

# 基于光流与多尺度上下文的图像序列运动遮挡检测

冯诚<sup>1</sup> 张聪炫<sup>1,2</sup> 陈震<sup>1</sup> 李兵<sup>2</sup> 黎明<sup>3</sup>

**摘要** 针对非刚性运动和大位移场景下运动遮挡检测的准确性与鲁棒性问题, 提出一种基于光流与多尺度上下文的图像序列运动遮挡检测方法. 首先, 设计基于扩张卷积的多尺度上下文信息聚合网络, 通过图像序列多尺度上下文信息获取更大范围的图像特征; 然后, 采用特征金字塔构建基于多尺度上下文与光流的端到端运动遮挡检测网络模型, 利用光流优化非刚性运动和大位移区域的运动检测遮挡信息; 最后, 构造基于运动边缘的网络模型训练损失函数, 获取准确的运动遮挡边界. 分别采用 MPI-Sintel 和 KITTI 测试数据集对所提方法与现有的代表性方法进行实验对比与分析. 实验结果表明, 所提方法能够有效提高运动遮挡检测的准确性和鲁棒性, 尤其在非刚性运动和大位移等困难场景下具有更好的遮挡检测鲁棒性.

**关键词** 图像序列, 遮挡检测, 深度学习, 多尺度上下文, 非刚性运动

**引用格式** 冯诚, 张聪炫, 陈震, 李兵, 黎明. 基于光流与多尺度上下文的图像序列运动遮挡检测. 自动化学报, 2024, 50(9): 1854-1865

**DOI** 10.16383/j.aas.c210324

## Occlusion Detection Based on Optical Flow and Multiscale Context

FENG Cheng<sup>1</sup> ZHANG Cong-Xuan<sup>1,2</sup> CHEN Zhen<sup>1</sup> LI Bing<sup>2</sup> LI Ming<sup>3</sup>

**Abstract** In order to improve the accuracy and robustness of occlusion detection under non-rigid motion and large displacements, we propose an occlusion detection method of image sequence motion based on optical flow and multiscale context. First, we design a multiscale context information aggregation network based on dilated convolution which obtains a wider range of image features through multiscale context information of image sequence. Then, we construct an end-to-end motion occlusion detection network model based on multiscale context and optical flow using feature pyramid, utilize the optical flow to optimize the performance of occlusion detection in areas of non-rigid motion and large displacements region. Finally, we present a novel motion edge training loss function to obtain the accurate motion occlusion boundary. We compare and analysis our method with the existing representative approaches by using the MPI-Sintel datasets and KITTI datasets, respectively. The experimental results show that the proposed method can effectively improve the accuracy and robustness of motion occlusion detection, especially gains the better occlusion detection robustness under non-rigid motion and large displacements.

**Key words** Image sequence, occlusion detection, deep learning, multiscale context, non-rigid motion

**Citation** Feng Cheng, Zhang Cong-Xuan, Chen Zhen, Li Bing, Li Ming. Occlusion detection based on optical flow and multiscale context. *Acta Automatica Sinica*, 2024, 50(9): 1854-1865

收稿日期 2021-04-15 录用日期 2021-07-02

Manuscript received April 15, 2021; accepted July 2, 2021

国家重点研发计划 (2020YFC2003800), 国家自然科学基金 (61866026, 61772255, 62222206), 江西省杰出青年人才计划 (20192BCB23011), 江西省自然科学基金重点项目 (20202ACB214007), 江西省优势科技创新团队 (20165BCB19007) 资助

Supported by National Key Research and Development Program of China (2020YFC2003800), National Natural Science Foundation of China (61866026, 61772255, 62222206), Outstanding Young Scientist Project of Jiangxi Province (20192BCB23011), Natural Science Foundation of Jiangxi Province (20202ACB214007), and Advantage Subject Team of Jiangxi Province (20165BCB19007)

本文责任编辑 桑农

Recommended by Associate Editor SANG Nong

1. 南昌航空大学测试与光电工程学院 南昌 330063 2. 中国科学院自动化研究所模式识别国家重点实验室 北京 100190 3. 南昌航空大学信息工程学院 南昌 330063

1. School of Measuring and Optical Engineering, Nanchang Hangkong University, Nanchang 330063 2. National Key Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 3. School of

运动遮挡是由图像序列中物体和场景相互运动引起的像素点遮挡现象. 通过定位遮挡像素点提取运动遮挡区域是图像处理、计算机视觉等领域的热点问题, 研究成果广泛应用于动作识别<sup>[1]</sup>、人体姿态识别<sup>[2]</sup>、光流估计<sup>[3]</sup>、人脸识别<sup>[4]</sup>、目标跟踪<sup>[5]</sup>和三维重建<sup>[6-7]</sup>等更高级视觉任务.

在过去的十多年中, 已有许多致力于研究光流估计和基于光流的运动遮挡检测工作<sup>[8-11]</sup>. 在传统方法中, 运动遮挡区域被定义为违反亮度守恒假设或前后一致性假设的异常区域. 传统的运动遮挡检测方法通常是利用光流前后一致性假设构建运动遮挡检测模型, 首先需要计算前后连续两帧图像的光

Information Engineering, Nanchang Hangkong University, Nanchang 330063

流, 然后通过前后向光流的一致性检测提取运动遮挡区域<sup>[12-15]</sup>. 然而, 基于光流一致性假设的运动遮挡检测方法存在较大的局限性: 一方面, 一致性阈值需要人工设定, 使得该类方法运动遮挡结果难以适用于不同场景; 另一方面, 该类方法需要前后向光流信息, 增加了运动遮挡检测的时间消耗, 导致模型的实时性较低.

针对基于光流一致性假设的运动遮挡检测方法精度低和实时性差的问题, 一些工作提出采用局部分层模型<sup>[16-17]</sup>来检测运动遮挡区域. 基于局部分层模型的运动遮挡检测方法通常利用连续两帧的前向光流以及参考帧作为输入, 采用分割算法将参考帧图像与光流分割为细密的局部区域, 建立概率模型对这些局部区域进行是否存在运动遮挡的判断. 基于局部分层模型的运动遮挡检测方法只需要计算一次光流, 使得其实时性更优. 同时通过引入参考帧分割图作为额外信息对运动遮挡检测进行引导, 使得该类方法运动遮挡检测准确性也更高.

考虑到运动遮挡的本质是由物体运动引起的场景中物体相互遮挡的现象, 运动往往是发生在三维空间中的, 图像所呈现的是三维空间的投影, 图像天然缺失场景的深度信息, 因此从二维图像中检测三维运动所引起的运动遮挡较为困难. 为更好地对运动遮挡进行准确的检测, 一部分研究提出利用场景流方法<sup>[18-19]</sup>重建三维运动, 在三维空间中进行运动遮挡检测. 基于场景流运动遮挡检测方法通常将前后帧三维点云作为输入, 通过对局部点云进行运动建模, 采用几何以及亮度信息进行约束, 最终将违反几何一致性和亮度守恒假设的区域定义为运动遮挡区域. 由于场景流相对于光流具有更多的空间运动信息, 因此基于场景流运动遮挡检测方法在三维空间进行运动遮挡检测, 再将其投影回二维平面. 与基于分层模型的运动遮挡检测方法相比, 基于场景流运动遮挡检测方法通过建模图像的局部深度关系获取遮挡信息, 能实现更准确的运动遮挡检测. 由于目前大多数高精度场景流方法需要点云作为输入, 而点云信息需要 RGB-D 摄像头或者激光雷达进行采集, 这对硬件要求较高. 另外, RGB-D 摄像头与激光雷达具有局限性, 例如, RGB-D 摄像头无法采集较远处的深度信息, 激光雷达则需要与 RGB 摄像头进行配准, 这些软硬件上的缺点限制了基于场景流运动遮挡检测方法的大范围推广应用.

