示,ReLU 函数具有几个明显的优点:首先,相比指数运算,ReLU 函数只需进行简单的大小判断,因此计算速度更快;其次,ReLU 函数在大约一半的情况下输出为 0,有效地增加了网络特征的稀疏性;最后,ReLU 函数具有分段特性,正半轴的导数始终为 1,避免了梯度消失的问题^[12]。这些优势使得 ReLU 函数在处理各种非线性问题时更加高效,并提高了神经网络的泛化能力。

$$f(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases} \quad (1)$$

在训练深度神经网络时,由于数据分布的不稳定性可能导致梯度消失的问题。为了解决这个问题,批标准化被引入。批标准化通过标准化网络层的输出值,使其具有正态分布。这样做的好处是避免了激活函数的输入过多靠近饱和区域,从而减弱了梯度消失的现象,并提高了网络的收敛速度^[13]。同时,批标准化还能够增加网络的稳定性和泛化能力。虽然批标准化会增加一定的计算复杂性,但它对于训练深度神经网络是非常有益的。对于输入数据集,其标准化的计算式为:

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (2)$$

$$\delta_B = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (3)$$

$$x_i' = \frac{x_i - \mu_B}{\sqrt{\delta_B^2 + \epsilon}} \quad (4)$$

3 改进的 Mobile-Unet 网络算法

3.1 数据集选择及数据预处理

3.1.1 数据集选择

本文使用的数据集是 PASCAL VOC 数据集¹⁾。

PASCAL VOC 是一种经典的图像识别和分割数据集,其中包含约 2.5 万张训练图像和约 1.4 万张测试图像。该数据集的标注包含像素级标注和边界框标注,并且数据集共有 4 大类,即 vehicle,household,animal,person,有 20 小类,具体分类图如图 3 所示。

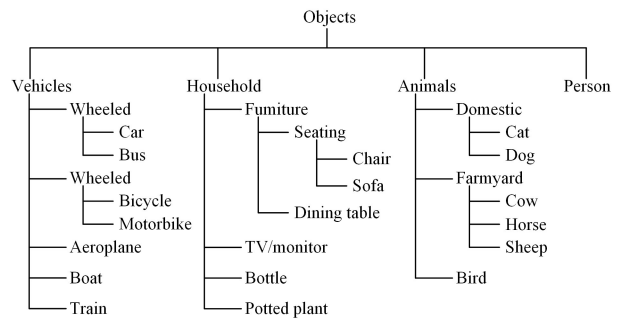


图 3 数据集分类

Fig. 3 Dataset classification

部分数据展示如图 4 所示。



图 4 部分数据展示

Fig. 4 Partial data display

3.1.2 数据预处理

本文对数据的预处理的步骤如下。

1)数据读取和预处理:将原始图像读入内存,并进行一些必要的预处理操作,如大小调整、归一化等。

2)数据集增强:为了提高数据集的多样性,防止过拟合,需要对训练数据集进行数据增强操作。数据增强包括随机裁剪、随机翻转、随机旋转等操作,如图 5 所示。



图 5 图像随机裁剪

Fig. 5 Random cropping of images

3)数据标注转换:将标注转换为模型可用的格式,如像素级标注转换为掩膜(mask)格式,如图 6 所示。

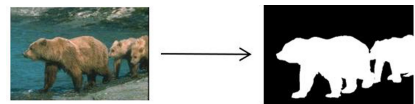


图 6 数据标注转换

Fig. 6 Data annotation conversion

4)批量处理:将预处理后的数据分为小批量进行处理,以便于模型进行训练。

本文选择了批量归一化方法,批量归一化方法是一种有效的逐层归一化方法。对于一个深层神经网络,令第 l 层的净输入为 $z(l)$,神经元的输出为 $a(l)$,即:

$$a^{(l)} = f(z^{(l)}) = f(Wa^{(l-1)} + b) \quad (5)$$

¹⁾ <http://host.robots.ox.ac.uk/pascal/VOC/>

其中, $f(\cdot)$ 是激活函数, W 和 b 为可学习参数。

在实际操作中, 由于无法对整个数据集进行统计量计算, 我们只能使用当前小批量数据的均值和方差来估计归一化所需的参数。然而, 直接将输入数据归一化到 $[-1, 1]$ 区间可能会限制激活函数在线性变化区间的取值范围。为了避免这种负面影响, 我们可以引入缩放和平移变换来调整神经网络的取值范围。通过这种调整, 我们可以保证归一化后的数据仍然具有足够的有效范围, 从而提高神经网络的表示能力^[14]。因此, 我们可以通过添加缩放和平移变换来对数据进行归一化处理, 该方式不会影响网络性能。

