第53卷第5期 南京航空航天大学学报 Vol. 53 No. 5
2021年10月 Journal of Nanjing University of Aeronautics & Astronautics Oct. 2021

DOI:10.16356/j.1005-2615.2021.05.011

基于全卷积孪生神经网络的复杂监控场景下前景提取方法

刘 峰[1,3],居 昊[1,2],干宗良[1,2]

(1.江苏省图像处理与图像通信重点实验室,南京 210003;2.南京邮电大学通信与信息工程学院,南京 210003;
3.南京邮电大学教育科学与技术学院,南京 210003)

摘要: 由于光照变化、相机抖动和动态背景等因素影响,现有基于传统图像处理方法的前景提取算法并不能在复杂场景下获得良好的分割效果。针对此类问题,本文提出了一种基于全卷积孪生神经网络的前景提取算法,仅需任意2帧图像即可准确提取运动前景。将输入的2帧图像分为背景图像与待提取图像,将其输入全卷积孪生神经网络得到二者的相似性度量图,该相似性度量图中包含待提取图像相对于背景图像的各像素变化情况信息;接着将相似性度量图与待提取图像融合,利用编解码网络以实现端到端的前景提取。在CDnet2014数据集上进行综合评估与测试,结果均证明了该方法的有效性。

关键词: 前景提取;度量学习;孪生神经网络;复杂监控场景

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1005-2615(2021)05-0743-08

Fully-Convolutional Siamese Networks for Foreground Subtraction in Complex Surveillance Videos

LIU Feng[1,3], *JU Hao*[1,2], *GAN Zongliang*[1,2]

(1. Jiangsu Key Laboratory of Image Processing and Image Communication, Nanjing 210003, China; 2. College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; 3. College of Education Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Due to factors such as illumination changes, camera jitter and dynamic background, existing foreground subtraction algorithms cannot achieve good segmentation results in complex scenes. To solve this kind of problems, this paper proposes a subtraction algorithm based on a fully convolutional siamese neural network, which can accurately segment the foreground with only two arbitrary frames. Specifically, the input two images are divided into the base image and the image to be segmented. The algorithm uses the fully-convolutional siamese network to get the similarity metric map of input frames. The similarity metric map contains information about changes in pixels of the image to be segmented relative to the base image. Then, the similarity metric map is fused with the image to be segmented, and the encoder-decoder network is used to achieve end-to-end foreground subtraction results. The paper evaluates the proposed algorithm on the CDnet2014 dataset to prove its effectiveness.

Key words: foreground subtraction; metric learning; siamese network; complex surveillance videos

将由静态或动态相机所捕捉到的图像或视频进行前背景分割是智能交通与视频监控领域一项

收稿日期: 2020-09-07;**修订日期:** 2020-11-07

通信作者: 刘峰,男,教授,E-mail:584257674@qq.com。

引用格式: 刘峰,居昊,干宗良. 基于全卷积孪生神经网络的复杂监控场景下前景提取方法[J]. 南京航空航天大学学报,2021,53(5):743-750. LIU Feng, JU Hao, GAN Zongliang. Fully-convolutional siamese networks for foreground subtraction in complex surveillance videos[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2021, 53(5):743-750.

重要的研究内容。前背景分割算法也常常作为预处理步骤在自动驾驶、机器人自主定位与导航、异常检测和识别中得到应用。解决这一计算机视觉任务的基本算法为背景减法，即将图像或视频序列中的当前帧与一个不断进行更新的背景模型相减，从而将运动对象与相对静止的背景场景进行分离。由于监控摄像头分布广泛，前背景分割需要对各种环境因素都具有良好的鲁棒性。传统的背景减法仅在特定类型的简单场景中表现良好，而对于场景照明变化、动态背景和相机运动等情况，传统的背景减法并不能得到准确的前背景分离效果。

过去几年中已有多种算法对该问题进行了广泛的研究，算法大约可以分为如下几类：(1)基于统计学的方法；(2)基于特征提取的方法；(3)基于神经网络的方法。基于统计学的方法以高斯混合模型(Gaussian mixed model, GMM)[1]为代表。该方法假设像素值在时间维度上服从混合高斯分布，以此为依据建立与更新背景模型，可以解决部分动态背景问题(如树叶与水面的晃动)。另外，基于统计学的非参数模型，如ViBe算法[2]与PBAS算法[3]处理了前背景相似与照明变化的问题。基于特征提取的方法利用变换域中所提取到的特征进行背景建模，通过背景模型提取运动前景。更进一步地，LBSP算法[4]、GOCM算法[5]等利用纹理特征以应对照明变化的情况。基于神经网络的方法利用神经网络将输入像素分类为背景或前景。Babaee等[6]采用固定的背景模型，利用卷积神经网络对图像像素进行前背景分类。而后神经网络方法多基于视频序列，将多帧图像输入神经网络直接生成前景掩模图。由于密集掩码预测任务所需的精确性，近年来几乎所有前背景分割算法都基于神经网络，尤其是全卷积神经网络。但该类算法依然存在一些问题：(1)对于人类而言，仅需2帧图像即可确定前景与背景。而现有算法多利用长短期记忆网络(Long short-term memory, LSTM)进行时序特征融合[7]，即网络输入端需要多帧图像，易出现信息的冗余。(2)部分基于单帧图像的方法[8]直接利用神经网络学习当前帧而得到决策边界，这种方法既不直观又极度依赖训练集，模型迁移能力差。

