

融合注意力机制的移动端人像分割网络

周鹏, 姚剑敏, 林志贤, 严群, 郭太良

引用本文:

周鹏, 姚剑敏, 林志贤, 等. 融合注意力机制的移动端人像分割网络[J]. 液晶与显示, 2020, 35(6): 547–554.

ZHOU Peng, YAO Jian-min, LIN Zhi-xian, et al. Mobile-based portrait segmentation network with attention mechanism[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(6): 547–554.

在线阅读 View online: <https://doi.org/10.3788/YJYXS20203506.0547>

您可能感兴趣的其他文章

Articles you may be interested in

基于局部期望最大化注意力的图像降噪

Local expectation-maximization attention network for image denoising

液晶与显示. 2020, 35(4): 350–359 <https://doi.org/10.3788/YJYXS20203504.0350>

基于注意力掩模融合的目标检测算法

Object detection algorithm based on attention mask fusion

液晶与显示. 2019, 34(8): 825–833 <https://doi.org/10.3788/YJYXS20193408.0825>

多尺度特征融合与极限学习机结合的交通标志识别

Traffic sign recognition based on multi-scale feature fusion and extreme learning machine

液晶与显示. 2020, 35(6): 572–582 <https://doi.org/10.3788/YJYXS20203506.0572>

基于U-Net的多图谱标签融合算法

Multi-atlaslabel fusion based on U-Net

液晶与显示. 2019, 34(11): 1091–1103 <https://doi.org/10.3788/YJYXS20193411.1091>

基于深度卷积生成对抗网络的图像识别算法

Image recognition algorithms based on deep convolution generative adversarial network

液晶与显示. 2020, 35(4): 383–388 <https://doi.org/10.3788/YJYXS20203504.0383>

文章编号:1007-2780(2020)06-0547-08

融合注意力机制的移动端人像分割网络

周 鹏¹, 姚剑敏^{1,2 *}, 林志贤¹, 严 群^{1,2}, 郭太良¹

(1. 平板显示技术国家地方联合工程实验室, 福州大学 物理与信息工程学院, 福建 福州 350108;

2. 晋江市博感电子科技有限公司, 福建 晋江 362200)

摘要: 现有的移动端人像分割网络存在分割精度差、分割边缘模糊等问题。为此, 提出了一种融合注意力机制的轻量化人像分割网络。首先, 利用 MobileNetV2 网络提取图像特征。然后对注意力模块 NLNet (Non-local neural networks) 进行轻量化处理, 随后将优化过的注意力模块嵌入到四层解码网络中。利用融合注意力机制的解码网络自适应地学习有效特征, 最后通过 SoftMax 层得到人像分割结果图。同时改进了损失函数, 引入多损失函数 (Multi-Loss), 使网络更容易收敛。解码网络融合注意力机制的方式使得轻量化网络可以在语义分割任务上取得较好的效果。实验结果表明, 模型在 550 张自采集的人像测试集上达到了 92.29% 的交并比 (MeanIOU), 单张图片在 Inter(R) Core i5 CPU 上的分割时间为 0.74 s。与传统的人像分割网络相比, 研究网络的分割精度和分割速度优势明显, 适合应用于移动端设备。

关 键 词: 人像分割; 注意力机制; 轻量化; 卷积神经网络

中图分类号: TP391.41 **文献标识码:** A **doi:** 10.3788/YJYXS20203506.0547

Mobile-based portrait segmentation network with attention mechanism

ZHOU Peng¹, YAO Jian-min^{1,2 *}, LIN Zhi-xian¹, YAN Qun^{1,2}, GUO Tai-liang¹

(1. *Nation & Local United Engineering Laboratory of Flat Panel Display Technology, College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China;*

2. *Jinjiang RichSense Electronic Technology Co., Ltd., Jinjiang 362200, China*)

Abstract: Existing mobile terminal segmentation network has problems such as poor segmentation accuracy and segmentation edge blurring. To this end, a lightweight portrait segmentation network that incorporates attention mechanisms is proposed. Firstly, the image features are extracted by MobileNetV2 network. Then, the attention module NLNet (Non-local neural networks) is lightened, and

收稿日期: 2019-09-12; **修订日期:** 2019-10-17.