为建立一种更为普适的运动遮挡检测方法, 一些研究提出可以通过对参考帧图像建立规则化网格化分层模型来实现运动遮挡检测. 相比于分层模型, 通过基于规则化网格<sup>[20-22]</sup>对参考帧进行划分可以获

得更精细的局部分割细节. 该类方法对网格顶点的亮度变化进行对比, 并采用阈值判断策略确定运动遮挡区域. 因此基于规则化网格的运动遮挡检测方法不仅具有更好的通用性, 也具有更高的精度. 但是精细化网格划分计算量较大, 消耗的时间更长, 无法在实时性和运动遮挡检测精度之间取得平衡.

依靠先验假设和判断策略的传统遮挡检测方法在计算速度和精度上都无法达到令人满意的水平, 随着计算机硬件性能的提升和卷积神经网络在计算机视觉领域的大规模应用, 基于深度学习的运动遮挡检测模型成为研究热点. Yu 等<sup>[23]</sup>率先提出一种基于亮度守恒和运动平滑假设的无监督遮挡检测方法, 该方法首先将卷积神经网络和光流一致性检测结合起来, 通过卷积神经网络计算光流, 大大减少了完成一次运动遮挡检测所需要的时间. 该方法虽然结合了卷积神经网络, 但其运动遮挡检测方法仍然是建立在传统方法中光流前后向一致性假设的基础上. 考虑到遮挡是运动的结果, 运动遮挡可以反哺光流的计算, Unflow<sup>[24]</sup>方法首次建立了光流-遮挡联合估计方法, 该方法使用无监督光流模型计算前向光流和后向光流, 采取传统前后一致性检测方法进行运动遮挡检测, 并在损失函数中同时对光流与遮挡检测进行约束, 实现了一个无监督的快速的光流-运动遮挡联合估计模型. 相比于 Yu 等<sup>[23]</sup>的方法, Unflow 方法将运动遮挡结合到光流估计方法中并通过损失函数加以约束, 实现了更准确的运动遮挡检测. Back2Future<sup>[25]</sup>方法实现了一个无监督全卷积光流-遮挡联合估计模型, 该模型中的运动遮挡检测部分不再依靠传统的前后向光流一致性假设, 而是采用卷积神经网络独立检测运动遮挡. 与 Unflow 方法相似, 该方法通过损失函数同时对光流和运动遮挡进行约束, 获得运动遮挡区域信息.

为弥补卷积神经网络训练样本较少的问题, IRR-PWC<sup>[26]</sup>方法提出具有运动遮挡真实值的合成数据集 FlyingChairsOCC, 引入运动遮挡真实值对模型进行监督学习, 并利用卷积神经网络模型提取运动遮挡特征估计运动遮挡. 相比于传统方法和基于卷积神经网络的无监督方法, IRR-PWC 方法显著提升了运动遮挡检测的精度. 在最新的研究中, MaskFlowNet<sup>[27]</sup>方法提出一种非对称特征匹配模块, 无需额外数据即可让模型学习到检测运动遮挡, 但是相比有监督方法, 精度仍显不足.

尽管现有的基于卷积神经网络运动遮挡检测方法在刚性运动场景下已经取得了较好的效果, 但当图像序列中包含非刚性运动和大位移等困难运动场景时, 运动遮挡检测的准确性和鲁棒性仍有待提升. 针对运动遮挡检测的准确性和鲁棒性问题, 本

文提出基于光流和多尺度上下文的运动遮挡检测模型. 首先设计多尺度上下文信息聚合网络; 然后采用特征金字塔将运动遮挡检测模型与光流估计模型相结合, 构建端到端的运动遮挡检测模型, 通过聚合多尺度上下文信息, 提升非刚性运动和大位移场景的遮挡检测准确性与鲁棒性; 最后, 使用本文提出的边缘损失函数训练模型, 进一步提升物体和运动边界区域的运动遮挡检测精度. 实验结果表明, 相比于传统方法、无监督方法和基于上下文的 IRR-PWC 方法, 本文方法能够有效提高运动遮挡检测的精度, 尤其针对非刚性运动和大位移等困难运动场景具有更好的鲁棒性.

## 1 基于光流与多尺度上下文的运动遮挡检测

### 1.1 多尺度上下文信息聚合网络

现有的运动遮挡检测方法通常耦合光流估计模型, 在模型计算过程中利用高精度光流优化运动遮挡检测结果. 但由于大位移和非刚性运动往往不符合光流估计中的基本假设, 使得多数模型设计的前提被破坏, 导致已有的方法针对非刚性运动和大位移场景遮挡检测准确性较差. 为提高运动遮挡检测的准确性和鲁棒性, 尤其在非刚性运动与大位移边界区域的可靠性, 本文从图像多尺度上下文信息入手, 设计基于扩张卷积的多尺度上下文信息聚合网络. 该网络的主要目的是聚合多尺度上下文信息, 以获取更大范围的图像特征信息, 令感受野覆盖不同尺寸的目标, 从而使模型能更好地对由大位移和非刚性运动导致的运动边界遮挡进行检测.

虽然扩张卷积已被多种基于卷积神经网络的视觉任务<sup>[28-29]</sup>证明其对于扩大模型感受野是十分有效的. 但是扩张卷积在扩大模型感受野的同时常常会使得模型在细节检测上出现更多平滑. 如图 1 所示, IRR-PWC 方法所采用的上下文网络模型是一种由具有不同扩张率的多个  $3 \times 3$  卷积所组成的串联结构. 由于该串联结构在不同尺度上的感受野是固定的, 无法对图像不同尺寸的目标都具有较好的检测效果, 从而导致 IRR-PWC 方法在大面积背景区域的运动遮挡检测效果较好, 但针对大位移和非刚性运动边界区域的检测效果仍有待提高.

近年来, 设计具备更大感受野的卷积层成为解决多尺度特征提取问题的一种常见方法. 如图 2 所示, 在图像分类任务中, GoogLeNet<sup>[30]</sup>模型在特征编码网络中平行设置了多个不同尺寸的卷积核 ( $1 \times 1$ 、 $3 \times 3$ 、 $5 \times 5$  等), 使感受野能够覆盖不同尺

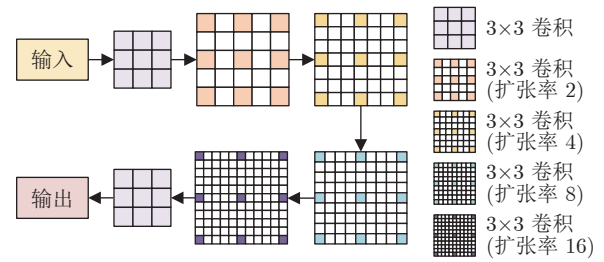


图 1 上下文网络结构示意图

Fig.1 Structure diagram of context network

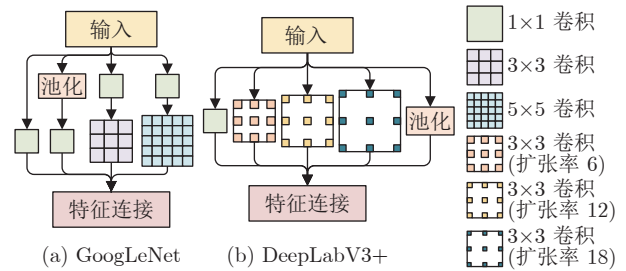


图 2 常见的感受野扩张网络结构示意图

Fig.2 Structure diagram of common receptive field expansion

寸的目标, 进而提取到更加丰富的特征信息, 显著提高了图像分类的准确性. 针对图像分割任务, DeepLabV3+<sup>[31]</sup>模型选择在特征提取网络中采用固定卷积核尺寸, 平行设置多个不同扩张率的  $3 \times 3$  卷积核覆盖不同尺寸的目标, 从而提升了不同大小目标的分割精度.

