

基于 U-Net 的复杂环境军事人员图像分割方法

陶志文, 牛福*

军事科学院系统工程研究院后勤科学与技术研究所, 北京 100071

摘要 针对当前多种复杂环境中军事人员图像分割的精度和效率问题, 提出一种基于改进的密集空洞卷积模块和串行注意力模块的编-解码网络。首先, 在 U 型编-解码网络中引入密集空洞卷积模块并对其进行改进, 提高网络分割多尺度目标的能力并降低参数量; 其次, 在 U 型编-解码网络中引入串行注意力模块, 使神经网络更加聚焦于图像中更重要的特征上; 最后, 对 U 型编-解码网络的编码结构中每次下采样后的卷积进行改进, 减少网络参数量。在多环境迷彩图像分割数据集上的实验结果表明, 提出的网络模型的平均交并比分别比 U-Net、SegNet、FCN-8s 高 2.27 个百分点、4.93 个百分点、10.46 个百分点, 有效降低了参数量, 改善了网络对多种复杂环境中军事人员的分割效果。

关键词 图像处理; 军事人员分割; 特征融合; 注意力机制; 复杂环境

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202259.0810005

Image Segmentation Method of Military Personnel in Multiple Complex Environments Based on U-Net

Tao Zhiwen, Niu Fu*

Institute of Logistics Science and Technology,

Academy of Systems Engineering, Academy of Military Sciences of Chinese PLA, Beijing 100071, China

Abstract Aiming at the problems of accuracy and efficiency of military personnel image segmentation in multiple complex environments, we propose an encoder-decoder network based on improved dense atrous convolution and serial attention modules. First, we add the dense atrous convolution module in the U-shaped encoder-decoder network to improve the network's ability to segment multiscale targets and reduce parameter amounts. Second, we add the serial attention module in the U-shaped encoder-decoder network, enabling the network to focus more on the important features in the image. Finally, convolution after each downsampling in the encoder structure of the U-shaped encoder-decoder network is improved to reduce the parameter amounts. The experimental results on the multiple environments camouflage dataset show that the mean intersection of the union of the proposed network is 2.27 percent, 4.93 percent, and 10.46 percent higher than U-Net, SegNet, and FCN-8s, respectively. The parameter amounts are significantly reduced, improving the effectiveness of the network for military personnel segmentation in multiple complex environments.

Key words image processing; military personnel segmentation; feature fusion; attention mechanism; complex environment

1 引言

人员搜索与救援是军队执行各类任务中一项

重要的工作, 高效的人员搜索与救援对有效保存军队的有生力量具有重要意义。图像分割作为人工智能的一个重要研究方向, 在人员搜索与救援中可

收稿日期: 2021-11-15; 修回日期: 2021-11-25; 录用日期: 2021-12-21

通信作者: *niufu@vip.sina.com

以发挥至关重要的作用,使用训练好的神经网络进行复杂环境中军事人员的分割对提高搜寻与救援的效率和成功率具有重要意义,对进一步进行图像内容的自动理解、图像场景的分析等任务具有重要作用。

在深度学习技术出现之前,主要利用数字图像处理等方面的知识进行图像分割研究,主要方法有基于阈值的图像分割方法^[1]、基于边缘检测的图像分割方法^[2]以及区域法^[3],传统的图像分割方法效果精度较低,且较为复杂。

深度学习技术出现后,图像分割逐渐成为了深度学习中的一个研究热点。Long 等^[4]提出了全卷积神经网络(FCN),FCN可以对图像进行像素级的分类,解决了像素级别的图像分割问题。由于FCN参数量较大,模型较为复杂,在FCN之后,Ronneberger 等^[5]提出了U-Net图像分割网络,U-Net是一种编-解码结构的分割网络,在图像分割中表现出众,得到了广泛应用。图像的全局上下文表示对神经网络的特征提取有重要作用,Zhao 等^[6]于2017年提出了金字塔场景解析网络(PSPNet),该网络可以更好地学习场景的全局上下文表示。为了在降低对计算资源的消耗的同时能提高分割效果,Badrinarayanan 等^[7]于2017年提出了SegNet图像分割网络,SegNet对计算机的资源消耗较低,预测分类的准确率相比FCN有一定提高。陈小龙等^[8]提出了一种用于语义分割的分组双注意力网络,该网络能够准确地捕获更加有效的图像上下文信息。李娜等^[9]提出了一种基于多尺度注意力网络的行人属性识别算法,提高了行人属性识别的准确

率。基于深度学习的图像分割技术相比传统的分割方法在分割精度上有了进一步的提升,但现有的针对复杂环境中的军事人员图像分割网络分割性能较低,难以进行较高精度的分割,针对复杂环境中中小目标较多的情况没有良好的分割效果。

