视频中稳定的跨场景前景分割

魏宗琪,梁 栋
(南京航空航天大学 计算机学院,江苏 南京 211100)

摘 要:通过训练单个模型进行跨场景前景分割是一项具有挑战性的任务,特别是对于大规模视频监控,因为现有的模型通常严重依赖特定场景的信息。光流是描述前景目标的运动信息,但现有的光流机制方法只能表示瞬时特征,对开放的环境变化不具有鲁棒性。为了通过细粒度的运动特征表示并适应场景实现前景分割,一种间隔光流的新模块被设计出来,并使用注意力模块将运动特征融合到模型中。基于这种互补机制,可以构建实现运动和外观特征交互的跨模态动态特征滤波器。与现有方法相比,提出的模块倾向于在前景和背景区域的运动模式之间学习更多的语义信息,从而获得更好的跨场景适应性和鲁棒性。此外,由于数据集偏差问题,在跨场景前景分割任务中小目标的分割结果不佳,因此进一步设计了一个类内尺度的焦点损失函数来平衡前景目标的大小多样性。提出的模块可以即插即用到任意视频监控识别框架中,提高了跨场景前景分割结果的质量。

关键词:前景分割;双重模态;注意力;跨场景;自适应
中图分类号:TP391.4 **文献标识码**:A **文章编号**:1673-629X(2022)12-0037-06
doi:10.3969/j.issn.1673-629X.2022.12.006

Stable Cross-scene Foreground Segmentation in Video

WEI Zong-qi, LIANG Dong
(School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China)

Abstract: Cross-scene foreground segmentation by training a single model is a challenging task, especially for large-scale video surveillance, because off-the-shelf models usually rely heavily on specific scene information. Optical flow is the motion information describing the foreground target. However, the existing optical flow mechanism methods can only represent the instantaneous motion and are not robust to open set. In order to achieve scene adaptation for foreground segmentation through fine-grained motion feature representation and interaction, interval optical flow was designed to combine fine-grained motion features with the attention module. Based on this, a cross-modal dynamic feature filter that realizes the interaction of motion and appearance features can be constructed. Compared with existing methods, the proposed module tends to learn more semantic information between the motion patterns of the foreground and background, so as to obtain better cross-scene adaptability and robustness. In addition, because the data set deviation usually misses small objects in cross-scene foreground segmentation tasks, a focus loss function of classification scale is further designed to balance the size diversity of foreground instances. The proposed module can be plug-and-played into any video surveillance recognition framework to improve the quality of the cross-scene foreground segmentation mask.

Key words: foreground segmentation; dual modality; attention; cross-scene; adaptive

0 引 言

视频前景分割旨在发现视频中视觉上显著的移动前景对象,并从背景中识别覆盖这些对象的所有像素。视频前景分割结果可以作为许多其他任务的重要预处理组件,例如图像和视频压缩[1]、视觉跟踪[2]和行人重新识别[3]。然而,在实际的应用时,仅训练一个用于大规模跨场景视频前景分割的深度模型仍然是一个具有挑战性的问题,因为现成的基于深度学习的分割模型依赖于场景特定的结构信息。模型训练去适应新场景需要额外费力的场景标注和从头开始训练或微调模型,否则前景尤其是微小的前景的分割结果会受到影响。

传统的无监督前景减法方法[4-6]侧重于建立统计模型来抑制动态背景的干扰,但它们在实现准确的背景更新方面存在瓶颈,同时还有使用卷积神经网络[7-11]代替背景减法的方法,但这些方法都是特定于

收稿日期:2021-11-20 **修回日期**:2022-03-22
基金项目:国家自然科学基金资助项目(61772268)
作者简介:魏宗琪(1996-),男,硕士研究生,通讯作者,研究方向为计算机视觉;梁 栋,副教授,研究方向为计算机视觉。

场景的,需要针对其他场景从头开始训练。深度背景减法模型(Deep Background Subtraction, DeepBS)[12]和时空注意力模型(Spatial Temporal Attention Model, STAM)[13]利用经过训练的卷积神经网络来实现跨视频场景的前景分割。跨场景分割往往比较粗糙,无法很好地保留物体和小物体的边界。由于卷积神经网络的发展,语义分割方法取得了显著进展。SOTA 方法包括 PSPNet[14]、DeepLabV3+[15]、BFP[16]和 CCL[17]。尽管语义分割方法可以为每一帧提供高级语义注释,但它们忽略了对视频前景分割非常重要的时间相关性和运动线索。

