DOI: 10.13382/j.jemi.B2407625

用于几何信息学习的图结构运动分割方法*

张纪友[1,2,3] 李 俊[1,2,3] 郭霏霏[4] 李琦铭[2,3]

(1. 福建农林大学机电工程学院 福州 350002;2. 中国科学院大学福建学院 泉州 362200;
3. 中国科学院海西研究院泉州装备制造研究中心 泉州 362200;4. 泉州职业技术大学 泉州 362000)

摘 要:针对现有运动分割方法在交通场景下实用性方面的不足,性能和验证时间难以平衡的问题,提出用于几何信息学习的图结构运动分割方法(GS-Net)。GS-Net 由点嵌入模块、局部上下文融合模块、全局双边正则化模块和分类模块组成。其中,点嵌入模块将原始关键特征点数据从低维线性难可分的空间映射到高维线性易可分的空间,有利于网络学习图像中运动对象之间的关系;局部上下文融合模块利用双分支图结构分别在特征空间和几何空间提取局部信息,随后将两种类型的信息融合得到更强大的局部特征表征;全局双边正则化模块则利用逐点和逐通道的全局感知来增强局部上下文融合模块得到的局部特征表征;分类模块将前面得到的增强局部特征表征映射回低维分类空间进行分割。GS-Net 在 KT3DMoSeg 数据集的误分类率均值和中值分别为 2.47%和 0.49%,较于 SubspaceNet 分别降低 8.15%和 7.95%;较于 SUBSET 分别降低 7.2%和 0.57%。同时,GS-Net 在网络推理速度相比 SubspaceNet 和 SUBSET 均提升两个数量级;GS-Net 在 FBMS 数据集召回率和 F-measure 分别为 82.53%和 81.93%,较于 SubspaceNet 分别提升 13.33%和 5.36%,较于 SUBSET 分别提升 9.66%和 3.71%。实验结果表明 GS-Net 能够快速、精确地分割出真实交通场景中的运动物体。

关键词: 运动分割;关键点提取;图结构;特征融合;深度学习;自动驾驶

中图分类号: TP183; TN911.73 **文献标识码**: A **国家标准学科分类代码**: 520.2040

Graph structure motion segmentation method for geometric information learning

Zhang Jiyou[1,2,3] Li Jun[1,2,3] Guo Feifei[4] Li Qiming[2,3]

(1. School of Mechanical and Electrical Engineering, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 2. University of Chinese Academy of Sciences, Fujian, Quanzhou 362200, China; 3. Quanzhou Institute of Equipment Manufacturing, Haixi Institute, CAS, Quanzhou 362200, China; 4. Quanzhou Vocational and Technical University, Quanzhou 362000, China)

Abstract: The graph-structured motion segmentation method (GS-Net) for geometric information learning is proposed to address the shortcomings of existing motion segmentation methods in terms of their practicality in traffic scenarios, and the difficulty in balancing performance and validation time. GS-Net consists of a point embedding module, a local context fusion module, a global bilateral regularization module, and a classification module. The point embedding module maps the original key feature point data from a low-dimensional linearly difficult-to-differentiate space to a high-dimensional linearly easy-to-differentiate space, which is conducive to the network learning the relationship between moving objects in the image; the local context fusion module utilizes the dual-branching graph structure to extract local information from both the feature space and the geometric space, and then fuses the two types of information to obtain a more powerful local feature representation, The global bilateral regularization module uses point-by-point and channel-by-channel global sensing to enhance the local feature representations obtained by the local context fusion module; the classification module maps the enhanced local feature representations back to the low-dimensional classification space for segmentation. GS-Net's mean and median misclassification rates on the KT3DMoSeg dataset are 2.47% and 0.49%, respectively, which are 8.15% and 7.95% lower than

收稿日期: 2024-06-26 Received Date: 2024-06-26

*基金项目:国家自然科学基金(62102394)、福建省科技计划(2023N3010)项目资助

those of SubspaceNet, and 7.2% and 0.57% lower than those of SUBSET. Meanwhile, GS-Net improves the network inference speed by two orders of magnitude compared to both SubspaceNet and SUBSET. GS-Net's recall and F-measure on the FBMS dataset are 82.53% and 81.93%, respectively, showing improvements of 13.33% and 5.36% compared to SubspaceNet, and 9.66% and 3.71% compared to SUBSET, respectively. The experimental results demonstrate that GS-Net can quickly and accurately segment moving objects in real traffic scenes.

Keywords: motion segmentation; key point extraction; graph structure; feature fusion; deep learning; autonomous driving

0 引　言

计算机视觉是近年来备受关注的研究领域之一。随着人工智能的飞速发展,它在图像处理、模式识别等多个交叉学科中发挥着重要作用。计算机视觉的一项重要任务是有效提取视频帧之间的上下文关系,并通过语义分割将所获取的信息进行表达,传递给计算机或机器人。运动分割在这一任务中起着关键作用,它旨在识别和分离视频序列中的不同运动模式,从而实现对个体对象运动的分析和理解。其在自动驾驶[1]、运动目标跟踪[2]、三维重建[3]等领域具有重要的实际应用价值。

运动分割方法大致分为两类:基于像素提取的方法[4-5]和基于关键特征点提取的方法[6-20]。基于像素提取的方法(视频对象分割)的主要目的是在保留运动目标边界的同时,准确提取运动目标,并在视频序列中建立物体像素之间的时间关联。相比之下,基于关键特征点提取的方法旨在将移动对象中的稀疏关键点划分为单个不重叠的聚类。后者使用稀疏关键点作为输入数据,网络在推理时间和计算量方面具有优势,更适合实时的运动分割。因此,本研究主要研究基于关键特征点提取的方法。

传统基于关键特征点提取的方法倾向于将运动分割视为模型拟合问题。例如,Raguram 等[6]提出的随机抽样一致的通用框架(universal framework for random sample consensus,USAC)运用"假设和测试"范式,持续采样假设,迭代拟合最佳模型并计算相应内点,直至找到所有模型实例。在内点比例较大的简单场景中,该类算法速度快、精度高;但在内点比例低的复杂场景下,其速度和精度会急剧下降。这一现象引起了大量研究者的关注,随之涌现出大量改进算法,主要包括改进采样器、降低对阈值的依赖、局部优化3种类型[7-9]。由于USAC是迭代式算法,若采样算法欠佳,会产生大量无效样本,进而导致大量无用计算。因此,文献[7]将局部与全局采样相结合,先进行局部均匀采样,再逐渐向全局化过渡。通过这种方式,该方法能更早发现局部结构,克服局部采样的不足,获得良好的拟合效果,同时加快了分割速度。文献[8]旨在降低算法对阈值的依赖,它基于边际化样本共识思想,引入新的模型质量(评分)函数,并采用迭代加权最小二乘法求解,进一步提高了算法的分割精度和速度。文献[9]结合了鲁棒模型拟合和能量最小化的方法,能够对空间相干点分布进行建模,并在局部优化中利用这一特性,取得了不错的分割效果。上述这些方法都遵循"拟合和删除"的流程,其本质是对单个运动实例的分割精度和速度进行优化,因此可能会因首个实例分割不准确而影响后续的分割效果。

除了上述方法,研究者们还设计了一些基于聚类[10-12]的传统分割算法,以Li等[10]提出的信息融合两层网络(information fusion on the two-layer network,IFTLN)为例,它通过两层网络结构融合模型假设和数据点信息,能够有效处理同类型多实例对象中存在的大量异常值和数据点分布不均衡的问题。尽管这类算法取得了一定成果,但仍存在无法区分各种模型实例(如单应性矩阵和基础矩阵)的问题。综上所述,这两类传统运动分割方法难以学习到不同运动对象之间的区别性几何信息,无法高效地实现多类多实例的同时分割。

