

## 视觉光流计算技术及其应用

崔毅博<sup>①</sup> 汤仁东<sup>②</sup> 邢大军<sup>③</sup> 王隽<sup>④</sup> 李尚生<sup>\*①</sup>

<sup>①</sup>(海军航空大学 烟台 264001)

<sup>②</sup>(军事医学研究院 北京 100850)

<sup>③</sup>(北京师范大学认知神经科学与学习国家重点实验室 北京 100875)

<sup>④</sup>(中国科学院自动化研究所 北京 100190)

**摘要:** 视觉光流计算是计算机视觉从处理2维图像走向加工3维视频的重要技术手段, 是描述视觉运动信息的主要方式。光流计算技术已经发展了较长的时间, 随着相关技术尤其是深度学习技术在近些年的迅速发展, 光流计算的性能得到了极大提升, 但仍然有大量的局限性问题没有解决, 准确、快速且稳健的光流计算目前仍然是一个有挑战性的研究领域和业内研究热点。光流计算作为一种低层视觉信息处理技术, 其技术进展也将有助于相关中高层级视觉任务的实现。该文主要内容是介绍基于计算机视觉的光流计算及其技术发展路线, 从经典算法和深度学习算法这两个主流技术路线出发, 总结了技术发展过程中产生的重要理论、方法与模型, 着重介绍了各类方法与模型的核心思想, 说明了各类数据集及相关性能指标, 同时简要介绍了光流计算技术的主要应用场景, 并对今后的技术方向进行了展望。

**关键词:** 计算机视觉; 光流; 深度学习; 综述

中图分类号: TN919.8; TP391.41

文献标识码: A

文章编号: 1009-5896(2023)08-2710-12

DOI: [10.11999/JEIT221418](https://doi.org/10.11999/JEIT221418)

## Visual Optical Flow Computing: Algorithms and Applications

CUI Yibo<sup>①</sup> TANG Rendong<sup>②</sup> XING Dajun<sup>③</sup> WANG Juan<sup>④</sup> LI Shangsheng<sup>①</sup>

<sup>①</sup>(Naval Aviation University, Yantai 264001, China)

<sup>②</sup>(Academy of Military Medical Sciences, Beijing 100850, China)

<sup>③</sup>(State Key Laboratory of Cognitive Neuroscience and Learning, Beijing 100875, China)

<sup>④</sup>(Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** Visual optical flow calculation is an important technique for computer vision to move from processing 2D images to processing 3D videos, and is the main way of describing visual motion information. The optical flow calculation technique has been developed for a long time. With the rapid development of related technologies, especially deep learning technology in recent years, the performance of optical flow calculation has been greatly improved. However, there are still many limitations that have not been solved. Accurate, fast, and robust optical flow calculation is still a challenging research field and a hot topic in the industry. As a low-level visual information processing technology, the implementation of related high-level visual tasks will also be contributed by the technological advances of optical flow calculation. In this paper, the development path of optical flow calculation based on computer vision is mainly introduced. The important theories, methods, and models generated during the technological development process from the two mainstream technology paths of classical algorithms and deep learning algorithms are summarized, the core ideas of various methods and models are being introduced and the various datasets and performance indicators are explained, the main application scenarios of optical flow calculation technology are briefly introduced, and the future technical directions are also prospected.

**Key words:** Computer vision; Optical flow; Deep learning; Overview

## 1 引言

光流计算作为计算机视觉的一个长期基本任务,其重要性显而易见。由于运动视觉处理的特殊性,光流作为后面高级视觉处理的输入,对其准确度、实时性都有着极高的要求,光流计算的会直接影响其后的高级视觉处理。光流计算技术在计算机视觉的各主要研究方向如检测、分割、导航、姿态估计、3维重建等领域中都有着重要的应用,其相关算法在更上层的应用场景如自动驾驶、气象预报、雷达信息处理、卫星及航空影像分析、同步定位与地图构建(Simultaneous Localization And Mapping, SLAM)、视觉神经科学相关领域以及军事应用领域等前沿热点方向更是有着重要的研究价值与应用价值。由此可见光流计算技术的发展对于计算机视觉领域的重要意义。

本文按照传统光流计算技术的主要发展过程(第2节)、基于深度学习的光流计算技术发展过程(第3节)、光流测试相关数据集和性能评价指标(第4节)、光流计算技术的具体应用(第5节)、总结及光流计算技术未来发展趋势展望(第6节)的顺序安排各节内容。

## 2 传统光流计算技术的主要发展过程

通过图像计算光流,自20世纪80年代兴起,其中具有代表性的经典算法为Horn等人<sup>[1]</sup>提出的HS(Horn-Schunck)光流算法与Lucas等人<sup>[2]</sup>提出的LK(Lucas-Kanade)光流算法,而后基于这两种算法的各种改进版本有许多。其中HS算法是基于变分法求解光流,LK算法是基于差分法求解光流,但二者都基于两个共同的假设:

**假设1 亮度恒定假设:**同一目标在不同帧间运动时,其亮度不会发生改变。

**假设2 小运动假设:**短时间像素的位置不会剧烈变化,即相邻帧之间像素距离变化较小。

用数学模型说明如下:若 $I(x, y, t)$ 表示 $t$ 时刻 $(x, y)$ 位置的像素在图像上的亮度,则根据假设1和假设2得到

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (1)$$

若 $u = dx/dt, v = dy/dt$ 分别为像素沿 $x$ 与 $y$ 轴的速度,对式(1)进行泰勒展开,忽略高阶无穷小后对 $t$ 求导,而后代入 $(u, v)$ ,则可得

$$I_x u + I_y v + I_t = 0 \quad (2)$$

式(2)就是由光流的基本假设推出的光流基本方程。在其基础上通过加入不同的约束、改变求解方式,得到像素的 $(u, v)$ 值即图像相邻帧之间的光流。

### 2.1 HS光流算法

HS光流算法是一种优化算法,通过在假设1、假设2的基础上加入全局平滑约束条件,即假设 $(u, v)$ 值小范围内变化很小,其加入的全局平滑约束项为

$$\xi_c^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \quad (3)$$

此约束项表征光流的连续性即平滑(其为0时代表光流在任意方向无变化),结合光流基本方程建立平滑约束下光流优化方程(也称能量函数)

$$E(u, v) = \iint [\alpha^2 \xi_c^2 + (I_x u + I_y v + I_t)^2] dx dy \quad (4)$$

其中, $\alpha$ 是平滑项的权重,根据假设最小化此方程,利用欧拉-拉格朗日(Euler-Lagrange)方程求解,经迭代至收敛得到光流信息 $(u, v)$ 。此类优化方法易陷入局部极值,其初始值很重要,而实际使用时往往无法获得初始值,这就导致此类算法在新场景下会有一段不稳定期。

### 2.2 LK光流算法

LK光流在假设1和假设2的基础上增加假设3:

**假设3 空间一致性:**某一个小窗口内的像素短时间内具有相同的运动(即 $(u, v)$ 相同)。

若小窗口为 $w \times w$ ,则其中有 $w^2$ 个像素,则可根据式(2)列出超定方程并用最小二乘法解此方程即可求出这个 $w \times w$ 窗口的光流值 $(u, v)$ 。由于此类方法在大窗口下无法保证相同运动方向和假设2这个微分条件的成立,需利用图像金字塔技术<sup>[3]</sup>把图像分层压缩到低分辨率,把大位移运动变成了高层金字塔的小位移运动,进而配合插值算法逐层应用此方法求解光流。