鉴于大位移和非刚性运动会导致连续帧图像间不同区域像素点运动的不连续性, 因此, 获取更大范围内不同尺度的图像特征信息, 令模型感受野能够覆盖不同尺寸的遮挡区域, 有望显著提高模型在大位移和非刚性运动边界区域的像素点遮挡检测性能<sup>[32]</sup>. 因此, 为解决传统上下文模型针对大位移和非刚性运动区域遮挡检测的不足, 本文受到 GoogLeNet 和 DeepLabV3+ 启发, 设计了一种并联形式的多尺度上下文模型, 该模型主要由 4 条针对不同尺度目标的并行扩张卷积分支组成, 在共享输入的前提下可以在 4 条分支上进行不同尺度的检测. 最终 4 条分支的信息进行汇总, 从而得到多尺度上下文信息. 如图 3 所示, 多尺度上下文信息聚合网络具有 4 条平行分支, 每条分支包含 3 个具有不同扩张率的  $3 \times 3$  卷积层, 对应的扩张率组合分别为 (1, 2, 1)、(2, 4, 1)、(4, 8, 1) 以及 (8, 16, 1), 所有分支均输出 64 通道的特征图以进行最终的信息聚合. 通过叠加每个分支的输出, 模型最终通过数个普通的卷积层聚合所有分支的上下文信息, 生成运动遮





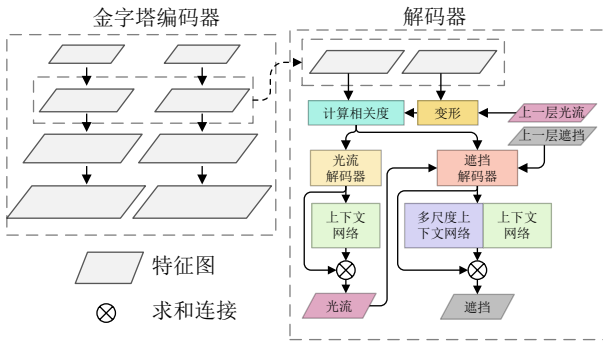


图 5 基于光流和多尺度上下文信息的遮挡检测模型结构  
Fig.5 The structure of the occlusion detection model based on optical flow and multiscale context information

模型的网络结构. 从图 5 中可以看到, 输入的连续帧图像由下采样因子为 0.5 的特征金字塔提取特征图, 并从分辨率最小的金字塔顶端开始计算光流与运动遮挡. 在金字塔的每一层, 首先计算第 1 帧图像特征图和第 2 帧经上一层光流变形后的特征图的相关度, 通过光流解码器生成该层的光流, 并使用上下文网络进行优化. 然后将特征图的相关度、光流以及上一层的运动遮挡检测结果作为输入, 通过运动遮挡检测网络计算遮挡区域. 该层光流和运动遮挡结果经过上采样后作为初始光流和运动遮挡送入金字塔下一层, 以进行下一层的光流估计和运动遮挡检测. 经过对应特征金字塔层数的迭代次数后, 模型输出一个原分辨率 1/4 大小的运动遮挡检测结果, 通过插值运算进行上采样, 将运动遮挡检测结果的分辨率恢复为输入图像的原分辨率.

图 6 展示了本文方法和 IRR-PWC 方法针对 MPI-Sintel 数据集 market\_5 图像序列的运动遮挡检测对比结果, 图中人的手臂和脚在两帧图像中发

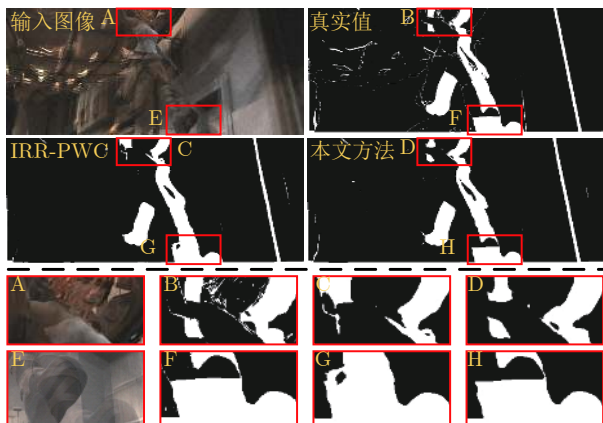


图 6 本文方法和 IRR-PWC 方法遮挡检测结果对比  
Fig.6 Comparison of occlusion detection results between our method and IRR-PWC method

生了明显的非刚性大位移运动. 从图 6 中对比结果可以看出, 本文方法与 IRR-PWC 方法相比, 能够更准确地对运动遮挡进行检测, 在大位移和非刚性运动区域具有更好的遮挡检测鲁棒性, 尤其在运动边界区域具有更好的遮挡检测精度与效果.

## 2 运动边缘损失函数

现有的运动遮挡检测方法在运动边界区域常常出现错误, 其主要原因是光流在运动边界区域存在模糊现象.

在现有的基于卷积神经网络的运动遮挡检测模型中, 由于光流隐式地参与了运动遮挡检测的过程, 因此光流的精度将直接影响到运动遮挡检测的准确性. 基于卷积神经网络的光流估计方法, 通常采用端点误差作为损失函数, 评价光流估计值和真实值之间的差异程度, 其计算方式为

$$L_{\text{flow\_epe}} = \sum_{x=0}^W \sum_{y=0}^H |\mathbf{Flow}_{\text{out}}(x, y) - \mathbf{Flow}_{\text{gt}}(x, y)|_F \quad (5)$$

式中,  $\mathbf{Flow}_{\text{out}}(x, y)$  和  $\mathbf{Flow}_{\text{gt}}(x, y)$  代表光流估计值和真实值在像素点坐标  $(x, y)^T$  处的值,  $|\cdot|_F$  代表  $L_2$  范数正则化运算. 这种经典的基于端点误差的损失函数, 虽然能在整体上监督光流估计网络的训练, 达到较好的平均精度, 但是容易忽略局部区域的光流性能. 因此, 仅使用基于端点误差的损失函数进行监督训练, 往往会使光流在物体和运动边界区域出现模糊现象.

在图像处理领域中, 浮雕算法能突出图像变化, 淡化图像中相同的颜色部分, 使图像表现出纵深感. 其计算方式可以视为图像与一个固定卷积核的卷积过程:

$$I_R = I * K, \quad K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (6)$$

式中,  $I$  是原始图像,  $K$  是卷积核,  $*$  是卷积运算符,  $I_R$  是  $I$  的浮雕图.