关于多种复杂环境下军事人员的语义分割的研究较少,梁新宇等^[10]针对迷彩伪装目标侦察识别的作战保障要求,构建了包含四种野外环境的迷彩伪装目标图像语义分割数据集;李鑫等^[11]基于无监督网络对军事目标检测算法进行了研究,通过生成对抗网络有效解决了数据集不足和精确度较低的问题。

为解决多种复杂环境中军事人员的分割精确率较低和网络模型参数量较大的问题,提出一种基于U-Net的分割网络,以提高分割精度,并降低参数量,提升在复杂环境中对军事人员的搜寻与救援的效率。

2 基于U-Net的复杂环境军事人员图像分割网络

2.1 DESI-U-Net模型

以U-Net为主干网络,基于改进的密集空洞卷积(DAC)^[12]模块、串行的高效通道注意力模块(ECAM)^[13]、空间注意力模块(SAM)^[14]以及分解卷积的方法^[15],提出了新的编-解码网络模型DAC-ECAM-SAM-INCEPTION-U-Net(简称DESI-U-Net)。DESI-U-Net的结构如图1所示,图例注明了不同模块所代表的含义。

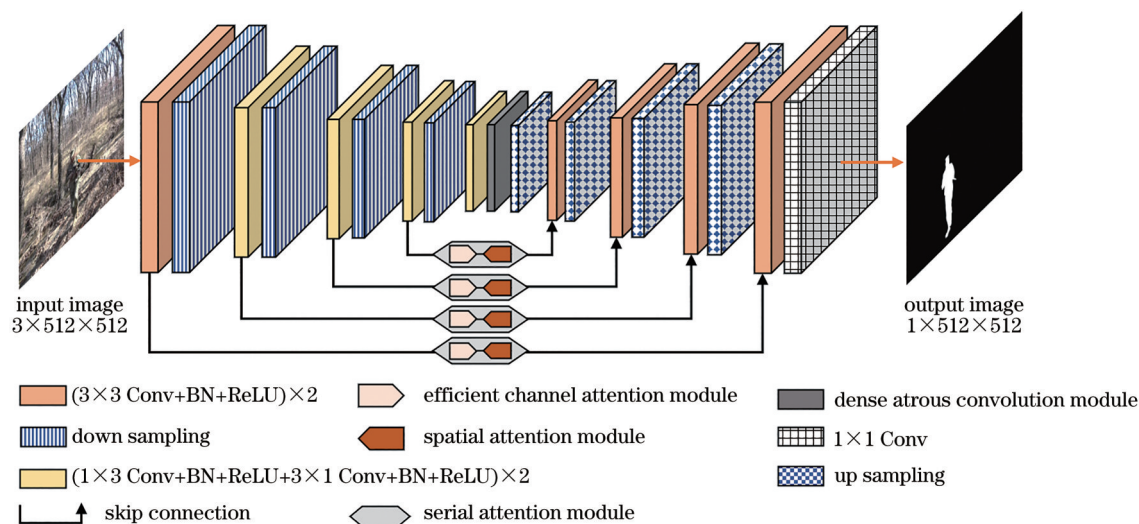


图1 DESI-U-Net结构

Fig. 1 Architecture of DESI-U-Net

为了提升网络对图像中不同尺度目标的分割性能,增强网络对多尺度特征的提取能力并降低参数量,在 U-Net 的编码结构与解码结构之间加入了改进的 DAC 模块;在 U-Net 的跳跃连接 (skip connection) 上按照串行的方式依次加入 ECAM 和 SAM,使得网络能同时关注特征在通道和空间两个方面的重要程度,让网络模型能更加聚焦于图像中要分割的目标,增强网络对重要目标的分割效果;最后将 U-Net 的编码结构中每次下采样后的 3×3 卷积分解为一个 1×3 卷积和一个 3×1 卷积,以减少网络参数量,提升计算效率^[15],增强网络的分割能力。

2.2 改进的密集空洞卷积模块

在多种复杂环境中,不同尺度的目标较多,为了提高网络分割多尺度目标的能力,扩大感受野,提高分割效果,在网络中引入了 Gu 等^[12]提出的 DAC 模块,并对其进行了改进。DAC 模块包含四个不同的分支,每个分支级联了一个或多个不同空洞率的卷积,每个分支的感受野为 3, 7, 9, 19, 可以用于提取图像中不同尺度的特征。改进后的 DAC 模块如图 2 所示,将 DAC 中原有的 3×3 空洞卷积替换为 3×3 的空洞深度可分离卷积(DSC)^[16],以此来达到提取多尺度特征并降低网络参数量的效果。