在以上两种算法的基础上,针对光流的不同问题有相应的改进方法,如:为解决算法易陷入局部极值问题而提出的前后向光流法<sup>[4]</sup>,为解决遮挡问题而融合卡尔曼滤波运动预测的LK改进算法<sup>[5]</sup>,以及各类通过改进定位角点准确度与匹配准确度来提升光流准确性的改进方法等。

随着深度学习的兴起,利用卷积神经网络来进行光流估计已经成为一种重要方法,与计算机视觉其他领域横向比较,基于深度学习的光流计算方法在准确度、鲁棒性、实时性等方面有着天然的优势。因此基于深度学习的光流计算通常被认为是区别于经典算法的一种新模式,也是一个极具发展前景的技术方向。

## 3 基于深度学习的光流计算技术发展

有别于传统人工设计的方法来求解光流,深度学习从数据角度出发,利用数据训练相关模型,

从而得到可以准确进行光流估计的模型,并利用此模型在应用场景中对光流进行估计。随着GPU以及计算机算力的不断发展,目前基于深度学习的光流计算无论在准确度还是实时性上都已经超过经典算法。

Dosovitskiy等人<sup>[6]</sup>在2015年首次提出的基于卷积神经网络(Convolutional Neural Networks, CNN)的FlowNet实现了利用卷积神经网络进行光流的估计(第1代监督模型),为训练所提出的模型同时开发了Flying Chairs数据集,FlowNet的提出表明了完全基于卷积神经网络端到端的架构有能力解决光流估计的相关问题。此后基于深度学习的模型逐步在性能上赶超经典算法。其中主要可以分为两个大类:基于监督学习的光流估计模型和自监督学习的光流估计模型。下面分别进行介绍。

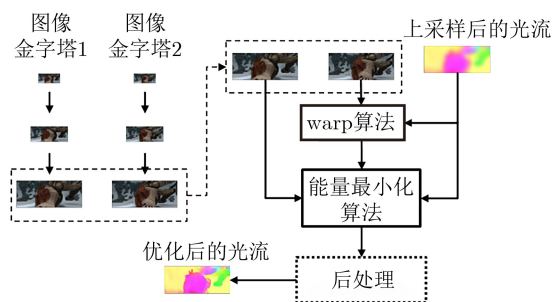
### 3.1 基于监督学习的光流估计模型

监督学习模型往往需要结合相关领域知识,利用监督学习的方式来对模型进行训练,其中结合领域知识方面主要有两种途径:

第1种是结合数据相关知识。普遍的做法就是通过把领域知识制作成光流相关数据集,使得神经网络可以利用这类数据集进行训练和评估,从而使得训练出的深度学习模型可以进行光流估计。相关数据集介绍详见第4节。

第2种是借鉴之前已有经典算法的约束条件与计算框架。结合卷积神经网络可以提取图像高维度特征以及可以进行并行计算的特点,进而得到光流在准确率和实时性上的提升,其中比较有代表性的就是DCflow<sup>[7]</sup>,上述两种领域知识结合方法在监督学习中可以同时使用。

DCflow参考FullFlow<sup>[8]</sup>提出的代价体(cost volume)以及由粗糙到精细的(Coarse-To-Fine, CTF)范式,这种CTF范式结构主要分为4步:特征提取(features)、构建代价体(cost volume)、代价体处理与光流后处理。



(a) 传统基于能量最小化的光流计算方法框架图

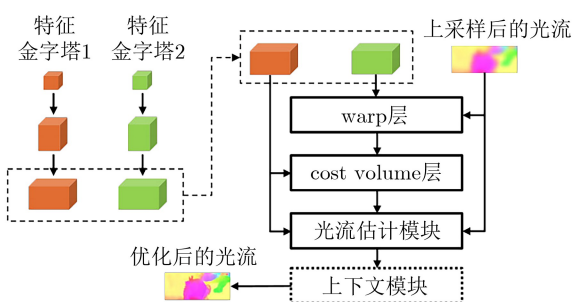
CTF范式的主要计算量在计算和优化代价体以及后处理上,代价体表示了每个位置提取出的高维特征之间的联系,在FlowNet2.0<sup>[9]</sup>(把FlowNetS/C相关模块进行了组合堆叠实验)中作者也支持显式地构建代价体,称其往往比隐式的效果要好。继FlowNet2.0之后,深度学习算法在实时性和精度上都开始超越经典算法,其中较为重要的借鉴经典算法结构的深度卷积神经网络模型是基于CTF范式的PWC-Net(Pyramid, Warping, and Cost volume, PWC)<sup>[10]</sup>(第2代监督模型),其框架如图1所示。

PWC-Net相较于FlowNet2.0,准确率更高且速度更快,改进了FlowNet2.0模型参数量大且训练繁琐的问题(速度快2倍参数量下降17倍)。其主要结构借鉴传统基于能量最小化的CTF光流计算框架(如图1),与传统方法不同的是,其先利用卷积层进行图像金字塔计算;再利用warp层模仿warp算法把第2帧图像利用上一帧光流扭曲(warp)到第1帧;然后对金字塔每一层利用卷积提取的特征(features)构建代价体(cost volume)找到特征之间的关系,把第1帧图像的特征、代价体、上一帧的光流输入到光流估计层得到光流估计;对应传统光流计算的后处理模块,最后用基于空洞卷积的上下文网络进行后处理,这个网络输入上一层的光流估计值和光流估计值的倒数第2层特征,可以优化并把光流放大到所需大小,类似经典算法结构的深度学习光流模型还有许多<sup>[11]</sup>。对PWC-Net进行改进效果较好的是利用迭代残差细化(Iterative Residual Refinement,IRR)方法的IRR-PWC模型<sup>[12]</sup>,这种方法可以在维持参数数量的情况下提高准确率,在附加去遮挡模块后,可以对遮挡情况下的光流预测更加准确。PWC-Net整体架构如图1所示。

图1(b)中warp层的数学表达式为

$$c_w^l(x) = c_2^l(x + \text{up}_2(w^{l+1})(x)) \quad (5)$$

其中,  $c_t^l$ 代表 $t$ 时刻图像t在图像金字塔第 $l$ 层的特征,  $x$ 是像素坐标(包括横纵坐标),  $w^{l+1}$ 是



(b) PWC-Net框架图

图1 传统框架与PWC-Net框架图对比图(图片改绘自文献<sup>[10]</sup>)

$l + 1$ 层的光流,  $\text{up}_2(\mathbf{w}^{l+1})(\mathbf{x})$ 是对应位置上采样的  $l + 1$ 层的光流, 式(5)表达的意思是如果光流估计准确, 则在  $\mathbf{x} + \text{up}_2\mathbf{w}^{l+1}(\mathbf{x})$ 处特征  $\mathbf{c}_2^l$ 应与在  $\mathbf{x}$ 处特征  $\mathbf{c}_w^l$ 相同。

代价体(cost volume)的数学表达式可以表示为

$$\mathbf{cv}^l(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{N}(\mathbf{c}_1^l(\mathbf{x}_1))^T \cdot \mathbf{c}_w^l(\mathbf{x}_2), |\mathbf{x}_2 - \mathbf{x}_1|_\infty < \mathbf{d} \quad (6)$$