随着人工智能的发展,深度学习在运动分割领域得到广泛应用。相较于传统运动分割算法,基于深度学习的方法具备显著优势,能够从数据中自动学习模式,无需人工设计特征模式,且分割性能出色。当前,多数基于深度学习的运动分割方法[13-16]是借助卷积神经网络学习各运动对象的几何信息,以Cavalli等[15]提出的神经过滤最小样本(neurally filtered minimal samples,NeFSAC)为例,该方法结合卷积神经网络为传统运动分割方法中的"拟合和删除"流程学习内点采样可能性,取得了一定的成果。然而,这些方法仍存在诸多局限,如无法实现同类型多个运动对象的同时拟合,神经网络仅能用于优化采样过程。对此,Wei等[17]提出可微采样器、可微求解器等组件,使算法能够学习整个随机鲁棒估计流程,为后续研究者使用神经网络进行端到端优化传统运动分割方法提供了可能,但仍未解决"拟合和删除"流程中的固有问题。Kluger等[18]利用神经网络预测样本和内点权重,并行处理多模型实例,有效解决了同类多个运动对象的计算瓶颈问题,但还无法实现多类型多实例的分割。针对上述问题,Xu等[19]提出子空间聚类网络(subspace clustering network,SubspaceNet),以聚类思想解决多类型多实例同时分割问题,但受卷积神经网络感受野限制,无法充分提取运动对象间的复杂几何信息。随着Vision Transformers技术的兴起,注意力机制模块在各类视觉任务中展现出媲美或优于卷积神经网络的性能,Li等[20]利

用 Transformer 学习运动对象的几何信息,提出(geometric information via transformer,GIET),但该方法训练耗时久、运算量大,难以满足实时运动分割需求。

在真实交通场景中,运动分割算法对速度和精度要求颇高,而现有方法难以平衡性能与验证时间,无法满足该需求。为此,本文聚焦于在点云分割和图像匹配领域表现优异的图神经网络。图神经网络[21-22]在各类基于二维和三维点云的视觉任务中展现出与现有卷积和注意力机制相媲美甚至更优越的性能。因此,本研究提出新型网络(graph structure network,GS-Net),利用图结构提取运动目标的复杂几何信息,以提高运动分割的速度和精度。GS-Net 包含点嵌入、局部上下文融合、全局双边正则化及分类模块。点嵌入模块将低维关键点映射至高维特征空间,用于强化网络对不同运动对象中相似关键点的区分能力,局部上下文融合模块融合特征与几何空间信息助力识别运动边界与局部变化,全局双边正则化模块过滤增强几何信息,分类模块预测得分用于分割。各模块参数基于实用性优化,在速度和精度间实现平衡,经公开数据集的实验证实了 GS-Net 的实用性。

鉴于现有运动分割方法在交通场景实用性方面存在不足,尤其是难以平衡性能与验证时间,本文提出一种用于几何信息学习的图结构运动分割网络框架。该框架通过有效运用图结构,增强了运动分割中几何信息的提取能力,进而提升了方法在工程应用中的适用性。局部上下文融合模块借助双分支图结构捕捉并融合几何信息,全局双边正则化模块运用全局感知细化局部特征表征。这两个模块致力于学习真实场景图像帧中运动对象内部点间的关系,以及它们与其他运动对象的交互,以此提高分割精度。在公开的真实交通场景运动分割数据集上进行验证,GS-Net 均获得了最优结果。GS-Net 不仅在准确性方面显著超越其他现有方法,而且在推理速度上也有大幅提高,充分证实了其在实际工程应用中的优越性与高效性。

1　GS-Net 分割算法

1.1　网络结构

给定一个包含 M 个真实场景图像帧的序列,M 的值根据具体的运动分割任务而有所不同。本研究的主要目标是通过去除噪声点并将属于相同运动的点进行分组,来执行基于关键特征点提取的运动分割。本文方法 GS-Net 网络结构如图 1 所示,输入数据由 N 个关键点组成,这些关键点可代表如车辆、行人等待分割的运动对象。每帧图像的关键点以二维点云坐标形式表示,总共 M 帧,因此网络输入维度是 $N \times 2M$ 。GS-Net 由 4 个模块组成:点嵌入模块、局部上下文融合模块、全局双边正则化模块以及分类头模块。点嵌入模块类似于自然语言处理中的词嵌入,将每个运动对象的关键点嵌入到高维空间,既能丰富关键点特征也能便于后续的特征融合。局部上下文融合模块将嵌入后的特征$\mathcal{F}$(feature)以及原始的关键点坐标$\mathcal{P}$(point cloud 2D)作为输入,利用图结构在特征空间和几何空间内提取每个关键点的 K 个邻居的几何信息,并进行特征融合得到 $\mathcal{F}_L$,有助于网络准确识别运动边界和局部变化。全局双边正则化模块利用逐点和逐通道的全局感知提取全局平均特征 $\mathcal{F}_G$,该特征可表述为关键点之间的重合特征,会模糊不同运动对象和背景间的界限。为增强关键点特征的区分性,本研究采用减法过滤策略,通过 $\mathcal{F}_L - \mathcal{F}_G$ 去除 $\mathcal{F}_L$ 中含有的冗余几何信息,突出关键特征,从而提高分割的准确性。最后,分类头模块接收全局双边正则化模块的输出,用于对所有关键点的分类得分进行预测,确定每个关键点的最高预测分数,维度是 $N \times C$,其中 C 代表运动类别的数量。在同一个分割任务中,运动类别数量是已知的,并且在所有图像帧中保持不变。

1.2　点嵌入模块

点嵌入模块由两个 LBR(linear, batch-norm, relu)层组成,类似于自然语言处理中的词嵌入,主要用于将原始图像中不同运动对象的关键点坐标嵌入到高维特征空间,其具有两个优点:第 1 是增强特征表示能力,在高维空间中,特征表示更加丰富,从而能够更好地描述运动对象的复杂性;第 2 是易于区分相似关键点,在低维空间中,不同运动对象的关键点可能会具有相似的特征,导致难以区分它们。在高维空间中,这些关键点会变得更加可分,从而提高分割精度。

1.3　局部上下文融合模块

该模块主要用于捕捉运动对象的局部几何信息,这对于理解运动对象的形状及其运动模式至关重要,因为稀疏关键点的邻居更有可能是属于同一运动类别的对象。现有的卷积神经网络和注意力机制难以提取复杂局部几何信息,针对该问题,本研究提出局部上下文融合模块,模块包括局部几何分支和局部特征信息分支,两个分支均由局部图构建层、LBR 层和最大池化层构成。

局部几何分支用于提取运动对象关键点在几何空间的邻居几何信息,让网络能够学习到原始图像中运动对象的形状与边界。具体过程包括 3 个步骤。

1)使用 EdgeConv[23] 生成最近邻 KNN,任何一个关键点 p_i 的 K 个近邻表示为 $\forall p_{i_k} \in N(p_i)$,p_i 代表第 i 个关键点的二维点云坐标。

2)构建所有关键点的局部几何图,针对每个关键点 p_i ,以它为中心与 K 个邻居构成的局部几何图表示为 $\widetilde{p_i} = [p_i; p_{i_k} - p_i]$,$\widetilde{p_i} \in \boldsymbol{R}^{4\times K}$,其中 $p_{i_k} - p_i$ 是关键点 p_i 与第