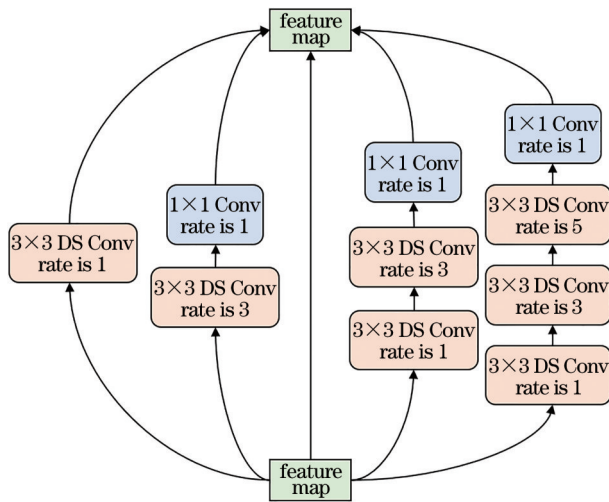


图 2 改进的 DAC 模块结构

Fig. 2 Structure of improved DAC module

由于在复杂环境中军事人员的迷彩和环境的相似度很高,因此提取更多的多尺度特征有助于网络模型提取到更丰富的复杂环境中的军事人员的特征信息。在复杂环境中的搜寻与救援任务中,常使用小型移动设备进行搜寻,而引入深度可分离卷积的 DAC 模块有助于降低网络模型的参数量和复

杂度,使得网络模型能更容易地嵌入到移动设备中,提高网络模型在复杂环境中的应用价值,因此对 DAC 模块进行了上述改进。

DSC 由深度卷积和 1×1 的逐点卷积组成。DSC 经深度卷积对特征图的每个通道进行特征提取后,再经逐点卷积对得到的特征图进行合并,最后得到输出的特征图。DSC 的计算公式为

$$C_{D(h,w)} = \sum_{k,k} K_{(k,k)} I_{(h-k,w-k)}, \quad (1)$$

$$C_{P(h,w)} = \sum_c K_{(1,1)} I_{(h,w,c)} \quad (2)$$

由(1)、(2)式可推导出 DSC 的计算公式,表达式为

$$C_{DS(h,w)} = C_P [K_{(1,1)}, C_D(K_{(k,k)}, I)], \quad (3)$$

式中: $C_{D(h,w)}$ 表示深度卷积; $C_{P(h,w)}$ 表示逐点卷积; K 表示卷积核; I 表示输入的特征图; c 表示输入特征图的通道数^[17]。

相比标准卷积, DSC 参数量有较大下降。设输入的特征图尺寸为 $C_1 \times H_1 \times W_1$, 输出的特征图尺寸为 $C_2 \times H_2 \times W_2$, 卷积核大小为 $k \times k$, 其中 C_1 表示输入特征图的通道数, C_2 表示输出特征图的通道数, H 和 W 分别表示特征图的高和宽。标准卷积的参数量 F 和 DSC 的参数量 F' 可分别表示为

$$F = k \times k \times C_1 \times C_2, \quad (4)$$

$$F' = C_1 \times k \times k + 1 \times 1 \times C_1 \times C_2. \quad (5)$$

DSC 的参数量和标准卷积的参数量比值 α 为

$$\alpha = \frac{F'}{F} = \frac{C_1 \times k \times k + 1 \times 1 \times C_1 \times C_2}{k \times k \times C_1 \times C_2} = \frac{1}{C_2} + \frac{1}{k^2}. \quad (6)$$

如果使用普通的标准卷积,网络参数量将较大,而采用深度可分离卷积后可以大幅降低网络的参数量,在对分割精度影响较小的情况下大幅降低模型参数量,从而降低计算成本。

2.3 串行注意力模块

注意力机制 (attention mechanism) 是神经网络中的一个重要概念,最早被提出并应用于机器翻译模型中^[18],现在在人工智能的各个研究方向中有着广泛的应用。深度学习中的注意力机制与人类视觉的注意力机制相似,主要思想是将有限的计算资源分配给更重要的信息,提高重要的信息权重,降低不重要的信息权重。在计算机视觉任务如图像分割中,注意力机制的主要思想是让神经网络着重学习图像中更重要的区域。

串行注意力模块 (SERAM) (为与空间注意力模块 SAM 相区别,故简称为 SERAM) 的结构如

图 3 所示。为了在神经网络中获得更好的训练效果,并且不会有太多的额外计算量,Wang 等^[13]提出了高效通道注意力模块(ECAM),Woo 等^[14]在 CBAM 中使用了空间注意力模块。受 CBAM 的思想启发^[14],将 ECAM 和 CBAM 中的 SAM 按顺序以串行的方式组合起来,得到串行注意力模块。

