

融合注意力机制的DeeplabV3 + 服装图像分割方法

肖雅慧, 张自力, 胡新荣, 彭涛, 张俊

引用本文

肖雅慧, 张自力, 胡新荣, 彭涛, 张俊. 融合注意力机制的DeeplabV3 + 服装图像分割方法[J]. 计算机科学, 2024, 51(6A): 230900153-7.

XIAO Yahui, ZHANG Zili, HU Xinrong, PENG Tao, ZHANG Jun. Clothing Image Segmentation Method Based on Deeplabv3+ Fused with Attention Mechanism [J]. Computer Science, 2024, 51(6A): 230900153-7.

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)

[基于SAMNV3的滚动轴承智能故障诊断方法](#)

Intelligent Fault Diagnosis Method for Rolling Bearing Based on SAMNV3

计算机科学, 2024, 51(6A): 230700167-6. <https://doi.org/10.11896/jsjcx.230700167>

[基于LSTM和注意力机制的远程会诊需求预测](#)

Forecasting Teleconsultation Demand Based on LSTM and Attention Mechanism

计算机科学, 2024, 51(6A): 230800119-7. <https://doi.org/10.11896/jsjcx.230800119>

[融入类别标签和主题信息的用户兴趣识别方法](#)

User Interest Recognition Method Incorporating Category Labels and Topic Information

计算机科学, 2024, 51(6A): 230500169-8. <https://doi.org/10.11896/jsjcx.230500169>

[基于注意力的多尺度蒸馏异常检测](#)

Attention-based Multi-scale Distillation Anomaly Detection

计算机科学, 2024, 51(6A): 230300223-11. <https://doi.org/10.11896/jsjcx.230300223>

[基于改进Deeplabv3 + 算法的滚珠丝杠驱动表面点蚀缺陷检测](#)

Detection of Pitting Defects on the Surface of Ball Screw Drive Based on Improved Deeplabv3+ Algorithm

计算机科学, 2024, 51(6A): 240200058-6. <https://doi.org/10.11896/jsjcx.240200058>

融合注意力机制的 DeeplabV3+ 服装图像分割方法

肖雅慧¹ 张自力^{1,2} 胡新荣^{1,2} 彭涛^{1,2} 张俊³

1 武汉纺织大学计算机与人工智能学院 武汉 430200

2 湖北省服装信息化工程技术研究中心 武汉 430200

3 武汉工程大学计算机科学与工程学院 武汉 430205

(2115363055@mail.wtu.edu.cn)

摘要 针对在服装图像语义分割中存在由服装颜色、纹理、背景以及多目标遮挡导致的边缘分割粗糙和分割精度低等问题,文中基于 Deeplabv3+ 框架,提出了一种图像语义分割算法(FFDNet)。首先,模型的骨干网络采用 ResNet101 网络,并添加通道空间注意力模块(Feature-Enhanced Attention Module, FEAM),通过对特征图加权来挖掘并增强特征信息,提高网络表达能力。其次引入特征对齐模块(Feature Align Module, FAM)作为一种新的上采样方式,解决不同尺度特征融合之间特征未对齐导致分割错误且效率低的问题,以此提高对服装图像分割的准确性和鲁棒性。最后,FFDNet 在 Deepfashion2 和 PASCAL VOC 2012 数据集上的平均交并比分别达到 55.2% 和 79.4%;在参数量方面,该模型相比原模型在 Deepfashion2 上仅增加了 0.61 MB。与其他现有经典模型对比,其分割性能更优,能有效捕获图像局部细节信息,减少像素分类错误。

关键词: 服装图像;语义分割;注意力机制;Deeplabv3+ 网络;特征对齐

中图分类号 TP391

Clothing Image Segmentation Method Based on Deeplabv3+ Fused with Attention Mechanism

XIAO Yahui¹, ZHANG Zili^{1,2}, HU Xinrong^{1,2}, PENG Tao^{1,2} and ZHANG Jun³

1 School of Computer Science and Artificial Intelligence, Wuhan Textile University, Wuhan 430200, China

2 Engineering Research Center of Hubei Province for Clothing Information, Wuhan 430200, China

3 School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China

Abstract Aiming at the problems of rough edge segmentation and low segmentation accuracy caused by color, texture, background and multi-object occlusion in clothing image segmentation, an image semantic segmentation method(FFDNet) based on Deeplabv3+ with attention mechanism is proposed. Firstly, the backbone network of the model uses the ResNet101 network. The feature-enhanced attention module(FEAM) is added at the end of it. The feature map is weighted from the two dimensions of channel and spatial to mine and enhance the feature information and optimize the segmentation edge to improve network clarity. Secondly, a feature align module(FAM) is introduced as a novel upsampling method to address the problem of segmentation errors and low efficiency caused by misalignment between features during the fusion of different scale features, so as to improve the accuracy and robustness of clothing image segmentation. Finally, the mean intersection over union of the proposed method reaches 55.2% and 79.4% on Deepfashion2 and PASCAL VOC2012, respectively. In terms of parameter size, the model only increases by 0.61MB compared to the original model on Deepfashion2. The segmentation performance of the FFDNet is superior to the existing state-of-the-art network models, which can effectively capture image local detail information and reduce pixel classification errors.

