

深度学习光流计算技术研究进展

张聪炫^{1,2}, 周仲凯¹, 陈震¹, 葛利跃¹, 黎明¹, 江少锋¹, 陈昊¹

(1. 南昌航空大学无损检测技术教育部重点实验室, 江西南昌 330063; 2. 中国科学院自动化研究所, 北京 100190)

摘要: 图像序列光流计算是图像处理与计算机视觉等领域的重要研究方向. 随着深度学习技术的快速发展, 以卷积神经网络为代表的深度学习理论与方法成为光流计算技术研究的热点. 本文主要对深度学习光流计算技术研究进行综述, 首先介绍了有监督学习、无监督学习和半监督学习的光流计算网络模型与训练策略, 然后重点阐述并分析了不同网络模型优化方法. 针对光流计算模型的评估问题, 分别介绍了 Middlebury、MPI-Sintel 和 KITTI 等数据库及评价基准, 并对不同类型深度学习和传统变分光流模型进行对比与分析. 最后, 总结了深度学习光流计算技术在模型复杂度与泛化性、光流估计鲁棒性、小样本训练准确性等方面的关键技术问题, 并指出了可能的解决方案与研究思路.

关键词: 光流计算; 深度学习; 卷积神经网络; 训练策略; 优化方法; 评价基准

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2020)09-1841-09

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.09.023

Research Progress of Deep Learning Based Optical Flow Computation Technology

ZHANG Cong-xuan^{1,2}, ZHOU Zhong-kai¹, CHEN Zhen¹, GE Li-yue¹, LI Ming¹, JIANG Shao-feng¹, CHEN Hao¹

(1. Key Laboratory of Nondestructive Testing, Ministry of Education, Nanchang Hangkong University, Nanchang, Jiangxi 330063, China;

2. Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Optical flow computation is an important research direction in image processing and computer vision. With the rapid development of the deep learning technology, the convolutional neural network based deep learning theories and methodologies have been the research focus of optical flow computation. This article mainly reviews the research progress of the deep learning based optical flow estimation technologies. First, the typical models and training strategies of the optical flow computing networks with supervised learning, unsupervised learning and semi-supervised learning are introduced. Second, the optimization methods of various network models are described and analyzed. Third, the evaluation benchmarks of Middlebury, MPI-Sintel and KITTI databases are summarized, and the experimental comparison results and analysis between the different deep-learning and variational optical flow methods are conducted. Finally, we discuss some issues of the deep learning based optical flow computation technology including the model complexity and generalization, the robustness of optical flow estimation and the accuracy of the small sample training. Afterwards, we point out several possible solutions and research ideas to address the above mentioned issues.

Key words: optical flow computation; deep learning; convolutional neural network; training strategy; optimization method; evaluation benchmark

1 引言

1981年, Horn 和 Schunck^[1]首次提出光流基本守恒假设和计算模型后, 光流计算及相关技术研究受到

广泛关注. 根据计算手段的不同, 早期的光流计算方法可大致分为以下四类^[2]: 基于变分模型的方法^[3]; 基于局部匹配的方法^[4]; 基于能量的方法^[5]以及基于相位的方法^[6].

收稿日期: 2019-08-26; 修回日期: 2019-12-23; 责任编辑: 梅志强

基金项目: 国家自然科学基金(No. 61866026, No. 61772255, No. 61866025); 江西省优势科技创新团队计划(No. 20165BCB19007); 江西省科技创新杰出青年人才计划(No. 20192BCB23011); 航空科学基金(No. 2018ZC56008); 中国博士后科学基金(No. 2019M650894); 江西省重点研发计划(No. 20171BBG70052); 江西省研究生创新专项资金项目资助(No. YC2018049)

相对于其他传统光流计算方法,变分方法在模型可延伸性和光流稠密性等方面具有显著优势,因此,变分方法长期占据图像序列光流计算技术研究的主导地位.虽然变分方法在光流计算的精度和鲁棒性等方面取得了巨大提升^[7-9].但是,该类方法针对大位移、非刚性运动、运动遮挡等困难场景的光流估计效果仍有待提高.同时,由于变分方法在优化过程中需要大量迭代运算,导致其时间消耗过大,限制了变分光流计算技术在实时视觉任务中的应用.

近年来,随着深度学习理论与技术的快速发展,卷积神经网络被广泛应用于光流计算技术研究,由于该类方法具有计算速度快、稳定性高等显著优点,因此成为光流计算研究领域的热点^[10-13],研究成果被广泛应用于无人机实时导航与避障^[14]、车辆前景检测与辅助驾驶^[15]、运动目标跟踪与识别^[16]以及视频压缩与传输^[17]等对计算实时性要求较高的领域.

如图1所示,深度学习光流计算方法可分为有监督

学习光流估计方法、无监督学习光流估计方法以及半监督学习光流估计方法.有监督学习光流计算模型首先利用卷积神经网络在多尺度卷积空间提取图像特征,然后根据图像特征建立相邻图像像素点的对应关系,最后根据像素对应关系计算稠密光流场.虽然有监督学习模型能够获取较高精度的光流估计结果,但该类方法需要大量标签数据训练模型且模型训练过程复杂,导致学习时间消耗过大,难以应用于不包含真实光流数据的现实场景.无监督学习光流计算模型使用辅助光流代替真实光流或利用不依赖于真实光流的损失函数进行网络参数训练,能够克服标签数据对模型的限制.但是由于模型训练并不准确,导致光流估计精度较低.半监督学习光流计算模型在综合有监督与无监督学习方法的基础上,联合标签数据和真实数据进行模型训练,能够有效克服网络模型训练对标签数据的依赖性.然而由于模型与训练数据的局限性,导致光流估计精度仍低于有监督学习方法.