由于 ECAM 参数量低且效率高的特点,因此采用 ECAM 作为串行注意力模块的组成部分,以实现

更高效的在通道上为特征图分配权重的作用。而 SAM 可以使得网络能关注到特征图在空间上的重要性,提升空间维度上高价值特征信息的权重。因此采用串行的 ECAM 和 SAM 的注意力模块,以达到提高分割结果准确率的同时不显著提高网络参数量的效果,并使得网络能同时关注特征在通道和空间两个维度上的权重,进一步提高所提网络模型的特征表达能力。

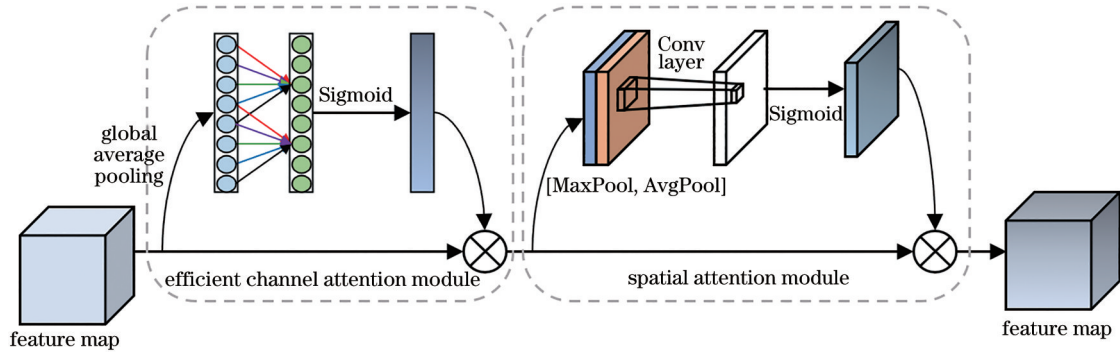


图 3 串行注意力模块的结构

Fig. 3 Structure of serial attention mechanism

ECAM 首先经全局平均池化,再经过一个卷积核大小为 k 的一维卷积,经过 Sigmoid 后与输入进行点乘,实现在通道上的权重分配。ECAM 是在 SE 注意力模块^[19]的基础上进行改进得到的,ECAM 在性能和参数量上都优于 SE 注意力模块。

SE 注意力模块是 ECAM 的基础。SE 注意力模块先对输入的特征图进行全局平均池化,并进行特征压缩(squeeze)。压缩的公式为

$$\mathbf{z} = F_{\text{squeeze}}(\mathbf{x}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \mathbf{x}(i, j), \quad (7)$$

式中: \mathbf{x} 表示输入的特征图^[19]。

对特征图压缩后进行激励(excitation),激励的公式为

$$s = F_{\text{excitation}}(\mathbf{z}, W) = \sigma[W_2 \delta(W_1 \mathbf{z})], \quad (8)$$

式中: W_1 为 $C \times \frac{C}{r}$, W_2 为 $\frac{C}{r} \times C$, r 表示缩放系数, C 表示通道数; σ 表示 Sigmoid 激活函数; δ 表示 ReLU 激活函数; s 表示经过激励后得到的权重^[19]。

将权重 s 融入原输入的特征图中,公式^[19]为

$$\bar{\mathbf{x}} = F_{\text{scale}}(\mathbf{x}, s) = s\mathbf{x}. \quad (9)$$

ECAM 对 SE 注意力模块进行了优化和改进,降低了模型的复杂度,只考虑邻近的通道之间的相互关系。ECAM 的权重计算公式为

$$\omega_m = \sigma \left(\sum_{p=1}^q \omega_m^p y_m^p \right), \quad y_m^p \in \Omega_m^q, \quad (10)$$

式中: y_m 表示特征图的通道; ω_m 表示通道的权重大小; Ω_m^q 表示 y_m 的 q 个邻近通道的集合。

ECAM 通过一维卷积来实现高效的权重分配和计算,一维卷积的卷积核大小为 k ,权重公式为

$$\omega = \sigma[C_{1D, k}(y)], \quad (11)$$

式中: C_{1D} 表示一维卷积; y 表示特征图的通道^[13]。

一维卷积的卷积核的大小 k 可自适应确定:

$$k = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}}, \quad (12)$$

式中: γ 的值为 2; b 的值为 1; $\lfloor t \rfloor_{\text{odd}}$ 表示离 t 最近的奇数^[13]。

SAM 首先分别经全局平均池化和全局最大池化,再对结果进行拼接,对结果进行一次卷积和经 Sigmoid 激活函数后再与输入进行相乘,得到分配权重后的特征图^[14]。SAM 的计算公式为

$$M_s(\mathbf{F}) = \sigma \left\{ f^{7 \times 7} \left(\left[\text{AvgPool}(\mathbf{F}); \text{MaxPool}(\mathbf{F}) \right] \right) \right\} = \sigma \left\{ f^{7 \times 7} \left(\left[\mathbf{F}_{\text{avg}}^s; \mathbf{F}_{\text{max}}^s \right] \right) \right\}, \quad (13)$$