其中,  $\mathbf{d}$ 为超参, 表征最大相关距离。这部分就是利用向量点积来进行特征相关性计算, 从而计算出特征之间的相似性, 代价体即表征这种特征之间相似性的映射。

在监督模型中, RAFT(Recurrent All-pairs Field Transforms for optical flow)<sup>[13]</sup>是十分重要的模型(第3代监督模型), 其框架如图2所示。这个模型实现了整个网络端到端训练的同时, 在效果上超过了PWC-Net和IRR-PWC, 且实现了模型的轻量化。其主要思路是利用卷积提取两帧图像的特征, 而后对特征做内积得到4D代价体(4D cost volumes)作为两个特征之间相似性的度量空间, 有别于其他算法, 光流会通过一个门控循环单元(Gated Recurrent Unit, GRU)的一个输出, 在4D 代价体空间内进行查询, 查询结果将用来更新GRU进行迭代细化, 这样就可以有效利用上下文信息, 通过GRU最终输出精细化后的光流, 最后通过利用周围像素上采样恢复光流到原图像分辨率。RAFT模型表明了端到端的深度学习模型在性能上可以超越人为模块化设计的模型, 并且截至本文成稿时间, 仍是深度学习模型性能进行比较的一个基准。

以PWC-Net, RAFT为代表的这种CTF范式的神经网络模型可以提高性能的主要途径是优化后3步, 以减少计算量、提高精度及其抗噪性能。当然还有借鉴其他范式的模型, 如经典算法EpicFlow<sup>[14]</sup>的由稀疏到稠密(Sparse-To-Dense, STD)范式, 其

主要步骤为: 计算稀疏匹配集、计算光流稀疏到稠密的插值、后处理以及优化得到光流。

另一种光流计算重要范式是STD。如PatchBatch<sup>[15]</sup>在经典EpicFlow的基础上利用神经网络提取高维度特征来计算稀疏匹配集, 提高了匹配集的质量, 结合EpicFlow的稠密插值最终得到光流。FTDM(Fully-Trainable Deep Matching)<sup>[16]</sup>通过训练一个u型拓扑的CNN来等效深度匹配(Deep Matching, DM)算法, 从而计算出稀疏匹配集, 而后利用EpicFlow稠密插值得到光流。此类方法主要改进方向在于计算稀疏匹配集、计算稀疏到稠密的插值这两个步骤上, 即如何在低算力的情况下找到质量更高的稀疏匹配集, 以及如何进行更好的插值计算。

以上两种框架都是基于CNN的, 而也有部分方法致力于改进CNN卷积模型本身性能, 使其更适用于光流估计任务, 从而在本质上提高光流估计的性能, 如模型PPAC-HD3<sup>[17]</sup>, 利用概率像素自适应卷积提高模型性能, 优化光流的边缘及精度。

随着深度学习模型的性能不断提升, 目前基于Transformer的模型在语言、图像以及多模态处理方面表现出了突出的性能。截至2022年10月KITTI数据集上表现最优的纯视觉光流估计模型是基于Transformer的GMFlow(Global Matching, GM)<sup>[18]</sup>作为第4代监督模型, 其框架如图3所示, 其改进版GMFlow+实现了多模态处理。GMFlow主要利用attention技术优化了利用卷积处理代价体自带的局部局限性问题, 可以做到全局匹配, 所以在处理大位移上优势明显, 其基于Transformer架构的处理方法是短期内的主流方向。

Transformer等利用self-attention技术的深度学习模型, 其核心公式表示为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (7)$$

其中,  $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 为基于图像的嵌入向量,  $d_k$ 是 $\mathbf{K}$ 的

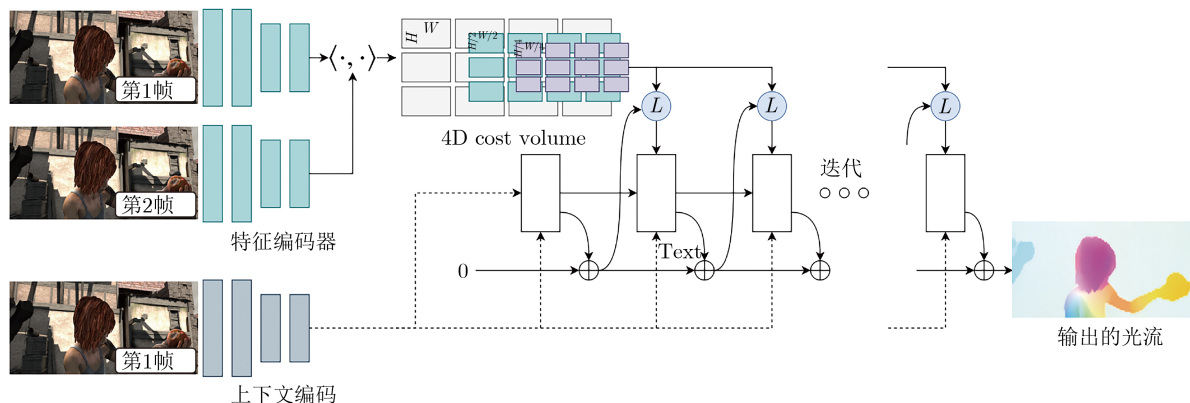


图 2 RAFT模型框架图(图片改绘自文献[13])

维度, self-attention其本质是找到相关特征, 并有针对性地强化高维特征向量, 从而使得所需相关特征突出出来, 与代价体本质一致但不会受到距离限制, 利用深度学习框架并行计算的优势在计算速度上相较于代价体更快。

除主要创新及作为行业基准的模型外, 每年还有大量基于以上模型的改进模型。近3年基于RAFT的改进居多, 效果也日益增强, 基于Transformer架构的模型也逐步体现出其优势, 结合图神经网络技术解决卷积固有的相关缺陷是未来发展的主要方向之一。各类模型主要解决的问题和特点详见表1,

其中浅灰色为CTF范式模型, 浅蓝色为STD范式模型, 加粗字体为行业广泛认可的基准模型。

由于域差(domain gap), 以及制作光流训练的数据集本身成本与技术难度很高, 除监督学习外深度学习光流模型的另一种主要模式是自监督学习模型, 不利用人工标注的数据集进行学习, 可以极大降低训练模型的成本, 若可以直接利用真实数据进行自监督训练, 则可有效避免数据集与真实数据之间域差问题。

### 3.2 基于自监督学习的光流估计模型

在2016年Long等人<sup>[23]</sup>利用简单的编码解码器

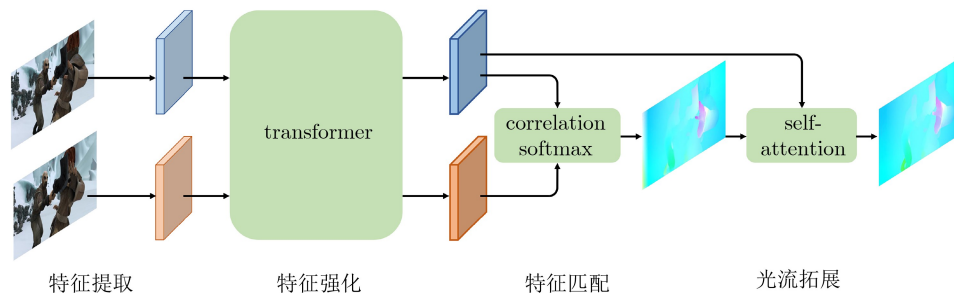


图3 GMFlow的框架图(图片改绘自文献[18])

