文章编号：2096-1618(2022)01-0067-05

# 采用双注意力机制 Deeplabv3+算法的服装图像分割

赵 乙，何 嘉
(成都信息工程大学计算机学院，四川 成都 610225)

**摘要**：近年来服装时尚行业经济发展迅速，为了让用户选择服装和服装的设计更方便快捷，提高服装图像的分割效率尤为重要。目前的方法大多属于传统的分割方法，或者基于深度卷积神经网络(DCNN)。针对服装图像分割时易受背景、颜色、纹理等的影响，且服装的边缘分割不准确，基于 Deeplabv3+算法提出了双注意力机制的方法识别分割服装图像，使用通道注意力机制和位置注意力机制构成名为 CPAM 的模块对 Deeplabv3+网络进行改进。特征图经过多次下采样后再经过通道和位置注意力模块(CPAM)与 ASPP 模块并行，最后通过上采样得到预测图像。实验证明对不同场景的服装图像分割，加入 CPAM 模块的模型能更准确地将服装分割出来。

**关 键 词**：服装图像分割；DeepFashion2；Deeplabv3+；语义分割
**中图分类号**：TP391.41 **文献标志码**：A
**doi**：10.16836/j.cnki.jcuit.2022.01.012

## 0 引言

近年来，由于时尚产业的巨大潜力，服装的视觉分析也成了研究的一大热门。但是服装图像通常因为场景姿势的问题，存在遮挡，以及消费者与商业图像在服装领域有所不同，在实际的应用中，消费者对于服装图像的理解还有一些挑战[1]。

对于服装设计师来说，服装图像的准确分割可以更好地获取消费者的穿着喜好信息，也可以提升自己的工作效率[2]。同时，消费者可以更方便了解服装信息。对于服装图像的处理和视觉分析，获得清晰的轮廓和更好的分类服装、织物和纤维图像[3]是后续处理的必要条件之一。

服装图像的分割由于服装本身属性繁多，面料款式纹理均有所不同，时常还受背景颜色的制约[4]。例如，图片中模特穿着上衣和短裙时，很容易被机器理解为连衣裙，因为这样的搭配外观上较为相似[4]。并且在实际场景下，每张服装图像尺度差异较大。因此，如何解决多尺度服装图像分割以及将服装图像进行特征融合是服装图像语义分割中亟待解决的问题。

文献[5]提出了一种自动推荐服装的方法，这种方法需要先识别人体的姿态，再确定出服装的区域，最后把服装部分分割出来。该方法实现了服装的自动推荐，但是这种模型分割的精确度非常低。文献[6]通过巨大的数据集检索来查看相似衣服以标记查询的图像。针对服装图像分割方法通常需要依赖于人体姿势的问题，文献[7]提出了在不考虑人体姿态的情况下改进的图像分割方法。该方法利用条件随机场模型，很大程度减轻了纹理复杂服装的过度分割以及估计不准确，但是对肤色类似的情况效果较差。白美丽等[8]提出了自监督学习框架，与 Deeplabv2 网络结合形成端到端的深度卷积网络框架，将人体姿态加入服装解析中，但服装分割的边界依旧不准确，且前景和背景无法很好地分离。

自 2012 年 AlexNet[9]成立以来，深度学习快速发展。深度卷积神经网络(DCNN)[10]与传统的机器学习方法相比，提取特征的功能比较强大。在图像分割领域，基于全卷积网络(FCN)[11]的最新方法得到了大力发展。

最近 Deeplabv3+[12]语义分割网络已经普遍和流行，王中宇等[13]通过在 Deeplabv3+中引入卷积块注意力模型得到了分割图像的简单方法，使边缘目标分割更加精细，最终将改进的网络运用于自动驾驶场景中。

基于以上提到的问题，文中将通道注意力机制和位置注意力机制引入 Deeplabv3+模型，将双注意力机制与 ASPP 模块进行并行，形成 CPAM 模块，并通过 DeepFashion2 数据集验证了改进模块后网络的效果。