从本质上讲,前景分割是一项与场景外观、运动和场景属性相关的分割任务。端到端模型训练为场景外观和运动特征的有效混合和融合提供了一条路径,可以获取运动前景区域和过滤场景中的复杂背景信息。光流是一种瞬时运动提示,但是鲁棒性较差且不足以描述像素级别的运动(运动目标整体)。针对现有的前景分割任务,该文试图解决以下问题:(1)如何更全面地描述场景中的前景;(2)即使是在新场景中使用,能否实现无需额外训练的即插即用的前景分割模型。通过集成来自不同模态(前景的运动和外观)的更多特征来解决这些问题,然后通过注意力模块引导的选择性连接结构消除没有前景代表性的特征。提出间隔光流注意力模型(Interval Optical Flow Attention Model,IOFAM),如图 1 所示。

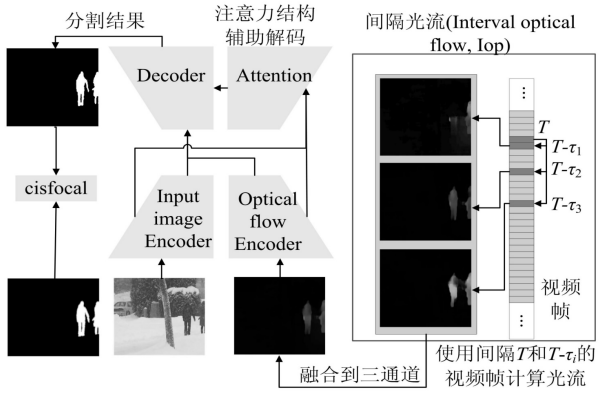

图 1 间隔光流注意力模型

1 研究现状

早期的研究集中在统计分布上来构建背景模型[5-6,18]。对视频数据中时空局部的描述[19-21]揭示了背景模型能够在保持时空依赖性上有显著的效果。上述统计建模方法通常计算成本低,有利于资源受限的视频监控系统。然而,为了消除光照变化和动态背景带来的影响,通常使用不精确的渐进背景更新解决方案[5]:(1)选择性更新,只有在将新样本归类为背景样本时才将新样本添加到模型中;(2)盲选更新,每个新样本都添加到模型中。选择性更新必须决定每个预测像素值是否是背景的一部分,利用分割结果作为更新标准可以看作是实现这一任务的一种简单方法,而无效的分割决策可能会导致之后的错误分割。盲选更新机制允许将不属于背景的强度值添加到模型中,但这会导致更多的假阴性,因为前景像素可能会错误地成为模型的一部分。必须对更新率进行权衡,该更新率调节更新背景模型的传播。由于对较小或临时变化的敏感性,高更新率会导致嘈杂的分割,而低更新率会产生过时的背景模型并导致错误分割。利用超像素[22-24]对背景更新,采用自适应阈值、颜色特征和图像纹理等对前景目标进行分割,将图像划分超像素块处理是分割中一种有效的方式。

基于深度神经网络的前景分割:

Brahamand[7]提出了第一种使用 CNN 进行背景减法的方法,该方法在给定的 N 个视频帧上执行时间特征维度的中值操作,然后通过图像帧、背景和地面实况像素的相应图像块来训练特定于场景的网络。MFC3-D[9]使用多尺度 3D 卷积来检测红外视频的前景对象。MSNet[10]使用生成对抗网络来生成背景。概率模型[11]将每个视频帧分成块,输入到用于去噪的自动编码器组中提取重要特征。分割模型[25]结合了边缘检测算法,在人体前景检测中对错误的分割背景进行过滤,使用边缘校正通道在深度分割网络中处理人体假阳性的问题。上面提到的所有方法都是特定于场景的,即如果将模型应用到其他新的场景,则需要从头开始训练。DeepBS[12]是第一种利用经过训练的卷积神经网络进行跨视频场景的前景分割任务的方法,但没有考虑运动信息。对于训练数据,它从 CDNet2014 数据集中随机选择 5% 的样本以及每个子集的相应地面实况。SAFF[26]融合了语义信息,在语义和表观特征的基础上进行前景分割,在目标的显著性和轮廓实现更精确的分割。为了解决前景背景颜色相近、物体遮挡等问题,基于双边网络[27]实现了视频像素级前景分割任务,将高维的特征空间通过降维至当前视频帧特征中,实现特征融合。为了应对光线因素对前景分割的影响,基于 ViBe[6]融合多帧差分法[28]的 RGB 图像及深度图像进行建模,然后利用选取基准(SC)融合策略和前景区域直方图信息优化目标结果。