3.2 模型改进

本文通过对 MobileNet 的 U-Net 模型进行改进, 加入了 SENet 通道注意力机制, 来提高特征提取的效果。在解码阶段, 首先对编码阶段提取到的有效特征层进行上采样处理, 以恢复细节信息。然后通过跳跃连接将两部分获取的特征层进行融合, 这样做的目的是增强模型对图像细节的还原能力和表现能力。模型的前三次跳跃连接使用了 SENet 通道注意力模块, 以确保模型持续关注视频序列中的关键特征, 从而提高对目标及其阴影分割的准确性和稳健性。最后, 融合后的特征层经过两次卷积处理, 并嵌入 SENet 通道注意力模块。这种改进能够全面提取图像特征并关注重要信息, 从而提高模型的性能和精度。具体结构如图 7 所示。

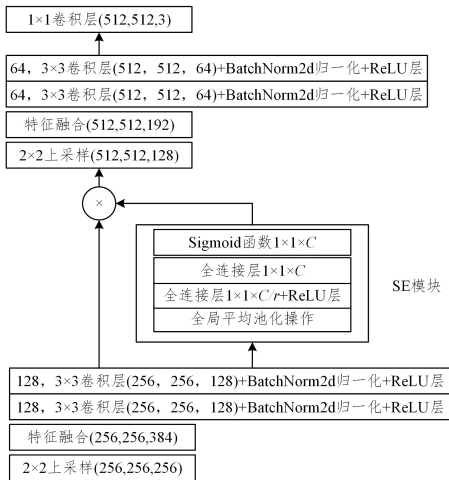


图 7 改进后的网络结构图

Fig. 7 Diagram of improved network structure

通过使用 SENet 通道注意力机制模块, 网络可以将注意力集中在输入特征层上, 从而更好地识别出最重要的信息。SE(squeeze-and-excitation)是一种压缩-激活网络, 它的工作原理如图 8 所示。首先, 模型对尺寸为 $C' \times H' \times W'$ 的输入张量 X 进行卷积运算, 生成尺寸为 $C \times H \times W$ 的特征图 U 。接下来, 对这个特征图进行了 3 个操作: 压缩操作、激活操作和缩放操作。

在 Squeeze 步骤中, 将输入的特征层进行一次卷积, 将特征图压缩为一个通道, 生成一个 $1 \times 1 \times C$ 的特征图。Excitation 步骤则是对 Squeeze 生成的特征图进行激活, 使用一个 sigmoid 激活函数将所有的元素处理为 $0 \sim 1$ 之间的值。在 Scale 步骤中, 将激活后的特征图乘上原始特征图, 按比例加权不同通道上的特征, 从而得到了最终的特征图。

1) Squeeze 操作

采用全局平均池化操作, 可以将输入特征图的每个通道

的空间特征进行压缩, 并得到一个通道的全局信息。这样做可以获得具有 C 个通道的 $1 \times 1 \times C$ 输出特征图, 其中包含了全局感受野的相关信息, 如式(6)所示:

$$Z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (6)$$

其中, F_{sq} 表示 squeeze 操作, H 和 W 分别表示图中 U 的宽和高, u_c 表示 U 的第 c 个通道的输出。

2) Excitation 操作

通过第 1 步的全局平均池化操作会得到一个结果 z , 然后对结果 z 进行 excitation 操作, 目的是获得输入特征层每一个通道的权值, 学习通道之间的相关性。该过程包含两个全连接层和两次激活操作。首先, 将输入特征图 z 与维度为 $C/r \times C$ 的 W_1 矩阵进行乘法运算, 其中 r 是通道压缩倍率。这样可以将 $1 \times 1 \times C$ 的输入特征图压缩为 $1 \times 1 \times C/r$ 。然后, 将压缩后的特征图输入到第一个全连接层, 经过全连接操作后, 再通过 ReLU 激活函数进行激活。在输出后进行第 2 个全连接层操作, 涉及到一个权重矩阵 W_2 (维度为 $C \times C/r$), 将输出值与 W_2 相乘, 然后通过 Sigmoid 函数进行激活, 将值限定在 0 到 1 之间, 这样得到的输出结果 s 的维度是 $1 \times 1 \times C$, 如式(7)所示:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (7)$$