## 1 网络与算法

### 1.1 Deeplabv3+网络介绍

Deeplabv1 版本提出空洞卷积操作，扩大感受野，获取更稠密的特征图。v2 版本在 v1 版本基础上提出 ASPP 结构，该结构使用空洞卷积对特征图进行采样操作，再连接条件随机场以便于获取更为精准的分割图像。v3 版本去除了全连接条件随机场(CRF)操作，将 ASPP 结构改为 3 个 3×3 卷积操作，空洞率分别为 6、12、18，再连接全局平均池化。Deeplabv3+网络在 v3

收稿日期：2021-10-14

基础上又添加了编码-解码结构,中间部分进行一次上采样,最后再进行一次上采样,其完整的网络结构如图 1 所示。改进后的网络利用卷积层对输入的服装图像进行特征提取,接着使用池化层降低特征图的维度,这个过程也叫作下采样。下采样的过程会导致信息丢失严重,这对语义分割任务不利。故而 Deeplabv3+在深度特征提取网络中加入 ASPP 模块,不仅可以增加感受野,还能减少下采样过程中的信息损失。此外,Deeplabv3+网络还能达到对多尺度目标的分割能力。

最后,为减少下采样过程中丢失信息对分割造成不利影响,Deeplabv3+采用 Encode-Decode 结构,使用 Encode 结构进行特征提取,使用 Decode 进行上采样。在上采样过程中融合较低层次的特征,恢复目标部分边界信息,上采样过程使用的是线性插值方法。

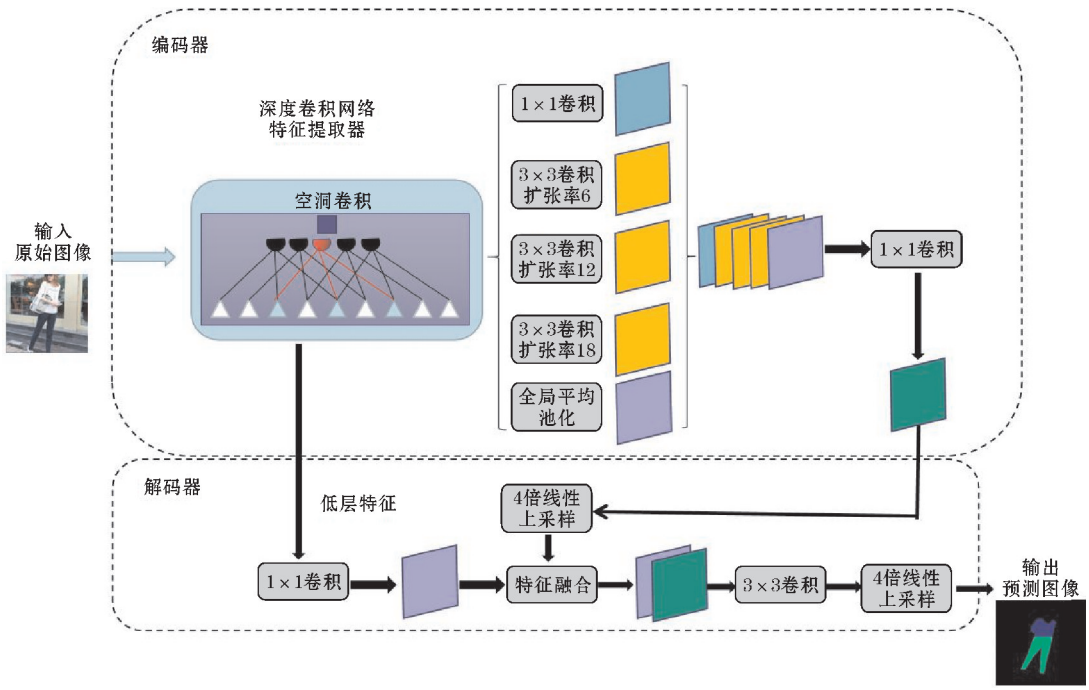


图 1 Deeplabv3+网络结构

## 1.2 改进的 Deeplabv3+算法

### 1.2.1 通道注意力模块

在深度学习的领域,通道注意力机制[14]得到更加频繁的使用。通道注意力模块用于语义分割时,不同的通道特征图存在紧密度不同的联系,提取不同通道的语义信息,可以使相互联系的特征图更加突出。通道注意力模块如图 2 所示。