本文提出算法利用两帧图像进行端到端地前背景分割。算法分为两步骤，将输入的2帧图像分为背景图像与待提取图像。第1步利用全卷积孪生神经网络生成2帧图像的相似性度量图。该部分工作基于文献[9-10]所提出的思想，与之不同的是，算法对2帧图像中前景重叠的情况进行进一步处理，该相似性度量图为后续网络输出提供基础，有助于后续网络更加关注输入图像对中发生变化的部分。以文献[10]为代表的算法需已知一张背景图像，这在实际情况中往往是不现实的。当输入的2帧图像的前景位置出现重叠时，该类型算法并不能得到准确的前背景分割结果。第2步将第1步所得到的相似性度量图与待提取图像进行融合，将融合结果输入前景提取网络。前景提取网络为包含编码器与解码器的全卷积神经网络，解码器部分采用转置卷积使网络输出与输入尺寸相匹配。前景提取网络整体采用U-net型结构[11]以提高网络性能。

1 本文算法

本文具体算法如图1所示。采用全卷积孪生神经网络对背景图像与待提取图像$(\text{image}_1,\text{image}_2)\in\mathbf{R}^{C\times H\times W}$进行特征提取，得到特征图$(\text{feature}_1,\text{feature}_2)\in\mathbf{R}^{c\times h\times w}$。每个特征图包含$h\times w$对特征向量$(f_1,f_2)\in\mathbf{R}^{c\times 1}$，计算每对特征向量之间的欧几里得距离，得到二者之间的相似性度量图。本文在CDnet2014数据集中随机选取背景图像与待提取图像[12]，并将二者所对应的ground truth进行叠加，合成新的ground truth，并将像素类型由前景，背景两类重新分为发生变化，未发生变化与前景重叠3类，丰富相似性度量图的语义特征，具体细节将在后续章节中讨论。该相似性度量图$\text{map}\in\mathbf{R}^{1\times h\times w}$中包含了图像对中发生变化像素的位置信息与边缘信息，可使后续前景提取网络更加关注发生变化的部分。因此将map进行上采

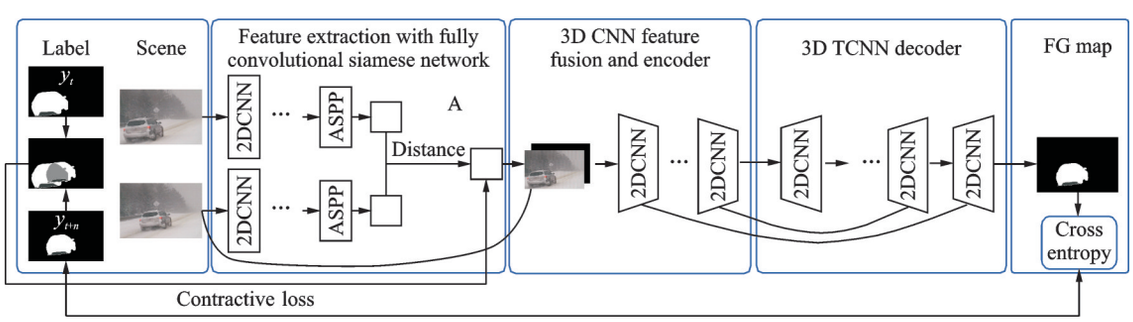


图1 本文算法框架流程图

Fig.1 Framework of the algorithm

样并与image_2进行融合，得到网络后续的输入图像块$\text{image}\in\mathbf{R}^{C\times H\times W}$。对于融合后的图像image，利用编解码网络进行前景提取。解码器采用转置卷积进行特征图尺寸扩张。为了提高前景提取准确度，在每一次转置卷积之后，将解码器中的浅层特征图与编码器中相对应的深层特征图在维度上拼接。