为了突出不同运动物体之间光流的不连续性, 锐化光流场边缘以及增强同一运动平面上像素点运动的平滑性, 以此促使模型在训练过程中更关注运动边界区域, 预测锐利的光流边缘, 为运动边界区域的遮挡检测提供更可靠的光流信息. 本文受到浮雕算法的启发, 提出一种运动边缘损失函数, 通过计算光流中一个像素点和其相邻右下角像素点光流的差值, 并与采用相同计算方式计算的真实值进行对比. 本文提出的运动边缘损失函数定义为

$$L_{\text{edge}} = \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} \left| (\mathbf{Flow}_{\text{gt}}(x, y) - \mathbf{Flow}_{\text{gt}}(x+1, y+1)) - (\mathbf{Flow}_{\text{out}}(x, y) - \mathbf{Flow}_{\text{out}}(x+1, y+1)) \right|_{\text{F}} \quad (7)$$

因为不同运动物体之间的光流通常具有很大的差异, 本文提出的边缘损失函数能够在物体和运动边界区域检验光流估计值和真实值之间的差异, 以此监督模型训练, 提高边界区域精度, 使运动遮挡检测准确性更高.

此外, 本文使用二分类交叉熵作为计算运动遮挡检测估计值和真实值之间差异的损失函数, 其计算方式为

$$L_{\text{bce}} = \sum_{x=0}^W \sum_{y=0}^H \left[ \text{Occ}_{\text{gt}} \cdot \log \text{Occ}_{\text{pred}} + (1 - \text{Occ}_{\text{gt}}) \cdot \log(1 - \text{Occ}_{\text{pred}}) \right] \quad (8)$$

式中,  $\text{Occ}_{\text{pred}}$  和  $\text{Occ}_{\text{gt}}$  分别代表遮挡检测预测值和遮挡真实值.

最后, 本文将基于端点误差的损失函数、边缘损失函数以及二分类交叉熵损失函数组合为多尺度损失函数, 以此对本文提出的基于光流和多尺度上下文的运动遮挡检测模型进行监督训练. 多尺度损失函数的伪代码计算方式如算法 1 所示.

### 算法 1. 计算多尺度损失函数

**输入.**  $\mathbf{Flo}_{\text{gt}}$  为光流真实值,  $\mathbf{Flo}_{\text{out}}$  为光流预测值,  $\text{Occ}_{\text{gt}}$  为遮挡检测真实值,  $\text{Occ}_{\text{out}}$  为遮挡检测预测值.

**输出.** 多尺度损失函数计算值.

1. **function** Loss( $\mathbf{Flo}_{\text{gt}}$ ,  $\mathbf{Flo}_{\text{out}}$ ,  $\text{Occ}_{\text{gt}}$ ,  $\text{Occ}_{\text{out}}$ )
2.  $epe \leftarrow L_{\text{epe}}(\mathbf{Flo}_{\text{gt}}, \mathbf{Flo}_{\text{out}})$
3.  $edge \leftarrow L_{\text{edge}}(\mathbf{Flo}_{\text{gt}}, \mathbf{Flo}_{\text{out}})$
4.  $bce \leftarrow L_{\text{bce}}(\text{Occ}_{\text{gt}}, \text{Occ}_{\text{out}})$
5. **if**  $epe > bce$
6. **then**  $w1 \leftarrow 1$ ,  $w2 \leftarrow epe/bce$
7. **else**  $w1 \leftarrow bce/epe$ ,  $w2 \leftarrow 1$
8. **if**  $edge > epe$
9. **then**  $w3 \leftarrow 0$
10. **else**  $w3 \leftarrow epe/edge$
11. **return**  $w1 \times epe + w2 \times bce + w3 \times edge$

算法 1 中,  $L_{\text{epe}}$  表示式 (5) 中的端点误差损失函数  $L_{\text{flow\_epe}}$ ,  $L_{\text{edge}}$  表示式 (7) 中的边缘损失函数  $L_{\text{edge}}$ ,  $L_{\text{bce}}$  表示式 (8) 中二分类交叉熵损失

函数  $L_{\text{bce}}$ ,  $w1$ ,  $w2$ ,  $w3$  分别表示不同损失函数的权重. 利用算法 1 中的多尺度损失函数算法, 对本文提出的运动遮挡检测模型进行监督学习, 可以产生边缘清晰的光流估计结果, 优化物体和运动边界区域的运动遮挡检测结果.

## 3 实验与分析

### 3.1 测量指标

本文采用 F1 分数 (F1 score)、漏检率 (Omission rate, OR) 以及误检率 (False rate, FR) 三种评价指标对本文方法和对比方法进行评价分析, 以检验运动遮挡检测结果的准确性和鲁棒性. F1 分数是分类问题上广泛使用的评价指标, 它同时兼顾准确率 (Precision) 和召回率 (Recall) 两种评价方法, 避免在样本不均匀情况下, 使用单一评价方法时会出现极端错误的问题. F1 分数的计算方式为

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (9)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (11)$$

式中,  $TP$  是正确检测的被遮挡点的数量,  $FP$  是未正确检测的被遮挡点数量,  $FN$  是误检为被遮挡点的非遮挡点个数.  $\text{precision}$  是准确率, 代表正确检测的被遮挡点数量占检测到的所有被遮挡像素点数量的比率.  $\text{recall}$  是召回率, 代表正确检测的被遮挡像素点数量占真实值中被遮挡像素点数量的比率.

漏检率和误检率的计算方式为

$$OR = \frac{TP}{N_{\text{sum}}} \times 100\% \quad (12)$$

$$FR = \frac{FP}{N_{\text{sum}}} \times 100\% \quad (13)$$

式中,  $N_{\text{sum}}$  代表图像像素点总数量.  $OR$  表示未检测到的被遮挡点数量占图像像素点总数的比率.  $FR$  表示误检测为被遮挡点的非遮挡点数量占图像像素点总数的比率.

### 3.2 对比方法

为验证本文方法遮挡检测的准确性与鲁棒性, 分别选取 Unflow<sup>[24]</sup>、Back2Future<sup>[25]</sup>、Mask-Flownet<sup>[27]</sup> 以及 IRR-PWC<sup>[26]</sup> 等传统和深度学习运动遮挡检测模型与本文方法进行综合测试分析. 其中, Unflow 方法首先采用无监督学习估计图像光流, 然后使用传统的前后一致性检测模型提取运动



遮挡区域,属于传统方法;Back2Future方法使用与Unflow方法相同的方式进行运动遮挡检测,通过输入多帧图像得到鲁棒性更好的结果,属于多帧无监督学习方法;MaskFlowNet提出一个基于变形卷积的可学习遮挡检测模块,无需对运动遮挡进行监督学习,依靠损失函数的约束直接学习运动遮挡检测方法,属于两帧无监督学习方法;IRR-PWC首先提出一个包含遮挡真实值的数据库,然后利用该数据库对遮挡检测模型进行监督学习,使得模型可以对运动遮挡区域进行推断,属于两帧有监督学习方法。

### 3.3 MPI-Sintel 图像序列实验

MPI-Sintel数据集<sup>[33]</sup>是一个由开源动画CG电影制作的合成数据集,包含大量非刚体运动与大位移等困难运动场景,且提供运动遮挡真实值。MPI-Sintel数据集包含clean和final两类子数据集,其中,clean数据集只经过简单动画渲染,而final数据集经过完整动画渲染,包含更多运动模糊、光照变化以及大气反射等。