3) Scale 操作

Scale 操作通过乘法运算, 将在第 2 步中得到的权重 s 逐通道地应用到之前的特征上。这样做的目的是加强重要的特征信息, 同时减弱无用的特征信息, 如式(8)所示:

$$x = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (8)$$

在 scale 操作中使用 F_{scale} 来表示该操作, 其中 s_c 表示在第 2 步中获得的权重, u_c 表示提取到的特征。将这些权重与特征相乘, 得到最终的结果 x 。

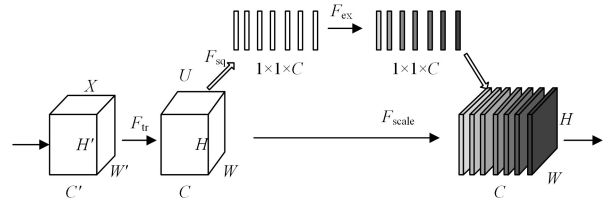


图 8 SENet 注意力机制模块^[1]

Fig. 8 SENet attention mechanism module^[1]

3.3 改进后的 Mobile-Unet 模型的使用

预测输出由数据增强、生成显著图、初步分割和背景透明化这 4 部分组成, 流程如图 9 所示。

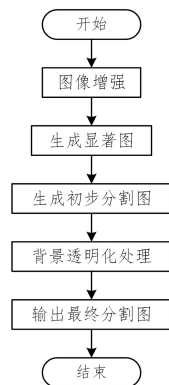


图 9 预测输出流程示意图

Fig. 9 Schematic diagram of prediction output process

3.3.1 图像增强

在使用模型进行分割前,通过调用 PIL. ImageEnhance 库,设置 brightness, contrast, color, sharpness 参数值将需要进行分割的图片进行锐化,使得图像中的显著性目标更加显著,便于显著图的生成,相当于给图片进行锐化,如图 10 所示。

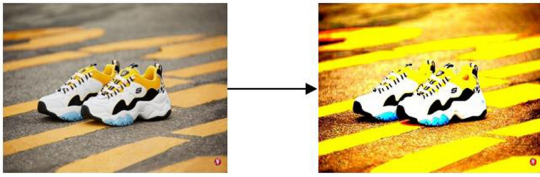


图 10 图像增强示意图

Fig. 10 Image enhancement diagram

3.3.2 生成显著图

首先对输入的图像进行重塑,将需要分割的图像输入到网络中根据权重进行切割,然后输出得到特征点矩阵(显著区域的像素点),根据特征点矩阵对空矩阵进行像素点上色,得到最终的显著图,如图 11 所示。

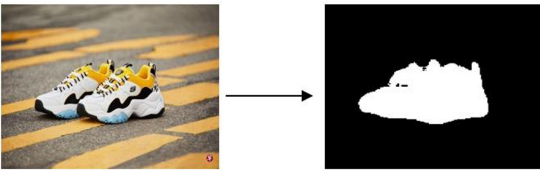


图 11 取模示意图

Fig. 11 Schematic diagram of mold taking

3.3.3 初步分割

首先将得到的显著图和原图通过 CV2. bitwise_or 进行图像算数运算中的按位或运算,得到原图中的显著部分,再根据显著图将原图的显著区域提取出来,得到初步分割图,如图 12 所示。

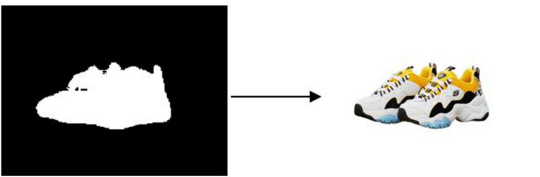


图 12 初步分割示意图

Fig. 12 Preliminary segmentation diagram

3.3.4 背景透明化

由于第四个通道即图像背景要修改的透明度,首先将三通道图像转换为四通道图像,并将第四通道的色彩值设置为大于浅灰色,即像素点变为透明块,最终可输出背景为透明的最终分割图,如图 13 所示。