基金项目: 国家重点研发计划课题 (No.2016YFB0401503); 广东省科技重大专项 (No.2016B090906001); 福建省科技重大专项 (No.2014HZ0003-1); 广东省光信息材料与技术重点实验室开放基金资助项目 (No.2017B030301007)

Supported by National Key R & D Project (No.2016yfb00401503); Guangdong Science and Technology Major Project (No.2016b09096001); Fujian Science and Technology Major Project (No.2014hz0003-1); Guangdong Optical Information Materials and Technology Key Laboratory Open Fund Support Project (No.2017b0301007)

* 通信联系人, E-mail: yaojm@fzu.edu.cn

then the optimized attention module is embedded in the four-layer decoding network. The decoding network with attention mechanism adaptively learns the effective features, and finally obtains the portrait segmentation result map through SoftMax. At the same time, the loss function is improved, and Multi-Loss (multi-loss function) is introduced to make the network easier to converge. The way of attention mechanisms fused in decoding networks makes the lightweight network achieve better results in the semantic segmentation task. Experimental results show that the model achieves 92.29% MeanIOU on the 550 self-collected portrait test set, and the split time of the single picture on the Inter(R) Core i5 CPU is 0.74 s. Compared with the traditional portrait segmentation network, the research network has obvious advantages in segmentation precision and segmentation speed, and is suitable for mobile devices.

Key words: portrait segmentation; attention mechanism; lightweight; convolutional neural network

1 引言

语义分割作为计算机视觉的重要课题之一,从像素级理解图像,旨在为图像中的每一个像素点分配一个分类类别,从而获得逐像素的特征分类图^[1]。现如今,越来越多的应用场景需要精确而高效的语义分割技术,如室内导航、无人驾驶、增强现实等。特别是在手机端相册的应用中,需要利用语义分割技术完成人像分割,从而满足人们对于拍摄图片中人像背景的替换、渲染、虚化等美图功能。

近些年随着深度学习算法的蓬勃发展^[2-4],大量研究员开始投入精力研究如何将神经网络应用于语义分割技术中。2012年,Farabet^[5]等人利用一种新颖的场景标注方法,首次将卷积神经网络应用于语义分割算法中。2014年,Long^[6]等人提出了FCN(Fully Convolutional Networks),利用卷积层代替全连接层的方式实现了端到端的图像分割。2017年,Chen^[7]等人提出了DeepLabV3模型,使用空洞卷积(Atrous convolution)来扩展感受野,该方式可以获取更多的空间上下文信息,有效提升了分割精度。2017年,Zhao^[8]等人提出PSPNet(Pyramid Scene Parsing Network),在FCN基础上,提出在空间维度上建立空间金字塔模块,该方式精细化了物体的边缘,可以更好地分割物体。2017年,Lin^[9]等人提出RefineNet(Multi-path refinement net-works),设计了多路径提取特征的结构,改善了特征提取效果,语义分割精度提升明显。

语义分割网络为了追求精度,需要占据庞大

的计算资源。移动端有限的硬件条件,始终阻碍着语义分割技术在移动端的部署。2018年,Sandler等人提出的MobileNetV2^[10]为解决这一难题带来了可能。大量研究员通过将语义分割网络的主干部分替换为MobileNetV2的方式,成功地将网络在移动端部署,但也因此带来了分割精度差、分割边界模糊等问题。

FCN^[6]、SegNet^[11](A deep convolutional encoder-decoder architecture)和UNet^[12]作为最早将编码器-解码器结构应用于语义分割领域的神经网络,证明了该结构可以有效地捕获特征细节,获得具有丰富上下文信息的特征地图。2018年,Zhang^[13]等人提出通道注意力模型,将常用于语音识别中的注意力机制应用到语义分割领域中,证明该方式可以利用丰富的上下文信息,有效地探索特征地图通道之间的关系,恢复分割物体的边界。

本文受此启发,将注意力机制融入到编码器-解码器结构中,有效地解决了移动端人像分割网络分割效果差的问题。为了解决引入注意力模型带来的额外计算量问题,对注意力模型进行了轻量化处理。最后为了降低训练难度,加速网络收敛,在训练阶段,将原有的损失函数改为Multi-Loss(多损失函数),有效提升了分割精度。实验证明本文提出的网络结构在速度和精度上达到了有效平衡,可以部署到移动端使用。