1.1　全卷积孪生神经网络

度量学习作为一种特征对相似性度量的方法，其目的是通过训练和学习，减小或限制同类样本特征向量之间的距离，同时增大非同类样本特征向量之间的距离。本文认为，前景提取问题是待提取图像与背景图像的像素相似度比较问题，同样使用度量学习的思想，使输入图像对$(\text{image}_1,\text{image}_2)\in\mathbf{R}^{C\times H\times W}$，待提取图像中背景像素位置处的特征向量之间距离减小，前景像素之间的距离增大。网络A部分即采用全卷积的孪生神经网络检测两帧图像像素之间的相似度，采用Deeplabv2[13]网络的特征提取部分作为骨干网络。考虑到相比于语义分割而言，前景提取中像素种类较少且多集中分布，因此在获得顶层特征$\text{feature}_{\text{top}}\in\mathbf{R}^{\frac{w}{4},\frac{h}{4},512}$之后，采用了类空间金字塔池化(Spatial pyramid pooling，SPP)的方法对特征进行多尺度信息提取，如图2所示，而并没有采用原Deeplabv2结构中的空洞空间金字塔池化(Atrous spatial Pyramid pooling，ASPP)。

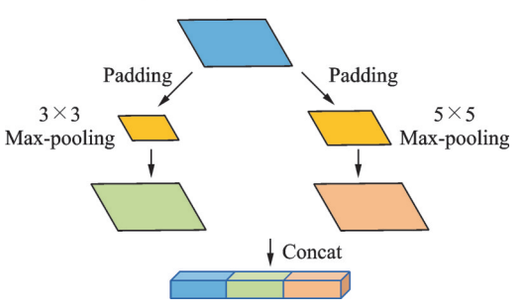


图2　空间金字塔池化结构图

Fig.2　ASPP framework

与文献[14]中的SPP方法不同，在得到顶层特征之后，将特征图进行填充以保证池化后特征图大小不变。为了应对输入图像对尺寸不一致的问题以及获得多尺度信息，采用3×3和5×5的不同池化尺寸，通过控制步长产生固定大小的输出特征图，再将特征图进行拼接后采用1×1的卷积核固定输出特征维度。由于池化与1×1的卷积所需要的总参数量仅为顶层特征图的维度$c=512$，因此所提出的结构相较于ASPP参数量有着明显的减少。算法将输入像素对分为发生变化，未发生变化和前景重叠3种类型。通过该方式可以使网络输出的相似性度量图中包含更多前景重叠部分的边缘信息与正确的前景位置信息，更易于后续网络进行前景提取。

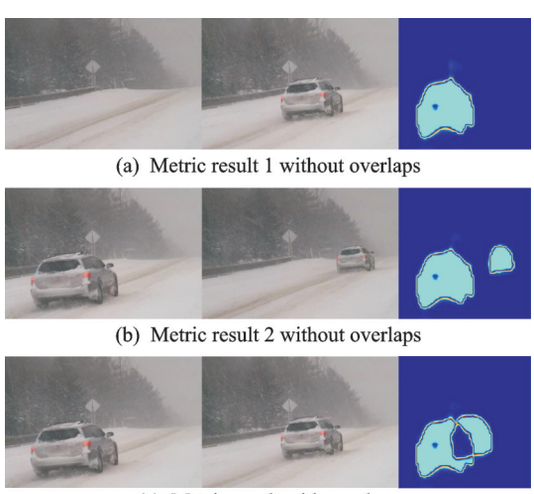
1.2　对比损失函数

全卷积孪生神经网络通常采用对比损失以使同类样本之间距离缩小，不用样本之间距离增大。对比损失公式为

$$L(W,(Y,X_1,X_2))=\frac{1}{2N}\sum_{n=1}^{N}YD_W+(1-Y)\max(m-D_W,0)^2 \tag{1}$$

式中：$D_W(X_1,X_2)$代表两个样本特征X_1和X_2的距离；Y代表样本标签；$Y=0$表示像素点为前景点，$Y=1$表示像素点为背景点；m为设定阈值；N为样本个数。当像素点为前景点时，损失函数$L_f=\max(m-D_W,0)^2$。在训练迭代过程中，输入图像对中代表前景点像素对之间的距离会趋于m。反之，当像素点为背景点时，损失函数$L_b=D_W{}^2$。即输入图像对中代表背景点像素对之间的距离会趋于0。从文献[12]中的工作得知，使用欧几里得距离相较于余弦相似度可以显著提升网络性能，所以令$D_W(X_1,X_2)=\|X_1,X_2\|_2$。