2 间隔光流注意力模型

2.1 网络结构

间隔光流注意力模型如图 1 所示。所提出的模型使用编码-解码结构,对静态视频帧外观特征和场景运动信息进行编码,并在解码过程中集成了注意力模块(Attention)以融合视频帧和光流两个编码器

(Encoder)和解码器(Decoder)的特征。

2.2　间隔光流

该文提出的间隔光流用于增强对场景中目标运动准确性的描述。光流作为瞬时运动描述特征,在表现运动方面缺乏稳定性和充分性。来自长间隔视频帧的光流具有物体的长期运动线索,但物体的轮廓不精确;短间隔视频帧计算的光流具有当前帧的准确运动线索,但有时不足以描述整个运动物体,例如图 1 中右侧框的第一个光流。间隔光流(IOF),如图 1 右,使用当前视频帧和不同长度的间隔帧计算 3 个光流,不同帧间隔计算得到不同特性的光流可以相互补充,实现充分运动特征和准备运动目标轮廓描述的特征综合。具体步骤:通过设置间隔当前帧的长度参数 τ_1、τ_2 和 τ_3,得到当前时刻 τ 的帧位置,以及 $T-\tau_1$、$T-\tau_2$ 和 $T-\tau_3$ 时刻的帧,最后计算 T 时刻的光流信息,记为 $\mathrm{Op}(\tau_1)$、$\mathrm{Op}(\tau_2)$ 和 $\mathrm{Op}(\tau_3)$。将具有不同间隔的三个光流合并到三个通道中作为间隔光流 $\mathrm{Iop}(T)$,直接使用已有光流模型直接计算光流。

2.3　注意力模块

该文提出一种新的注意力模块,旨在解码器阶段通过密集的注意力过程合并解码器和编码器特征,为解码过程提供更充分的时空特征。具体来说,首先提取高级特征用来提供全局信息,然后指导注意力模块加权适当的低级特征,即预测输入图像中的两种编码器特征融合为具有外观和运动信息的特征,通过解码器层对像素级特征重新加权并与后者连接。

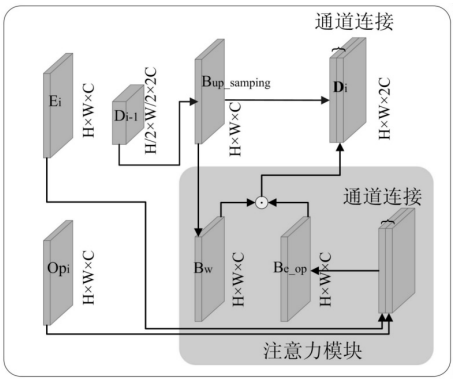

图 2　注意力模块

在图 2 中,解码过程是从前一个解码层 D_{i-1} 到下一层 D_i。输入特征包括对应编码层视频帧特征 E_i 和光流特征 Op_i 以及解码器中的前一层解码特征 D_{i-1},输出部分是解码器层特征 D_i。为了更清楚地解释 Attention 模块的运行机制,使用 B_w 和 $B_{\mathrm{e\_op}}$ 作为这一过程阶段的结果。具体过程如下:假设得到了两个特征图张量 $E_i \in R^{H\times W\times C}$ 和 $\mathrm{Op}_i \in R^{H\times W\times C}$($H$ 和 W 是单个特征图的高度和宽度,C 表示特征图通道数)。为了得到 D_i,首先在两个编码器中连接了两种对应的特征图 E_i 和 Op_i,进行拼接后,通道 C 变成原来通道的两倍 $2C$,然后通过卷积得到 $B_{\mathrm{e\_op}} \in R^{H\times W\times C}$:

$$B_{\mathrm{e\_op}} = \mathrm{conv}(\mathrm{Relu}(E_i \parallel \mathrm{Op}_i)) \tag{1}$$