式中: \mathbf{F} 表示输入的特征图; $\text{AvgPool}(\cdot)$ 表示全局平均池化; $\text{MaxPool}(\cdot)$ 表示全局最大池化; $f^{7 \times 7}$ 表示卷积核大小为 7×7 的卷积^[14]。

串行注意力模块使得网络同时关注特征在通道和空间两个方面的重要程度,并且参数量较小,让网络能更加聚焦于图像中要分割的目标,降低无关信息的权重,提高网络对军事人员的分割能力。

2.4 卷积分解

Szegedy 等^[15]在 Inception-v3 中对 $n \times n$ 的卷积进行分解,将其分解为一个 $1 \times n$ 卷积和一个 $n \times 1$ 的卷积,这种方法可以降低网络模型的参数量,节约计算成本,提高运算速度,从而提升网络的性能。应用 Inception-v3 中将 $n \times n$ 的卷积分解为一个 $1 \times n$ 卷积和一个 $n \times 1$ 的卷积的方法,将所提网络的编码结构中每次下采样后的 3×3 卷积分解为一个 1×3 和一个 3×1 卷积,以达到进一步降低网络模型参数量的目的。

虽然应用了改进的 DAC 模块后,网络模型的参数量已大幅下降,但在对分割精度影响非常小的情况下,应用卷积分解的方法可进一步降低网络参数量,从而进一步降低网络模型的复杂度,使得网络模型的参数量和分割精度达到更好的平衡。

设输入的特征图大小为 $C \times H \times W$,输出的特征图大小为 $C' \times H' \times W'$, $n \times n$ 卷积的参数量 L 的计算公式为

$$L = n \times n \times C \times C'. \quad (14)$$

卷积分解后,将 $1 \times n$ 卷积的输入和输出通道均设置为 C ;将 $n \times 1$ 卷积的输入通道设置为 C ,输出通道设置为 C' 。经卷积分解后的卷积参数量 L' 的计算公式为

$$L' = 1 \times n \times C \times C + n \times 1 \times C \times C'. \quad (15)$$

标准卷积的参数量和分解卷积后的参数量的比值 γ 为

$$\gamma = \frac{L}{L'} = \frac{n \times n \times C \times C'}{1 \times n \times C \times C + n \times 1 \times C \times C'} = \frac{n \times C'}{C + C'}. \quad (16)$$

按照 DESI-U-Net 的改进方式,对编码结构中每次下采样以后的 3×3 卷积进行分解。当 $n=3$,且当 $C=64$ 和 $C'=128$ 、 $C=128$ 和 $C'=256$ 、 $C=256$ 和 $C'=512$ 、 $C=512$ 和 $C'=1024$ 时, γ 均大于 1,证明了标准的 3×3 卷积的参数量大于一个 1×3 卷积串联一个 3×1 卷积的参数量。

引入改进的 DAC 模块和 SERAM 后,在网络模型每次下采样后采用不同类型卷积的网络的参数量如表 1 所示。从表 1 可以看出,经卷积分解后,网络的参数量降低了大约 18.8%。

表 1 采用不同类型卷积的网络的参数量

Table 1 Parameter amounts of network with different type of convolution

Convolution type	Parameter amount / 10^6
3×3 Conv	38.75
1×3 Conv + 3×1 Conv	31.45

3 实验与结果分析

3.1 损失函数

损失函数是提升模型分割性能非常重要的一个因素。实验所使用的多环境迷彩数据集 (MECD) 中存在正负样本不均衡的问题,在 MECD 中,军事人员目标占比小,而复杂环境占比大。参考 Zhu 等在文献[20]中使用的损失函数,采用混合 Focal Loss^[21]和 Dice Loss^[22]的损失函数作为网络训练的损失函数,以改善 MECD 中正负样本不平衡的情况。损失函数的公式为

$$L(y, \hat{y}) = L_F(y, \hat{y}) + L_D(y, \hat{y}). \quad (17)$$

Focal Loss 的公式为

$$L_F(y, \hat{y}) = -\alpha \hat{y}^{\gamma} (1 - \hat{y})^{\gamma} \log_e y - (1 - \alpha)(1 - \hat{y})^{\gamma} \log_e (1 - y), \quad (18)$$