Keywords Clothing image, Semantic segmentation, Attention mechanism, Deeplabv3+ network, Feature alignment

1 引言

近年来,服装图像语义分割被广泛应用于虚拟试衣、电子商务以及时尚行业等各大领域。对虚拟试衣技术来说,服装图像分割算法可帮助消费者在线选购合适、满意的服饰,节省在实体店选购服装的时间和精力;对时尚行业来说,准确的服装图像分割方法不仅能帮助设计师获取新颖的创意和灵感,同时还能给用户推荐更多个性化的时

尚产品并增加其消费购物的新鲜感。

国内外很多学者提出基于传统方法的服装图像分割方法, Han 等^[1]针对传统分水岭算法容易造成过分割现象,提出融合人工标记的分水岭分割算法,基本上满足了复杂背景下目标提取的基本要求,但该方法未解决多目标遮挡引起的分割精度低等问题。Gao 等^[2]针对传统的 K-means 聚类算法在图像分割中无法确定最佳 K 值的大小,提出一种改进的 K-means 算法,目的是将彩色图像转换为灰度图像并统计灰

基金项目:湖北省教育厅科学技术研究计划项目(B2017066)

This work was supported by Science and Technology Research Project of Education Department of Hubei Province(B2017066).

通信作者:张自力(zlzhang@wtu.edu.cn)

度直方图的峰值数作为聚类数 K ，之后对图像的像素点进行 K -means 聚类完成图像分割。虽然这些传统分割算法极大地提高了服装图像的分割效率且有效地降低了人工提取特征的开销和成本，但因为服装图像本身内容细节复杂、易受背景干扰，仍然存在漏检、漏分割和边缘效果差等问题。

近年来，随着深度学习的发展，越来越多的图像语义分割模型被提出，其中，Long 等^[3]提出的全卷积神经网络(Fully Convolutional Networks, FCN)虽然在不同的基准上取得了较好的效果，但由于 FCN 结构固定，缺点也很明显：上采样过程中丢失大量的细节信息，没有充分地考虑到像素与像素之间的关系。Ron-nebeger 等^[4]提出 U-Net，采用编解码结构以拼接的方式融合不同尺度信息，但无法充分整合全局上下文信息。Zhao 等^[5]提出了 PSPNet(Pyramid Scene Parsing Network)，其网络对不同区域的上下文信息进行聚合，提升了模型理解全局上下文语境的能力。谷歌团队提出的 Deeplab 系列模型^[6-9]通过引入空洞卷积来增大感受野进而获取更多的上下文信息。Yang 等^[10]提出将 Fast R-CNN 模型与 Grabcut 算法相结合，成功地对服装图像进行自动分割，很好地检测和提取了服装分割的轮廓边缘，极大地提高了图像分割的准确率，但此方法提取的语义信息不够充分，在复杂背景下其分割效果仍不够理想。

目前注意力机制已经普遍应用到了语义分割领域^[11-14]。Hu 等^[15]提出的注意力网络 SENet 在增加少量参数的基础上，压缩特征图以便在通道维度上对特征层进行特征提取，该模型能够即插即用，可以方便嵌入到各种 CNN 网络。CBAM^[16]模块在 SENet 的基础上通过大尺度卷积核的采样引入空间信息。后续许多研究人员提出注意力机制融合网

络^[17]和多尺度特征融合模块^[18-22]来提取和整合图像中需要重点关注的区域特征。Xu 等^[23]提出边界注意力机制模块指导高层上下文信息和空间细节信息融合，提升神经网络表达能力。

本文针对原 Deeplabv3+ 网络在提取特征信息时容易丢失细节信息的问题，提出一种融合注意力机制的 Deeplabv3+ 服装图像分割算法。首先针对网络特征提取不充分的情况，引入基于通道与空间注意力机制的 FEAM 模块，以增强网络对有用特征的提取能力和处理服装边缘分割的能力；其次针对 Deeplabv3+ 解码过程中低级信息与高级信息融合时的因特征未对齐产生分割精度偏低的问题，引入一种新颖的特征对齐模块 FAM，充分融合语义信息和空间信息。