表1 光流估计监督模型汇总

年份	模型名称	主要问题方向	突出特点
2015	<b>FlowNet<sup>[6]</sup></b>	表明端到端的CNN可以做光流估计	Encoder-decoder结构
2016	PatchBatch <sup>[15]</sup>	利用CNN提取高维特征做描述符计算匹配然后用Epic Flow(经典算法)稠密插值	使用STD结构
	<b>FlowNet2.0<sup>[9]</sup></b>	相较FlowNet提高了光流精度	用CNN提取高维描述符 对CNN结构组合进行探索
2017	<b>DCFlow<sup>[7]</sup></b>	提高了光流精度	提出4D 代价体以及CTF结构范式
	<b>PWC-Net<sup>[10]</sup></b>	相较FlowNet2.0减少了模型参数 提高了光流精度 提高了光流实时性	融合了特征金字塔 使用了扭曲(warp)层 以及标准CTF框架
2018	LiteFlowNet <sup>[11]</sup>	相较FlowNet2.0减少了网络参数	提出了对特征层 使用扭曲(warp) 结合残差网络减少参数数量 提出了级联流推断结构
	<b>IRR-PWC<sup>[12]</sup></b>	相较PWC-Net减少参数数量, 提高了精度	提出了轻量化的迭代残差细化模块 可集合遮挡预测模块提高预测精度
2020	<b>RAFT<sup>[13]</sup></b>	在CTF后处理阶段使用GRU结构进行处理, 可以有效使用上下文信息进行更新保持模型参数情况下, 有效提高了精度和处理速度	使用GRU处理4D 代价体
	<b>GMA<sup>[19]</sup></b>	通过互相关技术构建了Global Motion Aggregation(GMA)模块 优化了遮挡处理	应用互相关技术 提高了遮挡下模型的性能
2021	<b>SCV<sup>[20]</sup></b>	减少了4D 代价体的计算量 提高了计算速度与精度	通过KNN建立稀疏的4D 代价体 通过多尺度位移编码器得到稠密的光流
	AGFlow <sup>[21]</sup>	对场景整体运动的理解 遮挡问题	结合运动信息与上下文信息 结合图神经网络自适应图推理模块进行光流估计
2022	<b>GMFlowNet<sup>[22]</sup></b>	优化了大位移问题 优化了边缘	在RAFT基础上融合了基于attention机制的匹配模块
	<b>GMFlow<sup>[18]</sup></b>	基于attention技术优化了代价体与卷积固有的局部相关性限制 优化了大位移	基于transformer架构 采用全局匹配

(Encoder-Decoder)神经网络MIND来对前后两帧之间的关联关系进行学习, 并认为光流估计是这种求解关联关系的一个子问题, 其方法就是取视频流前后3帧图像, 利用第1帧和第3帧来对第2帧进行估计, 并用夏博尼尔损失(Charbonnier loss)( $\rho(x) = \sqrt{(x^2 + 2)}$ )作为损失函数(Loss)进行训练, 成功表明了自监督学习可以有效学习前后两帧之间的关联关系。

同年, Yu等人<sup>[24]</sup>提出了基于光度误差(photo-metric loss, 用以衡量经推测的光流扭曲(warp)后的第2帧和第1帧的差异)与平滑误差(smoothness loss, 衡量空间相邻光流预测之间的差异)类似FlowNet的端到端自监督训练模型Unsupervised-FlowNet, 并达到了当时KITTI数据集的最佳效果。UnFlow<sup>[25]</sup>在其误差计算的基础上, 利用交换前后两帧顺序, 送入CNN预测前后双向光流并计算误差的方法(理论上这两个光流方向是相反的), 进一步提高了训练模型的预测精度。DDFlow<sup>[26]</sup>在UnFlow的基础上, 采用了知识蒸馏结构, 其中老师网络与UnFlow类似, 但在学生网络中增加了对遮挡相关的损失函数(Loss), 可以对遮挡进行学习, 而非简单剔除。对于遮挡问题, SelFlow<sup>[27]</sup>在PWC-Net的基础上结合光度误差, 提出了一个自监督的模型, 其主要思想是首先训练一个无遮挡情况下光流预测的CNN, 而后对图像进行超像素分割(防止分割形式单一)并按分割随机分配噪声遮

挡, 用第1个CNN对光流的预测结果来指导第2个有遮挡情况下的CNN训练, 从而提高第2个模块的抗遮挡能力。

UFlow<sup>[28]</sup>综合了之前所提出的各类方法, 并对各类方法中所有组件进行了评估测试, 从而选择出最优组合, 并通过总结得到了4种模型优化方法: 代价体归一化、遮挡梯度停止、同级流分辨率下应用平滑度、调整图像大小用以自监督训练, 从而提高了模型性能。在UFlow的基础上, UPFlow<sup>[29]</sup>通过把传统上采样用到的双线性插值优化为可学习的双线性插值, 降低了以往CTF模式上采样造成的误差, 合并其提出的模型蒸馏损失, 使得光流预测的边缘更加清晰。直到目前较新的SMURF<sup>[30]</sup>实现了RAFT架构自监督的方法, 自监督光流性能已经超越之前监督模型的基准(PWC-Net和FlowNet2)。自监督模型相关进展总结如表2, 其中加粗字体为基准模型。

除监督与自监督模型之外, 还有半监督模型, 这类模型可以充分利用监督数据提高模型的精度, 同时拥有自监督模型训练数据易于获取的优点, 但从另一个角度看, 这类模型也同时有着这两类模型的缺点。其中比较有代表性的是SSFlow(Semi-Supervised, SS)<sup>[33]</sup>, 其主要思想是利用生成对抗神经网络(Generative Adversative Nets, GAN), 通过真实数据和合成数据同时训练, GAN的判断模块可以学习到合成数据与真实数据之间的域差(do-

表2 光流估计自监督模型汇总

年份	模型名称	主要问题方向	突出特点
2016	MIND <sup>[23]</sup>	自监督学习可以有效学习前后两帧之间的关联关系, 并用于初始化光流计算	自监督端到端
	<b>Unsupervised FlowNet<sup>[24]</sup></b>	得到与FlowNet近似的精度自监督模型	借鉴传统能量优化中的光度损失和平滑损失训练自监督模型
2017	DFDFlow <sup>[26]</sup>	提高了精度	利用全卷积的DenseNet自监督训练
	<b>UnFlow<sup>[25]</sup></b>	提高了精度	端到端 交换前后两帧顺序 对误差进行学习
2018	<b>Unsupervised OAFlow<sup>[31]</sup></b>	精度超越部分监督模型 优化遮挡问题	设计的对遮挡的mask层屏蔽错误匹配
	DDFlow <sup>[26]</sup>	优化遮挡问题	采用知识蒸馏, 老师网络类似UnFlow 学生网络增加针对遮挡的loss
2019	<b>SelFlow<sup>[27]</sup></b>	提高抗遮挡能力	基于PWC-Net框架
	<b>UFlow<sup>[28]</sup></b>	分析了之前自监督光流各个组件的性能以及优化组合 提高了光流估计精度	利用超像素分割随机遮挡增强数据 提出的改进方法是代价体归一化、用于遮挡估计的梯度停止、在本地流分辨率下应用平滑性, 以及用于自我监督的图像大小调整
2021	UPflow <sup>[29]</sup>	提高了精度 优化了边缘细节	用CNN改进了基于双线性插值的coarse to fine方法
	<b>SMURF<sup>[30]</sup></b>	提高精度超过PWC-Net和FlowNet2 RAFT自监督实现	金字塔蒸馏损失 基于RAFT模型 full-image warping预测真外运动 多帧自我监督, 以改善遮挡区域
2022	SEBFlow <sup>[32]</sup>	基于事件摄像机的自监督方法	采用新传感器