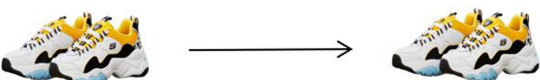


图 13 背景透明化

Fig. 13 Background transparency

3.3.5 输出结果

使用 6 张图片进行模型输出,得到的分割结果如图 14 所示(第二列为第一列的分割结果,第四列为第三列分割结果)。



图 14 输出结果

Fig. 14 Output results

3.3.6 Mobile-Unet 算法流程

Mobile-Unet 算法流程的伪代码如图 15 所示,算法输入为 PASCAL VOC 数据集,首先遍历数据集中的每张图片,分别进行图像归一化(第 2 行),图像数字化(第 5 行)和图像增强(第 8 行)的数据预处理操作,然后进行模型训练(第 12 行)并设置 learning rate=0.00, batch size=32,通过打乱的数据集更有利于训练,采用自底向上的图像显著性检测算法(第 16 行)来提取显著图,最后训练完毕得到 .h5 权重文件。

```

流程 1 基于 Mobile-Unet 网络的图像分割
Input: PASCAL VOC 数据集
Output: .h5 权重文件
1. while 依次遍历所有图片 do
2.   function 图像归一化
3.     采用批量归一化进行图像归一化操作
4.   end function
5.   function 图像数字化
6.     将图像使用像素点来表示
7.   end function
8.   function 图像增强
9.     随机裁剪、随机翻转、随机旋转
10.  end function
11. end while
12. function Mobile-Unet 模型训练
13.   learning_rate=0.001
14.   batch_size=32
15.   90% 训练,10% 预测估计
16.   将训练图像数据集输入模型
17. end function
18. 训练完毕得到 .h5 权重文件
    
```

图 15 Mobile-Unet 算法流程伪代码

Fig. 15 Pseudocode of Mobile-Unet algorithm process

3.3.7 模型处理图片流程

模型处理图片流程伪代码如图 16 所示,模型的输入为需要进行主体分割的图像,首先依次遍历所有需要分割的图像,然后进行数据增强(第 2 行),通过调用 PIL. ImageEnhance 库将图片进行锐化操作,然后利用训练好的权重文件进行图像的前景主体分割(第 5 行),最后进行图像的背景透明化(第 8 行),将第四通道的色彩值设置为大于浅灰色,最终输出主体图像。

```

流程 2 利用训练好的模型处理图像
Input: 需要进行主体分割的图像
Output: 透明背景主体图像
1. while 依次遍历所有需要分割的图像 do
2.   function 数据增强
3.     调用 PIL. ImageEnhance 库, 设置参数值
4.   end function
5.   function 主体分割
6.     使用. h5 权重文件进行图像前景主体分割
7.   end function
8.   function 背景透明化
9.     将图像第四通道色彩值设为大于浅灰色
10.  end function
11. end while
12. 得到最终主体图像
    
```

图 16 模型处理图片流程的伪代码

Fig. 16 Pseudocode of model processing image process

3.3.8 模型整体流程

如图 17 所示, 首先对训练集进行图像预处理, 包括归一化、数字化和增强操作。其次, 使用预处理后的数据集进行网络训练, 得到权重文件。然后, 利用权重文件对输入图像进行特征提取和显著性检测, 通过显著图融合获取最终的显著图。最后, 根据显著图进行初步切割, 并进行背景透明化处理, 得到分割结果。