2 Deeplabv3+ 网络模型

Deeplabv3+ 模型网络在 Deeplabv3 的基础上引入了编解码结构，其解码器结构是仿照 U-Net 网络结构实现了一个用来优化边缘分割精度的上采样模块，该模型网络的主干网络采取 Resnet101，骨干网络后面接空洞卷积金字塔模块 ASPP。ASPP 模块将输入特征图分别进行 1×1 卷积，执行空洞卷积扩张率为 6, 12, 18 的 3×3 卷积和全局平均池化操作。之后以通道维度将获取的多尺度特征图进行拼接，用以对输出的特征图进行特征提取，进行 1×1 卷积降维处理之后，将该特征送入解码器部分。

从图 1 可以看到，解码器部分主要对特征进行尺度融合后上采样，将主干网络 Resnet101 的 $1/4$ 低级特征图信息与 1×1 卷积后的高级特征图信息融合，然后进行双线性插值上采样得到输出结果。

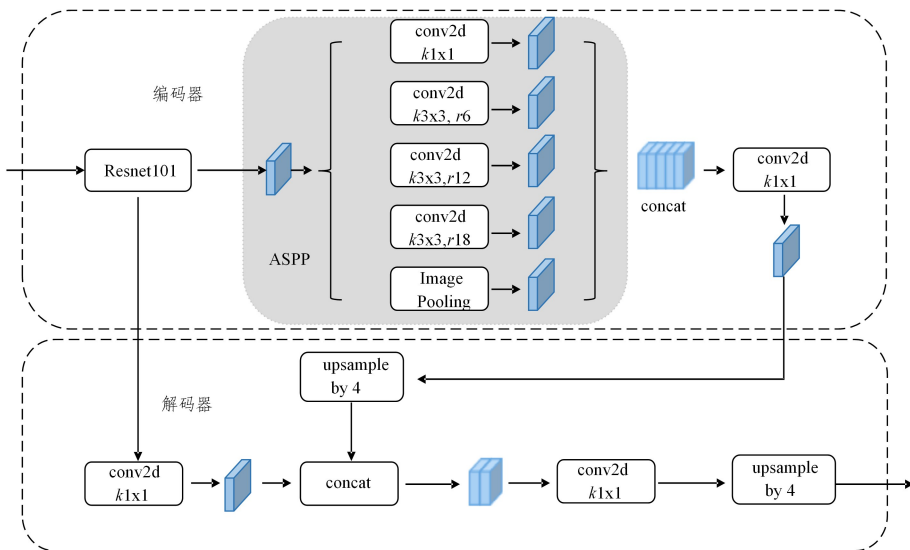


图 1 原 Deeplabv3+ 网络结构

Fig. 1 Structure of original Deeplabv3+ network

3 改进的 Deeplabv3+ 网络结构

利用 Deeplabv3+ 网络结构在服装数据集上进行图像分割实验。从实验结果可以发现，由于服装图像背景复杂、款式和风格多样，且因模特的姿态和拍摄角度的不同容易产生目标遮挡、皱褶等问题，这些都会对服装图像分割的精度造成一定的影响。

图 2 为改进后的 Deeplabv3+ 的模型结构图。Deeplabv3+ 网络由编码器和解码器两部分组成，在编码器中，首先将骨干网络 Resnet101 的 $1/16$ 的深层次特征送入改进后的 F_{ASPP} 加强特征提取网络中；然后在不同的并行空洞卷积中并联 FEAM，FEAM 将获得的深层次特征先通过 1×1 卷积降维，再通过单一的 FEAM 模块处理，使模型更关注重要特征，提升模型分割精度。最后与原始 ASPP 网络获得的特征层进行

相加融合,得到输出特征。

在解码器中,首先将骨干网络 Resnet101 的 1/4 低层次特征率先通过 FEAM 处理,让模型在学习过程中更注重空间特征,提升网络处理边缘特征的能力;其次与编码器部分获得的特征图经过上采样后的输出进行 FAM 模

块特征对齐,减少特征融合之间因像素不对齐而产生分类错误的问题;最后将经过特征对齐的特征图与编码器部分上采样后的输出进行拼接,在完成拼接后进行 1×1 卷积 4 倍以上采样,以恢复到原图尺寸并生成预测图像。