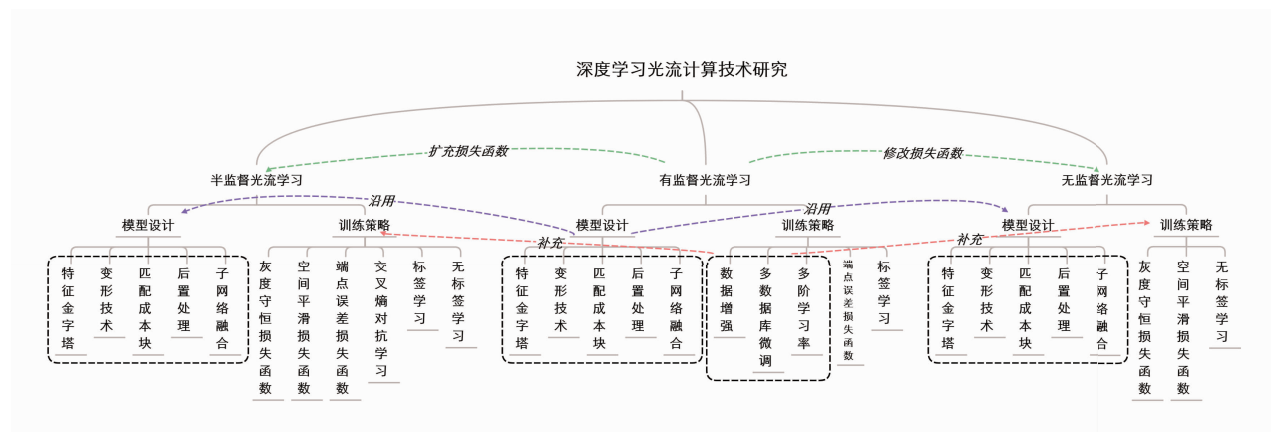


图1 深度学习光流计算技术研究体系与发展脉络图

本文首先对不同学习策略的深度学习光流计算技术进行介绍,重点分析了不同光流计算网络模型及训练策略的进展与代表工作;然后详细论述了深度学习光流计算技术中的网络模型优化方法;分别介绍了通用的光流测试数据库及评价基准,并对深度学习和变分光流模型进行对比与分析;最后,对深度学习光流计算技术现存的关键技术问题及后续研究方向进行详细分析.

2 卷积神经网络光流模型与训练策略

相对于变分光流方法,卷积神经网络光流模型不依赖于图像数据守恒假设,因此对复杂场景图像序列具有更好的适用性和延展性.此外,虽然网络模型的训练过程需要较长时间,但是当模型参数训练完成后便可对输入图像序列进行光流实时预测,计算效率相比变分方法具有显著优势.

2.1 有监督学习光流网络模型与训练策略

2.1.1 有监督学习光流网络模型

Dosovitskiy^[18]等人首先使用卷积神经网络搭建了基于有监督学习的光流估计模型 FlowNet,该工作首次证明了利用通用 U-Net CNN^[19]架构直接估计图像序列光流的可行性.

针对 FlowNet 模型光流估计精度较低的问题,Ilg^[20]等人提出的 FlowNet2.0 模型首先通过堆叠 FlowNet 模型加大网络的深度,然后针对性估计小位移运动光流,最后将堆叠网络与小位移光流融合获得光流预测结果.虽然 FlowNet2.0 模型的光流估计精度超越了传统的变分光流方法,但是由于 FlowNet2.0 模型是由 FlowNet 网络堆叠而来,因此面临结构复杂和网络训练难度剧增等问题.Ranjan 和 Black^[21]将经典的空间金字塔模型与卷积神经网络相结合,提出 SpyNet 网络模型,显著减小了模型的尺寸和参数量.但是,由于 SpyNet 网

络结构过于简单,导致光流估计精度较低. Sun^[22] 等人提出基于紧凑型卷积神经网络的光流估计模型 PWC-Net,首先使用特征金字塔从原始图像序列提取置信度较高的特征图作为输入,然后通过引入变形层和成本层减小模型尺寸和参数量,最后采用多层扩张型卷积网络估计光流,在减少模型参数量和消耗的同时显著提高了光流预测的精度.

针对运动遮挡光流估计的可靠性问题,文献[23]通过匹配前向与后向光流提取运动遮挡区域,然后利用遮挡信息修正光流预测结果,有效提高了运动遮挡光流估计的鲁棒性. 针对卷积操作易导致运动边缘过度平滑的问题, Hui^[24] 等人提出基于正则化约束的 LiteFlowNet 光流模型,首先将金字塔特征提取网络与光流估计网络分开处理以减小网络模型的尺寸,然后通过引入正则化约束项保护光流的边缘特征信息. 鉴于图像光流与目标分割存在共同性^[25], Cheng^[26] 等人提出基于双向结构的 SegFlow 网络模型,通过在光流估计网络中引入目标分割信息,并利用交替迭代进行网络训练,能够显著提高光流估计精度与鲁棒性.

2.1.2 有监督学习模型训练策略

现阶段, FlowNet^[18]、PWC-Net^[22] 等有监督光流模型通常采用改进梯度下降策略^[27] 进行端到端训练,首先采用短学习或长学习策略对模型在 FlyingChairs 数据库进行初始训练以获取预测光流的常规特征;然后采用微调学习策略对模型在 FlyingThings3D 数据库进行微调训练提高模型对旋转、缩放等运动的适用性与鲁棒性;最后采用微调学习或两段型微调学习策略对模型在 MPI-Sintel 和 KITTI 等评测数据库再次进行微调训练,以提升网络模型的光流预测精度与鲁棒性. 此外,文献[18]通过数据增强策略降低网络模型训练的过拟合风险,数据增强方式如下表 1 所示.