仅采用正负样本对时并不能处理前景重合时的情况，如图3所示。图3(a)中输入图像对中包含一幅背景图，经全卷积孪生神经网络后能够输出较准确的前景图。图3(b)中输入图像对均包含前景物体，两者并不重叠，经全卷积孪生神经网络后所输出的前景掩模中包含了前后两幅图中的前景物体，这显然是错误的。而在图3(c)中输入图像对中的前景物体产生重叠，输出的前景掩模图为输入图像对中前景像素点的并集，显然也并不能输出正确的前景提取结果。采用正负样本对无法表示两幅图像中像素的全部关系。针对于此，将样本像素类型分为3类。$Y=0$表示前景；$Y=1$表示前景



(a) Metric result 1 without overlaps

(b) Metric result 2 without overlaps

(c) Metric result with overlaps

图3　不同输入图像对下相似性度量图

Fig.3　Metric results of different image pairs

重叠；$Y=2$表示背景，对式(1)进行微调为

$$L(W,(Y,X_1,X_2))=\frac{1}{2N}\sum_{n=1}^{N}\frac{Y}{2}D_W+\left(1-\frac{Y}{2}\right)\max(m-D_W,0)^2+C(Y) \tag{2}$$

$$C(Y)=\frac{1}{\tau_{\text{const}}-Y}\qquad \tau_{\text{const}}\in(1,2) \tag{3}$$

显然，在训练迭代的过程中会使前景重叠像素对之间的距离与前背景进行区分。该相似性度量图效果类似注意力图，通过每个像素点的不同像素值帮助后续网络训练。$C(Y)$为以像素类型Y为自变量的正则函数，其功能为：在CDnet2014中，存在相机视角旋转的场景。在该场景下，输入图像对中同一像素位置所对应的真实场景位置并不相同，并不能将该类像素对的特征距离设置为0。本文参考文献[15]中的思想，在原对比损失函数后添加较小的正则项$C(Y)$，该正则项可以加快收敛速度并使网络表现出更好的性能。

1.3 特征融合

对得到的相似性度量图$\text{map}\in\mathbf{R}^{1\times h\times w}$中每个相似性值取整并对map进行上采样得到与原图像对同样尺寸的相似性度量图$\text{map}'\in\mathbf{R}^{1\times H\times W}$。将map′进行伪彩色处理进行可视化如图4所示。显然，能够从该相似性度量图中得到前景像素的位置与前景重叠像素的位置。此外，还可以清晰地得出其区分处的边缘信息。

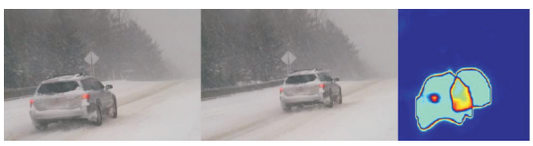


图4 前景重叠时的相似性度量图

Fig.4 Metric results with foreground overlaps

该相似性度量图map′与待提取图像image_2即可为前景提取提供充足的先验信息，与常见的RGB-D图像语义分割中对D通道的处理方法不同，由于map′本身的语义信息与待提取图像较为相似，因此图像融合前无需利用神经网络对相似性度量图进行二次特征提取。基于此想法，将map′与image_2进行简单的相加并进行归一化，得到编解码网络的输入$\text{image}\in\mathbf{R}^{3\times H\times W}$。

1.4 编解码网络

对$\text{image}\in\mathbf{R}^{3\times H\times W}$采用编解码网络结构进行端到端地前景提取，网络具体结构如图5所示。每层中包含参数$(k,s)[b,H,W,D]$分别表示卷积核大小、卷积步长、输入图像高度和宽度及输出特征图维度。