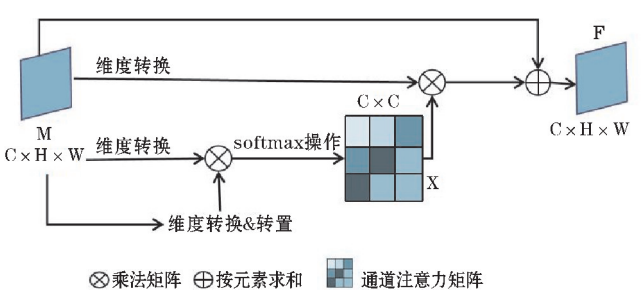


图 2 通道注意力模块

通道注意力直接通过 $M \in \mathrm{R}^{C\times H\times W}$ 计算出通道注意力图 $\boldsymbol{X} \in \mathrm{R}^{C\times C}$

$$X_{ji}=\frac{\exp[M_i^{'}\cdot(M^{'})_j^{\mathrm{T}}]}{\sum_{i=1}^{c}\exp[M_i^{'}\cdot(M^{'})_j^{\mathrm{T}}]} \tag{1}$$

式中,$M_i$ 代表第 $i$ 个元素值;$(M')^{\mathrm{T}}$ 表示转置矩阵的第 $j$ 个元素值;$X_{ij}$ 表示第 $i$ 个通道对第 $j$ 个通道的影响因子。对 $\boldsymbol{X}$ 转置后与 $M'$ 做矩阵乘法,与参数 $\beta$ 相乘后再与矩阵相加得到最终结果 $F$,如式(2)所示。

$$F_j=\beta\sum_{i=1}^{c}(X_{ij}M_i)+M_j \tag{2}$$

式中,$\beta$ 初始为 0。由式(2)可知,每个通道最终特征都是所有通道特征与原始通道特征的加权和。

### 1.2.2 位置注意力机制

位置注意力机制如图 3 所示。位置注意力模块旨在使用整个图像中的任意两点之间的关联增强其各自特征的表达式,还可以捕获全局空间上下文信息。

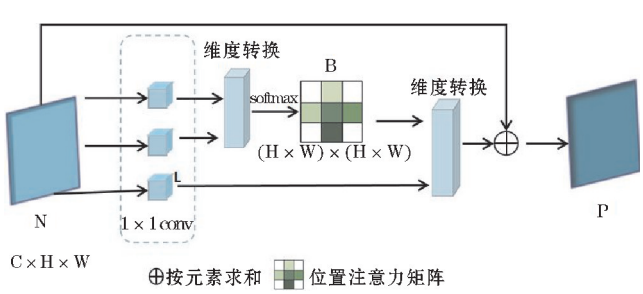


图 3 位置注意力模块

通过主干网络可以得到局部特征信息 $N \in R^{C\times H\times W}$,经过两次维度变换和一次转置操作,得到位置注意力 $\boldsymbol{B}$。如式(3)所示。

$$B_{ji}=\frac{\exp(O_i \cdot Q_j)}{\sum_{i=1}^{N}\exp(O_i \cdot Q_j)} \tag{3}$$

式中,$O_i$ 表示矩阵 $\boldsymbol{O}$ 第 $i$ 个位置的元素,$Q_j$ 同理。$N$ 为通道当中元素的个数,将 $L$ 进行维度变换后与 $\boldsymbol{B}$ 的转置矩阵进行乘法操作,最后再经过一次维度变换,转换成 $R^{C\times H\times W}$ 形状的 $\boldsymbol{P}$ 矩阵

$$P_j=\alpha\sum_{i=1}^{N}(B_{ji}L_i)+N_j \tag{4}$$

式中,$B_{ji}$ 为矩阵 $\boldsymbol{B}$ 的第 $i$ 个位置元素;$\alpha$ 为可学习的参数,初始值为 0。每个位置最后得到的特征 $\boldsymbol{P}$ 是所有位置和原始位置的特征进行加权得到的。因此,能够根据位置注意力图使上下文信息得到更好的聚合,同时相似的特征权重更高能够起到相互促进的作用。