main gap), 从而指导生成器生成的光流更加准确, 但GAN模型往往难以训练。

以上是基于深度学习的光流计算技术的大致发展过程, 基于深度学习的模型很大程度上解决了经典算法中不满足假设1和假设2的相关的问题, 例如遮挡或物体存在变形的情况下, 则亮度恒定假设无法保证; 物体位移较大的情况下, 光流变化无法满足假设2, 因此传统基于变分与优化的方法无法满足微分条件等。而且由于计算机分布式计算性能的提升, 基于深度学习的方法在实时性上往往优于经典算法, 所以在2017年后深度学习逐渐成为光流估计方面的主流算法。

#### 4 光流测试相关数据集和性能评价指标

数据集对于训练深度学习模型非常重要, 深度学习也可以认为是由数据驱动的一种算法, 数据集的质量直接影响训练出的模型的性能, 常用的用于训练和测试的公开数据集包括: 简单但实用、利用椅子模型和随机图像生成的合成光流数据集(FlowNet验证过的)FlyingChairs<sup>[6]</sup>, 主要用于车辆

自动驾驶方面包含激光雷达和真实场景光流的KITTI Flow数据集<sup>[34,35]</sup>, 城市自动驾驶数据集HD1K, 广为使用的合成动画数据集MPI-Sintel<sup>[36]</sup>, 密集小目标行人运动数据集Crowd-Flow<sup>[37]</sup>, 可以根据需求生成虚拟数据集的无人机模拟平台AirSim以及基于动画引擎Unreal Engine 4生成的高质量虚拟驾驶场景的模拟平台Carla。光流数据集的发展促进了光流相关算法的发展, 相关模型与算法的性能可以在数据集上得到验证, 在公开数据集上的测试结果往往作为评价光流算法与模型效果的重要参考。

常用的评估数据集, 也是公认的模型评价标准, 通常有以下几个(见表3)。

在光流计算性能指标上, 5个最重要的指标分别是端点误差(End-Point Error, EPE)、平均端点误差(Average End-Point Error, AEPE)、每秒帧速率(Frame Per Second, FPS)、角度误差(Angular Error, AE)和平均角度误差(Average Angular Error, AAE)。其中EPE为估计光流和真实光流之间的欧氏距离, 用来衡量光流估计的准确程度, AE常用于评估角度误差, 二者是互补的, AE对小

表3 光流估计模型评估公开数据集

名称	年份	特点	评价指标
KITTI2012 <sup>[34]</sup>	2012		EPE, AEPE, fps, AE, AAE
KITTI2015 <sup>[35]</sup>	2015	真实驾驶场景测量数据集, 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	异常光流占的百分比(FI) 背景区域异常光流占的百分比(bg) 前景区域异常光流占的百分比(fg) 异常光流占有所有标签的百分比(all)
MPI-Sintel <sup>[36]</sup>	2012	使用开源动画电影《Sintel》由开源软件动画制作软件blender渲染出的合成数据集, 场景效果可控, 光流值准确。数据集仅截取部分典型场景, 分为Sintel Clean和Sintel Final两组, 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	整幅图像上的端点误差(EPE, AEPE) 相邻帧中可见区域的端点误差(EPE -matched) 相邻帧中不可见区域的端点误差(EPE -unmatched) 距离最近遮挡边缘在0~10像素的端点误差(d0-10) 速度在0~10像素的帧区域端点误差(s0-10) 速度在10~40像素的帧区域端点误差(s10-40)
FlyingChairs <sup>[6]</sup>	2015	合成椅子与随机背景的组合的场景, 利用仿射变换模拟运动场景, FlowNet基于此合成数据集训练	EPE, AEPE, fps, AE, AAE
Scene Flow Datasets <sup>[38]</sup>	2016	包涵Driving、Monkaa、FlyingThings3D 3个数据集, 其中FlyingThings3D组合不同物体与背景, 模拟真实3维空间运动, 以及动画Driving、Monkaa共同构成此数据集。此合成数据集除光流外还提供场景流	EPE, AEPE, fps, AE, AAE
HD1K <sup>[39]</sup>	2016	城市自动驾驶真实数据集, 涵盖了以前没有的、具有挑战性的情况, 如光线不足或下雨, 并带有像素级的不确定性	EPE, AEPE, fps, AE, AAE
AutoFlow <sup>[40]</sup>	2021	基于学习的自动化数据集生成工具, 由可学习的超参数控制数据集的生成, 提供包括准确光流、数据增强、跨数据集泛化能力等, 可以极大提高模型训练效果	在训练时使用本数据集, 测试时利用其他数据集进行测试, 评价指标同测试数据集
SHIFT <sup>[41]</sup>	2022	目前最大合成数据集包含了雾天、雨天、一天的时间、车辆人群的密集和离散, 且具有全面的传感器套件和注释, 可以提供连续域数据	EPE, AEPE, fps, AE, AAE

幅度运动误差敏感，EPE对大幅度运动误差敏感。其定义分别为

$$\left. \begin{aligned} \text{EPE} &= \sum_{i=1}^N \sqrt{(u_{\text{est}}^i - u_{\text{gt}}^i)^2 + (v_{\text{est}}^i - v_{\text{gt}}^i)^2} \\ \text{AEPE} &= \frac{\text{EPE}}{N} \end{aligned} \right\} \quad (8)$$

$$\left. \begin{aligned} \text{AE} &= \sum_{i=1}^N \arccos \\ &\cdot \left( \frac{u_{\text{est}}^i \cdot u_{\text{gt}}^i + v_{\text{est}}^i \cdot v_{\text{gt}}^i + 1}{\sqrt{(u_{\text{est}}^i)^2 + (v_{\text{est}}^i)^2 + 1} \sqrt{(u_{\text{gt}}^i)^2 + (v_{\text{gt}}^i)^2 + 1}} \right) \\ \text{AAE} &= \frac{\text{AE}}{N} \end{aligned} \right\} \quad (9)$$

其中， $(u_{\text{est}}^i, v_{\text{est}}^i)$ 表示第*i*个像素的估计光流值， $(u_{\text{gt}}^i, v_{\text{gt}}^i)$ 表示第*i*个像素光流真值，*N*为总像素个数。AEPE与AAE越小说明估计的光流值与标准值误差越小，即越准确。FPS则是衡量实时性能的指标，其值越大代表方法的实时性越强，也常用1/FPS(即处理每一帧所用时间)来衡量此项性能。