表 1 光流网络模型训练数据增强方法

增强方法	具体增强方式
旋转变换	对训练集施加 $[-17^\circ, 17^\circ]$ 的旋转
平移变换	对训练集的长宽施加 $[-20\%, 20\%]$ 的平移
缩放变换	对训练集施加 $[0.9, 2.0]$ 的缩放
对比度增强	对训练集施加 $[-0.8, 0.4]$ 的对比度增强
高斯噪声	对训练集施加方差为 $[0, 0.04]$ 的高斯噪声
RGB 增强	对训练集的 RGB 通道施加乘数因子为 $[0.5, 2]$ 的通道增强
gamma 增强	对训练集施加 gamma 值取 $[0.7, 1.5]$ 的非线性变换
亮度增强	对训练集施加方差为 0.2 的高斯噪声颜色增强

2.2 无监督学习光流网络模型与训练策略

2.2.1 无监督学习光流网络模型

由于无监督学习光流网络无需包含真实光流的图像序列作为训练样本,因此可通过设计类似于变分光流能量泛函的损失函数作为网络收敛的判断条件. Jason^[28] 等人以有监督学习模型为基准首先提出无监督光流网络模型 UnsupFlowNet,该模型通过联合数据项与平滑项作为网络损失函数,使光流模型不依赖包含真实值的标签数据便可进行网络训练.

针对运动遮挡光流估计的准确性问题, Meister^[29] 等人提出的 Unflow 模型首先使用前景和背景双向光流的一致性匹配进行遮挡检测,然后设计基于遮挡感知的损失函数提高了运动遮挡场景下无监督学习光流估计模型的精度. Wang^[30] 等人首先使用背景光流进行前向变形得到遮挡映射信息,然后在损失函数中加入遮挡信息和边缘感知平滑约束,在提高光流估计精度的同时有效保护了光流的边缘结构特征. 针对长序列遮挡光流估计的可靠性问题, Janai^[31] 提出无监督多帧遮挡感知的光流估计模型 Back2FutureFlow,该模型损失函数由数据项、平滑项、恒定速度约束项和遮挡先验项构成,提高了长序列遮挡光流估计的鲁棒性. 近年来,随着动态视觉传感器的快速发展,其成像质量不受外部环境变化与相互运动的影响^[32],因此基于动态图像数据的损失函数能够使无监督光流网络模型更充分学习运动的一般特性^[33,34],有效提高模型的光流预测精度与可靠性.

2.2.2 无监督学习模型训练策略

相对于有监督学习光流估计模型,无监督学习网络的优势在于无需标签数据引导模型的最小化损失函数项,因此可直接利用真实场景训练样本进行网络训练. 在典型的无监督学习模型 Unflow^[29] 中,其网络结构与有监督学习模型类似. 不同之处在于:无监督学习网络通常使用灰度守恒损失函数、空间平滑损失函数或其他损失函数代替有监督学习模型中的端点误差损失函数. 此外,为增强无监督学习模型在真实场景下光流预测的鲁棒性,可采用真实场景图像序列对无监督光流模型进行额外微调,这是有监督学习模型所不具备的优势^[28].

2.3 基于半监督学习的光流网络模型与训练策略

2.3.1 半监督学习光流网络模型

为了充分利用无监督学习不依赖标签数据和有监督学习预测精度高的优势,结合无监督与有监督学习的半监督学习网络模型逐渐成为光流估计研究的热点^[35].

典型的半监督学习光流模型是利用标签数据有监督训练光流计算模型参数,并利用无标签数据无监督

地对模型进行微调^[28]. 针对无监督学习光流估计的适用性问题, 文献[36]提出引导光流学习模型, 通过使用变分光流作为标签数据引导网络模型的光流预测. 虽然该方法能够摆脱标签数据集的限制, 但由于变分能量泛函获取的光流标签数据存在较多噪声, 导致模型的光流预测精度较低. 针对标签数据的准确性问题, Yang 等人^[37]设计了一种基于数据驱动的标签数据学习策略, 利用 FlowNet2.0 有监督学习光流模型与无监督学习模型交替学习策略提高了半监督模型的光流计算精度. 文献[38]使用对抗生成网络将优化复杂的灰度守恒与空间平滑损失函数转为简单的二分类最大似然函数, 显著降低了半监督学习光流估计的复杂度.

2.3.2 半监督学习模型训练策略

半监督学习策略通过综合有监督与无监督学习的优势, 可在不明显降低光流模型预测精度的情况下, 提高网络模型针对现实场景图像序列的光流计算鲁棒性^[36]. 其训练步骤大致如下: 首先采用有监督学习策略, 利用标签数据监督网络模型学习预测光流的一般特性, 以提高模型光流估计的整体精度; 然后使用无标注数据对模型的参数进行正则化, 增强模型对现实场景的鲁棒性^[38]. 需要注意的是, 无监督与有监督学习的损失函数项在总体损失函数中的权重须合理设置, 以防止网络在训练过程中发生震荡, 导致光流模型的性能下降.

3 光流估计网络优化策略

3.1 特征金字塔

特征金字塔优化是采用可调学习参数的特征金字塔完成图像特征提取的方法, 目的是取出对亮度变化和纹理阴影具有不变性的鲁棒特征. 在 Ranjan 和 Black^[21]提出金字塔分层 SpyNet 光流网络模型后, 图像特征金字塔优化策略成为提高深度学习光流计算精度与鲁棒性的重要方法^[22-24]. 研究表明, 使用特征金字塔替代传统图像金字塔能够显著提升网络模型的光流预测精度, 并且特征金字塔中卷积部分的特征提取能力越强, 网络模型的光流预测精度越高^[22].