本文首先利用MPI-Sintel测试图像集对本文方法和其他对比方法进行定量对比与分析。表1和表2分别展示了各对比方法针对MPI-Sintel数据集运动遮挡检测结果数据对比。从表中可以看出,本文方法在F1分数评价指标上取得了最优表现,在漏检率和误检率上也取得了较好的成绩。对比传统方法和无监督学习方法,监督学习方法的精度更高,漏检率与误检率也更低。Back2Future作为多帧方法,相较于两帧方法的MaskFlowNet方法F1分数更高,漏检率更低,但是误检率更高。传统方法在该数据集上表现较差,对包含非刚性运动和大位移的困难场景效果不佳。

表1 MPI-Sintel数据集平均F1分数对比结果

Table 1 Comparison of average F1 score on MPI-Sintel dataset

对比方法	多帧	类型	clean	final
Unflow <sup>[24]</sup>		传统方法	0.28	0.27
Back2Future <sup>[25]</sup>	✓	无监督学习	0.49	0.44
MaskFlowNet <sup>[27]</sup>		无监督学习	0.37	0.36
IRR-PWC <sup>[26]</sup>		监督学习	0.71	0.67
本文方法		监督学习	0.75	0.72

为进一步验证本文方法针对非刚性运动和大位移场景遮挡检测的鲁棒性,本文分别选取alley\_2、ambush\_2、market\_6和temple\_2等图像序列进行实验测试。图7展示了本文方法和其他对比方法

表2 MPI-Sintel数据集平均漏检率与误检率对比结果(%)

Table 2 Comparison of average omission rate and false rate on MPI-Sintel dataset (%)

对比方法	clean		final	
	OR	FR	OR	FR
Unflow <sup>[24]</sup>	1.96	18.32	1.94	20.51
Back2Future <sup>[25]</sup>	5.03	2.75	5.08	2.96
MaskFlowNet <sup>[27]</sup>	5.77	1.37	5.76	1.72
IRR-PWC <sup>[26]</sup>	1.98	0.96	2.84	1.29
本文方法	1.85	0.83	2.31	1.08

针对上述图像序列的运动遮挡检测效果。从图7中可以看出,Unflow方法使用前后一致性检测计算运动遮挡区域,由于采用的前后向光流是由无监督模型计算,因此导致其运动遮挡检测效果较差,Back2Future方法和MaskFlowNet方法同属于无监督深度学习方法,由于Back2Future方法是多帧方法,其获得的图像信息多于MaskFlowNet方法,因此取得了更好的运动遮挡检测效果。但是无监督的方法漏检率仍然较高,对于非刚性运动和大位移场景的运动遮挡检测效果不佳。IRR-PWC方法和本文方法都属于有监督学习方法,通过引入对运动遮挡检测的监督学习,模型对运动遮挡检测的精度和鲁棒性相较传统方法和无监督学习方法更佳。不难看出,虽然IRR-PWC方法相对于其他对比方法精度和鲁棒性得到了很大的提升,但是在包含非刚体运动和大位移的困难场景下,效果仍然不佳。而本文方法则具有更好的鲁棒性,对比IRR-PWC方法,本文方法在非刚性运动和大位移区域的性能更优。

为定量分析各方法针对非刚性运动和大位移场景下运动遮挡检测的精度,表3展示了本文方法和对比方法针对上述图像序列的F1分数对比结果。从表3中不难看出,本文方法在不同图像序列均取得了最优或次优结果,说明本文方法针对非刚体运动和大位移图像序列具有更好的遮挡检测鲁棒性。

### 3.4 KITTI 图像序列实验

KITTI数据集<sup>[34]</sup>是真实交通环境中采集的道路场景,本文进一步在KITTI数据集上测试本文方法和各对比方法在真实道路场景下运动遮挡检测的性能表现。由于真实场景无法获得运动遮挡的真实值,无法进行定量对比,因此本文仅对运动遮挡可视化结果进行了定性对比。图8展示了各个方法以及本文方法在KITTI数据集部分图像序列上的运动遮挡检测结果。

由于KITTI数据集中包含较多的大位移运动,

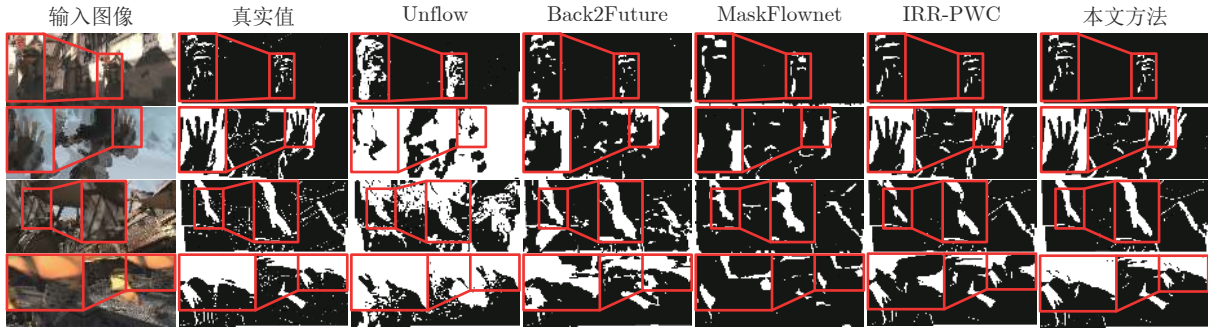


图7 MPI-Sintel 数据集非刚性运动与大位移序列遮挡检测结果对比图. 从左至右分别是: alley\_2、ambush\_2、market\_6 以及 temple\_2 序列

Fig.7 Comparison results of occlusion detection between non-rigid motion and large displacement sequences on MPI-Sintel dataset. From left to right are alley\_2, ambush\_2, market\_6, and temple\_2 sequence

表3 非刚性运动与大位移图像序列运动遮挡检测平均 F1 分数对比结果

Table 3 Comparison of average F1 scores of motion occlusion detection between non-rigid motion and large displacement image sequences

对比方法	clean				final			
	alley_2	ambush_2	market_6	temple_2	alley_2	ambush_2	market_6	temple_2
Unflow <sup>[24]</sup>	0.414 9	0.431 3	0.433 0	0.324 3	0.405 7	0.392 0	0.449 9	0.312 0
Back2Future <sup>[25]</sup>	0.681 6	0.588 8	0.629 0	0.271 2	0.675 6	0.519 9	0.623 9	0.268 3
MaskFlownet <sup>[27]</sup>	0.505 7	0.540 3	0.466 0	0.383 8	0.503 9	0.408 5	0.473 5	0.350 8
IRR-PWC <sup>[26]</sup>	0.870 9	0.917 2	0.815 5	0.740 4	0.877 0	0.780 9	0.802 3	0.690 5
本文方法	<b>0.881 1</b>	<b>0.921 6</b>	<b>0.830 4</b>	<b>0.774 7</b>	<b>0.876 4</b>	<b>0.795 9</b>	<b>0.810 6</b>	<b>0.710 3</b>

注: 加粗字体表示各列最优结果.

同一图像序列中第 1 帧图像和第 3 帧图像差异较大, 导致多帧方法 Back2Future 性能表现不佳. 同属于无监督学习的两帧方法 MaskFlownet 则展现出良好的鲁棒性.