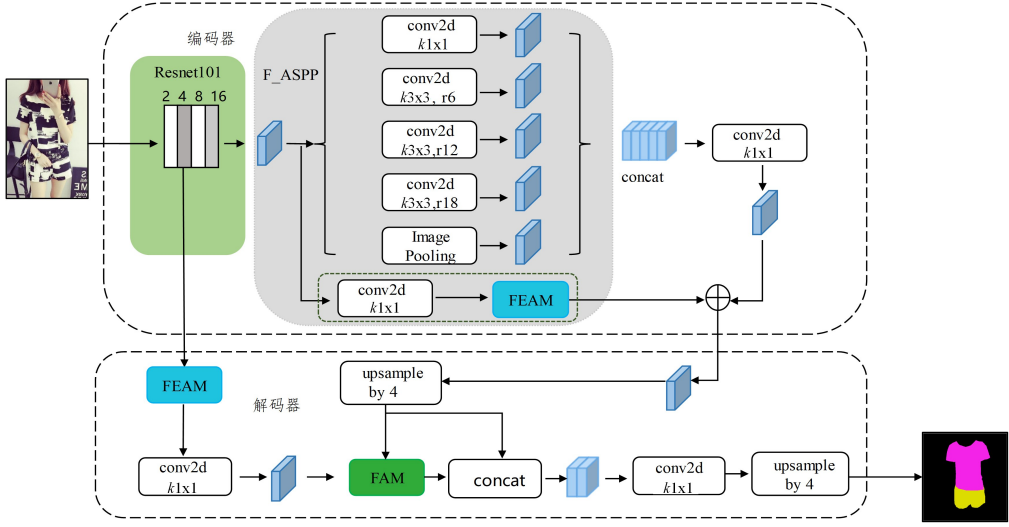


图2 融合注意力机制的 Deeplabv3+网络结构图

Fig. 2 Structure of Deeplabv3+ network fused with attention mechanism

3.1 FEAM 注意力机制

本文引入了一个轻量级的 FEAM 特征增强注意力模块,从通道和空间两个方向提高模型的信息挖掘能力和增强模型对多级特征的提取与整合能力,将通道注意力模块和空间注意力模块进行串联组合。通道注意力模块使用通道注意力权重调整对服装区域的关注度,重点关注特征图上有意义的内容,提高模型识别不同服装类别的能力;空间注意力模块使用空间注意力权重,重点关注特征的位置信息,提高关键特征信息的表征能力。如图 3 所示,FEAM 模块由通道注意力机制模块和空间注意力机制模块组成,对于通道注意力机制模块,应当对关键通道分配更大的权重。

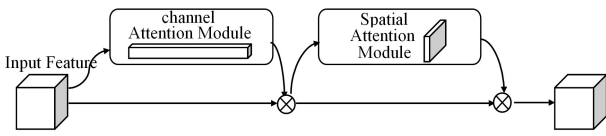


图3 FEAM 模块结构图

Fig. 3 Structure diagram of FEAM module

如图 4 所示,输入特征图大小为 $H \times W \times C$,可以看到,首先特征图 F 经过全局最大池化层后得到一个 $1 \times 1 \times C$ 的特征图,然后通过全连接层 f_{c1} ,获得通道方向大小为 $1 \times 1 \times C/16$ 的特征向量,继而通过全连接层 f_{c2} ,获得 $1 \times 1 \times C$ 的特征向量。

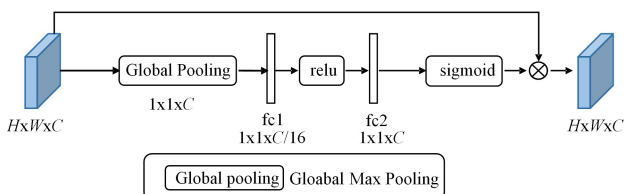


图4 通道注意力机制模块

Fig. 4 Channel attention mechanism module

入特征图 F 与权重系数 $W_c \in R^{1 \times 1 \times C}$ 相乘最终得到处理之后的新特征,如式(1)所示:

$$W_c(F) = \sigma\{f_{c2}\{\delta[f_{c1}(MaxPool(F))]\}\} \quad (1)$$

其中, $MaxPool$ 表示全局最大池化层, σ 表示 Sigmoid 激活函数, δ 表示 Relu 激活函数, f_{c1} 表示第一层全连接层, f_{c2} 表示第二层全连接层。

FEAM 的空间注意力机制将通过通道注意力机制模块处理后的特征图作为新的输入特征图,增强特征有用部分权重,抑制不必要的冗余信息。

对于空间注意力机制模块,如图 5 所示,首先对输入特征图进行通道维度上的全局最大池化层操作,得到一个 $H \times W \times 1$ 的 F_{max} ,然后经过一个 Sigmoid 函数(记为 σ)获得权重系数为 $W_s \in R^{H \times W}$,最后 W_s 与该特征图相乘得到空间注意力处理后的最终特征。

$$W_s(F) = \sigma(f^{7 \times 7} \cdot MaxPool(F)) = \sigma(f^{7 \times 7} \cdot F_{max}) \quad (2)$$

其中, $f^{7 \times 7}$ 表示 7×7 的卷积, $MaxPool$ 表示沿通道维度的全局最大池化层, F_{max} 表示经过最大池化之后的特征图。