3.2 匹配成本块

匹配成本块约束是提高深度学习光流预测精度的有效手段^[20-22]. 其匹配过程可分为两个步骤: 首先对输入图像进行特征提取, 然后对提取的特征进行特征比对从而构建匹配成本块^[39]. 因此, 匹配成本块可用来衡量图像序列中对应像素点的匹配成本代价^[40], 其定义如下所示:

$$c(x_1, x_2) = \sum_{o \in [-k, k] \times [-k, k]} \langle f_1(x_1 + o), f_2(x_2 + o) \rangle \quad (1)$$

式(1)中, $f_1(\cdot)$ 、 $f_2(\cdot)$ 表示输入图像序列的特征空

间, 符号 x_1, x_2 表示特征空间中匹配像素坐标, k 为匹配核尺寸. 通过式(1)可建立以 x_1, x_2 为中心点, 范围为 $o = 2k + 1$ 方形区域内的匹配模型, 并获取相应的特征匹配值.

3.3 图像变形技术

利用图像变形技术^[41]对第二帧图像进行运动补偿, 能够减弱像素点在图像帧间的形变与位移, 克服大位移与运动遮挡对光流计算的影响, 提高网络模型的精度与鲁棒性. 图像变形模型如下:

$$I_w(x) = I_2(x + w(x)) \quad (2)$$

式(2)中, $w(x)$ 表示像素点 x 的光流矢量, $I_2(x)$ 与 $I_w(x)$ 分别表示图像序列第二帧原始图像和变形图像在像素点 x 处的亮度值. 通过构建第一帧图像与变形图像之间的匹配成本块搜索光流特征, 可有效提高大位移与遮挡场景下模型的预测精度与可靠性.

3.4 后置处理

利用上下文信息对光流计算进行后置处理以提高估计精度是变分光流计算技术的常用手段. 常见的后置处理方法包括中值滤波^[42]、双边滤波^[43], 引导滤波^[44]等滤波优化策略. 深度学习光流网络模型亦可通过对输出光流进行后置处理提升光流预测的精度. FlowNet^[18]网络模型采用由粗到细迭代细化策略^[45]对网络模型预测结果进行后置处理, 显著提升了小位移场景下光流计算的准确性. PWC-Net^[22]网络模型采用扩张卷积^[46]作为后置处理, 通过拓展卷积过程的感受野进一步细化大位移场景下的光流计算结果, 显著提高了光流预测的精度.

3.5 子网络融合

当图像序列中包含弱小目标、小位移和复杂边缘等困难场景时, 由于卷积网络的平滑特性常常导致模型对弱小目标和小位移不敏感, 光流计算易产生边缘过度平滑现象^[18]. 针对此类问题, 子网络融合优化通过单独训练一个针对特定运动或场景的子网络, 并将子网络的输出光流与原始输出光流进行融合以提高模型的鲁棒性. 文献[47]通过联合立体匹配与光流估计, 在训练过程中利用特征复用同时提高了立体匹配与光流预测的精度. 此外, 文献[48, 49]通过融合静态场景的深度信息以及相机运动、姿势特性, 显著提高了光流计算的精度与鲁棒性.

4 光流测试数据库与评价标准

4.1 光流评价数据库

Middlebury 数据库^[50]: 由美国 Middlebury 学院计算机视觉实验室建立的 Middlebury 数据库是光流计算领域的首个标准测试数据库, 分别提供了 8 组训练图像序列和 8 组测试图像序列对光流计算模型进行评价排

名.但是由于 Middlebury 数据库训练与测试图像序列较少,且均为计算机合成或室内场景,因此其难以对不同光流计算模型的鲁棒性进行客观评价.

MPI-Sintel 数据库^[51]:美国华盛顿大学和佐治亚理工学院的研究人员通过截取电影《Sintel》中的不同场景制作了 MPI-Sintel 数据库,该数据库分为 Clean 和 Final 两个数据集,分别提供 1041 组训练图像集和 552 组测试图像集.其中,Clean 数据集包含大位移、弱纹理、非刚性大形变等困难场景,以测试光流计算模型的准确性.Final 数据集通过添加运动模糊、雾化效果以及图像噪声使其更加贴近于现实场景,用于测试光流估计的可靠性.

KITTI 数据库^[52]:由于 MPI-Sintel 数据库仅包含计算机合成图像集,难以反映光流模型针对现实场景的鲁棒性.德国卡尔斯鲁厄理工学院和丰田汽车研究中心通过车载摄像机和激光扫描仪制作了由真实街道场景组成的 KITTI 数据库.该数据库分为 2012 和 2015 两个数据集,共提供了 394 组训练图像集和 395 组测试图像集,其中 2012 测试图像集仅包含静止背景,2015 测试图像集则采用动态背景,显著增强了测试图像集的光流估计难度.

FlyingChairs 数据库^[18]:德国弗莱堡大学的模式识别与图像处理实验室通过将计算机合成的椅子模型与现实场景进行堆叠,并根据仿射变换模拟运动目标与背景的不同运动类型,制作了 FlyingChairs 数据库^[53].FlyingChairs 数据库提供了 20000 余组测试图像集供光流网络模型进行训练和评估,有效地缓解了标签数据对光流网络模型的限制.

FlyingThings3D 数据库^[54]:鉴于 FlyingChairs 数据库仅包含二维仿射变换,德国慕尼黑理工大学计算机视觉研究中心与弗莱堡大学模式识别与图像处理实验室利用三维建模技术联合开发了 FlyingThings3D 数据库.通过模拟三维旋转、空间平移、镜头移动等包含深度信息的复杂三维运动与空间场景,提供了 25000 多组包含真实光流值的图像序列.