式中: γ 为聚焦参数; α 为不同样本类别间的权重参数。

Dice Loss 的公式为

$$L_D(y, \hat{y}) = 1 - \frac{2|y \cap \hat{y}| + 1}{|y| + |\hat{y}| + 1}, \quad (19)$$

式中: y 表示实际标签值; \hat{y} 表示模型预测结果。

3.2 评价指标

在复杂环境人员搜寻与救援任务中,神经网络预测的正确率是非常重要的,因此使用平均交并比 (mIoU) 来评估神经网络的分割精度,其计算表达式为

$$P_{\text{mIoU}} = \frac{1}{n} \sum_{i=1}^n \left(\frac{|P \cap G_T|}{|P \cup G_T|} \right)_i, \quad (20)$$

式中: $|P|$ 表示图像分割神经网络预测的军事人员区域; $|G_T|$ 表示标注图像的实际军事人员区域。

3.3 数据集

为了验证网络的分割性能,采用建立的 MECD 对网络分割复杂环境中军事人员的性能进行评估。MECD 是为了满足多种复杂环境(包括城市废墟、山地、雪地等环境)中针对军事人员的搜寻与救援任务的需求而构建的数据集。MECD 的构建方法

参考了梁新宇等^[10]构建的迷彩伪装目标图像语义分割数据集的制作方式。MECD 含有多种角度、多种复杂环境的身着迷彩服及装备的军事人员图像。MECD 包含的迷彩种类丰富,含有多个国家的不同种类的迷彩服;含有多种复杂环境类型,包括城市废墟、雨林、丛林、山地、荒漠、雪地 6 种典型的复杂环境。

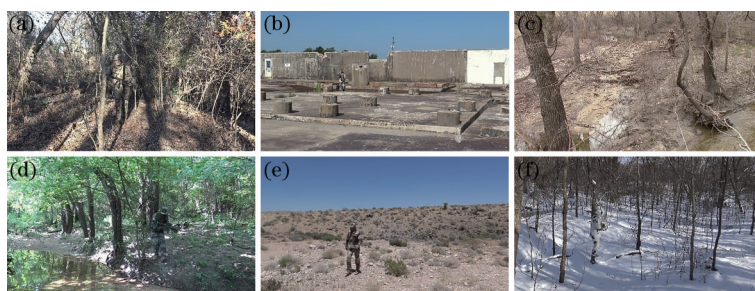


图 4 MECD 包含的 6 种复杂环境示例图像。(a)丛林;(b)城市废墟;(c)山地;(d)雨林;(e)荒漠;(f)雪地
Fig. 4 Six example images of complex environments in MECD. (a) Jungle; (b) city ruins; (c) mountain field; (d) rainforest; (e) desert; (f) snowfield

表 2 MECD 包含的迷彩类型信息

Table 2 Camouflage type information in MECD

Serial No.	Camouflage type	Country	Camouflage pattern
1	Phantomleaf WASP II Z4 (Urban)	Germany	City camouflage
2	Phantomleaf WASP. II . Z2	Germany	Jungle camouflage
3	Schneetarn	Germany	Snowfield camouflage
4	Partizan	Russia	Jungle camouflage
5	Snow KLMK	Russia	Snowfield camouflage
6	CCE	France	Rainforest camouflage
7	Lizard (Leopard)	France	Mountain camouflage
8	M04 Hellepuku	Finland	Universal camouflage
9	AOR II	USA	Jungle camouflage
10	A-TACS FG-X	USA	Rainforest camouflage
11	UCP (Universal Camouflage Pattern)	USA	Universal camouflage
12	SLOCAM	Slovenia	Rainforest camouflage
13	Chinese type 07 arid camouflage	China	Desert camouflage
14	Chinese type 07 jungle camouflage	China	Jungle camouflage

3.4 网络训练

实验使用的软硬件配置如下:操作系统为 Ubuntu 18.04、CUDA 版本 11.0、Python 3.8、深度学习框架使用 Pytorch 1.7、训练中通过 1 块显存为 31 GB 的 NVIDIA Tesla V100 显卡进行加速。训练模型中的 epoch 设置为 100,初始学习率设置为 0.0001,batch size 设置为 8,采用 Adam 优化方法优化模型。