### 1.3　基于双注意力机制的 Deeplabv3+网络

Deeplabv3+中的 ASPP 模块使用空洞卷积进行特征融合,使用的是空洞率不同的卷积操作。这是由于若空洞率过大则容易丢失部分特征,导致大尺度目标再分割时容易有棋盘格类型的空洞现象,这就会使大尺度目标分割的准确率降低。基于此,本文采用通道注意力和位置注意力模块与 ASPP 并行的方法,以弥补上述不足。并且,采用实验对比双注意力机制和单注意力机制的差距,改进后的 Deeplabv3+网络结构如图 4 所示。

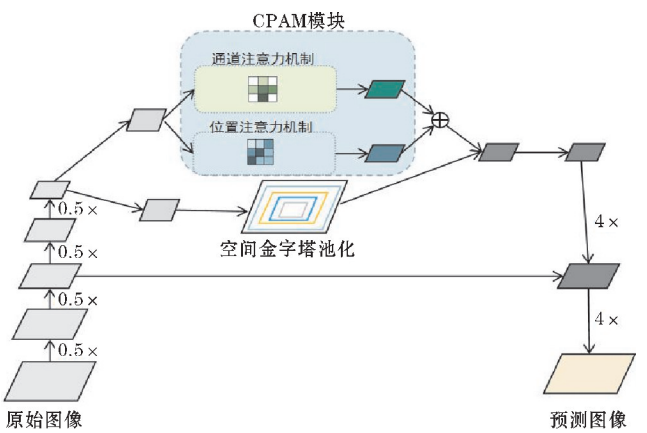


图 4　Deeplabv3+添加 CPAM 模块

图 4 是添加双注意力模块后的整体结构。可以看到卷积层的输出结果,会与空间金字塔池化[15]进行并行,并进行两次上采样得到预测图像。

## 2　实验

### 2.1　数据集

为验证本文所提方法的有效性,选择了 DeepFashion2[16]数据集。该数据集是时尚研究团队使用的最流行的数据之一,使用了高分辨率的服装图像。

实验选用了该数据集中带有细粒度标识的图像共 19.2 万张,这些图像都来自于不同平台的卖家和买家图片,差异度大,背景各有不同。实验使用 DeepFashion2 作为改进网络验证数据集,它的标注量是 Deepfashion 的 3.5 倍。其中有 49.1 万张图片,共有 4.38 万个服装标识,该数据集能够支持服装检测、分类、分割以及服饰检索。在实验中,使用到 DeepFashion2 数据集中 19.2 万张带有细粒度标签的服装图像。数据集设置了 13 个服装类别标签,short_sleeve_top,long_sleeve_top,long_sleeve_outwear,short_sleeve_outwear,vest,sling,shorts,trousers,skirt,short_sleeve_dress,long_sleeve_dress,vest_dress,sling_dress。实验中不再更改图像的格式。

### 2.2　实验参数与环境

为保证实验的公平性,使用了控制变量的方法将原网络和改进网络设置相同的超参数,4 个实验均将初始化学习率设置为0.0005,batch size 设置为 2,学习率使用多项式衰减方式,网络优化方法使用 momentun,动量为0.9,weight decay 设置为0.0005,并使用在 ResNet101 主干训练过的权值作为训练前的权值。实验在 Windows10 系统中执行,处理器为 AMD Ryzen 5 3600X 6-Core Processor 3.80 GHz,显卡型号是 NVIDIA GeForce RTX 2070 SUPER,显存为 8GB。深度学习框架为 pytorch1.6.0。

### 2.3　评价指标

语义分割有许多指标可以用来衡量算法的精度,实验选用准确率(accuracy)、平均交并比(mIoU)和频权交并比(FWIoU)3 种指标来评价网络的精度。以上 3 种标准值越高,代表预测的效果越好。

### 2.4　注意力模块嵌入实验

为验证双注意力机制方法在 Deeplabv3+结构中的有效性,设计了 4 个实验,分别是 Deeplabv3+网络、在 Deeplabv3+加入通道注意力机制的网络、在Deeplabv3+加入位置注意力机制的网络、同时在 Deeplabv3+加入通道注意力机制和位置注意力机制的网络进行对比,实验结果如表 1 所示。由表 1 可以看出,在Deeplabv3+加入注意力机制后的网络更加具有优势,其对于 Deeplabv3+的性能都具有促进作用。其中,同时加入通道注意力和位置注意力之后的网络效果最好,不加入的网