4.2 光流评价标准

Middlebury 基准:Middlebury 基准是首个光流计算客观评价标准,分别采用平均角误差^[55](Average Angle Error, AAE)和平均端点误差^[50](Average Endpoint Error, AEE)测量光流计算结果的角度偏移和距离误差:

$$AAE = \frac{1}{N} \arccos \left(\frac{\mathbf{u}_C \times \mathbf{u}_E + \mathbf{v}_C \times \mathbf{v}_E + 1}{\sqrt{(\mathbf{u}_C^2 + \mathbf{v}_C^2 + 1)(\mathbf{u}_E^2 + \mathbf{v}_E^2 + 1)}} \right) \quad (3)$$

$$AEE = \frac{1}{N} \sqrt{(\mathbf{u}_E - \mathbf{u}_C)^2 + (\mathbf{v}_E - \mathbf{v}_C)^2} \quad (4)$$

上式中, $(\mathbf{u}_C, \mathbf{v}_C)^T$ 和 $(\mathbf{u}_E, \mathbf{v}_E)^T$ 分别表示真实光流场和估计光流场,符号 N 表示图像像素点总数量.

MPI-Sintel 基准:为客观反映光流模型的计算精度

与鲁棒性,MPI-Sintel 基准利用端点误差定义了不同图像区域和不同位移的光流评价指标.如表 2 所示,MPI-Sintel 基准主要包括三种不同类别的评价指标,其中 EPE-all、EPE-matched 和 EPE-unmatched 分别反映了光流估计结果在全部图像区域、前后帧匹配像素点区域以及前后帧遮挡区域的计算精度;d0 ~ 10、d10 ~ 60 和 d60 ~ 140 分别反映光流估计结果在不同图像运动边界区域的准确性;s0 ~ 10、s10 ~ 40 和 s40 + 则分别反映光流估计结果在不同位移区域的可靠性.

表 2 MPI-Sintel 基准光流评价指标

评价指标	误差定义
EPE-all	图像中所有像素点的光流平均端点误差
EPE-matched	前后帧图像匹配像素点的光流平均端点误差
EPE-unmatched	前后帧图像未匹配像素点的光流平均端点误差
d0 ~ 10	距离边界 10 个像素以内的光流平均端点误差
d10 ~ 60	距离边界 10 ~ 60 个像素的光流平均端点误差
d60 ~ 140	距离边界 60 ~ 140 个像素的光流平均端点误差
s0 ~ 10	图像帧间位移小于 10 个像素的光流平均端点误差
s10 ~ 40	图像帧间位移 10 ~ 40 个像素的光流平均端点误差
s40 +	图像帧间位移大于 40 个像素的光流平均端点误差

KITTI 基准:为了反映光流模型在 KITTI 数据库的估计效果,KITTI 基准首先利用端点误差定义图像光流异常值,然后根据光流异常值检测结果定义 FL-all、FL-fg 和 FL-bg 三个评价指标对光流估计结果进行客观评价.如表 3 所示,KITTI 基准通过统计光流估计结果中异常值的比率反映光流模型在图像整体区域、前景区域和背景区域的计算准确性与鲁棒性.此外,通过调整异常值判断阈值的大小,可进一步评估光流模型在不同统计标准下的鲁棒性.

表 3 KITTI 基准光流评价指标

评价指标	误差定义
FL-all	图像整体区域中光流异常值的比率
FL-fg	图像前景区域中光流异常值的比率
FL-bg	图像背景区域中光流异常值的比率

4.3 模型测试对比与分析

为了验证深度学习方法相对于变分模型在计算精度与时间消耗等方面的优势,分别对 FlowNet2.0^[20]、UnFlow^[29]、Semi-OF^[38] 和 Classic + NL^[42] 等深度学习与变分光流方法的计算结果进行综合对比与分析.其中,FlowNet2.0 是有监督学习光流计算方法,其通过子网络融合与后置处理显著提升了光流预测的精度.UnFlow 是无监督学习计算模型,其利用灰度守恒假设与空间平滑构造无监督学习损失函数,提高了无标签数据光流估计的可靠性.Semi-OF 是半监督学习光流计算模

型,其利用对抗生成网络对损失函数进行建模,减低了半监督学习光流估计的复杂度.最后,Classic + NL 是变分光流计算模型,通过在金字塔分层变分估计模型中引入加权中值滤波优化策略,有效提高了变分光流估计的精度与鲁棒性.

表 4 分别展示了各对比方法在 Middlebury、MPI-Sintel 和 KITTI 测试数据库的光流估计误差与时间消耗.由于 Semi-OF 模型未在公共数据库发布其光流评测结果,表中 Semi-OF 的光流误差结果源自文献[38].从表 4 中不同光流估计模型的对比结果可以看出:(1)相对于传统的变分光流方法,深度学习光流估计模型在计算效率方面具有显著优势,能够满足实时计算

要求;(2)深度学习光流模型针对 Middlebury 等包含较少训练图像的数据库估计精度低于变分方法,但在 MPI-Sintel 和 KITTI 等包含足量训练图像的数据库计算精度优于变分模型;(3)有监督学习光流计算模型在 MPI-Sintel 与 KITTI 数据库的计算精度优于无监督和半监督光流计算模型,主要源于 MPI-Sintel 和 KITTI 数据库提供了较多的标签数据.(4)由于半监督学习光流模型折中考虑标签数量与计算精度损失,其在包含较少训练图像的 Middlebury 数据库计算精度优于有监督学习和无监督光流计算模型,但针对 MPI-Sintel 和 KITTI 等提供较多训练图像的数据库计算精度落后于有监督模型.