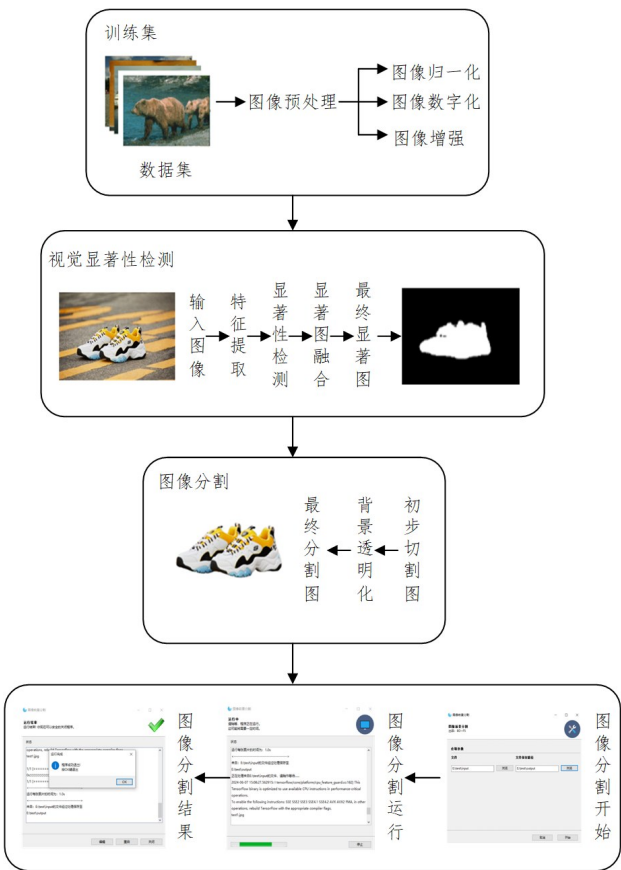


图 17 模型整体流程

Fig. 17 Overall model process

4 模型对比实验与评价指标

本文的实验环境如表 2 所列。

表 2 实验环境

Table 2 Experimental environment

处理器	Intel(R) Core(TM) i7-9750H CPU				
硬件配置	20GB 内存	512GB 固态硬盘			
操作系统	Windows 10 64 位				
软件平台	Pycharm				
深度学习框架	TensorFlow				
Python 版本	Python3.6				
Python 库	TensorFlow	TensorFlow-	NumPy	OpenCV-	Keras
	1.13.2	GPU 1.13.1	1.19.5	Python 4.5.1.48	2.1.5

4.1 改进后的 Mobile-Unet 模型训练

4.1.1 配置参数

1) 学习率 (learning rate)

在训练过程中, 学习率是一个重要的超参数, 控制着每次参数更新的大小。本文使用较小的学习率, 即 0.001, 并根据训练曲线进行调整。模型自带学习率为 0.01, 训练过程中发现训练速度较快, 但是网络不能很好地收敛, 在接近局部最优的附近震荡, 无法更新到局部最优处。因此, 经过对比实验, 将学习率设置为 0.001。

学习率为 0.01 时训练第二轮得到的权重文件。

如图 18 所示, 训练集损失值为 0.166, 测试集损失值为 0.439。

ep002-loss0.166-val_loss0.439.h5

图 18 权重文件 1

Fig. 18 Weight File 1

学习率为 0.001 时训练第二轮得到的权重文件。

如图 19 所示, 训练集损失值为 0.179, 测试集损失值为 0.385。

ep002-loss0.179-val_loss0.385.h5

图 19 权重文件 2

Fig. 19 Weight file 2

经过上述对比可知, 学习率为 0.001 时测试集的损失值小于学习率为 0.01 时测试集的损失值, 因此学习率为 0.001 更有利于网络更新到局部最优处。

2) 批量大小 (batch size)

批量大小指每次训练模型时使用的样本数量。由于 MobileNet 网络相对较轻, 因此可以选择较大的批量大小 (例如 32 或 64), 以提高训练速度。本文中批量大小为 32, 模型自带批量大小为 64, 训练过程中发现实验配置不高, 因此降低批量大小为 32, 增长了运行时间, 但提高了训练的稳定性。

3) 训练轮数 (number of epochs)

由于 MobileNet 网络相对较轻, 通常只需要较少的训练轮数。本文通过观察验证集的损失和准确率, 确定了最佳的训练轮数范围为 15~25 个 epochs。在这个范围内, 训练的模型表现出了较高的准确率和较低的损失率。

4) 打乱数据

打乱的数据更有利于训练, 使用 numpy 的 random 函数来随机打乱数据, 使得数据的 90% 用于训练, 10% 用于预测估计。

4.1.2 迁移学习

模型训练时使用了迁移学习,迁移学习的本质是通过搜索与目标领域相似的数据,然后增强和匹配这些数据,以在预训练模型的基础上实现参数和知识的迁移。迁移学习可以解决数据稀缺问题。学习速度问题和特征表示问题等。这种方法可以大大减少训练时间和数据量,在支持新任务方面具有重要意义^[15]。迁移学习可以通过基于实例的迁移、特征选择和映射,模型参数初始化以及多任务学习来实现。