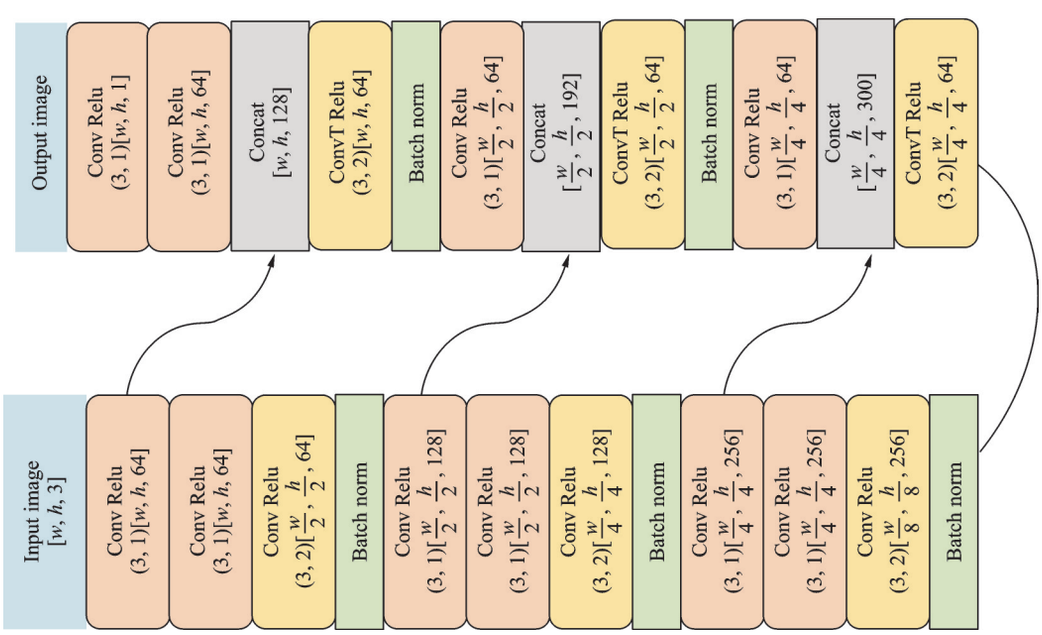


图5 编解码网络结构图

Fig.5 Encoder-decoder network

解码器对编码器所提取的特征图进行上采样以与原始输入图像尺寸相匹配。与文献[16]中采用的空洞卷积和文献[17]中采用的反池化不同，采用转置卷积使特征图大小加倍。训练时采用如式(4)所示的二进制交叉熵损失函数，即有

$$E=-\frac{1}{N}\sum_{n=1}^{N}\left[\,p_n\log\widehat{p_n}+(1-p_n)\log(1-\widehat{p_n})\,\right] \tag{4}$$

式中：$\widehat{p_n}=\sigma(x_n)\in[0,1]$为输出特征图经过Softmax层进行分数映射后所计算出的像素作为前景

的概率；p_n为前背景标签值。

2　实验结果与分析

2.1　数据准备与参数设置

本文算法将全卷积孪生神经网络与编解码网络分开训练，即先得到准确的相似性度量图之后再进行前景提取。对于全卷积孪生神经网络，采用预训练的Deeplabv2模型对网络的前10个卷积层进行权重初始化。选择随机梯度下降（Stochastic gradient descent，SGD）算法，设置 weight decay＝5×10^{-4}，learning rate＝1×10^{-5}，batch size＝4。训练时选取视频中30%的图像帧作为训练集，10%作为验证集。测试时，从视频中剩余的60%图像帧中选取1张包含较多前景像素的图像作为背景图像，其余图像作为待提取图像进行测试。

2.2　评价指标

本文采用RE，FPR，FNR，PWC，F_{measure}和Precision作为客观评价指标，定义如下

$$\text{RE}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{5}$$

$$\text{FPR}=\frac{\text{FN}}{\text{FP}+\text{TN}} \tag{6}$$

$$\text{FNR}=\frac{\text{FN}}{\text{TP}+\text{FN}} \tag{7}$$

$$\text{PWC}=\frac{100\times(\text{FN}+\text{FP})}{\text{TP}+\text{FN}+\text{FP}+\text{TN}} \tag{8}$$

$$F_{\text{measure}}=\frac{2\times\text{Precision}\times\text{Re}}{\text{Precision}+\text{Re}} \tag{9}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{10}$$

式中：TP表示预测结果为正确分类的前景像素个数；FP表示错误分类的前景像素个数；TN表示正确分类的背景像素个数；TP表示错误分类的背景像素个数。

2.3　实验结果

2.3.1　客观指标分析

本文使用了CDnet2014数据集所提供的评价工具对所提出算法进行评估。表1给出了本文算法在9个不同场景下的6项评价指标。一般而言，前景提取算法需要在取得较高召回率的同时尽可能不损失算法精度，所以F_{measure}可以准确评判算法优劣程度。可以看出，本文算法在所有场景中都取得了较高的F_{measure}值，尤其在全方位相机（Pan/Tilt/Zoom，PTZ）的场景下，由于对比损失函数中引入的较小正则项$C(Y)$，相机焦距变换仅对前景提取准确度有着微小的影响。但由于相机视角变换导致相似性度量图产生偏差，所以该情况下性能指标相对较低。