3.5 实验结果

实验中选择 FCN-8s^[4]、U-Net^[5]、SegNet^[7] 这

图 4 展示了 MECD 中包含的 6 种典型复杂环境。

表 2 为 MECD 中包含的 14 种不同种类的迷彩的详细信息。

MECD 中图像及其标注样本各有 1600 张。采取在训练中进行数据增强的方式扩增数据集。训练集、验证集和测试集按照 7:1:2 进行划分。

3 种基于深度学习的算法在 MECD 上进行实验,在实验条件、图像尺寸、显卡等相同的情况下,将它们得到的结果与 DESI-U-Net 的结果进行对比。实验结果如表 3 所示。

表 3 中的实验结果表明,DESI-U-Net 在 mIoU 上均优于 U-Net、SegNet、FCN-8s,mIoU 分别比 U-Net、SegNet、FCN-8s 高 2.27 个百分点、4.93 个百分点、10.46 个百分点,DESI-U-Net 在 MECD 上的分割精度更高,更加准确。

参数量是一个网络模型复杂度的一种体现,参

表 3 不同方法的 mIoU

Table 3 mIoU of different methods

Method	mIoU / %
FCN-8s	71.38
SegNet	76.91
U-Net	79.57
DESI-U-Net	81.84

数量过大会使得网络模型过于臃肿。由于在复杂环境中进行人员搜寻与救援通常采用的是无人机、机器人等小型搜救设备,因此轻量化的模型在嵌入小型的移动设备时更有优势,轻量级的模型有更广泛的应用前景和价值。表 4 比较了多种网络模型的参数量。由表 4 可知,DESI-U-Net 的参数量更低,更

表 4 不同网络模型的参数量

Table 4 Parameter amounts of different methods

Method	Parameter amount / 10^6
FCN-8s	134.27
SegNet	29.44
U-Net	35.05
U-Net+SERAM+DAC (without DSC)	63.89
U-Net+SERAM+DAC (with DSC)	38.75
DESI-U-Net	31.45

加轻量,且在分割效果上更有优势,有良好的应用前景。

图 5 展示了 FCN-8s、SegNet、U-Net 和 DESI-



图 5 四种网络在 6 种复杂环境中的部分分割结果。(a)原图;(b)标注图;(c) FCN-8s 分割结果;(d) SegNet 分割结果;(e) U-Net 分割结果;(f) DESI-U-Net 分割结果

Fig. 5 Partial segmentation results of four networks in 6 complex environments. (a) Original image; (b) ground truth; (c) segmentation results of FCN-8s; (d) segmentation results of SegNet; (e) segmentation results of U-Net; (f) segmentation results of DESI-U-Net

U-Net 在丛林、雨林、山地、雪地、城市废墟、荒漠 6 种环境中的部分分割结果。所提 DESI-U-Net 在多种复杂环境中对军事人员的分割有良好的效果提升。

从图 5 的分割结果中可以看到:FCN-8s 的分割结果较差,其分割结果与标注样本相去甚远,有明显欠缺;SegNet 的分割结果也较差,图像的分割结果均有较大的缺陷;U-Net 的分割结果相比 SegNet 和 FCN-8s 的分割结果较好,但在细节上仍有一定欠缺;DESI-U-Net 的分割结果较好,图像的分割结果都是最接近标注图像的,在细节与边缘区域也有良好的分割效果,对枪支装备等细节部分的分割效果较好。

在多种复杂环境下,DESI-U-Net 的分割结果明显好于 U-Net、SegNet 和 FCN-8s,无论是在丛林、雨林、山地还是雪地、城市废墟或者荒漠环境中,所提 DESI-U-Net 在 MECD 上都有良好的分割效果,且模型参数量小,复杂度低。

3.6 消融实验

为了验证改进的 DAC 模块、串行注意力模块、卷积分解的改进组合对提升复杂环境中军事人员分割性能的影响,进行了消融实验研究。在 U-Net 的基础上逐步添加上述模块,比较添加不同模块的网络在 MECD 上的 mIoU,消融实验结果如表 5 所示。

表 5 消融实验结果
Table 5 Results of ablation experiment

Method	mIoU / %
U-Net+DAC (without DSC)	80.69
U-Net+DAC (with DSC)	80.32
U-Net+SERAM+DAC (with DSC)	82.16
DESI-U-Net	81.84

表 5 中的消融实验结果表明:将未改进的 DAC 模块加入到 U-Net 中,mIoU 有一定提升,相比 U-Net 提高了 1.12 个百分点,但由表 4 可知其参数量较大;将改进后的 DAC 模块加入到 U-Net 中,mIoU 虽然有一定下降,但下降幅度较小,参数量却大幅降低了;将改进后的 DAC 模块、SERAM 加入到 U-Net 中后,效果提升较大,相比 U-Net,提高了 2.59 个百分点,但参数量还可以进一步降低;网络模型 DESI-U-Net 的 mIoU 虽然相比未进行卷积分解的网络模型仅降低了 0.32 个百分点,但由表 1 可知其参数量降低了 18.8%。在对最终分割结果影