2 模型设计

2.1 优化注意力模块

NLNet^[14](Non-local neural networks)是目

前常见的注意力模型之一,由全局上下文特征模块、特征信息转换模块、特征融合模块 3 个模块组成。通过全局上下文特征模块来获得融合了上下文信息的特征地图,利用特征信息转换模块对融合了上下文信息的特征地图进行通道校验,最后由特征融合模块将融合了上下文信息的特征地图整合到原特征地图中。本文在 NLNet 的基础上对模块结构进行了优化,在可接受的精度损失范围内,有效提升了模块的运行速度。模型结构对比图如图 1 所示。

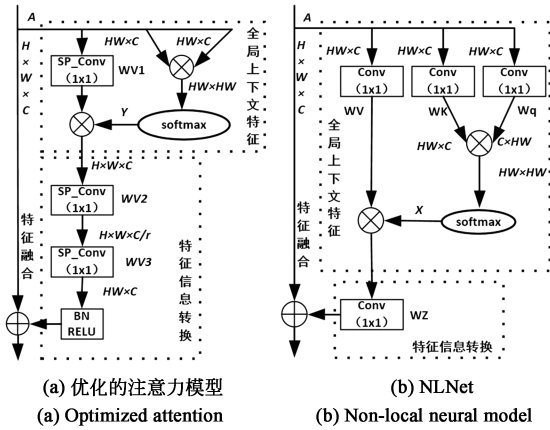


图 1 NLNet 与优化模型结构对比

Fig. 1 Structural comparison of NLNet and optimized model

图中 H 、 W 、 C 分别为输入图像的高、宽、通道数目; W_{v1} 、 W_{v2} 、 W_{v3} 为尺寸为 1 的深度可分离卷积; W_v 、 W_k 、 W_q 、 W_z 为尺寸为 1 的普通卷积核。

2.1.1 优化全局上下文特征模块

在全局上下文特征模块中,NLNet 将输入特征地图 A 分别经过两个 1×1 卷积核进行特征聚合,然后利用 Embedded Gaussian(嵌入式高斯)

对任意两个通道之间的成对关系进行建模,获得注意力地图 X ,见公式(1)。然后将特征地图 A 经过一个卷积核 W_v 进行特征聚合,随后与注意力地图 X 相乘获得融合丰富上下文信息的特征地图。但是对通道之间的成对关系建模前,对特征地图 A 进行特征聚合会破坏特征通道之间隐含的相互关系,同时会引入多余的计算量。设输入特征地图 A 的通道为 1 024,则两个 1×1 卷积核的参数达 $1\,024 \times 1 \times 1 \times 1\,024 \times 2$ (输入通道 $\times$ 卷积核尺寸 $\times$ 输出通道 $\times 2$) 之多。为此,本文将两个 1×1 的卷积核舍弃,直接对特征地图 A 进行 Gaussian(高斯)计算得到注意力矩阵 Y ,见公式(2)。该方式保留了通道之间隐含的相互关系,同时有效降低了计算量。另外,将 NLNet 中对特征地图进行特征聚合的普通卷积 W_v 换为深度可分离卷积^[15] W_{v1} ,该卷积核将普通的卷积操作分解为 depthwise(深度卷积)和 pointwise(点卷积)两个过程,可将计算量降为原有计算量的 $1/C_{out} + 1/(K_w + K_h)$ 。(C_{out} 为输出的通道数目, K_w 、 K_h 为卷积核的宽和高),推导过程见公式(3)。

$$X_{ij} = \frac{\exp(\langle W_q \cdot A_i, W_k \cdot A_j \rangle)}{\sum_{m=1}^C \exp(\langle W_q \cdot A_i, W_k \cdot A_m \rangle)} \quad (1)$$

$$Y_{ij} = \frac{\exp(\langle A_i, A_j \rangle)}{\sum_{m=1}^C \exp(\langle A_i, A_m \rangle)} \quad (2)$$

C 表示输入特征地图的通道数目。 i 是特征图中第 i 个通道的索引, j 为所有通道中的一个索引枚举。 X_{ij} 和 Y_{ij} 分别是注意力地图 X 和 Y 中的权值系数,表示特征地图中第 i 个通道对第 j 个通道的重要性。

$$\begin{aligned} \frac{\text{depthwise} + \text{pointwise}}{\text{conv}} &= \frac{H \times W \times C_{in} \times K_H \times K_W + H \times W \times C_{in} \times C_{out}}{H \times W \times C_{in} \times K_H \times K_W \times C_{out}} \\ &= \frac{1}{C_{out}} + \frac{1}{K_W \times K_H} \end{aligned} \quad (3)$$