表1　不同场景下本文算法评价指标

Table 1　Performance evaluation in different scenes

场景	RE	FPR	FNR	PWC	PRE	F_{measure}
Baseline	0.997 51	0.000 29	0.002 49	0.041 70	0.996 47	0.996 48
Dynamic background	0.992 71	0.000 32	0.007 29	0.056 48	0.991 36	0.992 03
Shadow	0.990 13	0.000 17	0.009 86	0.064 05	0.996 72	0.993 41
Bad weather	0.987 94	0.000 15	0.012 06	0.028 68	0.987 46	0.987 70
Low frame rate	0.986 79	0.000 16	0.013 21	0.004 20	0.992 00	0.989 39
Night videos	0.963 30	0.000 53	0.036 70	0.126 60	0.974 10	0.968 66
PTZ	0.979 16	0.000 17	0.020 84	0.041 10	0.985 37	0.982 26
Camera jitter	0.986 80	0.000 16	0.013 20	0.041 99	0.992 00	0.989 39
Intermittent	0.993 96	0.000 11	0.006 03	0.024 13	0.994 79	0.994 38
Overall	0.979 93	0.000 16	0.020 06	0.042 52	0.986 85	0.983 38

本文将实验结果与其他7种方法进行了比较，分别为Cascade CNN[18]、DeepBs[6]、FgSegNet_S[8]、FgSegNet[8]、IUTIS-5[19]、BSUV-Net[20]和SuBSENSE[21]。其中前3种方法利用了基于监督学习的神经网络模型，后2种方法采用了传统方法。

表2给出了每个模型在9个不同场景中各性能指标的平均值。表中红色表示评价指标排名第一，蓝色表示排名第二。可以看出所提出算法在各指标中都有明显的提高。为了直观比较这些算法，本文选取了Baseline、Dynamic background、Bad weather、PTZ和Intermittent motion中的典型场景测试算法性能。分割结果如图6所示，第1行为CD2014中部分场景图，第2行为Groundtruth，第3行为本文算法所得结果，后续依次为FgSegNet、Cascade-CNN、DeepBS、IUTIS-3和SuBSENSE所得结果。可以看出，本文算法所产生的分割结果更加接近Groundtruth。例如在第4列的PTZ场景

表 2 不同算法评价指标对比

Table 2 Performance evaluation of different methods

算法	RE	FPR	FNR	PWC	PRE	$F_{measure}$
本文算法	0.979 9	0.000 16	0.020 1	0.042 5	0.986 8	0.983 4
Cascade CNN[18]	0.950 6	0.003 2	0.049 4	0.405 2	0.899 7	0.899 7
DeepBs[6]	0.754 5	0.009 5	0.245 5	1.992 0	0.833 2	0.833 2
FgSegNet_S[8]	0.989 6	0.000 3	0.010 4	0.046 1	0.975 1	0.980 4
FgSegNet[8]	0.983 6	0.000 2	0.016 4	0.055 9	0.975 8	0.977 0
BSUV-Net[20]	0.820 3	0.005 4	0.179 7	1.140 2	0.811 3	0.786 8
SuBSENSE[21]	0.812 4	0.009 6	0.187 6	1.678 0	0.750 9	0.750 9
IUTIS-5[19]	0.784 9	0.005 2	0.215 1	1.198 6	0.771 7	0.808 7