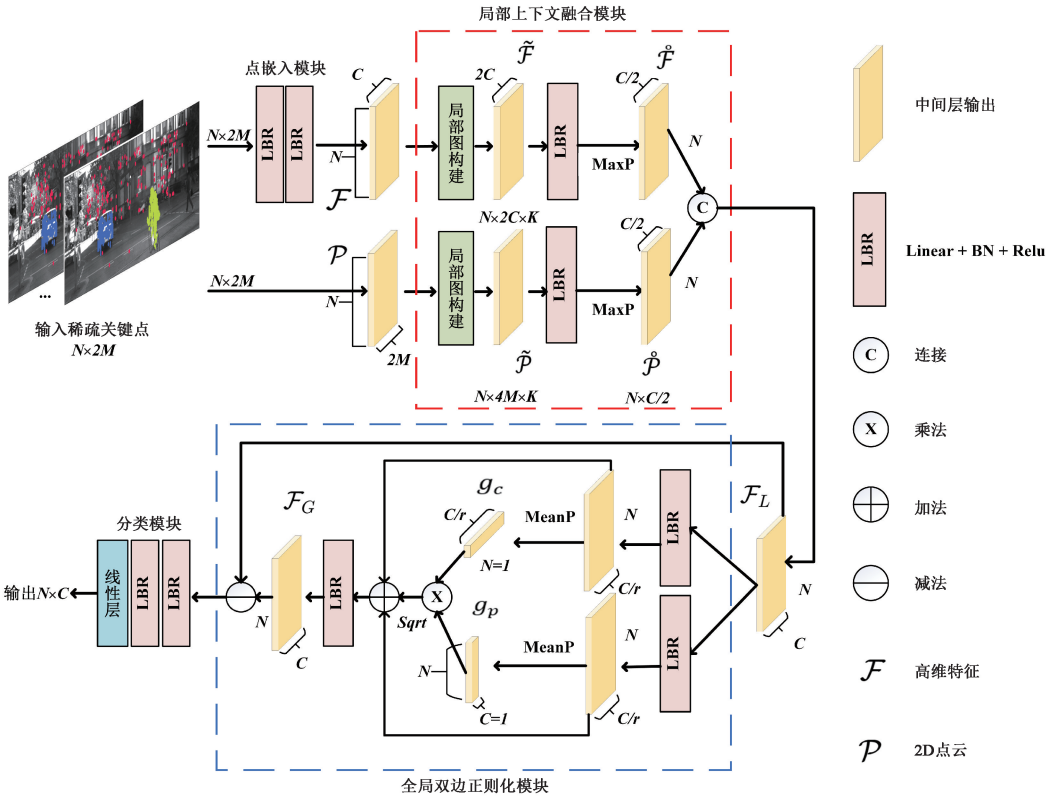

图 1　GS-Net 网络结构

Fig. 1　GS-Net network structure

K 个邻居在几何空间构成的边特征，聚合了重要的局部几何信息，一帧图像内所有关键点的局部几何图表示为 $\widetilde{\mathcal{P}}$。

3）利用 LBR 层对局部几何图进行编码，旨在进一步增强特征的表达能力。随后，在 K 个邻居维度上应用最大池化，以提取最具区分性的局部几何信息 $\mathring{\mathcal{P}}$。LBR 层中的线性层采用 1×1 卷积实现，目的是减少参数量并提高网络效率。批量归一化层和激活层用于保证数值稳定性并增加非线性。

$$\widetilde{\mathcal{P}} = [\widetilde{p_1};\cdots;\widetilde{p_i};\cdots;\widetilde{p_N}] \in \boldsymbol{R}^{N\times 4M\times K} \tag{1}$$

$$\mathring{\mathcal{P}} = \max_k(LBR(\widetilde{\mathcal{P}})), \mathring{\mathcal{P}} \in \boldsymbol{R}^{N\times\frac{C}{2}} \tag{2}$$

局部特征信息分支用于在点嵌入模块得到的高维空间中提取局部特征信息，f_i 表示第 i 个关键点的特征，详细步骤与局部几何分支类似。

针对每个关键点特征 f_i，以它为中心与 K 个邻居构成的局部特征图表示为 $\widetilde{f}_i = [f_i; f_{i_k} - f_i] \in \boldsymbol{R}^{2C\times K}$，其中 f_i 表示第 i 个关键点的特征，$f_{i_k} - f_i$ 是关键点 f_i 与第 K 个邻居在高维特征空间构成的边特征，聚合了重要的局部特征信息。1 帧图像内所有关键点的局部特征图表示为 $\widetilde{\mathcal{F}}$。经过 LBR 层和最大池化层后，得到最具区分性的局部特征信息。

$$\widetilde{\mathcal{F}} = [\widetilde{f}_1;\cdots;\widetilde{f}_i;\cdots;\widetilde{f}_N] \in \boldsymbol{R}^{N\times 2C\times K} \tag{3}$$

$$\mathring{\mathcal{F}} = \max_k(LBR(\widetilde{\mathcal{F}})), \mathring{\mathcal{F}} \in \boldsymbol{R}^{N\times\frac{C}{2}} \tag{4}$$

最终把局部几何信息 $\mathring{\mathcal{P}}$ 和局部特征信息 $\mathring{\mathcal{F}}$ 在通道维度上拼接得到局部上下文融合模块的最终输出 $\mathcal{F}_L$。

$$\mathcal{F}_L = concat(\mathring{\mathcal{P}}, \mathring{\mathcal{F}}), \mathcal{F}_L \in \boldsymbol{R}^{N\times C} \tag{5}$$

1.4　全局双边正则化模块

为了进一步增强局部几何信息 $\mathcal{F}_L$，得到更具区分性的关键点特征，本研究提出全局双边正则化模块。该模块从全局角度上提取图像中所有对象关键点的平均特征 $\mathcal{F}_G$，该特征可表述为所有运动对象和背景之间的冗余特征。基于此，本研究巧妙地运用了过滤策略，通过从 $\mathcal{F}_L$ 中减去全局冗余特征 $\mathcal{F}_G$，从而进一步增强 $\mathcal{F}_L$。

在运动分割任务中，全局信息至关重要，其可以帮助网络在全局角度上更好地区分不同的运动对象。尽管自注意力机制在捕获运动对象间的全局信息和建立长距离依赖关系方面具有优势，能够有效解决传统卷积神经网络在全局信息访问方面的问题。但其依赖点特征间长距

离依赖关系的计算方式往往伴随着高昂的内存消耗和计算负担。相比之下,本研究提出的方法在全局感知方面采用更为高效的基于全局通道和点描述符的方式,来计算特征图的元素级相互依赖关系,从而实现了对全局信息的捕捉。该模块结构简单,仅由 1×1 线性层、ReLU 激活函数和无参数的平均池化层构成,参数量少,在分割速度上具有优势,进一步增强了算法的实时性,后续将通过对比实验进行验证。

全局双边正则化模块分为通道信息分支和空间点信息分支。基于通道信息分支专注于学习点特征对每个通道的权重分配。为了得到全局通道描述符 $\boldsymbol{g}_c$,也就是通道信息分支的输出向量,本研究采用一个权重矩阵 $\boldsymbol{W}_c$,r 是缩减因子,用于缩减 $\mathcal{F}_L$ 的维度,旨在降低计算量,接着使用 ReLU 激活函数引入非线性,满足后续方程(10)的求解需求,最后沿空间轴对 N 个元素进行平均池化操作,以压缩空间信息,从而得到全局通道描述符 $\boldsymbol{g}_c$。μ_j 表示第 j 个通道的全局响应。

$$\boldsymbol{g}_c = \underset{N}{avg}(ReLU(\mathcal{F}_L\boldsymbol{W}_c)),\boldsymbol{g}_c \in \boldsymbol{R}^{1\times\frac{C}{r}} \tag{6}$$

$$\boldsymbol{W}_c \in \boldsymbol{R}^{C\times\frac{C}{r}},\boldsymbol{g}_c = [\mu_1,\cdots,\mu_j,\cdots,\mu_{\frac{C}{r}}],\mu_j \in \boldsymbol{R} \tag{7}$$