其中:depthwise 为深度卷积的计算量,pointwise 为点卷积的计算量,conv 为普通卷积核的计算量。 C_{in} 和 C_{out} 为输入输出的通道数, W 和 H 为输入图像的宽和高, K_w 、 K_h 为卷积核的宽和高。

2.1.2 优化特征信息转换模块

NLNet 采用一个 1×1 的卷积核作为特征信息转换模块来校准通道之间的依赖关系,帮助网络理解语义信息,但该方式引入了过多的计算量。设输入特征地图的通道为 C ,则 1×1 卷积核的参

数达 $C \times 1 \times 1 \times C$ (输入通道 $\times$ 卷积核尺寸 $\times$ 输出通道) 之多。为了适应移动端的硬件环境, 本文将单一卷积核替换成两个深度可分离卷积, 通过对特征地图的通道进行先压缩再扩张的方式来降低计算量。第一个深度可分离卷积的输入通道为 C , 输出通道为 $C/8$ (其中 8 为压缩系数), 对特征图进行通道压缩。第二个深度可分离卷积的输入通道为 $C/8$, 输出通道为 C , 对特征图进行通道扩张。该方式将参数数量由 $C \times C$ 降低为 $2 \times C \times (C/8)$, 有效地将计算量降低为原计算量的 $1/4$ 。

最后将特征信息转换模块得到的特征地图输出给特征融合模块。本文沿用了 NLNet 中的融合方式, 对具有丰富上下文信息的特征地图与原始特征地图进行像素级求和得到结果特征地图。

2.2 注意力机制融入到解码网络

常见的编码器-解码器结构由编码网络和解码网络两部分组成。利用堆叠的卷积和池化层作为编码网络来提取图像的特征信息, 然后将编码器中包含低等级特征信息的特征地图跳跃连接到解码网络中。解码网络由多个子模块组成, 每个子模块由连续的一两个卷积核堆叠而成, 负责对上一层子模块输出的特征地图与编码网络跳跃连接而来的特征地图进行特征聚合, 以此获得具有高、低等级特征信息的特征地图。

本文模型利用 MobileNetV2 作为编码网络。解码网络堆叠了 4 层解码子模块 (Decoder), 每个解码子模块嵌入了一个优化过的注意力模块和一个用来压缩特征地图通道数目、尺寸为 1 的深度可分离卷积。为了获得丰富的上下文信息, 选取编码网络中的第 2、4、7、14 层 (尺寸分别为原始特征地图的 $1/2, 1/4, 1/8, 1/16$) 的输出特征地图跳跃连接到解码网络中, 与解码子模块输出的特征地图在通道维度上进行堆叠, 再输出给下一层解码子模块。为了堆叠时特征地图尺寸的一致性, 在堆叠前对解码子模块输出的特征地图进行了上采样操作。重复上述操作, 最终将特征地图经过一个 1×1 的卷积核和 SoftMax 层得到分割结果图, 尺寸为输入图像的一半。模型具体结构见图 2。其中, ①: 编码网络的输出特征地图。②~⑤: 编码网络第 14、7、4、2 层的输出特征地图。⑥~⑨: 经过解码子模块并上采样后的特征地图。⑩: 输出特征地图。decoder- i ($i=1, 2, 3, 4$) 为解码子模块, 包含一个深度可分离卷积和一个轻量化

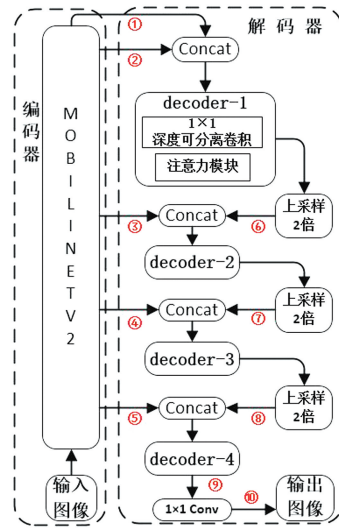


图 2 模型网络结构示意图

Fig.2 Schematic diagram of model network structure

注意力模块。表 1 给出了各特征地图的尺寸和通道数目。

表 1 模型各层输出特征地图尺寸

Tab.1 Size of the output feature map of each layer of the model

特征地图	尺寸(高 $\times$ 宽)	通道数目
①	14×14	320
②	14×14	96
③	28×28	32
④	56×56	24
⑤	112×112	16
⑥	28×28	192
⑦	56×56	64
⑧	112×112	48
⑨	112×112	8
⑩	112×112	2