迁移学习的主要步骤如下。

1)选择源任务和目标任务:首先需要选定一个与目标任务相关的源任务。源任务通常是已经完成的或者具有丰富学习经验的任务。

2)提取特征表示:从源任务和目标任务的数据中提取特征表示。通常通过使用预训练的深度学习神经网络模型,将数据映射到低维特征空间^[16]。

3)迁移知识:将源任务的学习经验应用于目标任务。常见的迁移方式包括特征迁移(使用源任务的特征表示)、参数迁移(使用源任务的参数初始化目标任务的模型)、示例迁移(使用源任务的样本数据)^[17]等。

4)自适应调整:对目标任务进行微调或调整,以适应特定的目标任务。该过程通常使用目标任务的有标签数据进行模型训练和优化^[18]。

本文通过对样本进行加权,赋予比较重要的样本较大的权重。这样就可以在小数据上实现迁移,使数据具有高可靠性。

4.1.3 权重文件

每个 epoch(训练轮数)之后都会生成一个, h5 格式的权重文件,如图 20 所示。保存的权重文件包含了训练好的神经网络各层的权值,加载权值后可以直接应用,无需再依赖于训练集进行训练。

```
ep002-loss0.330-val_loss0.345.h5
ep003-loss0.076-val_loss0.419.h5
ep003-loss0.385-val_loss0.409.h5
ep006-loss0.041-val_loss0.571.h5
ep011-loss0.022-val_loss0.842.h5
```

图 20 权重文件 3

Fig. 20 Weight file 3

4.2 分割效果对比实验

4.2.1 明亮条件下的分割效果对比

1)改进前的 Mobile-Unet 分割效果图,如图 21 所示。



图 21 明亮条件下改进前的 Mobile-Unet 分割效果图

Fig. 21 Mobile Unet segmentation effect before improvement under bright conditions

效果评价:模型识别主体较为准确,但受到背景影响,主体分割时主体有部分缺失。

2)改进后的 Mobile-Unet 分割效果图,如图 22 所示。



图 22 明亮条件下改进后的 Mobile-Unet 分割效果图

Fig. 22 Mobile Unet segmentation effect after improvement under bright conditions

效果评价:模型识别主体准确,边缘分割效果好,分割精确度高。

3)FCN 模型分割效果图,如图 23 所示。



图 23 明亮条件下 FCN 模型分割图

Fig. 23 FCN model segmentation diagram under bright conditions

效果评价:模型识别主体较为准确,对于物体边缘的分割效果不如改进后的 Mobile-Unet。

4.2.2 昏暗条件下的分割效果对比

1)改进前的 Mobile-Unet 分割效果图,如图 24 所示。



图 24 昏暗条件下改进前的 Mobile-Unet 分割效果图

Fig. 24 Mobile Unet segmentation effect before improvement under dim conditions

效果评价:模型能够识别到主体附近,但对于昏暗背景下的部分主体没有做到准确切割,并且边缘切割效果欠佳。

2)改进后的 Mobile-Unet 分割效果图,如图 25 所示。



图 25 昏暗条件下改进后的 Mobile-Unet 分割效果图

Fig. 25 Mobile Unet segmentation rendering after improvement under dim conditions

效果评价:模型识别主体准确,对于昏暗条件下的主体分割精确度高,边缘分割效果好。

3)FCN 模型分割效果图,如图 26 所示。



图 26 昏暗条件下 FCN 模型分割图

Fig. 26 FCN model segmentation diagram under dim conditions

效果评价:模型识别主体较为准确,但是对于昏暗条件下的主体分割效果不佳,并且物体边缘的分割效果不如改进后的 Mobile-Unet。

4.3 评价指标

本文分别采用了 7 种评价指标。1) MIOU (Mean Intersection over Union): 均交并, 指预测结果与真实标签的重叠面积除以二者并集面积的平均值, 该指标越高越好; 2) MSE (Mean Square Error): 均方差, 指衡量预测值与实际值之间的平均误差的平方值, 该指标越低越好; 3) Average_time: 处理每张图像的平均时间, 该指标越低越好; 4) loss: 训练集的损失值, 该指标越低越好; 5) Val_loss: 测试集的损失值, 该指标越低越好; 6) Acc (accurate): 训练集的准确值, 该指标越高越好; 7) Val_acc: 测试集的准确值, 该指标越高越好。经过实验得到改进后的 Mobile_Unet 模型的对应指标, 具体指标如表 3 所列。