空间点信息分支与通道信息分支一致,它使用另外一个权重矩阵 $\boldsymbol{W}_p$,沿通道轴对 C/r 个通道进行平均池化操作,以压缩通道信息,得到全局点描述符 $\boldsymbol{g}_p$。

$$\boldsymbol{g}_p = \underset{C}{avg}(ReLU(\mathcal{F}_L\boldsymbol{W}_p)),\boldsymbol{g}_p \in \boldsymbol{R}^{N\times1} \tag{8}$$

$$\boldsymbol{W}_p \in \boldsymbol{R}^{C\times\frac{C}{r}},\boldsymbol{g}_p = [\lambda_1,\cdots,\lambda_j,\cdots,\lambda_N],\lambda_j \in \boldsymbol{R} \tag{9}$$

与 Kim 等[24]提出的在向量之间使用 Hadamard 乘积不同,本文通过计算 $\boldsymbol{g}_c$ 和 $\boldsymbol{g}_p$ 外积的平方根来捕获低秩全局双线性响应。

$$G = sqrt(\boldsymbol{g}_p \otimes \boldsymbol{g}_c),\mathcal{G} \in \boldsymbol{R}^{N\times\frac{C}{r}} \tag{10}$$

$$\eta_{ij} = \sqrt{\lambda_i\mu_j},\eta_{ij} \in R \tag{11}$$

这种方法的核心在于如下两点:第 1 是利用低秩矩阵来降低几何信息的复杂性,从而在减少计算量和内存消耗的同时,保留关键的全局信息;第 2 是使用全局双线性响应捕获通道和空间信息。全局通道描述符 $\boldsymbol{g}_c$ 用于捕获真实交通场景图像帧之间的通道依赖性;全局逐点描述符 $\boldsymbol{g}_p$ 用于捕获运动对象的原始形状。本文通过计算两个全局描述符的双线性组合以完全保留两种类型的全局信息。对于每个元素 η_{ij} 而言,λ_i 和 μ_j 分别是第 i 个关键点和第 j 个通道的算术平均值,因此 η_{ij} 作为它们的几何平均值,提供了重要的空间几何信息和通道信息。遵循文献[24]的建议,使用 MLP 和两个残差连接,用于恢复通道维度并生成全尺寸的全局感知图 $\mathcal{F}_G$。这种方法不仅保证了全局信息的完整性,也增强了特征的表达能力。

$$\mathcal{F}_G = MLP(\mathcal{G} + \mathcal{F}_L\boldsymbol{W}_c + \mathcal{F}_L\boldsymbol{W}_p),\mathcal{F}_G \in \boldsymbol{R}^{N\times C} \tag{12}$$

在式(6)和(8)中使用的平均池化操作,分别用于在空间轴和通道轴两个维度上捕获全局信息。平均池化得到的平均值通常被视为总体特征,缺乏足够的区分性,适合用于过滤。而在运动分割任务中,学习具有代表性和区分性的关键点特征,能够有效提高网络的分割精度。基于此,本文采用过滤策略进一步增强局部上下文融合模块输出的几何信息 $\mathcal{F}_L$,通过从 $\mathcal{F}_L$ 中减去全局冗余特征 $\mathcal{F}_G$,以突出 $\mathcal{F}_L$ 中的独特几何信息,减少全局均值带来的模糊性。此外,本研究在最终的输出特征图中引入了激活函数 Relu,以增加更多的非线性,从而提升网络的表达能力。

$$\mathcal{F}_{out} = Relu(\mathcal{F}_L - \mathcal{F}_G),\mathcal{F}_{out} \in \boldsymbol{R}^{N\times C} \tag{13}$$

1.5　分类模块

分类模块旨在有效地将真实交通图像帧中的每个运动对象的几何信息嵌入到低维空间中,从而精确地表达多个运动对象在表示空间中的语义亲和力。分类模块包括两级 LBR 与一级通用线性变换层。两级 LBR 的主要作用在于结合线性和非线性操作,使得网络能更好地捕捉并学习目标运动对象的深层次信息,从而更精确地拟合目标函数。一级线性变换将每个运动对象的关键点特征中的类别信息精确地嵌入到类别空间中,以实现更精确的运动分割。

1.6　损失函数

本文将带 Logits 的二元交叉熵损失函数(binary cross-entropy loss with logits,BCEWithLogitsLoss)作为网络的损失函数(L),衡量网络预测结果与真实结果之间的差异。它由 Sigmoid 和二元交叉熵损失函数(binary cross-entropy loss,BCE Loss)组合而成,对数和指数的特性使得该函数比普通交叉熵损失函数在数值上更稳定。损失函数为:

$$\sigma(Z_n) = \frac{1}{1+e^{-Z_n}} \tag{14}$$

$$L = -w_n[Y_n\log\sigma(Z_n) + (1-Y_n)\log(1-\sigma(Z_n))] \tag{15}$$

式中:w 是损失函数的权重;Z 是每个关键点的预测值;Y 是每个关键点的真实标签;n 是批次中的样本数量。

1.7　测试过程

为了实现高效的运动分割,本研究将关键点的几何坐标作为输入用于训练 GS-Net 网络,该网络输出底层运动子空间的相应特征表示,以便进行聚类。随后,为了获得每个点的离散聚类结果,本文采用高效的 K-means 算法对网络输出进行处理。最终,将具有最高分数的类标签分配给每个点,从而获得最终的分割结果。

2　实验

2.1　实验设置

1)数据集介绍

为了公平评估所有方法在真实交通场景的分割性能。本文使用两个公开的基于关键点的真实场景运动分割数据集(KT3DMoSeg[25]和FBMS[26])进行测试。KT3DMoSeg数据集是使用KITTI[27]自动驾驶数据集构建的,由22个序列组成,每个序列包含10~20帧,其中含有2~5个刚性运动。每个运动目标的几何模型可以是单应性、基本矩阵和仿射变换,它们之间没有明确的边界。与SubspaceNet相同,本研究利用留一法交叉验证来评估KT3DMoSeg中每个序列的前5帧,这被称为"普通"设置。此外,由于每个序列有10~20帧,本文可以用剩余的5帧进一步增强训练数据,这称之为"增强"设置。FBMS数据集包含59个视频序列,其中包含720个带注释的帧。本文在训练集中的29个视频序列上训练网络,并使用FBMS中的前10帧设置来评估剩余30个测试视频序列的性能。

2)实验设置和训练细节

本文使用PyTorch框架训练、验证和测试。实验设备是搭载4块12 GB NVIDIA GeForce GTX TITAN X GPUs的Linux服务器。针对两个数据集,网络参数都是随机初始化,没有预训练模型。实验使用Adam优化器对网络进行300个epoch的训练和优化,学习率始终为0.001。对于KT3DMoSeg和FBMS数据集,网络输入的帧数M分别设置为5和10。最后,网络输出分类$S \in \boldsymbol{R}^{N\times C}$,其中$C$表示数据集中潜在运动对象的最大数量,KT3DMoSeg和FBMS分别设置为5、10。

3)评价指标

为了公平评估所有方法的分割性能,本研究遵循SubspaceNet的实验设置,在KT3DMoSeg数据集上使用误分类率的平均值和中值作为评价指标,在表格中分别以平均误差(mean error, MeanErr)和中值误差(median error, MedErr)表示,所有误分类率以%为单位;FBMS数据集则使用文献[26]的精度(precision)、召回率(recall)和F-measure作为评价指标,F-measure是精度和召回率的一种综合评价指标,它平衡了精度和召回率对模型性能评估的贡献。