在上述指标中EPE, AEPE是最常用的比较算法准确度的性能指标，在所有数据集上通用，可以对模型整体性能进行评估，但无法衡量在某单一方面的性能。目前模型评估普遍利用的是KITTI2012, KITTI2015, Sintel Clean, Sintel Final 4个数据集。由于不同数据集数据分布以及对光流的衡量指标不一致，同一模型在不同数据集的性能也不同，但整体与表格所列顺序正相关。详细信息以及各类模型与算法准确度排名各类数据集官网都有实时更新，仅以本文介绍的部分模型及其改进型为例，准确率以Sintel与KITTI数据集官网截至成稿时公布数据为准，其性能如表4所示。

## 5 光流计算技术的具体应用

光流计算技术在实际应用时，常作为一个单独模块来使用，也可以和其他模块组合使用，准确的光流可以提供物体有效的运动、结构等信息。

在视频处理方面，常常直接利用光流信息。如在视频检测领域，常利用光流信息来进行特征聚合

以及特征在非关键帧之间传播；在视频跟踪领域，比较经典的跟踪-学习-检测(Tracking-Learning-Detection, TLD)算法<sup>[42]</sup>分为跟踪模块、学习模块、检测模块3大部分，其中的跟踪模块最早就是基于光流计算技术的经典LK算法来估计物体运动的，随着光流计算技术的发展，此框架下的光流模块也可以被更先进的模块替换；基于双流框架的视频检测与行为识别也常利用光流作为输入之一，以光流提供的物体运动信息来辅助进行行为识别；基于光流的运动信息，也可以为人面部表情识别、手势识别、动作识别等相关技术提供有效信息等。

光流在视频跟踪技术的具体应用场景也有很多，在交通监控上，对车辆、行人的异常行为进行检测，用以检测可疑滞留物、人群异常聚集、突发火情等；在体育比赛中，可以利用此项技术进行运动员跟踪；在军事领域，目标的锁定与跟踪应用更加广泛，各类基于视频的武器平台与弹药的导引头都需要视频跟踪技术，尤其是目前许多国家出现了智能化的无人武器平台，其中利用视频进行检测跟踪锁定目标已经是此类武器平台的重要组件之一。

基于光流的动作识别技术的应用场景也很广泛，如在安防监控领域，可以利用动作识别来预防公共场所的突发事件，若利用人工监控，则往往成本高、效率低；在视频检索中应用动作识别，根据视频动作分析其视频行为，进而对相关视频进行检索与推荐；在人机交互领域用动作识别技术完成人机对话，目前已经在许多游戏中广泛使用等。

在实时定位与地图构建(SLAM)中，可以利用光流信息配合相机模型的3维运动约束，经过优化算法，可以从光流中得到相机的3维位姿进而确定与其连接的物体的位姿，这种基于光流的位姿测量技术常被用于无人机、自动驾驶等领域，许多基于视觉导航的机器人也是利用此项技术进行实时导航，尤其是在纹理丰富的室内场景较为常用，与基于惯导与GPS的导航不同，基于计算机视觉的导航无需接受其他任何信号且没有惯导那种累积误差，在军事领域的具体应用场景有无人飞行器自动着陆、导弹精确导航、基于光流的目标锁定跟踪技

表4 部分模型在Sintel及KITTI数据集上的性能(截至2023年3月)

模型名称	MPI Sintel Clean(EPE-ALL)	MPI Sintel Clean排名	MPI Sintel Final(EPE-ALL)	MPI Sintel Final排名	KITTI 2012 (Avg-All)	KITTI 排名
GMFlow+	1.028	2	2.367	18	1.0	1
RATF	1.609	57	2.855	61	-	-
SMURF	3.152	143	4.183	121	1.4	20
PWC-Net	4.386	256	5.042	184	1.7	39

术、爆炸云分析、与陀螺仪结合进行弹体高度估计等应用也比较普遍等。视觉SLAM定位及点云生成效果如图4所示。

与此技术相关的还有基于光流的3维重建技术，其基本原理也是通过光流解算相机位姿，而后基于不同视角的相机位姿利用几何约束生成点云以实现物体的3维重建等，此项技术的具体应用场景包括文物3D数据录入、3D动画建模、医疗影像、3D光流(场景流)、军事战场测绘等领域。

此外，在气象预报方面，基于雷达数据的雷暴识别追踪和外推预报技术，可以利用光流替代交叉相关法，对云团等进行外推预报，提高天气预报的准确度；在医学上，3维光流可以用于器官运动估计，以及基于光流场的图像配准等应用；由于光流相关算法部分基于变分优化，在红外图像配准方面也有相关算法的应用；在军事应用方面基于光流的目标锁定跟踪技术、爆炸云分析、与陀螺仪结合进行弹体高度估计等应用也比较普遍；在火灾烟雾预防检测等领域，光流法也有着重要的应用；自动驾驶领域光流与激光雷达的组合达到了KITTI数据集光流最佳效果，这种多传感器融合也是光流计算技术应用的一个重要方面等等。

光流计算技术的进展与脑科学视觉运动感知等领域既相互交叉也相互启发，大脑对视觉进行编码的过程在某些方面与深度神经网络类似。如Mountcastle (1957), Hubel(1962)以及Wiesel (1963)等人研究发现，在大脑皮层若干区域，反应特性上表现出相似选择性的细胞聚集在一起，这与卷积神经网络中卷积核提取某一高维特征的特点类似。通过和其他动物实验数据比对发现，不同物种和皮质区域的组织类型的相似性表明存在着将方向和旋转域映射到皮层表面的普遍原则(即运动感知)，而光流计算技术可以认为是对这种映射原则的一种模拟。

除上述外光流的应用场景可以覆盖基于计算机视觉技术的大部分领域，相关算法在其他场景的应用也很多。作为计算机视觉的基本问题之一，光流

计算技术的发展有助于其所覆盖相关领域的技术进步与性能提高。

## 6 总结与发展趋势展望

本文介绍了光流相关的基础知识，总结了光流计算技术主要的技术发展路线，对技术发展过程中比较典型的算法与模型进行了简要的阐述，对相关算法的核心创新点与思路进行了归纳，对光流评估数据集与指标方面做了简要分析，并对光流的应用场景进行了简要总结。

光流计算技术结合深度学习是目前光流计算的主要发展趋势，其主要的技术方向是获得一种可以适用于任何场景快速且精确的光流估计方法，个人认为其中主流技术的发展趋势是利用更加先进的深度学习架构如图神经网络、transformer架构、3D卷积模型等，提高模型预测的准确度，强化模型的泛化能力和推理能力，解决诸如遮挡、小目标、大位移、光照、边界模糊、形变、噪声等方向的光流计算问题。

从目前光流计算的发展来看，利用深度学习模型来进行更准确、实时性更好的光流预测依旧是光流计算发展的长期目标；在基于现有深度学习的模型上进行优化，在保证精度的同时加强实时性与降低模型参数的规模，这也是基于深度学习的光流模型可以继续优化的方向；2D光流相关算法可以结合深度信息，向3D光流(场景流)方向发展，加强动态环境中对3D运动的理解；随着脑科学的发展，人类视觉机制将被进一步解析，利用仿生模拟人类视觉过程，如利用神经动力学模型结合深度学习对物体运动进行预测，也是目前重要的发展方向之一；生物视觉具有高稳定性、高适应性和低功耗等特点，是下一代人工智能算法开发的重要参考借鉴对象，生物视觉系统中存在专门处理运动信息的神经元和神经环路<sup>[44,45]</sup>，相关神经机制也得到了初步解析<sup>[46,47]</sup>，如何借鉴相关机制，开发出性能更优的类脑算法，将是光流计算领域极具潜力的发展方向之一。