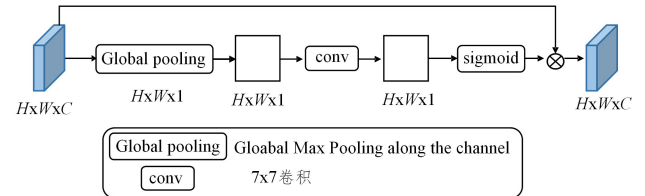


图5 空间注意力机制模块

Fig. 5 Space attention mechanism module

3.2 特征对齐模块

本文提出的特征对齐模块 FAM 用于改进多尺度分辨率融合的效果,该模块的具体架构如图 6 所示。传统的 Encoder-Decoder 结构一般是通过预训练的网络进行一个初步的特征提取,然后通过解码器中进行 Skip-Connection 操作将相邻尺度不同分辨率的特征图进行完全融合。然而编解码

最后用 Sigmoid 激活函数求得权重系数,记作 W_c ,将输

过程中低分辨率特征图利用插值算法恢复分辨率易造成语义不对齐的问题,相较而言,主干网络采用残差连接更易引起不对齐的问题,所以多尺度特征图如何高效融合,以及如何建立低分辨率特征图和高分辨率特征图的映射关系是十分必要

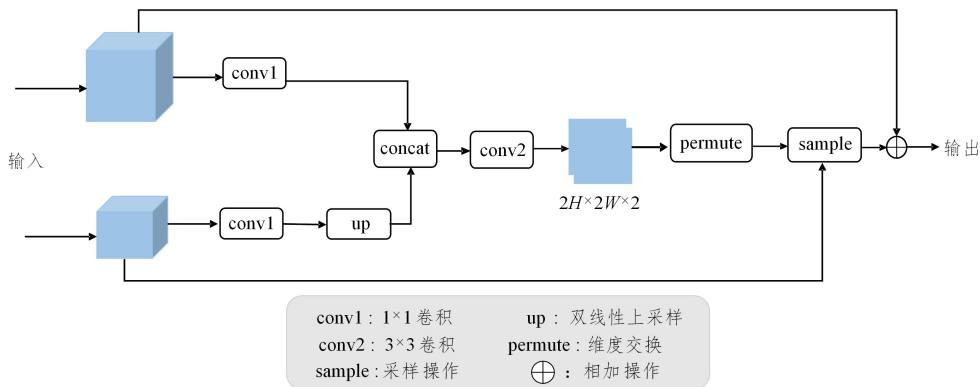


图6 特征对齐模块

Fig. 6 Feature alignment module

为了高效准确地融合相邻尺度特征,本文引入一种新颖的特征对齐模块(FAM),核心是初步融合低分辨率和高分辨率特征图,然后将融合后的特征图作为低分辨率上采样的位置参照,进而上采样低分辨率特征图实现高效融合。首先将低分辨率特征图与高分辨率特征图都进行 1×1 卷积降维操作且合并,之后通过 3×3 卷积和 permute 操作,与高分辨特征图进行融合,通过采样操作学习到一个低分辨率到高分辨率的特征映射,如式(3)所示:

$$G = \text{conv2}(\text{cat}(\text{up}(\text{conv1}(F_l)), \text{conv1}(F_h))) \quad (3)$$

然后通过学习到的特征映射关系 G 获得低分辨率特征图在初步融合特征图上对应的像素位置,接着对低分辨率上的各像素点周围的4个邻居(左上、右上、左下、右下)进行双线性插值,获得上采样后的高分辨率特征图 P_h ,如式(4)所示:

$$P_h(a, b) = P_l(\lambda G(a, b)) \quad (4)$$

其中, a, b 表示特征映射关系 G 中第 a 行第 b 列的像素点, λ 表示采样操作中对应的偏移量因子。

$$F = \sum_{p \in N(P_l)} \omega_p F_l(p) \quad (5)$$

其中, $N(P_l)$ 表示低分辨率周围像素点的集合, ω_p 表示4个相邻像素点相对中心像素点的权重。最终特征图 F_l 的表达式如式(6)所示:

$$F_l = F + F_h \quad (6)$$

4 实验结果分析

4.1 实验数据集

本文采用的数据集是 Deepfashion2^[27] 和 PASCAL VOC2012。

Deepfashion2 是一个大型公开的基准数据集,总共有 49.1 万张图片,包含了 13 种服装类别,支持服饰检测和分类、姿态检测和分割检索等其他任务。由于数据集过于庞大且考虑到实验硬件环境的因素,因此本文选取数据集的 77848 张图片作为训练集,10492 张作为验证集,10568 张则作为测试集。

样本的分布情况如图7所示,0:背景,1:短袖上衣,2:长袖上衣,3:短袖外衣,4:长袖外衣,5:背心,6:吊带,7:短裤,

的。Li等^[24]受相邻视频帧之间运动对齐的光流的启发,提出流对齐模块学习相邻特征图之间的语义流。Huang等^[25]提出特征对齐分割网络精准聚合多分辨率特征图。Huang等^[26]提出特征对齐金字塔网络学习像素的变换偏移。