其中,conv 表示卷积核 3×3,公式 1 用于提取外观特征和减少通道,$\parallel$ 用于通道连接,Relu 是激活函数。在解码层 $D_{i-1} \in R^{H/2\times W/2\times 2C}$,做上采样卷积得到 $B_{\mathrm{up\_sampling}} \in R^{H\times W\times C}$。然后通过卷积和激活操作得到加权系数特征 $B_w \in R^{H\times W\times C}$(系数值在 0 和 1 之间)。

$$B_w = \mathrm{BN}(\sigma(\mathrm{conv}(\mathrm{Relu}(B_{\mathrm{up\_sampling}})))) \tag{2}$$

其中,σ 是 Sigmoid 激活函数,conv 表示卷积核 3×3,BN 是批量归一化(Batch Normalization)。然后 B_w 与特征图 $B_{\mathrm{e\_op}}$ 通过矩阵对位相乘得到加权特征图(Atten 结果),这一步是 Attention 模块中解码器的加权操作。批量归一化后,从 $B_{\mathrm{up\_sampling}}$ 中得到原始解码器特征,在原来的 Decoder 特征中加入了 Dropout(dpt)操作,每个节点在训练过程中都有 50% 的概率被抑制,在推理过程中去掉这个操作,将加权编码器特征图和原始解码器特征连接起来,得到当前解码层 i 中的 $D_i \in R^{H\times W\times 2C}$。

$$D_i = (B_w \cdot B_{\mathrm{e\_op}}) \parallel \mathrm{BN}(\mathrm{dpt}(B_{\mathrm{up\_sampling}})) \tag{3}$$

其中,·是矩阵的对位点乘。

2.4　损失函数

Focal Loss[29] 的提出是为了解决模型训练中的正负不平衡以及难易样本的问题,用于基于二元交叉熵函数的对象检测。结合前景分割任务,为了解决小目标分割结果不好的问题,定义了一帧 $S(fg)$ 中前景和背景的面积比,然后在前景类内定义一个平衡系数 β,如下所示:

$$\beta = t_3 \min(1/S(fg), 50) \tag{4}$$

其中,t_3 是一个超参数。设置 β 取 $S(fg)$ 和 50 最小值的原因是为了防止潜在场景没有目标的情况,防止无穷大,其中 50 是训练场景中小物体采样后设置的值。为了改善小目标结果,基于调整面积的参数提出用于平衡前景类别内部的类内尺度焦点损失(Class in Scale Focal loss,cisfocal):

$$L_{\mathrm{cisfocal}} = \begin{cases} -\beta\alpha(1-p)^{\gamma}\log(p), y=1 \\ -(1-\alpha)p^{\gamma}\log(1-p), y=0 \end{cases} \tag{5}$$

其中,p 表示模型预测的概率,前景标签 $y=1$,背景标签 $y=0$。α 是前景和背景像素样本的平衡参数,γ 是调节难易样本的参数,对于困难样本,它将获得较低的权重。β 是用于平衡前景中不同尺寸的目标参数,对于小目标,为了让模型更关注它,损失将适当调大。为了稳定地训练模型,在训练过程中加入曼哈顿距离 l1 loss 作为正则化。它是在预测的 p 和真实值 y 之间测量的,$L_{l1} = \| p - y \|_1$。最终的损失函数可以表示

如下：

$$L = t_1 L_{\text{cisfocal}} + t_2 L_{l1} \tag{6}$$

3 实 验

3.1 数据集及预处理

在两个数据集(CDNet 2014[30]和 LIMU[31])上评估所提出的前景分割模型的分割效果。按照 DeepBS[12]中的训练设置,对于训练数据,从 CDNet 2014 中的 5 万张数据集随机选择 5% 的样本及不同场景特点的子集的标注来训练模型。CDNet 2014 中剩下的 95% 的样本用于测试模型,没有任何训练集重叠。模型基于 CDNet 2014 数据集训练,为了验证模型的跨场景能力,在没有经过训练的 LIMU 数据集进行直接的推理,分为 CameraParameter (CP)、Intersection (ITS)和 LightSwitch (LS)三个具有不同特点的场景,分割前景无需任何后处理即可获得。