表 3 改进后的 Mobile_Unet 模型评价指标

Table 3 Improved Mobile_unet model evaluation index

MIOU	MSE	Average_time	Loss	Val_loss	Acc	Val_acc
0.96	0.056	1.7~2.5 s	0.154	0.165	0.971	0.965

4.4 模型性能指标对比

表 4 列出了改进前的 Mobile_Unet 模型, 改进后的 Mobile_Unet 模型与 FCN 模型各指标的对比。其中, MIOU (Mean Intersection over Union) 为均交并; MSE (Mean Square Error) 为均方差; Average_time 为处理每张图像的平均时间; loss 为训练集的损失值; Acc 为训练集的准确值。

表 4 模型性能指标对比

Table 4 Comparison of model performance indicators

	改进前 Mobile-Unet	改进后 Mobile-Unet	FCN
loss	0.156	0.154	0.161
MIOU	0.95	0.96	0.93
ACC	0.969	0.971	0.966
Average_time/s	1.9	1.7	2.1
MSE	0.059	0.056	0.064

表 4 中, 改进后的 Mobile-Unet 与改进前的 Mobile_Unet 相比, loss 值低了 0.002, MIOU 提高了 0.01, ACC 提高了 0.002, 平均处理时间缩短了 0.2 s, MSE 降低了 0.003。

改进后的 Mobile-Unet 与 FCN 相比较, loss 值低了 0.007, MIOU 提高了 0.03, ACC 提高了 0.005, 平均处理时间缩短了 0.4 s, MSE 降低了 0.008。

各模型性能对比柱状图如图 27 所示。

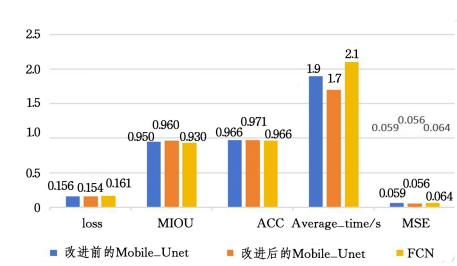


图 27 各模型性能对比柱状图

Fig. 27 Bar chart of performance comparison of different models

综上, 改进后的 Mobile-Unet 模型分割效果更好, 分割准确率更高, 模型更加轻量化, 处理图片的速度更快。

5 算法工程化

项目工程化, 可以用 Python 进行部署, 用 PyInstaller 工

具将 python 解析器和脚本打包成离线型可执行文件。个性化设计 GUI 界面, 有单张图像分割或同时多张图像分割操作供用户选择, 模型大小不超过 100 MB, 不占内存空间, 方便操作。

6 结论与展望

本文的主要创新点如下: (1) 在模型的解码阶段引入 SENet 通道注意力机制, 以提高特征提取的效果; (2) 对数据集进行数据增强, 使用随机旋转剪切等方法来提高数据集的多样性, 防止过拟合; (3) 在对输入图片进行模型预测阶段, 对输入图片进行锐化操作, 使得图像主体更加明显。

本文通过引入 SENet 通道注意力机制来改进 Mobile-Unet 图像分割模型。该改进使网络能够集中注意力于输入特征层, 以更好地识别和提取最重要的信息。在进行了明亮条件和昏暗条件下的图像主体分割对比实验后, 得知改进后的模型在图像边缘分割效果方面表现出色, 并且在 MSE、MIOU 和损失的数值上均优于改进前的 Mobile-Unet 模型。与 FCN 网络相比, 改进后的 Mobile-Unet 模型性能更好, 处理图像的速度也更快。最后, 利用 PyInstaller 工具将模型打包成离线可执行文件。总体而言, 改进后的模型能够有效地处理非特定类别图像的分割任务, 并提高图像边缘分割的精确程度, 同时具备快速处理图像的能力, 有望应用于视频图像的前景主体分割。