8:长裤,9:裙子,10:短袖连衣裙,11:长袖连衣裙,12:背心裙,13:吊带裙。

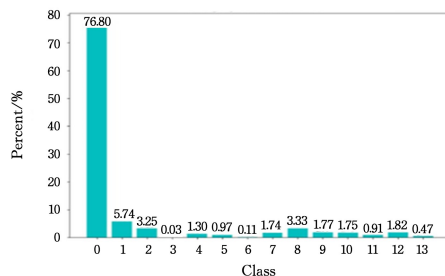


图7 数据集类别分布图

Fig. 7 Dataset class distribution

PASCAL VOC 2012 数据集常用于图像分类、语义分割和图像检测3类视觉目标识别任务,其中包括人、地面、动物和背景等21个类别,整体数据集由训练集和验证集组成,训练集有10582张图片,验证集有1449张,且输入图像大小均为 $512 \text{ pixel} \times 512 \text{ pixel}$ 。

4.2 实验配置环境

实验配置环境如表1所列。

表1 实验配置环境

Table 1 Experimental environment configurations

实验环境	配置
操作系统	Ubuntu 18.04
处理器	Intel(R) Xeon(R) Gold 6330
GPU	RTX 3090(24 GB)
CUDA 版本	cuda 11.0
Python 版本	python 3.8
深度学习框架	Pytorch
Torch 版本	Torch 1.7.0

4.3 实验评价指标

本文实验采用的评价指标有平均交并比(Mean Intersection over Union, MIoU)、像素准确率(Pixel Accuracy, PA)和平均像素准确率(Mean Pixel Accuracy, MPA)。IoU指实际样本和预测样本的交集和并集之比,而MIoU指每一类交并比求和平均的结果;MPA则是计算每个类被正确分类像素数的比例,其中 p_{ii} 表示预测结果为 i 类,其真实结果也是 i 类的

像素点,同理, p_{ji} 表示预测结果是 j 类,但真实结果为 i 类的像素点。

$$PA = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \quad (7)$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (8)$$

$$MPA = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (9)$$

4.4 消融实验

为了验证 FEAM 注意力机制和 FAM 特征对齐模块设计的合理性和嵌入到 Deeplabv3+ 网络的有效性,本文就 FEAM 的有效性设计了如表 2 所列的消融实验,用来验证 FEAM 是否聚焦目标特征。首先将嵌入到主干网络 1/4 低分辨率上的特征图记为 FEAM_l,同时并行 ASPP 嵌到高分辨率上的特征图记为 FEAM_h,Text1 为原 Deeplabv3+ 网络,Text 2,Text3 和 Text4 分别在 Text1 的基础上添加 FEAM_l 模块、FEAM_h 模块以及两者。从表 2 可以看到,实验 2 仅在骨干网络上添加 FEAM_l,MIoU 提升了 0.001;实验 3 仅在空洞卷积 ASPP 上添加 FEAM_h 模块,MIoU 增加了 0.002;实验 4 同时添加两类模块,MIoU 达到了 0.54。该实验证明通道空间注意力机制具有聚焦目标边缘特征、提高图像分割准确率的能力。

表 2 FEAM 消融实验

Table 2 FEAM ablation experiment

Test	FEAM_l	FEAM_h	ValMIoU/%
1	—	—	53.5
2	✓	—	53.6
3	—	✓	53.8
4	✓	✓	54.0

本文设计了 3 组对比消融实验,如表 3 所列,分别是原 Deeplabv3+、嵌入 FEAM 的 FEAM_DeepLabv3+ 以及嵌入 FEAM 和 FAM 的 FFDNet 依次进行对比的消融实验。

表 3 在 Deepfashion2 上添加不同模块的效果比较

Table 3 Effect comparison of Deepfashion2 with different modules

方法	Accuracy	MIoU	MPA	参数量
Deeplabv3+	0.902	0.535	0.656	59.34×10^6
FEAM_DeepLabv3+	0.904	0.540	0.667	59.87×10^6
FFDNet	0.906	0.552	0.672	59.95×10^6

实验结果如表 3 所列。由表 3 可看出,首先在主干网络 Resnet101 的首尾添加注意力机制模块 FEAM,相比原 Deeplabv3+ 在 MIoU 指标上有了 0.005 的提升,说明 FEAM 注意力机制模块使得服装位置和边缘轮廓分割得更仔细,语义信息提取更充实;其次在 FEAM 注意力机制模块的基础上继而添加特征对齐模块,可以看到总体有了 0.017 的提升,说明该模块增强了拥有细粒度特征的底层信息和拥有丰富语义特征的高层信息的关联度,提高了解码能力,进一步提高了分割精度。

为了能更进一步地直观感受模块带来的分割性能变化,本文对原图和分割图进行各 0.5 的混合,将背景信息置为黑色,便于直观展示出有效的服装信息。从图 8 可以看到,嵌入 FEAM 注意力机制和特征对齐模块 FAM 后服装边缘分割更