2.3 Multi-Loss

语义分割网络因为网络过深, 存在训练难度大、分割精度难以提升等问题。为此, 本文在损失函数上做了一些修改, 采用 Multi-Loss 对各层输出结果进行监督。将标签尺寸大小改变为原图的 $1/16, 1/8, 1/4, 1/2$, 然后分别与解码网络中的 decoder-1, decoder-2, decoder-3, decoder-4 生成

的特征地图做交叉熵操作得到各层的损失值 Loss1, Loss2, Loss3, Loss4, 最终的损失函数为各层交叉熵的系数和。

$$\text{total_Loss} = \sum_{i=1}^4 \beta_i \cdot \text{Loss}_i \quad (4)$$

其中 β 为损失权值系数。随着网络的进行, 损失逐渐降低, 为了权衡各层损失对整体损失的影响, 将 β_1, β_2 与 β_3, β_4 的比值设为 1 : 2。图 3 展示了整体损失与各层的损失的收敛情况。

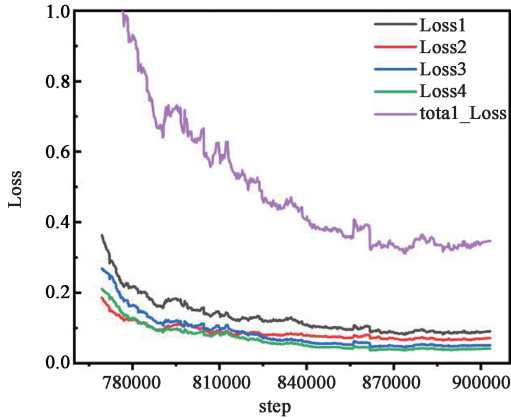


图 3 总损失与编码层各层损失收敛图

Fig.3 Convergence graph of total loss and loss of each layer of the decoder layer

3 实验及结果分析

实验环境: 实验采用开源的深度学习框架 TensorFlow, 并利用该框架下的 TensorBoard 对精度和损失的收敛情况进行观察。计算机配置为 Window10 操作系统, 8 G 内存, 处理器为 Inter (R)Core i5。为了更好地匹配移动端条件, 整个实验在 CPU 环境中进行测试。

数据集处理: 实验在 PASCAL、COCO 等 4 个公开数据集中挑选了 38 470 张与人像有关的图片作为预训练样本对模型进行预训练。然后利用自采集的 6 166 张人像图片作为数据集对模型进行微调(Finetune)。验证集为自采集的人像图片 550 张。利用左右翻转和随机裁取操作进行数据增强。所有图片的标注均采用语义分割标注工具 Labelme。

训练策略: 实验采用的梯度优化器为 Adam 算法。学习率采用多项式学习率迭代策略, 初始学习率为 0.1, 每 2 万步衰减到原学习率的 1/10,

训练总步长为 90 万步。输入图像尺寸为 224×224 。

评价指标: 图片单张分割时间(s)、平均像素精度 MPA 与交并比 MIOU。见公式(5)、(6)。

$$\text{MPA} = \frac{1}{k+1} \times \sum_{i=2}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (5)$$

$$\text{MIOU} = \frac{1}{K+1} \cdot \sum_{j=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (6)$$

其中: k 为分类的类别, P_{ii} 为真正的数量; P_{ij} 为假正的数量; P_{ji} 为假负的数量。

3.1 注意力模块的优化效果

为了验证轻量化注意力模块的效果, 实验分别将优化了全局上下文特征模块的注意力模型、优化了特征信息转换模块的注意力模型、两者同时优化后的注意力模型与原模型 NLNet 分别嵌入到解码网络中, 对模型的分割情况进行测试, 结果见表 2。