表 4 深度学习光流计算模型与变分光流计算方法对比结果

对比方法	Middlebury		MPI-Sintel			KITTI			运行时间 (单位:s)
	AEE	AEPE	EPE-all	EPE-matched	EPE-unmatched	FL-all	FL-fg	FL-bg	
FlowNet2.0 ^[20]	5.87	0.52	5.739	2.752	30.108	10.41 %	10.75 %	8.75 %	0.09
UnFlow ^[29]	8.77	0.76	10.219	6.061	44.110	11.11 %	10.15 %	15.93 %	0.12
Semi-OF ^[38]	—	0.32	7.31	—	—	31.01%	—	—	0.10
Classic + NL ^[42]	2.90	0.32	9.153	4.814	44.509	35.61 %	34.24 %	42.46 %	270

5 总结与展望

截至目前,基于深度学习的光流计算技术在计算精度与鲁棒性等方面已取得了显著的进展^[56].但受制于卷积神经网络模型自身的限制和模型训练策略以及标签数据的匮乏,深度学习光流计算技术仍有许多相关工作有待进一步深入研究,主要包括:

(1)根据通用近似定理,神经网络具有很强的学习能力,然而要记住更多信息势必会使网络模型变得更复杂.虽然局部连接、权重共享和池化等优化操作可以简化神经网络模型、缓解模型复杂度和表达能力之间的矛盾.但是,添加过多的卷积以及优化操作势必增加训练网络模型的难度.近期,通过研究人脑处理信息过载方式而提出的注意力机制被证明是解决卷积神经网络学习、记忆能力不足的有效手段,因此在后续研究中应考虑利用注意力机制提高光流网络模型的计算能力.

(2)现实世界是千变万化、错综复杂的,当图像中包含非刚性大形变、物体的遮挡与闭塞等复杂场景时,如何建立具有泛化性能的光流预测网络模型是深度学习光流计算技术研究的难点.传统的变分光流计算技术在处理大位移、遮挡等方面已积累丰富的手段与方法,因此,在后续的研究中应考虑如何结合深度学习与变分方法的各自优势,在保证深度学习光流计算效率的同时显著提高光流估计的精度与鲁棒性.

(3)随着网络的不断加深,现有的数据样本集已逐渐难以满足有监督学习光流估计模型的训练任务需

求.因此如何利用有限的标签数据探索高精度、强鲁棒性的光流计算网络学习策略是有监督学习光流计算技术的关键问题.此外,有监督学习光流网络模型的代价函数是影响网络模型光流估计效果的重要因素,在后续的研究中应考虑如何设计具有自适应能力的网络代价函数,以提高网络模型的泛化能力.

(4)现有的无监督学习光流计算网络模型通常是以图像数据守恒假设为基础进行网络训练,而光照变化、大位移以及运动遮挡等困难场景会导致像素点亮度突变,导致模型的光流预测精度较差.近年来,对抗生成网络是深度学习领域的研究热点,其能够较好地拟合生成和判别函数.在今后的研究中,可以考虑利用对抗生成网络优化无监督学习光流计算模型的参数训练,提高无监督学习光流估计的鲁棒性.

(5)无监督学习光流模型未利用标签数据作为训练样本,导致难以获得高精度预测效果.而有监督学习模型需要大量标签数据进行网络训练,因此针对现实场景的光流估计鲁棒性仍有待进一步提高.为了充分利用无监督学习和有监督学习的优势,在今后的研究中应重点考虑半监督学习光流网络模型,利用有限的标签数据和大规模无标签数据提高网络模型的泛化能力,显著增强网络模型的光流估计精度与鲁棒性.

参考文献

- [1] Horn B K P, Schunck B G. Determining optical flow[J]. *Artificial Intelligence*, 1980, 17(1-3): 185-203.

- [2] 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[J]. 电子学报,2016,44(12):3044-3052.
ZHANG Cong-xuan, CHEN Zhen, LI Ming. Review of the 3D reconstruction technology based on optical flow of monocular image sequence[J]. Acta Electronica Sinica,2016,44(12):3044-3052. (in Chinese).
- [3] 陈震,张聪炫,晏文敬,吴燕平. 基于图像局部结构的区域匹配变分光流算法[J]. 电子学报,2015,43(11):2200-2209.
CHEN Zhen, ZHANG Cong-xuan, YAN Wen-jing, WU Yan-ping. Region matching variational optical flow algorithm based on image local structure[J]. Acta Electronica Sinica,2015,43(11):2200-2209. (in Chinese).
- [4] Bailer C, Taetz B, Stricker D. Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation[A]. International Conference on Computer Vision[C]. Boston: IEEE,2015. 4015-4023.
- [5] Heeger D J. Optical flow using spatiotemporal filters[J]. International Journal of Computer Vision,1988,1(4):279-302.
- [6] Fleet D J, Jepson A D. Computation of component image velocity from local phase information[J]. International Journal of Computer Vision,1990,5(1):77-104.
- [7] Zhang C X, Chen Z, Wang M R, Li M, Jiang S F. Robust non-local TV-L1 optical flow estimation with occlusion detection[J]. IEEE Transactions on Image Processing,2017,26(8):4055-4067.
- [8] 葛利跃,张聪炫,陈震,黎明,陈昊. 相互结构引导滤波 TV-L1 变分光流估计[J]. 电子学报,2019,47(3):707-713.
GE Li-yue, ZHANG Cong-xuan, CHEN Zhen, LI Ming, CHEN Hao. Mutual-structure guided filtering based TV-L1 optical flow estimation[J]. Acta Electronica Sinica,2015,43(11):2200-2209. (in Chinese).
- [9] 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[J]. 电子学报,2019,47(6):1316-1323.
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[J]. Acta Electronica Sinica,2019,47(6):1316-1323. (in Chinese).
- [10] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[A]. International Conference on Learning Representations[C]. San Diego: ICLR,2015. 1-14.
- [11] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Rabinovich A. Going deeper with convolutions[A]. International Conference on Computer Vision and Pattern Recognition[C]. Boston: IEEE,2015. 1-9.
- [12] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition[A]. International Conference on Computer Vision and Pattern Recognition[C]. Las Vegas: IEEE,2016. 770-778.
- [13] Huang G, Liu Z, Van DerMaaten L, Weinberger K Q. Densely connected convolutional networks[A]. International Conference on Computer Vision and Pattern Recognition[C]. Hawaii: IEEE,2017. 4700-4708.
- [14] Jung S, Hwang S, Shin H, Shim D H. Perception, guidance, and navigation for indoor autonomous drone racing using deep learning[J]. IEEE Robotics and Automation Letters,2018,3(3):2539-2544.
- [15] Gomaa A, Abdelwahab M M, Abo-Zahhad M, Minematsu T, Taniguchi R I. Robust vehicle detection and counting algorithm employing a convolution neural network and optical flow[J]. Sensors,2019,19(20):4588.
- [16] Zhu Z, Wu W, Zou W, Yan J. End-to-end flow correlation tracking with spatial-temporal attention[A]. International Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City: IEEE,2018. 548-557.
- [17] Djelouah A, Campos J, Schaub-Meyer S, Schroers C. Neural inter-frame compression for video coding[A]. International Conference on Computer Vision[C]. Long Beach: IEEE,2019. 6421-6429.
- [18] Dosovitskiy A, Fischer P, Ilg E, Hausser P, Hazirbas C, Golkov V, Brox T. FlowNet: Learning optical flow with convolutional networks[A]. International Conference on Computer Vision and Pattern Recognition[C]. Boston: IEEE,2015. 2758-2766.
- [19] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[A]. International Conference on Medical Image Computing and Computer-Assisted Intervention[C]. Munich: Springer,2015. 234-241.
- [20] Ilg E, Mayer N, Saikia T, Dosovitskiy A, Brox T. FlowNet 2.0: Evolution of optical flow estimation with deep networks[A]. International Conference on Computer Vision and Pattern Recognition[C]. Hawaii: IEEE,2017. 1647-1655.
- [21] Ranjan A, Black M J. Optical flow estimation using a spatial pyramid network[A]. International Conference on Computer Vision and Pattern Recognition[C]. Hawaii: IEEE,2017. 4161-4170.
- [22] Sun D, Yang X, Liu M Y, Kautz J. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume[A]. International Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City: IEEE,2018. 8934-8943.
- [23] Ilg E, Saikia T, Keuper M, Brox T. Occlusions, motion and depth boundaries with a generic network for disparity, optical flow or scene flow estimation[A]. European