3.2 实验环境与设置

在实验过程中提前做了很多超参数调优的实验,对比了很多不同的设置。最后对于实验中的间隔光流,设置 $\tau_1 = 1$, $\tau_2 = 5$ 和 $\tau_3 = 10$。在损失函数中,最后设置 $t_1 = 0.8$, $t_2 = 0.2$, $t_3 = 0.25$, $\alpha = 0.75$, $\gamma = 0$。训练批次数据个数大小为 16,总共训练了 160 个 epoch。用 Adam 作为优化器,其 $\text{beta}_1 = 0.95$, $\text{beta}_2 = 0.999$,学习率设置为 5×10^{-5} 的小值。实验基于两张 1080Ti 卡的环境下进行。

3.3 评价指标

使用 Recall = TP/(TP+FN)、Precision = TP/(TP+FP)和 F-measure(F1) = 2×Recall×Precision/(Recall+Precision)作为实验的评价指标,对像素级的分割结果的评价,TP、FP 和 FN 表示前景结果的正检、错检和漏检,Recall 表示完整性,Precision 表示边缘准确性,F-measure(F1)则是综合指标。

3.4 消融实验

在消融实验中,验证了间隔光流、注意力结构和类内尺度焦点损失,综合上述的模块得到的结果最优,在综合指标 F-measure(F1)达到 0.977 6。如表 1 所示,对比第 1、2、3 和 8 行的结果,结合间隔光流的模型具有显著的提升。对比第 1 和 9 行,验证注意力结构,在综合指标 F1 中提升 9.85 个百分点。对比第 1、4、5、6 和 7 行的结果,最好的损失函数的组合为 cisfocal loss 和 l1 loss 的组合。

表 1 在 CDNet 2014 数据集上的消融实验

	Op(τ_1)	Op(τ_2)	Op(τ_3)	Att	cisfocal	focal	l1	F1/%
1	√	√	√	√	√		√	0.977 6
2	√	√		√	√		√	0.971 4
3	√			√	√		√	0.963 2
4	√	√	√	√		√	√	0.972 1
5	√	√	√	√	√			0.973 2
6	√	√	√	√		√		0.971 3
7	√	√	√	√			√	0.966 5
8				√	√		√	0.902 9
9	√	√	√		√		√	0.879 0

3.5 对比实验

在对比实验中,对比的模型分为两类:(1)跨场景的深度神经网络模型;(2)基于具体场景的背景减法模型。DeepBS[12]和 STAM[13]和提出的 IOFAM 采用相同的训练策略。对具体场景训练的模型,对比了基于深度神经网络的 FgSegNet[32]和基于背景减法的 GMM[33]、CPB[18]和 SubSENSE[34]。通过不同模型的实验结果说明方法的鲁棒性和有效性。在跨场景实验中,模型还对比了两个语义分割模型 PSPNet[14]和 DeepLabV3+[15]。

文中提到的模型都是在 CDNet 2014 数据集中训练的,表 2 中的实验结果对比突出说明所提模型的跨场景能力,以及使用单个模型的简洁性与有效性。表 2 显示 IOFAM 在 Recall、Precision 和 F-measure(F1)综合指标都达到了 SOTA 的结果。对需要在具体场景单独训练的模型 FgSegNet、GMM、CPB 和 SubSENSE,只有一个模型的 IOFAM 在综合指标 F1 的对比中仍然是最优的。IOFAM 对比单个模型训练的 STAM[13],在 F1 指标中提高了 1.25 个百分点。对比去掉注意力结构的 $\text{IOFAM}_{\text{noAtt}}$ 和去掉光流特征的 $\text{IOFAM}_{\text{noOp}}$,并结合表 1 中的消融实验说明注意力和光流在模型训练的重要性。

为了验证模型的跨场景能力,在 LIMU 数据集的三个典型场景中进行了测试,结果如表 3 所示。为了更好地说明模型的跨场景能力,在对比实验中加入了两个语义分割模型 PSPNet[14]和 DeepLabV3+[15]。通

过综合指标 F-measure(F1),在 CP 的子场景中,PSPNet 作为语义分割的结果更好,F1 为 0.865 6,但在另外两个子场景中的结果较差,实验也说明视频前景分割任务和语义分割任务的不同。在 LIMU 数据集的跨场景实验中,IOFAM 在整体的 F1 综合指标达到 SOTA 为 0.798 1。