表 2 注意力模块优化效果对比

Tab.2 Comparison of the optimization effect of the attention module

嵌入模块	MIOU/%	分割时间/s
NLNet	92.54	1.05
优化全局上下文特征模块	92.29	0.90
优化特征信息转换模块	92.32	0.87
两者同时优化	92.11	0.74

观察可知, 与嵌入原模块 NLNet 相比, 优化全局上下文特征模块后, MIOU 损失了 0.25%, 速度提升了 0.15 s。优化特征信息转换模块后, MIOU 损失了 0.22%, 速度提升了 0.18 s。两者同时优化后, MIOU 仅仅损失了 0.43%, 速度提升了 0.31 s。在可接受的精度损失范围内, 速度提升明显。证明了文章中提出的注意力模块优化方案具有可行性。

3.2 解码网络层数的设计

解码网络可以有效地捕捉上下文信息, 配合注意力模块可以提升分割精度, 但解码网络的层数对网络运行速度有一定影响。实验就融合注意力机制的解码网络层数进行实验, 实验结果见表 3。

表 3 加入不同层数解码网络的分割指标

Tab.3 Segmentation indicator of decoding network with different layers

添加层数	MIOU/%	单张图片分割时间/s
未添加	88.26	0.61
一层	90.23	0.66
两层	91.07	0.70
三层	91.72	0.72
四层	92.29	0.74

分析表 3 可知,随着解码网络层数的增加,MIOU 得到提升,但单张图片分割消耗的时间相应增加。未添加解码网络时 MIOU 为 88.26%,时间为 0.61 s。添加 1 层、2 层、3 层、4 层后 MIOU 分别提升了 1.97% 和 2.81%,3.46% 和 4.03%。但时间相应增加了 0.05,0.09,0.11,0.13 s。相对于精度的提升,带来的速度影响在可接受范围内,因此网络设计最终选择了 4 层解码网络。

3.3 解码网络中融合注意力机制对分割效果的影响

解码网络中融合注意力机制可以有效地恢复物体的分割边界,提升分割细节。实验将解码网络中融合注意力模块和未融合注意力模块产生的分割图片做对比实验。部分对比数据见图 4。

观察图 4 可知,解码网络未融合注意力模块时,对于人体的边缘和细节部分分割效果较差,多

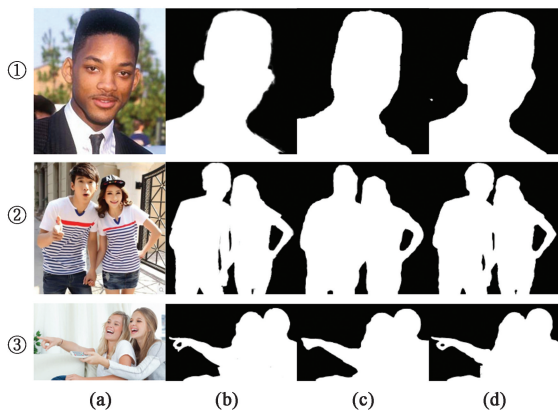


图 4 解码网络中融合注意力模块对分割效果的影响。(a)原图;(b)标签;(c)未融合;(d)融合后。

Fig.4 Effect of the fusion attention module in the decoding network on segmentation. (a) Original; (b) Label; (c) Unfused; (d) Fused.

分割漏分割情况严重。如例 1 中不能分割出人的耳朵。例 2 中只能大致分割出人体的轮廓,对于胳膊、小臂等部位存在漏分割情况。例 3 中无法正确分割出人的手指。在解码网络中融合注意力模块后,对于细节的分割效果明显提升。恢复了未添加注意力模块时漏分割的耳朵、手指部分,对于人体的轮廓边缘分割较为清晰。证明了解码网络融合注意力模块可以有效地修复分割边界,提升分割的效果。