- Conference on Computer Vision [C]. Munich; Springer, 2018. 614 – 630.
- [24] Hui T W, Tang X, Loy CC. LiteFlowNet: A lightweight convolutional neural network for optical flow estimation [A]. International Conference on Computer Vision and Pattern Recognition [C]. Salt Lake City; IEEE, 2018. 8981 – 8989.
- [25] Sevilla-Lara L, Sun D, Jampani V, Black M J. Optical flow with semantic segmentation and localized layers [A]. International Conference on Computer Vision and Pattern Recognition [C]. Las Vegas; IEEE, 2016. 3889 – 3898.
- [26] Cheng J, Tsai Y H, Wang S, Yang M H. Segflow: Joint learning for video object segmentation and optical flow [A]. International Conference on Computer Vision [C]. Venice; IEEE, 2017. 686 – 695.
- [27] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization [EB/OL]. <http://arxiv.org/abs/1412.6980>, 2014.
- [28] Jason J Y, Harley A W, Derpanis K G. Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness [A]. European Conference on Computer Vision [C]. Amsterdam; Springer, 2016. 3 – 10.
- [29] Meister, Simon, Junhwa Hur, Stefan Roth. UnFlow: Unsupervised learning of optical flow with a bidirectional census loss [A]. Thirty-Second AAAI Conference on Artificial Intelligence [C]. Hawaii; Springer, 2018. 682 – 691.
- [30] Wang Y, Yang Y, Yang Z, Zhao L, Wang P, Xu W. Occlusion aware unsupervised learning of optical flow [A]. International Conference on Computer Vision and Pattern Recognition [C]. Salt Lake City; IEEE, 2018. 4884 – 4893.
- [31] Janai J, Guney F, Ranjan A, Black M, Geiger A. Unsupervised learning of multi-frame optical flow with occlusions [A]. European Conference on Computer Vision [C]. Munich; Springer, 2018. 690 – 706.
- [32] Zhu A Z, Yuan L, Chaney K, Daniilidis K. Unsupervised event-based learning of optical flow, depth, and egomotion [A]. International Conference on Computer Vision and Pattern Recognition [C]. Long Beach; IEEE, 2019. 989 – 997.
- [33] Paredes-Vallés F, Scheper K Y W, De Croon G C H E. Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 2(8): 2051 – 2064.
- [34] Shedligeri P, Mitra K. Live Demonstration: Joint estimation of optical flow and intensity image from event sensors [A]. International Conference on Computer Vision and Pattern Recognition [C]. Long Beach; IEEE, 2019. 1 – 2.
- [35] Rasmus A, Berglund M, Honkala M, Valpola H, Raiko T. Semi-supervised learning with ladder networks [A]. Advances in Neural Information Processing Systems [C]. Montréal; MIT, 2015. 3546 – 3554.
- [36] Zhu Y, Lan Z, Newsam S, Hauptmann A G. Guided optical flow learning [A]. International Conference on Computer Vision and Pattern Recognition [C]. Hawaii; IEEE, 2017. 1 – 5.
- [37] Yang G, Deng Z, Wang S, Li Z. Masked label learning for optical flow regression [A]. International Conference on Pattern Recognition [C]. Beijing; IEEE, 2018. 1139 – 1144.
- [38] Lai W S, Huang J B, Yang M H. Semi-supervised learning for optical flow with generative adversarial networks [A]. Advances in Neural Information Processing Systems [C]. Long Beach; MIT, 2017. 354 – 364.
- [39] Weinzaepfel P, Revaud J, Harchaoui Z, Schmid C. Deepflow: large displacement optical flow with deep matching [A]. International Conference on Computer Vision and Pattern Recognition [C]. Portland; IEEE, 2013. 1385 – 1392.
- [40] Hosni A, Rhemann C, Bleyer M, Rother C, Gelautz M. Fast cost-volume filtering for visual correspondence and beyond [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(2): 504 – 511.
- [41] Lin T Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection [A]. International Conference on Computer Vision and Pattern Recognition [C]. Hawaii; IEEE, 2017. 2117 – 2125.
- [42] Sun D, Roth S, Black M J. A quantitative analysis of current practices in optical flow estimation and the principles behind them [J]. International Journal of Computer Vision, 2014, 106(2): 115 – 137.
- [43] Xu L, Jia J, Matsushita Y. Motion detail preserving optical flow estimation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 34(9): 1744 – 1757.
- [44] Zhang C, Ge L, Chen Z, et al. Refined TV-L1 optical flow estimation using joint filtering [J]. IEEE Transactions on Multimedia, 2019.
- [45] Revaud J, Weinzaepfel P, Harchaoui Z, Schmid C. Epicflow: edge-preserving interpolation of correspondences for optical flow [A]. International Conference on Computer Vision and Pattern Recognition [C]. Boston; IEEE, 2015. 1164 – 1172.
- [46] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [EB/OL]. <http://arxiv.org/abs/1511.07122> V2, 2015.
- [47] Lai H Y, Tsai Y H, Chiu W C. Bridging stereo matching and optical flow via spatiotemporal correspondence [A]. International Conference on Computer Vision and Pattern Recognition [C]. Long Beach; IEEE, 2019. 1890 – 1899.
- [48] Ranjan A, Jampani V, Balles L, Kim K, Sun D, Wulff J, Black M J. Competitive collaboration: joint unsupervised learning of depth, camera motion, optical flow and motion segmentation [A]. International Conference on Computer