表 2　在 CDNet 2014 数据集上的实验结果　%

算法	Recall	Precision	F1
IOFAM	0.966 1	0.989 3	0.977 6
$IOFAM_{noAtt}$	0.836 8	0.926 7	0.879 5
$IOFAM_{noOp}$	0.928 9	0.878 1	0.902 5
DeepBS	0.754 4	0.833 6	0.754 4
STAM	0.945 7	0.985 2	0.964 7
FgSegNet	0.607 3	0.623 1	0.609 2
GMM	0.684 5	0.602 3	0.570 6
CPB	0.704 7	0.622 1	0.632 4
SubSENSE	0.812 3	0.751 4	0.740 7

表 3　在 LIMU 数据集上的 F1 指标实验结果　%

算法	CP	ITS	LS	Overall
IOFAM	0.797 9	0.785 1	0.849 3	0.798 1
$IOFAM_{noAtt}$	0.699 7	0.735 4	0.796 5	0.729 1
$IOFAM_{noOp}$	0.705 3	0.729 3	0.698 3	0.713 0
DeepBS	0.670 2	0.524 5	0.633 1	0.607 4
STAM	0.774 3	0.673 9	0.716 2	0.734 3
FgSegNet	0.637 1	0.642 3	0.674 9	0.652 0
GMM	0.637 9	0.641 3	0.674 3	0.651 4
CPB	0.654 3	0.676 8	0.663 2	0.665 4
SubSENSE	0.674 4	0.652 0	0.693 1	0.675 7
PSPNet	0.865 6	0.131 3	0.651 5	0.750 2
DeepLabV3+	0.774 0	0.675 6	0.333 6	0.698 3

4　结束语

针对前景分割中的跨场景问题提出了一种间隔光流注意模型(IOFAM),以实现具有实际应用价值的跨场景前景分割任务。与最先进的跨场景深度模型、特定场景深度模型、背景减法模型在未训练数据集 LIMU 的实验结果对比,表明在无需任何额外训练的情况下具有良好的场景泛化能力。虽然采用双输入,但该框架实现了单一模型和端到端的训练,不需要场景适应等额外的微调操作。未来的工作将是使用自监督学习来探索特定训练场景的注意力模型。

参考文献:

[1] ITTI L. Automatic foveation for video compression using a neurobiological model of visual attention[J]. IEEE Transactions on Image Processing,2004,13(10):1304-1318.

[2] HU W,LI X,ZHANG X,et al. Incremental tensor subspace learning and its applications to foreground segmentation and tracking[J]. International Journal of Computer Vision,2011,91(3):303-327.

[3] CHEN X,FU C,ZHAO Y,et al. Salience-guided cascaded suppression network for person re-identification[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). Seattle:IEEE,2020:3300-3310.

[4] WREN C R, AZARBAYEJANI A, DARRELL T, et al. Pfinder:real-time tracking of the human body[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1997,19(7):780-785.

[5] ELGAMMAL A,DURAISWAMI R,HARWOOD D,et al. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance[J]. Proceedings of the IEEE,2002,90(7):1151-1163.

[6] BARNICH O,VAN DROOGENBROECK M. ViBe:a universal background subtraction algorithm for video sequences[J]. IEEE Transactions on Image Processing,2010,20(6):1709-1724.

[7] BRAHAM M, VAN DROOGENBROECK M. Deep background subtraction with scene-specific convolutional neural networks[C]//2016 international conference on systems, signals and image processing (IWSSIP). Bratislava:IEEE,2016:1-4.

[8] WANG Y,LUO Z,JODOIN P M. Interactive deep learning method for segmenting moving objects[J]. Pattern Recognition Letters,2017,96:66-75.

[9] WANG Y,ZHU L,YU Z. Foreground detection for infrared videos with multiscale 3-d fully convolutional network[J]. IEEE Geoscience and Remote Sensing Letters,2018,16(5):712-716.

[10] PATIL P W,MURALA S. Msfgnet:a novel compact end-to-end deep network for moving object detection[J]. IEEE Transactions on Intelligent Transportation Systems,2018,20(11):4066-4077.