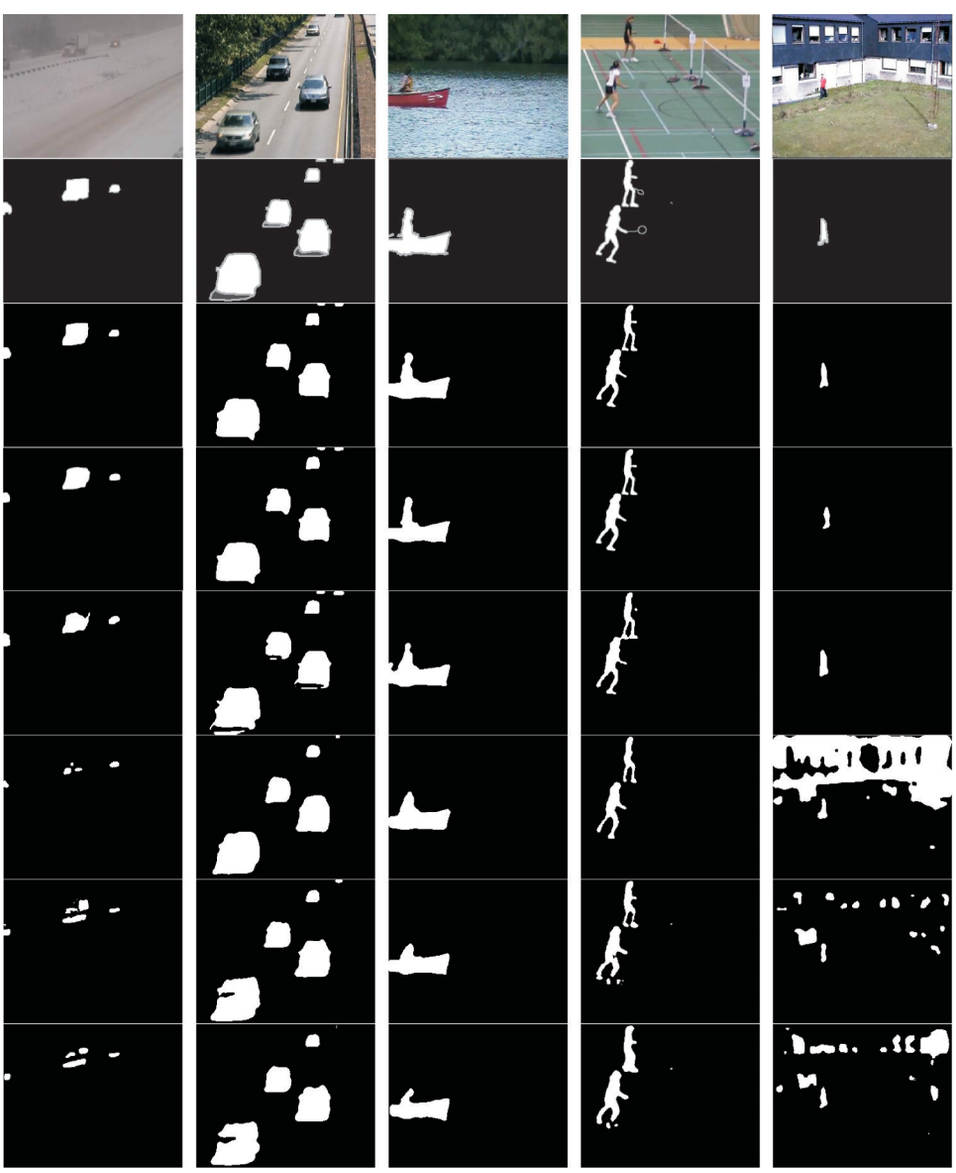


图 6 不同复杂场景下算法前景提取结果

Fig.6 Foreground subtraction results of different methods

中，人形掩模具有更丰富的细节，这主要得益于全卷积孪生神经网络对于变化像素的粗定位与后续编解码网络细化分割。

2.3.2 迁移能力分析

在考虑算法输出结果准确性的同时本文兼顾了模型的迁移能力。将本文算法与 FgSegNet、BSUV-Net 算法在 CDNet2014 数据集中 Baseline 场景下的训练结果应用于 SBI2015[22] 数据集中的 Highway 场景，所得结果的评价指标如表 3 所示。一些典型分割结果如图 7 所示。可以看出，在仅利

表 3　SBI2015数据集中不同算法指标比较

Table 3　Performance evaluation of different methods in SBI2015

算法	RE	PRE	$F_{measure}$
本文算法	0.749 4	0.870 5	0.805 4
FgSegNet	0.928 3	0.355 6	0.514 2
BSUV-Net	0.874 6	0.517 4	0.650 1

用CDNet2014数据集上的训练模型时，其余两种算法所检测出的前景像素点较多，所以在RE与FNR两项指标上较好，但在检测出的前景点中多数为误检。相比较而言，本文算法虽然RE与FNR较低，但检测较为准确，作为模型优劣评估标准的$F_{measure}$值最高，由此可见本文算法有良好的迁移能力。

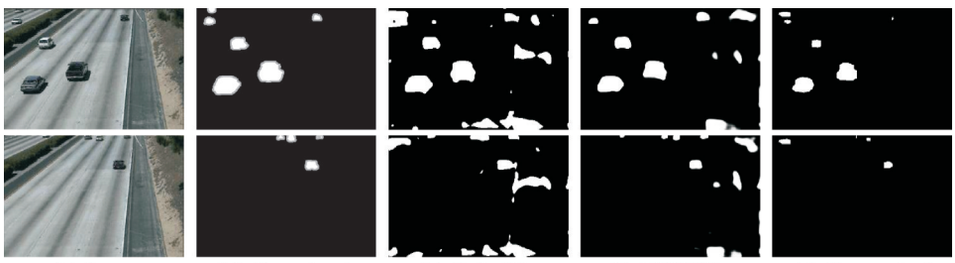


图 7　SBI2015数据集中不同算法分割结果

Fig.7　Segmentation results of different methods in SBI2015

3　结　　论

本文将全卷积孪生神经网络与编解码网络相结合进行端到端的前景提取，采用全卷积孪生神经网络产生图像对之间的相似性度量图使前景提取网络更加关注前景像素，以获得更加准确的前景提取结果。实验证明，本文算法在保证前景提取结果准确性的同时兼顾了迁移能力，利用数据集上的训练模型即可较为准确地对其余数据集中的相似场景进行前景提取。由前文可知，相似性度量图本身已包含前景的位置与边缘信息，理论上仅需将其与待提取图像结合，通过简单的图像处理方法即可获得前景提取结果。但本文算法中数据集的评价指标采用较为复杂的编解码网络进行前景提取，并且训练时所需训练集的占比较大，难以进行实际应用，这也是未来算法的改进方向。

参考文献：

[1] ZIVKOVIC Z. Improved adaptive gaussian mixture model for background subtraction[C]//Proceedings of the 17th International Conference on Pattern Recognition. [S.l.]: IEEE, 2004.