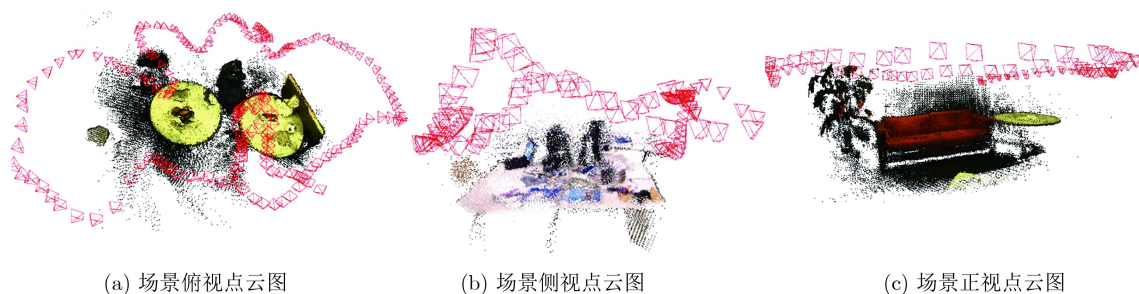


图4 光流SLAM效果图(图片出自文献[43])

## 参 考 文 献

- [1] HORN B and SCHUNCK B G. Determining optical flow[C]. SPIE 0281, Techniques and Applications of Image Understanding, Washington, USA, 1981. doi: [10.1117/12.965761](https://doi.org/10.1117/12.965761).
- [2] LUCAS B D and KANADE T. An iterative image registration technique with an application to stereo vision[C]. The 7th International Joint Conference on Artificial Intelligence, Vancouver, Canada, 1981: 674–679.
- [3] BOUGUET J Y. Pyramidal implementation of the lucas kanade feature tracker. Intel Corporation, Microprocessor Research Labs, 1999. doi: [10.1.1.185.585](https://doi.org/10.1.1.185.585).
- [4] BERTSEKAS D P. Constrained Optimization and Lagrange Multiplier Methods[M]. Boston: Academic Press, 1982: 724.
- [5] SONG Xiaojing. A Kalman filter-integrated optical flow method for velocity sensing of mobile robots[J]. *Journal of Intelligent & Robotic Systems*, 2014, 77(1): 13–26.
- [6] DOSOVITSKIY A, FISCHER P, ILG E, *et al.* FlowNet: Learning optical flow with convolutional networks[C]. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015: 2758–2766. doi: [10.1109/iccv.2015.316](https://doi.org/10.1109/iccv.2015.316).
- [7] XU Jia, RANFTL R, and KOLTUN V. Accurate optical flow via direct cost volume processing[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017: 5807–5815. doi: [10.1109/cvpr.2017.615](https://doi.org/10.1109/cvpr.2017.615).
- [8] CHEN Qifeng and KOLTUN V. Full flow: Optical flow estimation by global optimization over regular grids[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 4706–4714. doi: [10.1109/CVPR.2016.509](https://doi.org/10.1109/CVPR.2016.509).
- [9] ILG E, MAYER N, SAIKIA T, *et al.* FlowNet 2.0: Evolution of optical flow estimation with deep networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017: 1647–1655. doi: [10.1109/cvpr.2017.179](https://doi.org/10.1109/cvpr.2017.179).
- [10] SUN Deqing, YANG Xiaodong, LIU Mingyu, *et al.* PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8934–8943. doi: [10.1109/CVPR.2018.00931](https://doi.org/10.1109/CVPR.2018.00931).
- [11] HUI T W, TANG Xiaou, and LOY C C. LiteFlowNet: A lightweight convolutional neural network for optical flow estimation[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8981–8989. doi: [10.1109/cvpr.2018.00936](https://doi.org/10.1109/cvpr.2018.00936).
- [12] HUR J and ROTH S. Iterative residual refinement for joint optical flow and occlusion estimation[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019: 5747–5756. doi: [10.1109/CVPR.2019.00590](https://doi.org/10.1109/CVPR.2019.00590).
- [13] TEED Z and DENG Jia. RAFT: Recurrent all-pairs field transforms for optical flow[C]. 16th European Conference on Computer Vision, Glasgow, UK, 2020: 402–419. doi: [10.1007/978-3-030-58536-5\\_24](https://doi.org/10.1007/978-3-030-58536-5_24).
- [14] REVAUD J, WEINZAEPFEL P, HARCHAOUI Z, *et al.* EpicFlow: Edge-preserving interpolation of correspondences for optical flow[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 1164–1172. doi: [10.1109/cvpr.2015.7298720](https://doi.org/10.1109/cvpr.2015.7298720).
- [15] GADOT D and WOLF L. PatchBatch: A batch augmented loss for optical flow[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 4236–4245. doi: [10.1109/cvpr.2016.459](https://doi.org/10.1109/cvpr.2016.459).
- [16] THEWLIS J, ZHENG Shuai, TORR P H S, *et al.* Fully-trainable deep matching[C]. Proceedings of the British Machine Vision Conference 2016, York, UK, 2016.
- [17] WANNENWETSCH A S and ROTH S. Probabilistic pixel-adaptive refinement networks[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2020: 11639–11648. doi: [10.1109/cvpr42600.2020.01166](https://doi.org/10.1109/cvpr42600.2020.01166).
- [18] XU Haofei, ZHANG Jing, CAI Jianfei, *et al.* GMFlow: Learning optical flow via global matching[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022: 8111–8120. doi: [10.1109/cvpr52688.2022.00795](https://doi.org/10.1109/cvpr52688.2022.00795).
- [19] JIANG Shihao, CAMPBELL D, LU Yao, *et al.* Learning to estimate hidden motions with global motion aggregation[C]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 2021: 9752–9761. doi: [10.1109/iccv48922.2021.00963](https://doi.org/10.1109/iccv48922.2021.00963).
- [20] JIANG Shihao, LU Yao, LI Hongdong, *et al.* Learning optical flow from a few matches[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021: 16587–16595. doi: [10.1109/cvpr46437.2021.01632](https://doi.org/10.1109/cvpr46437.2021.01632).
- [21] LUO Ao, YANG Fan, LUO Kunming, *et al.* Learning optical flow with adaptive graph reasoning[C]. Thirty-Sixth AAAI Conference on Artificial Intelligence, Vancouver Canada, 2021.
- [22] ZHAO Shiyu, ZHAO Long, ZHANG Zhixing, *et al.* Global matching with overlapping attention for optical flow estimation[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022: 17571–17580. doi: [10.1109/CVPR52688.2022.01707](https://doi.org/10.1109/CVPR52688.2022.01707).