3.4 Multi-Loss 对分割精度的影响

Muli-Loss 可以减小训练的难度,提升分割像素精度。实验对比了引入 Multi-Loss 前后的平均像素精度 MPA 的收敛情况,实验结果见图 5。

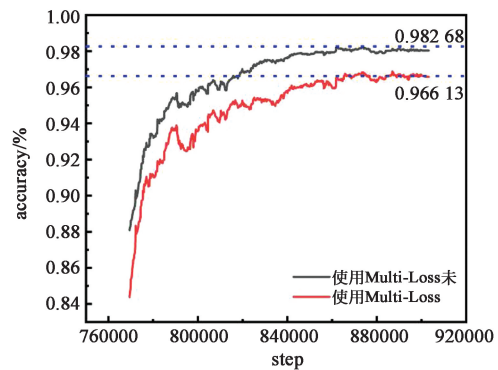


图 5 引入 Multi-Loss 的精度对比

Fig.5 Accuracy comparison of introducing Multi-Loss

观察精度收敛情况可知,引入 Multi-Loss 训练策略后,收敛精度一直优于未引入 Multi-Loss 训练策略,精度最终收敛到 98% 左右。相较于未引入 Multi-Loss 的最终精度 96%,精度提升了 2% 左右,证明了 Multi-Loss 训练策略能够有效提升分割精度,减小网络训练难度。

3.4 与其他分割方法对比

为了验证本文研究模型的分割性能,挑选了 FCN、PSPNet、SegNet 和 Unet 共 4 种前沿的语义分割网络作对比,它们的主干网络均为 MobileNet-V2。统计了各模型单张图片分割时间与 MIOU 指标,结果如表 4 所示。

由表 4 可知,本文研究模型达到了较为可观的速度与精度。与分割效果较差的 FCN 相比,MIOU 超出了 6.03%,与 SegNet 和 Unet 相比,MIOU 也有着 2.61% 和 2.23% 的优势。PSPNet

表4 不同分割模型的分割指标

Tab.4 Segmentation indicator of different segmentation models

模型	MIOU/%	单张图片分割时间/s
FCN	86.26	1.29
PSPNet	91.99	1.14
SegNet	89.68	1.17
Unet	90.06	1.23
本研究模型	92.29	0.74

与本模型分割精度相当,但分割单张图片,PSPNet需要1.14 s,本文研究模型仅需要0.74 s,快了0.4 s左右。与其他模型相比,本文研究模型在分割单张图片的时间上也都快了0.5 s左右,速度优势明显。

为了更直观对本文研究网络的分割性能进行评估,对各方法的分割结果图进行了可视化。部分数据结果如图6所示。观察可知:本研究模型可以清晰地分割出鼻子、耳朵的轮廓边界。分割边界几乎没有锯齿状。对于手指、胳膊等难分割

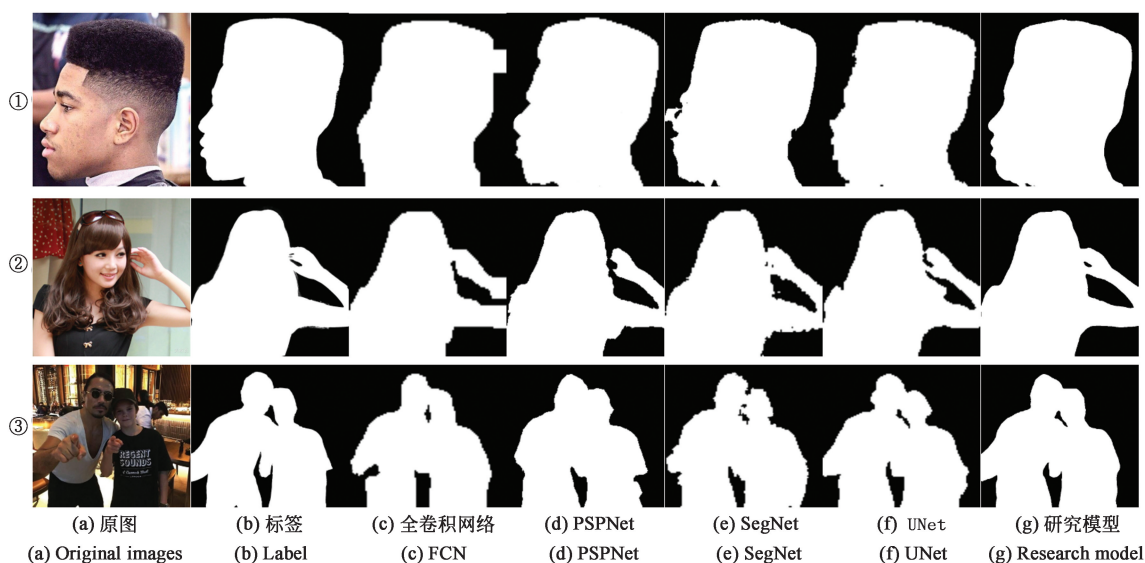


图6 不同分割方法对比图

Fig.6 Comparison of different segmentation methods

部位分割表现良好,几乎没有漏分割、多分割现象。相较之下FCN分割效果较差,分割边界较模糊。SegNet分割锯齿状严重。Unet与PSPNet存在对于胳膊、手指的多分割、漏分割现象。对比之下证实了本文模型分割性能的优越性。