[2] BARNICH O, DROOGENBROECK M V. ViBe: A universal background subtraction algorithm for video sequences[J]. IEEE Transactions on Image Processing, 2011, 20(6):1709-1724.

[3] KRYJAK T, GORGON M. Real-time implementation of the ViBe foreground object segmentation algorithm[C]//Proceedings of Computer Science & Information Systems. [S.l.]: IEEE, 2013.

[4] ST-CHARLES P L, BILODEAU G A, BERGEVIN R. Flexible background subtraction with self-balanced local sensitivity[C]//Proceedings of Computer Vision & Pattern Recognition Workshops. [S.l.]: IEEE, 2014.

[5] MANFREDI M, VEZZANI R, CALDERARA S, et al. Detection of static groups and crowds gathered in open spaces by texture classification[J]. Pattern Recognition Letters, 2014, 44(C):39-48.

[6] BABAEE M, DINH D T, RIGOLL G. A deep convolutional neural network for background subtraction [J]. Pattern Recognition, 2017, 76(17):8-36.

[7] AKILAN T, WU Q J, SAFAEI A, et al. A 3D CNN-LSTM based image-to-image foreground Segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(3):959-971.

[8] LIM L A, KELES H Y. Foreground segmentation using convolutional neural networks for multiscale feature encoding [EB/OL]. (2018-10-22) [2020-09-07]. https://arxiv.org/abs/1810.09111.

[9] GUO E, FU X, ZHU J, et al. Learning to measure change: Fully convolutional siamese metric networks for scene change detection [EB/OL]. (2018-08-22) [2020-05-07]. Https://arxiv.org/abs/1810.09111.

[10] SAKURADA K, OKATANI T. Change detection from a street image pair using CNN features and superpixel segmentation [C]//Proceedings of British Machine Vision Conference. [S.l.]: BMVC, 2015.

[11] RONNEBERGER O, FISCHER P, BROX T. U-

Net: Convolutional networks for biomedical image segmentation[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. [S.l.]: Springer International Publishing, 2015.

[12] WANG Yi, KONRAD J, ISHWAR P, et al. CDnet 2014: An expanded change detection benchmark dataset[C]//Proceedings of Computer Vision & Pattern Recognition Workshops. [S.l.]: IEEE, 2014.

[13] 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, 2018, 40(4):834-848.

[14] HE K, ZHANG X, REN S, et al. Spatial Pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904-1916.

[15] BAILER C, VARANASI K, STRICKER D. CNN-Based patch matching for optical flow with thresholded hinge embedding loss[C]//Proceedings of Computer Vision & Pattern Recognition. [S.l.]: IEEE, 2017.

[16] HU Z, TURKI T, PHAN N, et al. A 3D atrous convolutional long short-term memory network for background subtraction[J]. IEEE Access, 2018, 6:43450-43459.

[17] LIM K, JANG W D, KIM C S. Background subtraction using encoder-decoder structured convolutional neural network[C]//Proceedings of 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). [S.l.]: IEEE, 2017.

[18] WANG Y, LUO Z, JODOIN PM. Interactive deep learning method for segmenting moving objects[J]. Pattern Recognition Letters, 2017, 9(96):66-75.

[19] BIANCO S, CIOCCA G, SCHETTINI R. Combination of video change detection algorithms by genetic programming[J]. IEEE Transactions on Evolutionary Computation, 2017, 21(6):914-928.

[20] TEZCAN M O, ISHWAR P, KONRAD J. BSUV-Net: A fully-convolutional neural network for background subtraction of unseen videos[C]//Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). [S.l.]: IEEE, 2020.

[21] ST-CHARLES P, BILODEAU G, BERGEVIN R. SuBSENSE: A universal change detection method with local adaptive sensitivity[J]. IEEE Transactions on Image Processing, 2014, 24(1):359-373.

[22] MADDALENA L, PETROSINO A. Towards Benchmarking scene background initialization[C]//Proceedings of International Conference on Image Analysis and Processing. [S.l.]: Springer International Publishing, 2015, 7:469-476.

(编辑:刘彦东)