2.2　消融实验

为深入探究GS-Net各部分的最优设置,本研究针对模型嵌入的维度、局部上下文融合模块的邻居数量K,以及池化和正则化策略的选择展开了消融实验。在此基础上,为验证上述设置的合理性,本研究进一步针对处于上述"最优"设置下的两大核心组件开展了更为深入的消融实验。此外,为了能够客观地呈现GS-Net网络结构所具备的优越性,本研究将其与领域内的经典算法进行了深度对比实验。需说明的是,上述所有的消融实验均是在KT3DMoSeg数据集上开展的,并且在实验结果中,以红色标注的数值代表着最佳结果。

1)网络嵌入维度

嵌入维度本质为关键点的特征信息量,对于网络的学习至关重要。为了深入探究嵌入维度对网络性能的影响,本文分别使用不同嵌入维度构造网络进行实验。从表1能够看出,当嵌入维度设定为256时,网络的均值误分类率和中值误分类率分别达到了2.78%和0.49%,这两项结果均为次佳水平。相较于嵌入维度为128的网络,其均值误分类率降低了0.34%,这种提升效果是十分显著的。同时,参数量仅增加了269.61×10^3,由此带来的收益较高。然而,对于嵌入维度为512的网络而言,与256维度的网络相比,其性能提升极为有限。其均值误分类率仅仅降低了0.16%,但参数量却增加了$1\ 067\times10^3$,这样的收益较低。

综合以上情况,本文最终选择将网络的嵌入维度设置为256。在该维度下,网络的推理时间最短,具备最快的分割速度。所以,256维度下的网络能够较好地在速度和精度之间实现平衡。

表1　嵌入维度对GS-Net性能的影响

Table 1　Effect of embedding dimension on GS-Net performance

评价指标/维度数	32	64	128	256	512
MeanErr/%	3.46	3.18	3.12	2.78	2.62
MedErr/%	1.55	0.68	0.45	0.49	0.45
Time/ms	9.60	10.16	10.18	9.31	9.68
Params/($\times10^3$)	6.91	24.79	93.60	363.21	1 431.01

2)构建图的邻居数量

邻居数量K对从最近邻居获取的局部信息量起着决定性作用,这对于局部特征嵌入而言至关重要。为深入探究局部上下文融合模块的邻居数量对网络性能的影响,本研究分别采用不同的邻居数量K开展实验。如表2所示,当邻居数量小于5时,网络性能欠佳。原因在于此时提取的局部几何信息量不足,无法充分表征运动对象的复杂性,进而导致网络难以区分不同运动对象之间以及运动对象和背景点之间的边界。随着邻居数量的增加,局部几何信息量也相应增加,并在邻居数量为5时达到最佳性能。不过,当邻居数量超过5时,性能开始下降。以边界点为例,若邻居数量过多,关键点从邻居处提取的特征信息量会增大,不同关键点的特征重叠情况加剧,这会模糊不同运动对象之间的界限,最终致使网络分

割精度降低。本文以实现分割精度和速度的最佳平衡为目标,因此将邻居数量设置为 5。在此设置下,推理时间也取得了次优结果。

表 2　构建图的邻居数量对 GS-Net 性能的影响

Table 2　Effect of the number of neighbors in the constructed graph on the performance of GS-Net

评价指标/邻居数	1	3	5	7	9	11
MeanErr/%	3.87	3.48	2.47	2.63	2.78	2.71
MedErr/%	1.67	1.19	0.49	0.56	0.63	0.54
Time/ms	10.86	9.33	9.39	10.38	9.51	10.04

3)池化和正则化策略

为了探讨池化和正则化策略的选择对网络性能的影响,本文重点分析了全局双边正则化模块中池化操作的选择及局部上下文融合模块的输出 $\mathcal{F}_L$ 和全局感知图 $\mathcal{F}_G$ 之间的运算组合的选择。本文在两个分支的操作保持一致,其可以确保在不同分支中提取的特征具有相同的尺度和空间分布,从而在后续的特征融合和正则化过程中减少不必要的误差和噪声,提高分割的准确性,因此式(6)和(8)利用的池化操作是相同的,本文设置 6 种不同的组合方式进行实验比较,其中表 3 中 Max-pooling(x2)代表式(6)和(8)均使用最大池化。

具体实验结果如表 3 所示,平均池化加减法的组合表现出最佳的整体性能。虽然最大池化加减法的组合在中值误差方面达到了最优水平,但其均值误差相较于平均池化加减法的组合高出了 0.27%,中值误差仅低 0.14%。这表明在两个分支上使用平均池化要优于使用最大池化,原因在于平均值能够更全面地体现整体特征。换言之,此关键点特征属于难以区分的冗余特征,可以用于去除 $\mathcal{F}_L$ 中存在的冗余特征,实验结果也对这一点进行了验证,即减法能够有效地去除 $\mathcal{F}_L$ 中的冗余特征。因此,本文最终选择了平均池化加减法的组合作为网络的最终配置。

表 3　池化和正则化策略对 GS-Net 性能的影响

Table 3　Impact of pooling and regularization strategies on GS-Net performance

组合序号	Operation(in Equation 6,8)	Regularization	MeanErr/%	MedErr/%
1	Max-pooling(x2)	⊙	3.26	0.76
2	Max-pooling(x2)	⊕	3.18	0.49
3	Max-pooling(x2)	⊖	2.74	0.35
4	Avg-pooling(x2)	⊙	2.67	0.82
5	Avg-pooling(x2)	⊕	3.10	0.44
6	Avg-pooling(x2)	⊖	2.47	0.49

4)两大核心模块对网络性能的影响

局部上下文融合模块(local context fusion module, LCFM)和全局双边正则化模块(global bilateral regularization module,GBRM)是 GS-Net 的两大核心模块。为了验证两个核心模块的重要性,本文开展了消融实验,实验结果如表 4 所示。当网络仅使用卷积(即无两大核心模块)进行训练时,其分割性能远低于带有一个或多个核心模块的网络,这表明图结构提取复杂几何信息的能力优于卷积神经网络。此外,为进一步对比卷积神经网络和图结构提取局部信息的能力,本文增设了两组实验(表 4 中 LCFM+卷积、GBRM+卷积),具体方式是用卷积层分别替换其中一个核心模块。实验结果显示,与仅含一个核心模块的网络相比,添加卷积层并未带来显著的性能提升。

究其原因,一方面,相对于图结构而言,卷积操作难以提取更深层次的几何信息;另一方面,核心模块内部本身就含有一定的卷积层,这就使得额外添加卷积层对实验结果的影响不大。不仅如此,从表 4 和 6 的实验结果还能看出,仅含有 LCFM 的网络性能已经远远高于 SubspaceNet。这一结果有力地证明了本文提出的图结构和信息融合在提取几何信息方面的有效性。在此基础上,当加入 GBRM 后,分割性能得到了进一步的提升,这验证了本文所提出的过滤策略的合理性和有效性。此时,具备两个核心模块的 GS-Net 改进效果最为显著,始终优于仅含一个核心模块的网络,这也体现出两个核心模块之间相互协作、相辅相成的关系,共同推动了运动分割性能的提升。

表 4　核心模块 LCFM 和 GBRM 对 GS-Net 性能的影响

Table 4　Impact of core modules LCFM and GBRM on GS-Net performance　(%)

评价指标	MeanErr	MedErr
仅卷积	9.94	7.58
LCFM+卷积	3.25	0.85
GBRM+卷积	5.12	1.99
仅 LCFM	3.18	0.87
仅 GBRM	5.08	1.92
GS-Net	2.47	0.49