4 结 论

提出了一种轻量化人像分割网络,通过在解

码网络中融合注意力机制,有效恢复了物体的分割边界,提升了分割精度。为了进一步轻量化网络,提升分割速度,对注意力模块进行了轻量化处理。同时将损失函数改为Multi-Loss,加速网络收敛。实验证明,本研究网络模型达到了92.29%的MeanIOU,单张图片分割时间为0.74 s。相较于其他分割网络,速度精度优势明显,适合在移动端应用。

参 考 文 献:

- [1] 张健,李宏升.基于图论阈值算法的图像分割研究[J].液晶与显示,2014,29(4):592-597.
ZHANG J, LI H S. Image mosaic research based on wavelet and rough set algorithm [J]. *Chinese Journal of Liquid Crystals and Displays*, 2014, 29(4): 592-597. (in Chinese)
- [2] 孙志军,薛磊,许阳明,等.深度学习研究综述[J].计算机应用研究,2012,29(8):2806-2810.

- SUN Z J, XUE L, XU Y M. Overview of deep learning [J]. *Application Research of Computers*, 2012, 29(8): 2806-2810. (in Chinese)
- [3] 周筑博, 高俊, 张巍, 等. 基于深度卷积神经网络的输电线路可见光图像目标检测[J]. *液晶与显示*, 2018, 33(4): 317-325.
- ZHOU Z B, GAO J, ZHANG W, *et al.* Object detection of transmission line visual images based on deep convolutional neural network [J]. *Chinese Journal of Liquid Crystals and Displays*, 2018, 33(4): 317-325. (in Chinese)
- [4] 梁华, 宋玉龙, 钱锋, 等. 基于深度学习的航空对地小目标检测[J]. *液晶与显示*, 2018, 33(9): 793-800.
- LIANG H, SONG Y L, QIAN F, *et al.* Detection of small target in aerial photography based on deep learning [J]. *Chinese Journal of Liquid Crystals and Displays*, 2018, 33(9): 793-800. (in Chinese)
- [5] Farabet C, Couprie C, Najman L, *et al.* Learning hierarchical features for scene labeling[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 35(8): 1915-1929.
- [6] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA: IEEE, 2015: 3431-3440.
- [7] CHEN L C, PAPANDREOU G, SCHROFF F, *et al.* Rethinking atrous convolution for semantic image segmentation [J]. *ArXiv*, 2017:1706.05587.
- [8] ZHAO H S, SHI J P, QI X J, *et al.* Pyramid scene parsing network [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE, 2017: 6230-6239.
- [9] LIN G S, MILAN A, SHEN C H, *et al.* RefineNet: multi-path refinement networks for high-resolution semantic segmentation [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE, 2017: 5168-5177.
- [10] SANDLER M, HOWARD A, ZHU M L, *et al.* MobileNetV2: inverted residuals and linear bottlenecks [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 4510-4520.
- [11] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [12] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich, Germany: Springer, 2015: 234-241.
- [13] ZHANG Y L, LI K P, LI K, *et al.* Image super-resolution using very deep residual channel attention networks [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer, 2018: 294-310.
- [14] WANG X L, GIRSHICK R, GUPTA A, *et al.* Non-local neural networks [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 7794-7803.
- [15] HOWARD A G, ZHU M L, CHEN B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications [J]. *ArXiv*, 2017:1704.04861.

作者简介:



周 鹏(1994—),男,安徽芜湖人,硕士研究生,2017 年于武夷学院获得学士学位,主要从事计算机视觉、深度学习等方面的研究。E-mail: 946375455@qq.com



姚剑敏(1978—),男,福建莆田人,博士,副研究员,2005 年于中科院长春光学精密机械与物理研究所获得博士学位,主要从事人工智能、图像处理、信息显示技术等方面的研究。E-mail: yaojm@fzu.edu.cn