结束语 未来我们可以进一步利用庞大的数据集进行多次训练实验, 以找到最佳的参数配置, 来提高图像分割的准确性和速度。同时, 我们将继续改进网络中的特征提取算法, 以实现更准确地分割细小的图像前景对象。基于 Mobile-Unet 的图像分割算法具有广泛的应用场景, 包括无人驾驶、医疗影像和虚拟现实等领域。未来, 我们将继续将该算法应用于各种应用场景, 并探索其在视频图像前景分割方面的潜力, 实现背景去除和背景替换等操作。

参考文献

- [1] LIU Y, YANG S. Shadow segmentation of indoor moving objects based on improved UNet network [J]. Computer System Application, 2022, 31(12): 412-419.
- [2] MENG M Z, LI L, HE G Y, et al. Preliminary study on the degradation effect of MobileNetV2 on four types of breast X-ray BI-RADS lesions [J]. Journal of Clinical Radiology, 2022, 41(10): 1868-1873.
- [3] GUOFENG M. A Small Target Detection Method Based on the Improved FCN Model [J]. Advances in Multimedia, 2022, 2022.
- [4] YIMING L, ZHENGLE W, RUJIA W, et al. Flooding-based MobileNet to identify cucumber diseases from leaf images in natural scenes [J]. Computers and Electronics in Agriculture, 2023, 213.
- [5] CHEN L C, PAPANDEOV 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): 834-838.

- [6] WU D,ZHAO J,WANG Z. AM-PSPNet:Pyramid Scene Parsing Network Based on Attentional Mechanism for Image Semantic Segmentation[C]//ICPCSEE Steering Committee. Abstracts of the 8th International Conference of Pioneering Computer Scientists,Engineers and Educators(ICPCSEE 2022). Part I. Springer,2022:444.
- [7] ZHANG X,YAO Q A,ZHAO J,et al. Overview of Full Convolutional Neural Network Image Semantic Segmentation Methods [J]. Computer Engineering and Applications, 2022,58(8): 45-57.
- [8] HUANG P,ZHENG Q,LIANG O. Overview of Image Segmentation Methods [J]. Journal of Wuhan University(Science Edition),2020,66(6):519-531.
- [9] HUANG S P,LIU H N,ZHOU K S,et al. Zebra crossing segmentation based on improved Unet [J]. Intelligent Computer and Applications,2020,10(11):61-64,69.
- [10] SUN L X. Coal gangue image recognition based on convolutional neural network [J]. Computer Knowledge and Technology, 2020,16(21):16-18,22.
- [11] LI Y Q. Research and Implementation of Image Saliency Detection Algorithm Based on Attention Mechanism and U-Net [D]. Beijing:Beijing Jiaotong University,2020.
- [12] WANG P,GAO C,ZHU L,et al. Ischemic stroke lesion segmentation algorithm based on 3D deep residual network and cascaded U-Net [J]. Computer Application, 2019, 39(11): 3274-3279.
- [13] DING R J,GAO F F,XING L. An Intelligent Routing Strategy for the Internet of Things Based on Deep Reinforcement Learning [J]. Journal of the Internet of Things,2019,3(2):56-63.
- [14] TIAN X,WANG L,DING Q. Overview of Image Semantic Segmentation Methods Based on Deep Learning [J]. Journal of Software Science,2019,30(2):440-468.
- [15] ZHANG M Y. Research on Image Segmentation Based on Deep Learning [D]. Jilin:Jilin University,2017.
- [16] CHEN H X. Image Semantic Segmentation Based on Convolutional Neural Networks [D]. Hangzhou: Zhejiang University, 2016.
- [17] LIU S T,YIN F L. Image segmentation methods based on graph cuts and their new progress [J]. Journal of Automation,2012, 38(6):911-922.
- [18] HE J,GE H,WANG Y F. Overview of Image Segmentation Algorithms [J]. Computer Engineering and Science,2009,31(12): 58-61.



CHEN Xianglong, born in 2001, post-graduate, is a member of CCF (No. Q5078G). His main research interests include computer vision and data mining.



LI Haijun, born in 1968, Ph.D, associate professor, master's supervisor, is a member of CCF (No. F7747M). His main research interests include computer vision and data mining.