[11] GARCIA-GONZALEZ J,ORTIZ-DE-LAZCANO-LOBATO J M,LUQUE-BAENA R M,et al. Foreground detection by probabilistic modeling of the features discovered by stacked denoising autoencoders in noisy video sequences[J]. Pattern Recognition Letters,2019,125:481-487.

[12] BABAEE M,DINH D T,RIGOLL G. A deep convolutional neural network for video sequence background subtraction[J]. Pattern Recognition,2018,76:635-649.

[13] LIANG D,PAN J,SUN H,et al. Spatio-temporal attention model for foreground detection in cross-scene surveillance videos[J]. Sensors,2019,19(23):5142.

[14] ZHAO H,SHI J,QI X,et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vi-

sion and pattern recognition (CVPR). Honolulu: IEEE, 2017:2881-2890.
[15] 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). Munich: Springer, 2018:801-818.
[16] DING H, JIANG X, LIU A Q, et al. Boundary-aware feature propagation for scene segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision (ICCV). Seoul: IEEE, 2019:6819-6829.
[17] DING H, JIANG X, SHUAI B, et al. Context contrasted feature and gated multi-scale aggregation for scene segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). Salt Lake City: IEEE, 2018:2393-2402.
[18] WREN C R, AZARBAYEJANI A, DARRELL T, et al. Pfinder: real-time tracking of the human body[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7):780-785.
[19] ZHOU W, KANEKO S, HASHIMOTO M, et al. Foreground detection based on co-occurrence background model with hypothesis on degradation modification in dynamic scenes[J]. Signal Processing, 2019, 160:66-79.
[20] LIANG D, HASHIMOTO M, IWATA K, et al. Co-occurrence probability-based pixel pairs background model for robust object detection in dynamic scenes[J]. Pattern Recognition, 2015, 48(4):1374-1390.
[21] LIAO S, ZHAO G, KELLOKUMPU V, et al. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes[C]//2010 IEEE computer society conference on computer vision and pattern recognition (CVPR). San Francisco: IEEE, 2010:1301-1306.
[22] 张巧荣,徐国愚,张俊峰. 利用视觉显著性的前景目标分割[J]. 兰州大学学报:自然科学版,2019,55(6):833-840.
[23] 薛 萍. 基于超像素特征表示的图像前景背景分割算法[J]. 西安科技大学学报,2017,37(5):731-735.
[24] 翟 玲,朱 敏,戴李君. 基于超像素与特征改进的 Grabcut 前景分割[J]. 微型电脑应用,2015,31(11):48-50.
[25] 李 磊,孙佳伟. 神经网络与边缘检测相结合的人体前景分割算法[J]. 计算机与数字工程,2020,48(4):940-945.
[26] 李 熹,马惠敏,马洪兵,等. 融合语义—表观特征的无监督前景分割[J]. 中国图象图形学报,2021,26(10):2503-2513.
[27] 陈亚当,郝川艳. 动态双边网格实现的视频前景分割算法[J]. 计算机辅助设计与图形学学报,2018,30(11):2101-2107.
[28] 牛 杰,卜雄洙,钱 堃. 基于前景分割的目标实时检测方法[J]. 计算机应用,2014,34(5):1463-1466.
[29] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision (ICCV). Venice: IEEE, 2017:2980-2988.
[30] GOYETTE N, JODOIN P M, PORIKLI F, et al. Changedetection. net: a new change detection benchmark dataset[C]//2012 IEEE computer society conference on computer vision and pattern recognition workshops (CVPRW). Providence: IEEE, 2012:1-8.
[31] University Kynshu, LIMU[DB/OL]. 2008. https://limu.ait.kyushu-u.ac.jp/dataset/en/.
[32] LIM L A, KELES H Y. Foreground segmentation using convolutional neural networks for multiscale feature encoding[J]. Pattern Recognition Letters, 2018, 112:256-262.
[33] STAUFFER C, GRIMSON W E L. Adaptive background mixture models for real-time tracking[C]//Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (CVPR) (Cat. No PR00149). Ft. Collins: IEEE, 1999:246-252.
[34] ST-CHARLES P L, BILODEAU G A, BERGEVIN R. Subsense: a universal change detection method with local adaptive sensitivity[J]. IEEE Transactions on Image Processing, 2014, 24(1):359-373.