尽管 KITTI 数据集缺失了遮挡检测的真实值, 使得 IRR-PWC 方法和本文方法无法对运动遮挡检测进行监督学习, 但是 IRR-PWC 方法和本文方法仍然表现出较高的精度和鲁棒性, 证明由合成数据集训练出的遮挡检测模型在真实场景同样具有较高的可靠性. 从图 8 可以看到, 在真实场景下, 本文方法的运动遮挡检测精度相比于 IRR-PWC 方法仍然较高, IRR-PWC 方法仍然存在局部的误检和漏检的问题, 对于路边的指示牌等小物体以及行驶至图像外的车辆的运动遮挡检测效果不佳, 本文方法则较为完整地对手车等物体进行了正确的运动遮挡检测, 在真实场景运动遮挡检测中表现出较高的准确性和良好的鲁棒性.

### 3.5 运行时间对比

本文在相同的硬件环境下 (CPU: I7-6700K、GPU: GTX 1080Ti), 在 KITTI 数据集上测试了本

文方法和对比方法完成一次运动遮挡检测所需的平均运行时间, 具体的运行时间如表 4 所示.

从表 4 中可以看出, 无监督学习方法的计算速度是最快的, Back2Future 作为多帧方法, 需要计算三帧图像的前后向光流, 因此与 MaskFlownet 方法相比增加了一些运算时间. IRR-PWC 方法和本文方法由于需要先后计算同一对图像的前向和后向光流, 因此需要花费更多时间. 本文提出的多尺度上下文网络使用了较多的扩张卷积运算, 所以计算时间相较于其他方法更长. 随着运动遮挡检测方法模型规模的增大以及复杂性的增加, 多数方法往往牺牲计算速度换取更高的精度和鲁棒性, 而本文方法在不显著增加时间消耗的基础上大幅提升了遮挡检测的精度与鲁棒性, 综合性能最优.

### 3.6 消融实验

为阐明本文方法中不同组成部分对遮挡检测精度的影响, 本文进行了独立且完整的消融实验. 首先分别从完整的模型中单独移除多尺度上下文网络和边缘损失函数, 然后使用相同的训练方法, 分别重新训练了移除改进之后的模型.



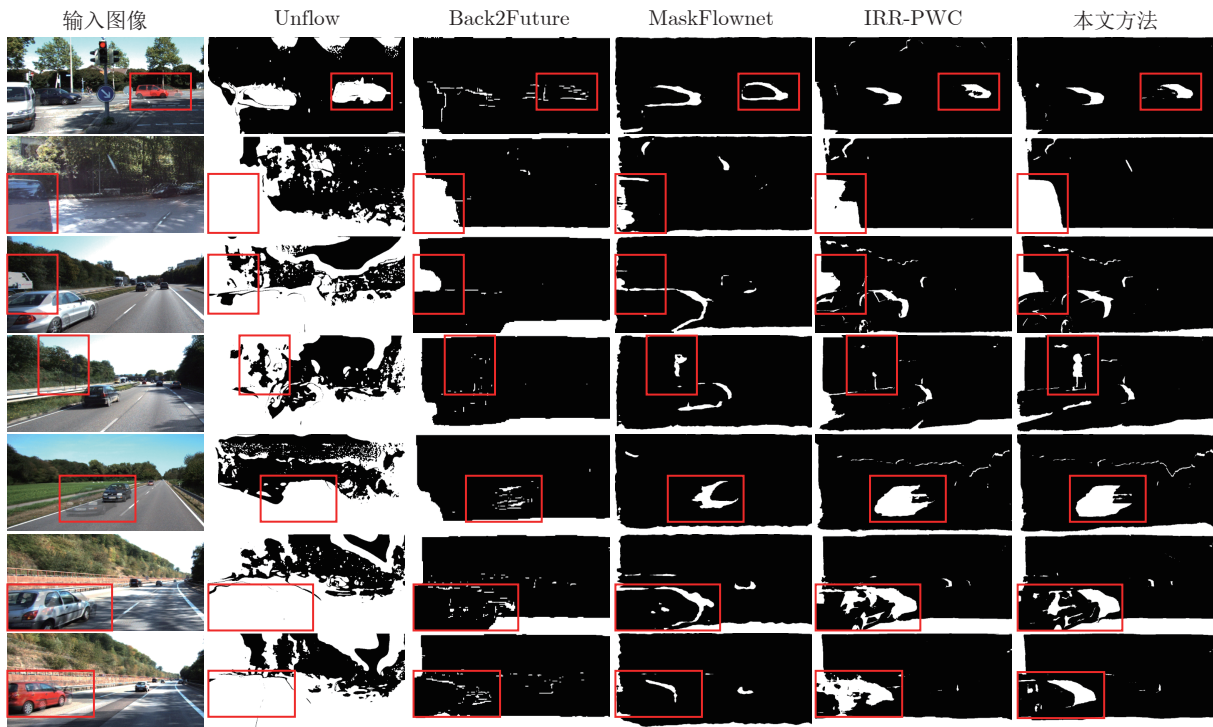


图 8 各个遮挡检测方法在 KITTI 数据集上的遮挡检测结果对比图. 从左至右分别是输入图像和 Unflow、Back2Future、MaskFlowNet、IRR-PWC 以及本文方法的运动遮挡检测图

Fig. 8 Comparison of occlusion detection results of each occlusion detection method on KITTI dataset. From left to right are the input image, Unflow, Back2Future, MaskFlowNet, IRR-PWC and our method

表 4 不同方法的时间消耗对比

Table 4 Comparison of time consumption of different methods

对比方法	多帧输入	类型	运行时间 (s)
Unflow <sup>[24]</sup>		传统方法	0.13
Back2Future <sup>[25]</sup>	√	无监督学习	0.13
MaskFlowNet <sup>[27]</sup>		无监督学习	<b>0.10</b>
IRR-PWC <sup>[26]</sup>		监督学习	0.18
本文方法		监督学习	0.19

注: 加粗字体表示评价最优值.

利用 MPI-Sintel 训练数据集对消融实验模型进行评价, 消融实验定量评价结果如表 5 所示. 消融实验结果表明, 相比于单独去除多尺度上下文网络的模型和单独去除边缘损失函数的模型, 全模型在 F1 分数评价指标取得了最好的成绩, 同时表 5 中的消融实验数据也显示了模型通过多尺度上下文网络和边缘损失函数取得了显著的遮挡检测效果提升, 其中上下文网络对模型的提升是最大的. 对比各消融模型运行时间, 由于增加了多尺度上下文网络, 因此全模型相比于去除多尺度上下文网络模型在运行时间和训练时间上需要更长时间, 虽然训练时间增加了 1 天, 但是对运行时间影响较小. 边缘

表 5 MPI-Sintel 全图像序列平均 F1 分数对比

Table 5 Comparison of average F1 scores of whole image sequence on MPI-Sintel

模型类型	MPI-Sintel 训练数据集			
	clean	final	运行时间 (s)	训练时间 (d)
全模型	<b>0.75</b>	<b>0.72</b>	0.19	13
去除多尺度上下文网络	0.72	0.68	0.18	12
去除边缘损失函数	0.74	0.71	0.19	13

注: 加粗字体表示评价最优值.

损失函数仅在训练时起作用且计算量较小, 不影响模型的运行时间和训练时间.