响很小的情况下,采用卷积分解的方式更好地达到了分割精度和参数量的平衡。

综上所述,DESI-U-Net 在分割精度和参数量上做到了较好的平衡,能实现更加鲁棒和高效的复杂环境中的军事人员分割。

4 结 论

针对多种复杂环境中军事人员图像分割精度较差且模型参数量大的问题,基于 U-Net 提出了分割网络 DESI-U-Net。首先在 U-Net 的编码结构后加入了改进的 DAC 模块,提升了网络对 MECD 中多尺度目标的分割能力的同时降低了模型的参数量;接着在 U-Net 的跳跃连接上加入了串行注意力模块,使得网络能同时关注到特征在通道和空间两方面的权重,提升了网络对特征重要程度的判断能力,增强了分割效果;最后,将编码结构中每次下采样后的 3×3 卷积分解为 1×3 卷积和 3×1 卷积,一定程度上降低了模型的复杂度,减少了参数量。在 MECD 上的实验结果表明,提出的 DESI-U-Net 相比 U-Net、SegNet 和 FCN-8s 有更好的分割效果,mIoU 达到了 81.84%,且参数量降低到了大约 31.45×10^6 ,在分割精度和模型参数量上都有较大优势。DESI-U-Net 能更加精确地分割多种复杂环境中的军事人员,改善了分割效果且参数量更少,在小型移动设备上有良好的应用价值,在针对多种复杂环境中军事人员的搜救任务上有良好的应用前景。

参 考 文 献

- [1] Otsu N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.
- [2] Davis L S. A survey of edge detection techniques[J]. Computer Graphics and Image Processing, 1975, 4(3): 248-270.
- [3] Adams R, Bischof L. Seeded region growing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(6): 641-647.
- [4] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [5] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]// Navab N, Hornegger J, Wells W M, et al. Medical

- image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [6] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.
- [7] 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.
- [8] Chen X L, Zhao J, Chen S Y, et al. Grouped double attention network for semantic segmentation[J]. Laser & Optoelectronics Progress, 2021, 58(22): 2210007.
陈小龙, 赵骥, 陈思溢, 等. 分组双注意力语义分割网络[J]. 激光与光电子学进展, 2021, 58(22): 2210007.
- [9] Li N, Wu Y Y, Liu Y, et al. Pedestrian attribute recognition algorithm based on multi-scale attention network[J]. Laser & Optoelectronics Progress, 2021, 58(4): 0410025.
李娜, 武阳阳, 刘颖, 等. 基于多尺度注意力网络的行人属性识别算法[J]. 激光与光电子学进展, 2021, 58(4): 0410025.
- [10] Liang X Y, Lin H K, Yang H, et al. Construction of semantic segmentation dataset of camouflage target image[J]. Laser & Optoelectronics Progress, 2021, 58(4): 0410015.
梁新宇, 林浩坤, 杨辉, 等. 迷彩伪装目标图像语义分割数据集的构建[J]. 激光与光电子学进展, 2021, 58(4): 0410015.
- [11] Li X, Wang S Q, Li A. A military target recognition algorithm based on unsupervised network[J]. Electronics Optics & Control, 2021, 28(10): 36-39.
李鑫, 王晟全, 李昂. 基于非监督网络的军事目标识别算法的研究[J]. 电光与控制, 2021, 28(10): 36-39.
- [12] Gu Z W, Cheng J, Fu H Z, et al. CE-net: context encoder network for 2D medical image segmentation [J]. IEEE Transactions on Medical Imaging, 2019, 38(10): 2281-2292.
- [13] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [14] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [15] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2818-2826.
- [16] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2021-06-01]. <https://arxiv.org/abs/1704.04861>.
- [17] Li C H, Lu Y. Facial expression recognition based on depthwise separable convolution[J]. Computer Engineering and Design, 2021, 42(5): 1448-1454.
李春虹, 卢宇. 基于深度可分离卷积的人脸表情识别[J]. 计算机工程与设计, 2021, 42(5): 1448-1454.
- [18] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [EB/OL]. (2014-09-01) [2021-05-04]. <https://arxiv.org/abs/1409.0473>.
- [19] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [20] Zhu W T, Huang Y F, Zeng L, et al. AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy[J]. Medical Physics, 2019, 46(2): 576-589.
- [21] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [22] Milletari F, Navab N, Ahmadi S A. V-net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV), October 25-28, 2016, Stanford, CA, USA. New York: IEEE Press, 2016: 565-571.