为平滑,从第一排短袖连衣裙的斜跨包红框处和第二排背心连衣裙的手臂处可以清晰地看出来,尤其是第二排背心连衣裙的手臂处,将模特手和服装边界分开的效果更好,更接近于真实的分割效果。综上所述,FFDNet 的分割更为精确和流畅,更适合于数据集 Deepfashion2 的服装分割。

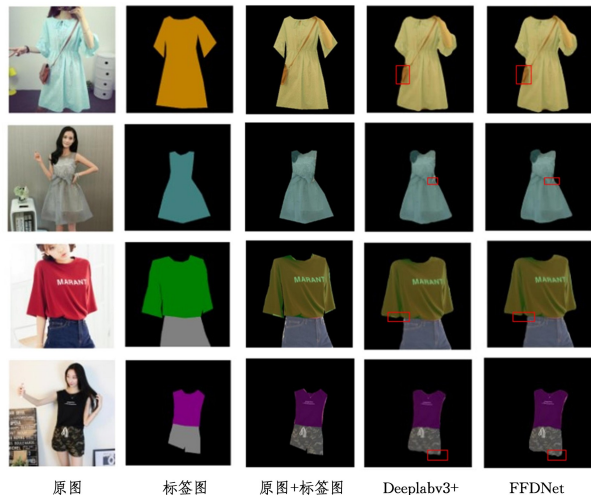


图 8 在 Deepfashion2 上的分割结果可视化

Fig. 8 Visualization of segmentation results on Deepfashion2

4.5 分割性能对比

1) 为了更进一步验证本文方法的有效性,选取了经典的分割网络 Swin-Unet^[28], PSPNet^[5], Deeplabv3+^[8] 和 FastFCN^[29] 与本文的 FFDNet 进行对比实验,实验量化指标采用 MIoU 和 MPA。结合表 4 可以看到,本文提出的 FFDNet 网络模型相比 Swin-Unet 和 PSPNet 在 MIoU 上有一定的提升,在原 Deeplabv3+ 网络和 FastFCN 上有 1.7% 和 0.7% 的提升。

表 4 与其他分割网络的性能进行比较

Table 4 Performance comparison with different segmentation networks

网络结构	MIoU	MPA
Swin-Unet ^[28]	0.392	0.446
PSPNet ^[5]	0.471	0.588
Deeplabv3+ ^[8]	0.535	0.648
FastFCN ^[29]	0.545	0.662
FFDNet	0.552	0.672

2) 为了说明模型的泛化能力,本文选取 Pascal VOC 2012 数据集来验证本文算法的合理性。由评价指标 MIoU 可以看到,本文提出的网络在非服装数据集上相比经典网络有所提升;由表 5 的分割评估结果可知,改进后的 Deeplabv3+ 网络 FFDNet 达到 79.4%。实验结果表明,基于 Deeplabv3+ 改进的网络 FFDNet 对 PASCAL VOC2012 数据集的图像分割是有效的。

表 5 VOC2012 数据集上的对比实验

Table 5 Comparative experiments on Voc2012 dataset

网络结构	MIoU
PSPNet	0.745
Deeplabv3+	0.773
FFDNet	0.794

为了进一步直观地看出本文算法用于 PASCAL VOC

2012 数据集的效果,图 9 给出了对比算法和本文算法 FFDNet 的分割图。第一行为原图,第二行为真实标签图,第三行和第四行为主干网络 Resnet101 的 PSPNet 和 Deeplabv3+ 网络分割图,第四行为改进后的 FFDNet 网络效果图。从图 9 的第二列可以看出,本文算法较好地分割出了奶牛的头和角,同理可以看出第三列旋转椅的椅子腿分割的效果较好,而

PSPNet 和 Deeplabv3+ 算法则没很好地分割出来。通过比较第四列公交车上人的分割结果,可以更清楚地看到本文算法 FFDNet 将小目标人的轮廓分割得更为细致。总体而言,虽然改进后的算法并不能完整地每一类准确地分割出来,但整体来看,改进后的 FFDNet 相比原始 Deeplabv3+ 算法有一定的提升。



图 9 PSACAL VOC2012 上的分割结果对比图

Fig. 9 Segmentation results comparison on PSACAL VOC2012

结束语 首先在主干网络 Resnet101 的 1/4 特征图上进行 FEAM 的注意力嵌入,优化图像边缘信息,随后在 Resnet101 的末尾进行卷积降维后送到 FEAM 模块中进行处理,同时,并行将 Resnet101 最后一层的输出特征图传到空洞卷积特征金字塔进行处理,继而将两个特征图进行相加融合,之后将获得具有丰富语义信息的特征图进行 1×1 降维并上采样,同时与经过 FEAM 处理的低级信息进行特征对齐融合,减少多尺度特征图融合过程中像素位置未对齐导致的分割精度不高的问题。实验结果证明,相比与原 Deeplabv3+ 网络,本文设计的网络是具有合理性和有效性的,并具有更好的分割效果,且 FEAM 和 FAM 只增加了少许参数,分割准确率并没有大幅降低。