为验证本文所提各项改进方法在运动边界区域的性能表现, 本文利用光流真实值生成运动边界遮挡掩膜对遮挡检测精度进行定向评估. 生成运动遮挡掩膜的步骤如下: 首先, 计算垂直和水平方向的光流场梯度. 然后, 将光流梯度大于整个光流场最大梯度 5% 的像素标记为边界像素, 其他像素标记为非边界像素, 在不丢失运动细节的前提下尽可能提取更多的运动边界. 最后, 将距离边界像素  $N$  个像素距离内的像素点标记为边界像素, 生成运动遮挡掩膜图, 图 9 展示了部分运动遮挡掩膜图. 表 6 列出了利用运动遮挡掩膜在不同运动边界区域对各



图 9 利用光流真实值生成的运动遮挡掩膜部分示例图 ( $N = 3$ )

Fig.9 Examples of motion occlusion mask generated by ground truth of optical flow ( $N = 3$ )

消融模型进行定量对比结果, 表中,  $N = 1$ 、 $N = 3$ 、 $N = 5$ 、 $N = 10$  分别代表在距离运动边界 1、3、5、10 个像素点距离范围内, 遮挡检测的平均 F1 分数. 由表 6 的消融实验结果可知, 在距离运动边界较近的区域 ( $N = 1$ ,  $N = 3$ ), 去除边缘损失函数导致遮挡检测精度明显下降. 在距离运动边界较远的区域 ( $N = 5$ ,  $N = 10$ ), 去除边缘损失函数对遮挡检测精度的影响逐步变小. 此外, 当去除多尺度上下文网络后, 本文方法在不同运动边界区域的遮挡检测效果均明显下降.

为了对多尺度上下文网络和边缘损失函数在提升运动遮挡检测性能表现上进行直观的对比与分析, 图 10 展示了不同消融模型针对 MPI-Sintel 数据集 alley\_1 图像序列的遮挡检测可视化结果. 从图 10 中可以看到, 在去除多尺度上下文网络后, 前景区域的运动遮挡检测效果大幅下降. 而在去除边

缘损失函数后, 运动边界附近的遮挡检测结果缺失了较多的细节信息. 因此, 由消融实验的定量和定性对比结果可知, 本文提出的多尺度上下文网络和边缘损失函数均能够显著提升运动遮挡检测的精度与效果. 其中, 多尺度上下文网络能有效提升前景区域的运动遮挡检测精度, 边缘损失函数能进一步完善运动边界区域的遮挡检测细节.

## 4 结束语

本文提出一种基于多尺度上下文与光流相结合的运动遮挡检测模型, 通过将多尺度上下文模型和原上下文模型结合组成运动遮挡检测模型, 并通过与光流估计模型相结合能有效提升非刚性运动和大位移场景下运动遮挡检测的精度与鲁棒性. 为保证遮挡检测网络模型的训练准确性, 本文设计了一种新的边缘损失函数, 并将边缘损失函数与端点误差损失函数以及二分类交叉熵损失函数相结合组成多尺度损失函数, 使网络同时对光流和运动遮挡进行监督学习, 提高了物体和运动边界区域的运动遮挡检测的鲁棒性. 分别采用 MPI-Sintel 和 KITTI 数据集测试图像集对本文方法和 Unflow、Back2-Future、MaskFlowNet 以及 IRR-PWC 等不同类型运动遮挡检测方法进行了综合实验对比. 实验结果表明, 本文方法具有较高的运动遮挡检测精度和鲁棒性, 尤其在包含非刚性运动和大位移等复杂运动场景具有更显著的优势. 最后, 通过消融实验验证了各模块的有效性, 实验结果表明, 本文所提出的多

表 6 MPI-Sintel 全图像序列在不同运动边界区域内的平均 F1 分数对比  
Table 6 Comparison of average F1 scores of whole image sequence in different motion boundary regions on MPI-Sintel

模型类型	MPI-Sintel 训练数据集							
	clean				final			
	$N = 1$	$N = 3$	$N = 5$	$N = 10$	$N = 1$	$N = 3$	$N = 5$	$N = 10$
全模型	<b>0.63</b>	<b>0.67</b>	<b>0.69</b>	<b>0.71</b>	<b>0.59</b>	<b>0.62</b>	<b>0.64</b>	<b>0.67</b>
去除多尺度上下文网络	0.59	0.62	0.65	0.67	0.55	0.59	0.61	0.63
去除边缘损失函数	0.60	0.64	0.67	0.69	0.56	0.60	0.62	0.64

注: 加粗字体表示评价最优值.

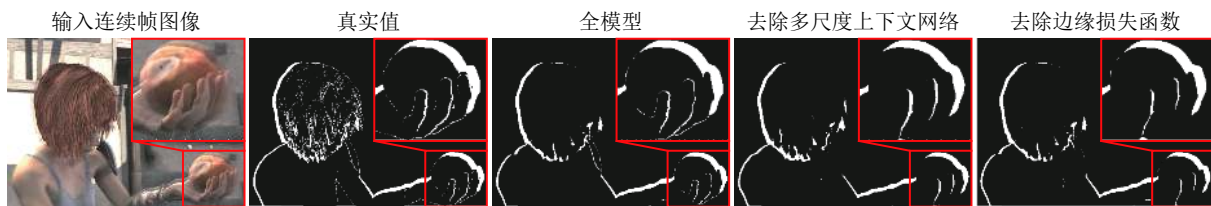


图 10 各消融模型可视化结果对比图

Fig.10 Comparison of visualization results of each ablation model

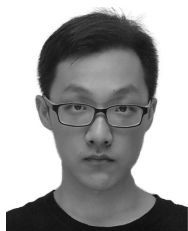
尺度上下文信息聚合模块和结合边缘损失的多尺度损失函数能有效提升模型对运动遮挡检测的精度。