- [23] LONG Gucan, KNEIP L, ALVAREZ J M, *et al.* Learning image matching by simply watching video[C]. 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 434–450. doi: [10.1007/978-3-319-46466-4\\_26](https://doi.org/10.1007/978-3-319-46466-4_26).
- [24] YU J J, HARLEY A W, and DERPANIS K G. Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness[C]. European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 3–10. doi: [10.1007/978-3-319-49409-8\\_1](https://doi.org/10.1007/978-3-319-49409-8_1).
- [25] MEISTER S, HUR J, and ROTH S. UnFlow: Unsupervised learning of optical flow with a bidirectional census loss[C]. Thirty-Sixth AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018. doi: [10.1609/aaai.v32i1.12276](https://doi.org/10.1609/aaai.v32i1.12276).
- [26] LIU Pengpeng, KING I, LYU M R, *et al.* DDFlow: Learning optical flow with unlabeled data distillation[C]. Thirty-Sixth AAAI Conference on Artificial Intelligence, Honolulu, USA, 2019: 8770–8777. doi: [10.1609/aaai.v33i01.33018770](https://doi.org/10.1609/aaai.v33i01.33018770).
- [27] LIU Pengpeng, LYU M, KING I, *et al.* SelFlow: Self-supervised learning of optical flow[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019: 4566–4575. doi: [10.1109/cvpr.2019.00470](https://doi.org/10.1109/cvpr.2019.00470).
- [28] JONSKOWSKI R, STONE A, BARRON J T, *et al.* What matters in unsupervised optical flow[C]. 16th European Conference on Computer Vision, Glasgow, UK, 2020: 557–572. doi: [10.1007/978-3-030-58536-5\\_33](https://doi.org/10.1007/978-3-030-58536-5_33).
- [29] LUO Kunming, WANG Chuan, LIU Shuaicheng, *et al.* UPFlow: Upsampling pyramid for unsupervised optical flow learning[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021: 1045–1054. doi: [10.1109/cvpr46437.2021.00110](https://doi.org/10.1109/cvpr46437.2021.00110).
- [30] STONE A, MAURER D, AYVACI A, *et al.* SMURF: Self-teaching multi-frame unsupervised RAFT with full-image warping[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021: 3886–3895. doi: [10.1109/cvpr46437.2021.00388](https://doi.org/10.1109/cvpr46437.2021.00388).
- [31] WANG Yang, YANG Yi, YANG Zhenheng, *et al.* Occlusion aware unsupervised learning of optical flow[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 4884–4893. doi: [10.1109/cvpr.2018.00513](https://doi.org/10.1109/cvpr.2018.00513).
- [32] SHIBA S, AOKI Y, and GALLEGGO G. Secrets of event-based optical flow[C]. 17th European Conference on Computer Vision, Tel Aviv, Israel, 2022. doi: [10.1007/978-3-031-19797-0\\_36](https://doi.org/10.1007/978-3-031-19797-0_36).
- [33] LAI Weisheng, HUANG Jiabin, and YANG M H. Semi-supervised learning for optical flow with generative adversarial networks[C]. The 31st International Conference on Neural Information Processing Systems, Long Beach, USA, 2018: 892–896.
- [34] GEIGER A, LENZ P, and URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA, 2012: 3354–3361. doi: [10.1109/cvpr.2012.6248074](https://doi.org/10.1109/cvpr.2012.6248074).
- [35] KENNEDY R and TAYLOR C J. Optical flow with geometric occlusion estimation and fusion of multiple frames[C]. The 10th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, Hong Kong, China, 2015: 364–377. doi: [10.1007/978-3-319-14612-6\\_27](https://doi.org/10.1007/978-3-319-14612-6_27).
- [36] BUTLER D J, WULFF J, STANLEY G B, *et al.* A naturalistic open source movie for optical flow evaluation[C]. The 12th European Conference on Computer Vision, Florence, Italy, 2012: 611–625. doi: [10.1007/978-3-642-33783-3\\_44](https://doi.org/10.1007/978-3-642-33783-3_44).
- [37] MAYER N, ILG E, HÄUSSER P, *et al.* A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]. The 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 4040–4048. doi: [10.1109/CVPR.2016.438](https://doi.org/10.1109/CVPR.2016.438).
- [38] SCHRÖDER G, SENST T, BOCHINSKI E, *et al.* Optical flow dataset and benchmark for visual crowd analysis[C]. 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018: 1–6. doi: [10.1109/AVSS.2018.8639113](https://doi.org/10.1109/AVSS.2018.8639113).
- [39] KONDERMANN D, NAIR R, HONAUER K, *et al.* The HCI benchmark suite: Stereo and flow ground truth with uncertainties for urban autonomous driving[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, USA, 2016: 19–28. doi: [10.1109/CVPRW.2016.10](https://doi.org/10.1109/CVPRW.2016.10).
- [40] SUN Deqing, VLASIC D, HERRMANN C, *et al.* AutoFlow: Learning a better training set for optical flow[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021: 10088–10097. doi: [10.1109/cvpr46437.2021.00996](https://doi.org/10.1109/cvpr46437.2021.00996).
- [41] SUN Tao, SEGU M, POSTELS J, *et al.* SHIFT: A synthetic driving dataset for continuous multi-task domain adaptation[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022: 21339–21350. doi: [10.1109/CVPR52688.2022.02068](https://doi.org/10.1109/CVPR52688.2022.02068).
- [42] KALAL Z, MIKOLAJCZYK K, and MATAS J. Tracking-learning-detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1409–1422. doi: [10.1109/TPAMI.2011.239](https://doi.org/10.1109/TPAMI.2011.239).

- [43] YE Weicai, YU Xingyuan, LAN Xinyue, *et al.* DeFlowSLAM: Self-supervised scene motion decomposition for dynamic dense SLAM[EB/OL]. <https://arxiv.org/abs/2207.08794>, 2022.
- [44] SANES J R and MASLAND R H. The types of retinal ganglion cells: Current status and implications for neuronal classification[J]. *Annual Review of Neuroscience*, 2015, 38(1): 221–246. doi: [10.1146/annurev-neuro-071714-034120](https://doi.org/10.1146/annurev-neuro-071714-034120).
- [45] GALLETTI C and FATTORI P. The dorsal visual stream revisited: Stable circuits or dynamic pathways?[J]. *Cortex*, 2018, 98: 203–217. doi: [10.1016/j.cortex.2017.01.009](https://doi.org/10.1016/j.cortex.2017.01.009).
- [46] WEI Wei. Neural mechanisms of motion processing in the mammalian retina[J]. *Annual Review of Vision Science*, 2018, 4: 165–192. doi: [10.1146/annurev-vision-091517-034048](https://doi.org/10.1146/annurev-vision-091517-034048).
- [47] ROSSI L F, HARRIS K D, and CARANDINI M. Spatial connectivity matches direction selectivity in visual cortex[J]. *Nature*, 2020, 588(7839): 648–652. doi: [10.1038/s41586-020-2894-4](https://doi.org/10.1038/s41586-020-2894-4).
- 崔毅博：男，博士生，研究方向目标检测与跟踪、运动感知等。  
汤仁东：男，博士，副研究员，研究方向为非人灵长类视觉神经机制解析。  
邢大军：男，博士，研究员，研究方向为视觉认知与计算。  
王 隽：女，博士，副研究员，研究方向为图像质量增强、视觉目标检测和强化学习。  
李尚生：男，硕士，教授，研究方向为雷达成像等。
- 责任编辑：余 蓉