对于当前服装分割网络,除了本文提到的特征对齐上采样模块,单一的直接融合编码器的高层语义信息和低层细节信息会导致细节特征容易被上下文特征所淹没,以致于小范围目标区域和边缘分割仍存在欠缺,未来可以进一步优化解码层,提高分割效率和准确率。

参考文献

- [1] HAN D Y, QIN X, LIU B S. Colored clothing image segmentation based on watershed region merging[J]. Journal of Guangxi Normal University (Natural Science Edition), 2015, 32(4): 75-81.
- [2] GAO Y P, SONG D, WANG Y J, et al. A modified K-means clustering clothing image segmentation algorithm[J]. Journal of Hunan College of Engineering(Natural Science Edition), 2021, 31(2): 54-59.
- [3] LONG J, SHELHAMER E, DARRELL T. Fullyconvolutional

- networks for semantic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [4] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C] // 18th International Conference Medical Image Computing and Computer Assisted Intervention (MICCAI 2015). Munich, Germany, Part III 18. Springer International Publishing, 2015: 234-241.
- [5] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.
- [6] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv: 1412. 7062, 2014.
- [7] 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.
- [8] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv: 1706. 05587, 2017.
- [9] 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.
- [10] YANG S, XU Z B, CHEN C. Automatic clothing image segmentation based on Faster R-CNN and Grabcut algorithm[J]. Intelligent Computers and Applications, 2020, 10(7): 306-310.

- [11] HUANG Z, WANG X, HUANG L, et al. Ccnet: Criss cross attention for semantic segmentation [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 603-612.
- [12] LI X, ZHONG Z, WU J, et al. Expectation maximization attention networks for semantic segmentation [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9167-9176.
- [13] ZHONG Z, LIN Z Q, BIDARTR, et al. Squeeze and attention networks for semantic segmentation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 13065-13074.
- [14] SONG Q, LI J, LI C, et al. Fully attentional network for semantic segmentation [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(2): 2280-2288.
- [15] HU J, SHEN L, SUNG. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [16] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [17] SONG Q, MEI K, HUANG R. AttaNet: Attention augmented network for fast and accurate scene parsing [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2021: 2567-2575.
- [18] ZHANG C, LIU X P, YUAN H, et al. Milling machine elastic image segmentation method based on improved DeepLabV3+ [J]. Journal of Optoelectronics, Lasers, 2023(5): 489-497.
- [19] GUO J, XIN Y L, XIE Q Q. Improved building segmentation in remote sensing image based on DeepLabV3+ [J]. Laser Journal, 2023: 1-10.
- [20] GU M H, LIU J, LI L Y, et al. Combining with the characteristics of learning and attention mechanism of clothing image segmentation [J]. Journal of Textile, Lancet, 2022(11): 163-171.
- [21] ZHAO Y, HE J. Clothing image segmentation using Deeplabv3+ algorithm with dual attention mechanism [J]. Journal of Chengdu University of Information Engineering, 2022(1): 67-71.
- [22] DENG F, FENG H, LIANG M, et al. FEANet: Feature enhanced attention network for RGB thermal real time semantic segmentation [C] // 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2021: 4467-4473.
- [23] XU J, XIONG Z, BHATTACHARYYA S P. PIDNet: A Real Time Semantic Segmentation Network Inspired by PID Controllers [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 19529-19539.
- [24] LI X, YOU A, ZHU Z, et al. Semantic flow for fast and accurate scene parsing [C] // European Conference on Computer Vision, Cham: Springer, 2020: 775-793.
- [25] HUANG Z, WEI Y, WANG X, et al. Alignseg: Feature aligned segmentation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(1): 550-557.
- [26] HUANG S, LU Z, CHENG R, et al. FaPN: Feature-aligned pyramid network for dense image prediction [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 864-873.
- [27] GE Y, ZHANG R, WANG X, et al. Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5337-5345.
- [28] CAO H, WANG Y, CHEN J, et al. Swin unet: Unet like pure transformer for medical image segmentation [J]. arXiv: 2105.05537, 2021.
- [29] WU H K, ZHANG J G, HUANG K Q, et al. FastFCN: rethinking dilated convolution in the backbone for semantic segmentation [J]. arXiv: 1903.11816, 2001.



XIAO Yahui, born in 1999, postgraduate, is a member of CCF (No. Q0221G). Her main research interests include machine learning and image processing.



ZHANG Zili, born in 1981, Ph.D, lecturer. His main research interests include machine learning and image processing.