## References

- Zhang Shi-Hui, He Qi, Dong Li-Jian, Du Xue-Zhe. Dynamic occlusion avoidance approach by means of occlusion region model and object motion estimation. *Acta Automatica Sinica*, 2019, **45**(4): 771–786  
(张世辉, 何琦, 董利健, 杜雪哲. 基于遮挡区域建模和目标运动估计的动态遮挡规避方法. *自动化学报*, 2019, **45**(4): 771–786)
- Yu C, Bo Y, Bo W, Yan W D, Robby T. Occlusion-aware networks for 3D human pose estimation in video. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019. 723–732
- Zhang Cong-Xuan, Chen Zhen, Xiong Fan, Li Ming, Ge Li-Yue, Chen Hao. Large displacement motion optical flow estimation with non-rigid dense patch matching. *Acta Electronica Sinica*, 2019, **47**(6): 1316–1323  
(张聪炫, 陈震, 熊帆, 黎明, 葛利跃, 陈昊. 非刚性稠密匹配大位移运动光流估计. *电子学报*, 2019, **47**(6): 1316–1323)
- Yao Nai-Ming, Guo Qing-Pei, Qiao Feng-Chun, Chen Hui, Wang Hong-An. Robust facial expression recognition with generative adversarial networks. *Acta Automatica Sinica*, 2018, **44**(5): 865–877  
(姚乃明, 郭清沛, 乔逢春, 陈辉, 王宏安. 基于生成式对抗网络的鲁棒人脸表情识别. *自动化学报*, 2018, **44**(5): 865–877)
- Pan J Y, Bo H. Robust occlusion handling in object tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Minneapolis, USA: IEEE, 2007. 1–8
- Liu Xin, Xu Hua-Rong, Hu Zhan-Yi. GPU based fast 3D-object modeling with Kinect. *Acta Automatica Sinica*, 2012, **38**(8): 1288–1297  
(刘鑫, 许华荣, 胡占义. 基于 GPU 和 Kinect 的快速物体重建. *自动化学报*, 2012, **38**(8): 1288–1297)
- Zhang Cong-Xuan, Chen Zhen, Li Ming. Review of the 3D reconstruction technology based on optical flow of monocular image sequence. *Acta Electronica Sinica*, 2016, **44**(12): 3044–3052  
(张聪炫, 陈震, 黎明. 单目图像序列光流三维重建技术研究综述. *电子学报*, 2016, **44**(12): 3044–3052)
- Bailer C, Taetz B, Stricker D. Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, **41**(8): 1879–1892
- Wolf L, Gadot D. PatchBatch: A batch augmented loss for optical flow. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 4236–4245
- Li Y S, Song R, Hu Y L. Efficient coarse-to-fine patch match for large displacement optical flow. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 5704–5712
- Menze M, Heipke C, Geiger A. Discrete optimization for optical flow. In: Proceedings of the German Conference on Pattern Recognition (GCPR). Aachen, Germany: Springer Press, 2015. 16–28
- Chen Q F, Koltun V. Full flow: Optical flow estimation by global optimization over regular grids. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 4706–4714
- Guney F, Geiger A. Deep discrete flow. In: Proceedings of the Asian Conference on Computer Vision (ACCV). Taipei, China: Springer Press, 2016. 207–224
- Hur J, Roth S. Joint optical flow and temporally consistent semantic segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands: Springer, 2016. 163–177
- Ince S, Konrad J. Occlusion-aware optical flow estimation. *IEEE Transactions on Image Process*, 2008, **17**(8): 1443–1451
- Sun D Q, Liu C, Pfister H. Local layering for joint motion estimation and occlusion detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, USA: IEEE, 2014. 1098–1105
- Sun D Q, Sudderth E B, Black M J. Layered image motion with explicit occlusions, temporal consistency, and depth ordering. In: Proceedings of the 24th International Conference on Neural Information Processing Systems (NIPS). Vancouver, Canada: Curran Associates Inc., 2010. 2226–2234
- Vogel C, Roth S, Schindler K. View-consistent 3D scene flow estimation over multiple frames. In: Proceedings of the European Conference on Computer Vision (ECCV). Zurich, Switzerland: Springer Press, 2014. 263–278
- Zanfir A, Sminchisescu C. Large displacement 3D scene flow with occlusion reasoning. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 4417–4425
- Zhang C X, Chen Z, Wang M R, Li M, Jiang S F. Robust non-local TV-L<sup>1</sup> optical flow estimation with occlusion detection. *IEEE Transactions on Image Process*, 2017, **26**(8): 4055–4067
- Zhang Cong-Xuan, Chen Zhen, Wang Ming-Run, Li Ming, Jiang Shao-Feng. Motion occlusion detecting from image sequence based on optical flow and Delaunay triangulation. *Acta Electronica Sinica*, 2018, **46**(2): 479–485  
(张聪炫, 陈震, 汪明润, 黎明, 江少锋. 基于光流与 Delaunay 三角网格的图像序列运动遮挡检测. *电子学报*, 2018, **46**(2): 479–485)
- Kennedy R, Taylor C J. Optical flow with geometric occlusion estimation and fusion of multiple frames. In: Proceedings of International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR). Hong Kong, China: IEEE, 2015. 364–377
- Yu J J, Harley A W, Derpanis K G. Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness. In: Proceedings of the European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands: Springer, 2016. 3–10
- Meister S, Hur J, Roth S. UnFlow: Unsupervised learning of optical flow with a bidirectional census loss. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI). San Francisco, USA: AAAI, 2017. 7251–7259
- Janai J, Güney F, Ranjan A, Black M, Geiger A. Unsupervised learning of multi-frame optical flow with occlusions. In: Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 713–731
- Hur J, Roth S. Iterative residual refinement for joint optical flow and occlusion estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 5747–5756
- Zhao S Y, Sheng Y L, Dong Y, Chang E I C, Xu Y. Mask-Flownet: Asymmetric feature matching with learnable occlusion mask. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Virtual Event: IEEE, 2020. 6277–6286
- Yang M K, Yu K, Zhang C, Li Z W, Yang K Y. DenseASPP for semantic segmentation in street scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE, 2018. 3684–3692
- Mehta S, Rastegari M, Caspi A, Shapiro L, Hajishirzi H. ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 561–580
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 1–9



- 31 Chen L C, Zhu Y K, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 833–851
- 32 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [Online], available: <https://arxiv.org/abs/1511.07122>, Apr 30, 2016
- 33 Butler D J, Wulff J, Stanley G B, Black M J. A naturalistic open source movie for optical flow evaluation. In: Proceedings of the European Conference on Computer Vision (ECCV). Florence, Italy: Springer, 2012. 611–625
- 34 Menze M, Geiger A. Object scene flow for autonomous vehicles. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 2061–3070

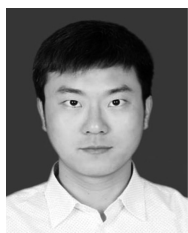


**冯 诚** 南昌航空大学测试与光电工程学院硕士研究生。主要研究方向为计算机视觉。

E-mail: fengcheng00016@163.com

**(FENG Cheng** Master student at the School of Measuring and Optical Engineering, Nanchang Hangkong

University. His main research interest is computer vision.)



**张聪炫** 南昌航空大学测试与光电工程学院教授。2014 年获得南京航空航天大学博士学位。主要研究方向为图像处理与计算机视觉。本文通信作者。

E-mail: zcxdsg@163.com

**(ZHANG Cong-Xuan** Professor at the School of Measuring and Optical

Engineering, Nanchang Hangkong University. He received his Ph.D. degree from Nanjing University of Aeronautics and Astronautics in 2014. His research interest covers image processing and computer vision. Corresponding author of this paper.)



**陈 震** 南昌航空大学测试与光电工程学院教授。2003 年获得西北工业大学博士学位。主要研究方向为图像处理与计算机视觉。

E-mail: dr\_chenzhen@163.com

**(CHEN Zhen** Professor at the School of Measuring and Optical

Engineering, Nanchang Hangkong University. He received his Ph.D. degree from Northwestern Polytechnical University in 2003. His research interest covers image processing and computer vision.)



**李 兵** 中国科学院自动化研究所模式识别国家重点实验室研究员。2009 年获得北京交通大学博士学位。主要研究方向为视频内容理解, 多媒体内容安全。E-mail: bli@nlpr.ia.ac.cn

**(LI Bing** Professor at the National Key Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of

Sciences. He received his Ph.D. degree from Beijing Jiaotong University in 2009. His research interest covers video understanding and multimedia content security.)



**黎 明** 南昌航空大学信息工程学院教授。1997 年获得南京航空航天大学博士学位。主要研究方向为图像处理, 人工智能。

E-mail: liming@nchu.edu.com

**(LI Ming** Professor at the School of Information Engineering, Nanchang

Hangkong University. He received his Ph.D. degree from Nanjing University of Aeronautics and Astronautics in 1997. His research interest covers image processing and artificial intelligence.)