5)GS-Net 与三大经典算法的对比分析

为了更深入地阐释 GS-Net 网络结构所具备的优越性,本文针对 GS-Net 展开了详尽的剖析,并与领域内经典算法开展了结构分析消融实验。本研究高度重视算法的实用性,即着重考量算法在实现分割精度与速度之间达成良好平衡的能力。除了常规的性能指标之外,还引入了网络参数量以及网络进行一次推理所需的时间成本这两个指标。为了有效规避因特殊情况而可能导致的数据偏差,以下所呈现的各项结果均为 300 次实验后所获

取结果的平均值。

具体的实验结果如表5所示。从表5数据可以清晰地看出:就网络进行一次推理的时间成本这一指标而言,GS-Net所产生的结果仅为0.009 s,相较于其他网络,此速度优势极为显著。而从网络参数大小的角度来观察,GS-Net所具有的参数数量是最少的。在分割性能方面,GS-Net同样表现出色,其平均误分类率仅为2.47%,取得了最优的结果。

表5　GS-Net与三大经典算法的结构对比分析

Table 5　Comparative analysis of the structure of GS-Net with the three classical algorithms

模型名称/评价指标	MeanErr/%	MedErr/%	参数大小/MB	推理时间/s
Corres-Net[28]	12.05	12.35	3.21	0.06
Subspace-Net	10.62	8.44	9.37	1.85
GIET	3.27	0.36	0.77	0.2
GS-Net	2.47	0.49	0.35	0.009

2.3　在不同数据集对比领域内相关算法

1)KT3DMoSeg数据集

本研究在KT3DMoSeg数据集上进行了大量多类运动分割任务的实验。GS-Net与一些非深度学习和基于深度学习的方法进行了比较。实验结果如表6所示,无论在“普通”设置还是“增强”设置下,本方法在22个KT3DMoSeg序列中均展现出超过现有方法的性能,其中“普通”和“增强”设置下的平均误分类率分别为2.47%和0.88%,中值误分类率分别为0.49%和0.14%,表6中两种设置的实验结果以斜杠分隔显示,其中MVC是由Xu等[25]在原文中提出的首个应用于数据集KT3DMoSeg的方法。

表6　KT3DMoSeg数据集上的运动分割性能

Table 6　Motion segmentation performance on the KT3DMoSeg dataset

方法		MeanErr/%	MedErr/%	推理时间/s
非深度学习	MVC	10.99	6.57	143.52
	SUBSET[29]	8.08	0.71	22.20
	CMFO[30]	6.73	3.82	–
	MMC[31]	5.78	2.89	3 230.1
	GMF[32]	4.58	1.10	–
	HMFMS[33]	4.48	0.69	0.83
深度学习	NMI[34]	16.91	11.65	1.85
	SubspaceNet	10.62/5.83	8.44/3.58	1.85
	GIET	3.27/1.23	0.36/0.11	0.2
	GS-Net	2.47/0.88	0.49/0.14	0.009

与传统方法相比,GS-Net表现出强大的分割性能,超过了领域内最先进的(heterogeneous model-fitting based motion segmentation method,HMFMS)。这是因为传统的运动分割方法主要依赖于模型拟合,需要指定基本模型类型(如基本矩阵或单应性矩阵)。然而,现实场景常涵盖多种模型类型,且当数据涉及多个模型实例时,指定单一类型的基本模型可能会遗漏关键信息。相比之下,本文提出的GS-Net不需要指定任何特定模型,而是通过网络直接学习真实图像中运动对象之间的关系进行运动分割。

与基于深度学习的方法相比,在采用“普通”设置的情况下,GS-Net的分割精度超过了除GIET“增强”设置之外的所有方法。这一优势主要归因于GS-Net简洁的图结构网络架构设计和不依赖大量训练数据的特性。在“增强”设置下,GS-Net的表现尤为出色,均值误分类率为0.88%,显著优于SubspaceNet的5.83%和GIET的1.23%,这一结果表明,在运动分割领域,图结构在提取复杂几何信息方面的性能要优于卷积神经网络和注意力机制。

表6通过测量GS-Net和其他对比方法处理5帧真实图像帧的推理时间来评估其分割速度。值得注意的是,大多数非深度学习方法是在基于CPU的MATLAB中实现的,并且需要迭代优化步骤来采样点以估计运动模型,计算时间相对较长。而深度学习方法则是基于GPU的PyTorch框架实现的。文献[19]评估的基于深度学习的方法,包括一些复杂的卷积层和结构,需要大约1.85 s来处理这些序列。而GS-Net由于采用了简单但有效的网络结构,只需0.009 s即可完成5帧测试数据的处理。实验结果展示了GS-Net在实时性方面的显著优势,这对于算法的实际部署至关重要。

为进一步验证GS-Net在交通场景中的分割性能,本文将子集约束多模型谱聚类算法(subset constrained multi-model spectral clustering,SUBSET)、SubspaceNet以及GS-Net在真实交通场景下的分割结果进行可视化。如图2所示,图中第1行至第4行分别表示真实图像帧“Seq059 Clip01”、“Seq095 Clip01”、“Seq009 Clip01”和“Seq009 Clip03”,图中不同物体能够以不同颜色的标注为最优分割,以图2(a)的图像真值(ground-truth)作为衡量分割精度的标准,旨在确保评估的公正性和准确性,SUBSET是一种传统的基于模型拟合的方法,采用多模型谱聚类框架,结合不同类型的模型来分割运动对象。相比之下,SubspaceNet是一种近期提出的基于深度学习的方法,构建了多模型多类型拟合网络,用于对交通场景下不同运动对象进行分割。鉴于此,本研究选取这两种具有代表性的方法作为综合比较的基准,以此展示GS-Net在处理复杂运动分割任务时的性能表现。

从图2 A的“Seq059 Clip01”可以观察到,SUBSET和SubspaceNet都无法很好地将汽车和背景区分开,GS-Net虽然能够对汽车和背景进行正常分割,但也存在少量误

分类点,这与汽车和背景的点数不平衡以及点与点之间距离太小有关。在图 2 B 的“Seq095 Clip01”视频序列中,SUBSET 虽然正确分割了移动汽车的大部分点,然而,但它错误地将部分背景点标记为移动汽车,未能有效区分背景与移动汽车。SubspaceNet 则将右侧汽车 1/2 关键点和左边汽车部分关键点错误地划分为背景点。相比之下,GS-Net 能较好地分割出两个运动物体,这表明 GS-Net 具备强大的边界识别能力,这种能力得益于 GS-Net 中图结构和过滤策略的高效性,它们有效解决了不同运动对象间数据点分布不均衡和边界模糊的问题。

在图 2C 的“Seq009 Clip01”视频序列中,GS-Net 和 SubspaceNet 能够较好地分割出运动汽车,而 SUBSET 错误地将部分背景点标记为运动汽车。在最后一行的“Seq009 Clip03”视频序列中,SUBSET 和 GS-Net 都能够正确分割出 3 个运动物体,而 SubspaceNet 无法正确分割右侧的汽车。综上所述,本文方法 GS-Net 在每帧图像中都取得最佳的分割效果,这体现了其网络设计的合理性,也彰显了图结构在几何信息挖掘方面的高效性。