络,总体准确率是 84%,mIoU 是 34%,FWIoU 是 79%。由此可见,加入双注意力的结构性能优于其他结构。

表 1 注意力机制的消融实验 单位:%

| Method | PA | CA | accuracy | mIoU | FWIoU |
|---|---|---|---|---|---|
| Deeplabv3+ | | | 84 | 34 | 79 |
| 通道注意力机制 | | √ | 89.8 | 52.08 | 80.29 |
| 位置注意力机制 | √ | | 90.02 | 52.08 | 80.93 |
| 通道和位置注意力机制 | √ | √ | 92.29 | 52.41 | 83.94 |

表 2 比较了经典分割网络 FCN-8S、Deeplab 系列网络中的 Deeplabv2 以及其他作者在这个系列网络上改进的网络,共 5 个实验。通过比较本文设计的双注意力模块结构与其他图像分割网络在 DeepFashion2 中的结果,得到如下结论:加入 CPAM 模块的网络在 Deepfashion2 的测试集上的分割精度相较于其他比较的网络都有所提高,这证明了改进网络在分割精度上有很大的提高。

表 2 本文方法与其他分割网络结果对比 单位:%

| Method | accuracy | mIoU | FWIoU |
|---|---|---|---|
| FCN-8S | 71.28 | 28.69 | 61.28 |
| SegNet | 69.10 | 18.37 | 59.34 |
| Deeplabv2 | 82.89 | 41.56 | 73.87 |
| Deeplabv2-SSL | 83.37 | 42.46 | 76.65 |
| 双注意力机制网络 | 92.29 | 52.41 | 80.93 |

实际环境中,服装图像会受各种外在因素的影响,例如服装花纹和场景相似、服装纹理复杂、服装款式复杂等因素,使得服装图像的分割很容易将不同目标分割成同一个,且分割的精度不高。另外,当服装显示不完全或者服装图像比例过大时,可能会有分类不准确的情况出现。由于 Deeplabv3+具有多尺度特征,可以将小标签的物体分割出来,且通过加入双注意力模块,网络能够较好地融合上下文信息,且对相似特征有更好的促进作用。

图 5 展示了本文提出的 CPAM 改进 Deeplabv3+的模型与传统分割方法和 Deeplabv3+对比实验的效果图。从图 5 显示的效果来看,文献[17]使用传统方法分割的图像较为粗糙,缺少大量的边缘细节,且服装边界参差不齐,瑕疵较多。剩下的两种都基于 Deeplab 系列网络,实验结果可知,单独的 Deeplabv3+网络不如本文改进的网络结构效果好,尤其是在大的服装类型方面。本文提出的 CPAM 模块加入 Deeplabv3+后,较 Deeplabv3+的分割精度提高明显,且在细节部位优化的效果更好,整体的分割精度也优于其他方法。

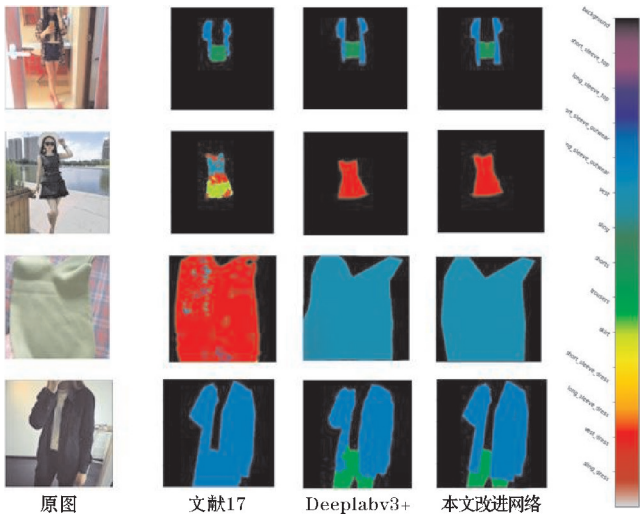


图 5 不同方法可视化结果

## 3 结束语

采用 CPAM 模块的 Deeplabv3+深度语义分割网络对服装图像进行分割,并利用 DeepFashion2 训练集对模型进行训练和测试,得到的主要结论为:

服装图像的分割需要服装语义和上下文线索的更高级别的信息。针对服装图像语义分割准确度较低的问题,提出一种基于 Deeplabv3+的双注意力机制深度神经网络来解决服装图像分割问题,结合两种注意力机制与 ASPP 模块并行,实现了端到端的深度卷积框架,该方法可以获取通道和位置上的上下文信息,注重对重点通道和重点位置的特征。实验结果表明,基于 Deeplabv3+改进的模型能够提高服装图像的分割效果,能够有效地将前景与背景分开,减少背景对整体服装分割的影响。

由于网络模型参数较多,对计算机的运算能力是一个挑战,且由于服装款式众多,以及图片中服装位置的不同,可能对服装的分割有很大的影响。当服装部分占全图大部分内容时,网络对服装的分割准确率较低。后期将考虑运用模型压缩和枝剪对算法进行优化,以期做到在网络性能不变的情况下,轻量化模型。

## 参考文献:

[1] Jouanneau W, Bugeau A, Palyart M, et al. Where Are My Clothes? A Multi-Level Approach for Evaluating Deep Instance Segmentation Architectures on Fashion Images[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:3951-3955.

[2] Castro H, Ramirez M. Segmentation task for fashion and apparel[EB/OL]. arXiv preprint arXiv:2006:11375.

[3] 张艳红,杨思,徐增波. 图像分割技术在服装领域的应用[J]. 软件导刊,2020,19(4):238-241.
[4] Inacio A D S, Lopes H S. EPYNET: Efficient Pyramidal Network for Clothing Segmentation[J]. IEEE Access,2020,8:187882-187892.
[5] Silvestre L. Regularity of the obstacle problem for a fractional power of the Laplace operator[J]. Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences,2007,60(1):67-112.
[6] 黄冬艳,刘骊,付晓东,等. 基于 HOG 和 E-SVM 的服装图像联合分割算法[J]. 计算机工程与应用,2017,53(18):199-203.
[7] 李冬艳,陈文雄. 上半空间高次分数阶 Laplace 方程解的不存在性[J]. 纺织高校基础科学学报,2017,30(1):18-22.
[8] 白美丽,万韬阮,汤汶,等. 一种改进的用于服装解析的自监督网络学习方法[J]. 纺织高校基础科学学报,2019,32(4):385-392.
[9] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems,2012,25:1097-1105.
[10] Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition,2013:3476-3483.
[11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE conference on computer vision and pattern recognition,2015:3431-3440.
[12] 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.
[13] 王中宇,倪显扬,尚振东. 利用卷积神经网络的自动驾驶场景语义分割[J]. 光学精密工程,2019,27(11):2429-2438.
[14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition,2018:7132-7141.
[15] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence,2015,37(9):1904-1916.
[16] 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.
[17] Yamaguchi K, Hadi Kiapour M, Berg T L. Paper doll parsing: Retrieving similar styles to parse clothing items[C]. Proceedings of the IEEE international conference on computer vision,2013:3519-3526.

# Garment Image Segmentationn Using Dual Attention Mechanism Deeplabv3+ Algorithm

ZHAO Yi　HE Jia

(College of Computer science, Chengdu University of Information Technology, Chengdu 610225, China)

**Abstract**: In recent years, the garment fashion industry economy has developed rapidly . In order to make the user´s choice of clothing and clothing design more convenient and fast, it is particularly important to improve the efficiency of clothing segmentation. Most of the present methods are traditional segmentation methods or based on deep convolutional neural network (DCNN). For the clothing image segmentation task is easily affected by background, color, texture, etc., and the clothing edge segmentation is not accurate, this paper proposes a method of dual-attention mechanism based Deeplabv3+ algorithm to identify and segment clothing images. Channel attention mechanism and location attention mechanism are used to form a module named CPAM to improve Deeplabv3+ network. After downsampling for several times, the feature image is parallel to the channel and position attention module (CPAM) and The ASPP module, and then the prediction image is obtained by upsampling. Finally, the experiment proves that the model with CPAM module can segment the clothing image more accurately in different scenes.

**Keywords**: clothing image segmentation; DeepFashion2; Deeplabv3+; semantic segmentation