- Vision and Pattern Recognition [C]. Long Beach: IEEE, 2019. 12240 – 12249.
- [49] Yin Z, Shi J. Geonet: Unsupervised learning of dense depth, optical flow and camera pose [A]. International Conference on Computer Vision and Pattern Recognition [C]. Salt Lake City: IEEE, 2018. 1983 – 1992.
- [50] Baker S, Scharstein D, Lewis J P, Roth S, Black M J, Szeliski R. A database and evaluation methodology for optical flow [J]. International Journal of Computer Vision, 2011, 92(1): 1 – 31.
- [51] Butler D J, Wulff J, Stanley G B, Black M J. A naturalistic open source movie for optical flow evaluation [A]. European Conference on Computer Vision [C]. Florence: Springer, 2012. 611 – 625.
- [52] Geiger A, Lenz P, Stiller C, Urtasun R. Vision meets robotics: The KITTI dataset [J]. International Journal of Robotics Research, 2013, 32(11): 1231 – 1237.
- [53] Aubry M, Maturana D, Efros A A, Russell B C, Sivic J. Seeing 3D chairs: Exemplar part-based 2d-3d alignment using a large dataset of cad models [A]. International Conference on Computer Vision and Pattern Recognition [C]. Columbus: IEEE, 2014. 3762 – 3769.
- [54] Mayer N, Ilg E, Häusser P, Cremers D, Dosovitskiy A, Brox T. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation [A]. International Conference on Computer Vision and Pattern Recognition [C]. Las Vegas: IEEE, 2016. 4040 – 4048.
- [55] Fleet D J, Jepson A D. Computation of component image velocity from local phase information [J]. International Journal of Computer Vision, 1990, 5(1): 77 – 104.
- [56] Ilg E, Cicek O, Galesso S, Klein A, Makansi O, Hutter F, Brox T. Uncertainty estimates and multi-hypotheses networks for optical flow [A]. European Conference on Computer Vision [C]. Munich: Springer, 2018. 652 – 667.

作者简介



张聪炫 男, 1984年7月出生于河南省焦作市。分别于2007和2014年在南昌航空大学和南京航空航天大学获得学士和博士学位。现为南昌航空大学副教授, 硕士生导师, 中国科学院自动化研究所博士后。主要研究方向为图像检测与智能识别。

E-mail: zcxds@163.com



周仲凯 男, 1994年3月出生于江苏省常州市。现为南昌航空大学测试与光电工程学院硕士研究生。主要研究方向为图像检测与智能识别。

E-mail: jsczzk123@163.com



陈震(通讯作者) 男, 1969年11月出生于江西省九江市。分别于1993、2000和2003年在西北工业大学获得学士、硕士和博士学位。现为南昌航空大学教授, 博士生导师。主要研究方向为计算机视觉、图像处理与模式识别。

E-mail: dr_chenzhen@163.com



葛利跃 男, 1992年10月出生于安徽省蚌埠市。现为南昌航空大学测试与光电工程学院硕士研究生。主要研究方向为图像检测与智能识别。

E-mail: lygeah@163.com



黎明 男, 1965年2月出生于江西省樟树市。分别于1985年、1990年和1997年在上海交通大学和南京航空航天大学获得学士、硕士和博士学位。现为南昌航空大学教授, 博士生导师。主要研究方向为图像处理与模式识别、智能计算。

E-mail: limingnialt@hotmail.com



江少锋 男, 1978年5月出生于江西省鹰潭市。分别于1999年和2008年在北京航空航天大学 and 南方医科大学获得学士和博士学位。现为南昌航空大学副教授, 硕士生导师。主要研究方向为医学图像处理、深度学习。

E-mail: jiangshaofeng@nchu.edu.cn



陈昊 男, 1982年9月出生于山东省平度市。分别于2004年和2011年在青岛大学和南京航空航天大学获得学士、博士学位。现为南昌航空大学副教授, 硕士生导师, 主要研究方向为智能进化算法及应用, 模式识别与人工智能。

E-mail: chenhaoshl@163.com