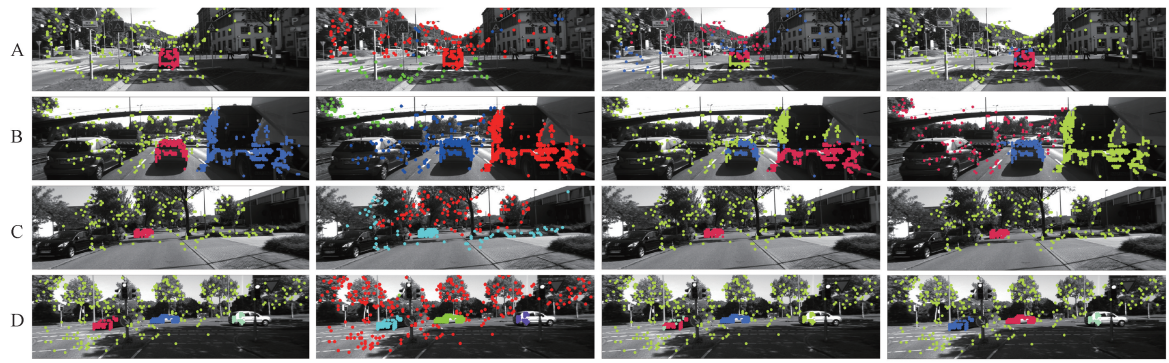

图 2　SUBSET、SubspaceNet 和 GS-Net 从 KT3DMoSeg 数据集获得的分割结果

Fig. 2　Segmentation results obtained by SUBSET, SubspaceNet and GS-Net from the KT3DMoSeg dataset

2) FBMS 数据集

遵循文献[19]的实验设置,本文将 GS-Net 与一些经典分割方法在 FBMS 数据集上展开了对比实验。从表 7 可知,GS-Net 在召回率和 F-measures 上分别达到了 82.53%和 81.93%,取得最佳结果。需要注意的是,在精确度方面,GS-Net 的表现虽算不上突出,但是综合 3 个指标考虑,GS-Net 整体表现最优,由于 F-measure 是对精度和召回率的综合评价指标,它能够平衡二者对模型性能评估的影响,所以 GS-Net 在综合指标 F-measure 上取得最佳性能这一结果,充分表明了图结构在提取几何信息方面的高效性。

表 7　FBMS 前 10 帧的性能

Table 7　Performance on the first 10 frames of FBMS

(%)

方法	Precision	Recall	F-measure
CGHMS[35]	74.23	63.07	64.97
CCC[36]	83.17	74.65	78.68
SubspaceNet	85.7	69.2	76.57
UA[37]	88.17	68.96	77.40
SUBSET	84.41	72.87	78.22
HOMC[38]	83.20	74.34	78.52
GS-Net	81.34	82.53	81.93

为了更直观地验证这一点,本研究对算法 SUBSET、SubspaceNet 以及 GS-Net 在真实交通场景下的分割结果进行了可视化处理。如图 3 所示,从“Cars5_01”可以看出,图 3(c)的方法 SUBSET 和图 3(b)的方法 SubspaceNet 都无法有效地分割两辆运动汽车。其中,SUBSET 将两辆汽车识别为一辆,并且错误地将部分背景点标记为运动物体,SubspaceNet 也存在类似的误分类问题。而在“Cars10_01”和“Cars4_01”中 3 种算法均展现出了较好的分割效果,但是都有着不同程度的误分类情况。在“Cars1_01”中,汽车关键点数量众多,对算法的全局和局部几何信息提取能力提出了更高要求,在这种情况下,SUBSET 和 SubspaceNet 的分割结果出现了显著问题,它们错误地将数量众多的汽车关键点分割成了不同类别或者将汽车与背景混淆。这再次凸显了在处理复杂交通场景时,算法的全局和局部信息提取能力的重要性。相比之下,本文方法 GS-Net 凭借其两大核心模块,在提取局部和全局几何信息方面表现出色,取得了最佳的分割效果,最接近于图 3(d)的图像真值(ground-truth),这充分证实了其在真实交通场景下强大的分割性能。

此外,为进一步展现 GS-Net 的高效性,本文中 GS-Net 与几种光流估计方法的性能对比情况如表 8 所示。

图 3　FBMS 数据集的性能可视化

Fig. 3　Performance visualization of the FBMS dataset

除 GS-Net 之外,所列的所有基于深度学习的方法都需要完整的 RGB 图像序列作为输入数据,这些方法主要是为光流估计这一特定目的而设计的。它们在处理数据时,依赖于图像中丰富的色彩、纹理等信息来计算光流,以实现对物体运动的估计。然而,GS-Net 与之不同,它具有独特的优势。GS-Net 专门针对仅涉及抽象轨迹的分割任务进行设计,能够在这种相对信息匮乏的情况下有效运作。在分割任务中,GS-Net 输入数据信息量远远低于光流方法的输入信息量。但即便如此,GS-Net 利用其图结构提取复杂信息的特性,在有限的几何信息输入中充分挖掘并提取有效信息。

表 8　GS-Net 与一些光流估计方法之间的性能比较

Table 8　Performance comparison between GS-Net and some optical flow estimation methods

方法	Deep	F-measure/%
Deepflow[39]	√	80.18
FlowNet2[40]	√	79.92
AD-Net[41]	√	81.20
PCSA[42]	√	83.10
MAM-Net[43]	√	88.30
GS-Net	√	81.93

通过这种方式,GS-Net 最终实现了较优的 F-measure 值,并且在性能上超过了一些经典的光流方法,如 Deepflow[41],FlowNet2[42],AD-Net[43])。这些实验结果强有力地证明了 GS-Net 在基于关键点的运动分割任务中具有出色的鲁棒性。这种鲁棒性意味着 GS-Net 在面对基于关键点的运动分割这种特定类型的任务时,能够稳定且有效地发挥作用,不受输入信息有限等因素的干扰,始终保持良好的性能表现。

3　结　论

本文提出用于几何信息学习的图结构运动分割方法(GS-Net),旨在解决现有运动分割方法在交通场景下实用性方面的不足。网络设计的难点在于,在提取复杂几何信息时,需维持较低计算量与内存消耗,以实现速度和精度的平衡。为此,研究针对网络各个模块进行最优设计,充分利用图结构提取运动目标复杂的几何信息,同时运用过滤策略进一步强化关键点特征,进而提升运动分割的速度与精度,以满足真实交通场景的需求。其方法核心在于局部上下文融合模块和全局双边正则化模块。前者解决了现有基于深度学习的运动分割方法难以有效提取局部几何信息的问题;后者采用高效的基于全局通道和点描述符的方式计算特征图的元素级相互依赖关系,减少了内存消耗和计算负担,并保证了全局信息的获取,解决了基于注意力机制的算法在提取全局信息时带来的高内存消耗和高计算量的问题。经真实交通场景的实验验证,GS-Net 在速度和精度方面均达到了最优水平,有力证实了图结构在运动分割领域的通用性和高效性。然而,鉴于真实运动场景复杂多变,轻微移动和边界

信息的学习问题仍未得到彻底解决。未来的研究方向可探索将时间信息融入网络,充分利用时间序列数据的特性,以解决运动对象随时间动态变化的问题,这将有助于实现更为准确、快速的运动分割,期望借此解决更多工业和生活中的实际问题。

参考文献

[1] 樊博, 高玮玮, 单明陶, 等. 融合注意力机制与重影特征映射的无人机交通场景目标轻量级语义分割[J]. 电子测量与仪器学报, 2023, 37(3): 21-28.
FAN B, GAO W W, SHAN M T, et al. Lightweight semantic segmentation of UAV traffic scene objects combining attention mechanism and ghost feature mapping [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(3): 21-28.

[2] 陈瑞东, 秦会斌. 多特征融合与卡尔曼预测的车辆跟踪算法[J]. 电子测量技术, 2023, 46(7): 32-38.
CHEN R D, QIN H B. Vehicle tracking algorithm based on multi-feature fusion and Kalman prediction [J]. Electronic Measurement Technology, 2023, 46(7): 32-38.

[3] 冯洲, 续欣莹, 郑宇轩, 等. 动态场景下基于实例分割和三维重建的多物体单目 SLAM[J]. 仪器仪表学报, 2023, 44(8): 51-62.
FENG ZH, XU X Y, ZHENG Y X, et al. Multi-object monocular SLAM based on instance segmentation and 3D reconstruction in dynamic scene[J]. Chinese Journal of Scientific Instrument, 2023, 44(8): 51-62.

[4] KEUPER M, TANG S, ANDRES B, et al. Motion segmentation & multiple object tracking by correlation co-clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(1): 140-153.

[5] ARRIGONI F, RICCI E, PAJDLA T. Multi-frame motion segmentation by combining two-frame results [J]. International Journal of Computer Vision, 2022, 130(3): 696-728.

[6] RAGURAM R, CHUM O, POLLEFEYS M, et al. USAC: A universal framework for random sample consensus[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 2022-2038.

[7] BARATH D, IVASHECHKIN M, MATAS J. Progressive NAPSAC: Sampling from gradually growing neighborhoods[J]. ArXiv preprint arXiv:1906.02295, 2019.

[8] BARATH D, NOSKOVA J, IVASHECHKIN M, et al. MAGSAC++, a fast, reliable and accurate robust estimator[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 1304-1312.

[9] BARATH D, MATAS J. Graph-cut RANSAC: Local optimization on spatially coherent structures[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(9): 4961-4974.

[10] LI Q, LAN X, LI J. Information fusion on the two-layer network for robust estimation of multiple geometric structures [J]. Information Sciences, 2020, 530: 148-166.

[11] BARATH D, ROZUMNYI D, EICHHARDT I, et al. Finding geometric models by clustering in the consensus space [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2023: 5414-5424.

[12] YAN Y, LIU M, CHEN S, et al. A novel robust model fitting approach towards multiple-structure data segmentation [J]. Neurocomputing, 2017, 239: 181-193.

[13] BRACHMANN E, ROTHER C. Neural-guided RANSAC: Learning where to sample model hypotheses [C]. Proceedings of the IEEE International Conference on Computer Vision, 2019: 4322-4331.

[14] KLUGER F, BRACHMANN E, ACKERMANN H, et al. Consac: Robust multi-model fitting by conditional sample consensus[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 4634-4643.

[15] CAVALLI L, POLLEFEYS M, BARATH D. NeFSAC: Neurally filtered minimal samples[C]. Proceedings of the European Conference on Computer Vision, 2022: 351-366.

[16] WEI T, MATAS J, BARATH D. Adaptive reordering sampler with neurally guided MAGSAC[C]. Proceedings of the IEEE International Conference on Computer Vision, 2023: 18163-18173.

[17] WEI T, PATEL Y, SHEKHOVTSOV A, et al. Generalized differentiable RANSAC[C]. Proceedings of the IEEE International Conference on Computer Vision, 2023: 17649-17660.

[18] KLUGER F, ROSENHAHN B. PARSAC: Accelerating robust multi-model fitting with parallel sample consensus [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(3): 2804-2812.

[19] XU X, ZHANG L, CHEONG L F, et al. Learning clustering for motion segmentation [J]. IEEE

Transactions on Circuits and Systems for Video Technology, 2022, 32(3): 908-919.

[20] LI Q, CHENG J, GAO Y, et al. Learning geometric information via transformer network for key-points based motion segmentation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(9): 7856-7869.

[21] LIAO T, ZHANG X, XU Y, et al. SGA-Net: A sparse graph attention network for two-view correspondence learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(12): 7578-7590.

[22] YE Y, JI S. Sparse graph attention networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 35(1): 905-916.

[23] WANG Y, SUN Y B, LIU Z W, et al. Dynamic graph cnn for learning on point clouds[J]. ACM Transactions on Graphics, 2019, 38(5): 1-12.

[24] KIM J H, ON K W, LIM W, et al. Hadamard product for low-rank bilinear pooling[J]. ArXiv preprint arXiv: 1610.04325, 2016.

[25] XU X, CHEONG L F, LI ZH W. Motion segmentation by exploiting complementary geometric models [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2859-2867.

[26] OCHS P, MALIK J, BROX T. Segmentation of moving objects by long term video analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(6): 1187-1200.

[27] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: The kitti dataset[J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.

[28] XU X, CHEONG L F, LI ZH W. Learning for multi-model and multi-type fitting[J]. ArXiv preprint arXiv: 1901.10254, 2019.

[29] XU X, CHEONG L F, LI ZH W. 3D rigid motion segmentation with mixed and unknown number of models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(1): 1-16.

[30] XI ZH, LIU J, LUO B, et al. Multi-motion segmentation: Combining geometric model-fitting and optical flow for RGB sensors[J]. IEEE Sensors Journal, 2022, 22(7): 6952-6963.

[31] HUANG Y, ZELEK J. Motion Segmentation from a Moving Monocular Camera[J]. ArXiv preprint arXiv: 2309.13772, 2023.

[32] JIANG Y B Y, XU Q, MA K, et al. What to select: Pursuing consistent motion segmentation from multiple geometric models [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(2): 1708-1716.

[33] LIN SH Y, YANG AN J, LAI T T, et al. Multi-motion segmentation via co-attention-induced heterogeneous model fitting[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(3): 1786-1798.

[34] OH SONG H, JEGELKA S, RATHOD V, et al. Deep metric learning via facility location[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5382-5390.

[35] BIDEAU P, ROYCHOWDHURY A, MENON R R, et al. The best of both worlds: Combining CNNs and geometric constraints for hierarchical motion segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 508-517.

[36] KEUPER M, TANG S Y, ANDRES B, et al. Motion segmentation & multiple object tracking by correlation co-clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(1): 140-153.

[37] KARDOOST A, KEUPER M. Uncertainty in minimum cost multicuts for image and motion segmentation[C]. Uncertainty in Artificial Intelligence, 2021: 2029-2038.

[38] LEVINKOV E, KARDOOST A, ANDRES B, et al. Higher-order multicuts for geometric model fitting and motion segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 608-622.

[39] WEINZAEPFEL P, REVAUD J, HARCHAOUI Z, et al. DeepFlow: Large displacement optical flow with deep matching[C]. Proceedings of the IEEE International Conference on Computer Vision, 2013: 1385-1392.

[40] ILG E, MAYER N, SAIKIA T, et al. Flownet 2.0: Evolution of optical flow estimation with deep networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2462-2470.

[41] YANG ZH, WANG Q, BERTINETTO L, et al. Anchor diffusion for unsupervised video object segmentation[C]. Proceedings of the IEEE International Conference on Computer Vision, 2019: 931-940.

[42] GU Y CH, WANG L J, WANG Z Q, et al. Pyramid constrained self-attention network for fast video salient

object detection [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 10869-10876.

[43] ZHAO X, LIANG H, LI P, et al. Motion-aware memory network for fast video salient object detection[J]. IEEE Transactions on Image Processing, 2024, 33: 709-721.

作者简介

张纪友,2022 年于厦门大学嘉庚学院获得学士学位,现就读于福建农林大学攻读硕士学位,主要研究方向为计算机视觉、深度学习、运动分割。

E-mail: 2408041956@qq.com

Zhang Jiyou received his B. Sc. degree from Xiamen University Tan Kah Kee College in 2022. Now he is a M. Sc. candidate at Fujian Agriculture and Forestry University. His main research interests include computer vision, deep learning, and motion segmentation.

李琦铭(通信作者),2016 年于厦门大学获博士学位,现为中国科学院海西研究院副研究员,主要研究方向为计算机视觉、目标检测及跟踪、机器学习及人机交互等。

E-mail: qimingli@fjirsm.ac.cn

Li Qiming (Corresponding author) received his Ph. D. degree from Xiamen University in 2016. Now he is an associate researcher in the Haixi Research Institute of Chinese Academy of Sciences. His main research interests include computer vision, target detection and tracking, machine learning and human-computer interaction, etc.