徐煦,马鹏飞,司建军,等.基于视频帧间局部相关信息的光流估计网络[J].辽宁工程技术大学学报(自然科学版),2025,44(1):120-128.doi:10.11956/j.issn.1008-0562.20230491
XU Xu,MA Pengfei,SI Jianjun,et al.Optical flow estimation via fusing sequence image intensity correlation information[J].Journal of Liaoning Technical University(Natural Science),2025,44(1):120-128.doi:10.11956/j.issn.1008-0562.20230491

基于视频帧间局部相关信息的光流估计网络

徐　煦，马鹏飞，司建军，高国军

（国能宝日希勒能源有限公司 设备维修中心,内蒙古 呼伦贝尔 021500）

摘　要：为解决光流估计网络在目标边缘分割、运动速度和运动方向不准确的问题，提出基于视频帧间局部相关信息的光流估计网络。运用特征编码器从图像中提取出编码特征，通过上下文网络获取图像的上下文特征。采用下采样处理减小特征尺寸提高计算效率。根据连续两帧光流图像位移较小的特性，提出一种分区计算视觉相似度的方法，构建更为精细的4D相关体。采用残差滤波器和相似卷积块的方法，分别针对相关体和光流信息进行操作，更有效地保留局部微小位移信息。研究结果表明：采用基于视频帧间局部相关信息的光流估计网络进行计算，端点误差分别实现了8.0%和5.7%的优化，显著提升了光流估计的准确性，对复杂场景下光流信息提取更准确。研究结果可为自动驾驶、智能安防等领域提供参考。

关键词：计算机视觉；光流估计；深度学习；区域匹配；迭代更新

中图分类号：TP391.4　　**文献标志码：**A　　**文章编号：**1008-0562(2025)01-0120-09

Optical flow estimation via fusing sequence image intensity correlation information

XU Xu, MA Pengfei, SI Jianjun, GAO Guojun

(Equipment Maintenance Center, Guoneng Baorixile Energy Company Limited, Hulun Buir 021500, China)

Abstract: To address the challenges associated with inaccurate target edge segmentation, motion speed, and motion direction, this paper introduces an optical flow estimation network that leverages local correlation information between video frames. Initially, the network employs a feature encoder to extract encoding features from the image and capture contextual information through a context network. Subsequently, the feature size is reduced through downsampling to enhance computational efficiency. Given the minute displacement of the optical flow image across consecutive frames, a partition-based visual similarity computation method is proposed to construct a more refined 4D correlation volume. Residual filters and similar convolution blocks are utilized for processing the correlation volume and optical flow information, respectively, ensuring the preservation of local small displacement details. The research results show that the optical flow estimation network based on the local correlation information between video frames has achieved optimizations of 8.0% and 5.7% respectively in the optical flow estimation evaluation metric (endpoint error, EPE). This significantly improves the accuracy of optical flow estimation and effectively alleviates the problem of inaccurate optical flow information extraction in complex scenarios. The research conclusions provide references for fields such as autonomous driving and intelligent security.

Key words: computer vision; optical flow estimation; deep learning; regional matching; iterative update

收稿日期：2023-11-17　**修回日期：**2024-03-09　**接受日期：**2024-03-21　**责任编辑：**朱含露
基金项目：国家自然科学基金项目（61601213）
作者简介：徐　煦（1983-），男，辽宁 葫芦岛人，硕士，高级工程师，主要从事机电设备检修、管理和智能化研究、矿山领域智能化等方面的研究。E-mail：xuxuxu510@163.com

0　引言

光流估计是计算机视觉领域的关键问题之一，广泛应用于动作识别[1]、视频理解[2]、自动驾驶[3]等领域。在处理光流估计问题时，传统方法基于亮度一致性和空间平滑度能量最小化的方式获得光流估计结果。HORN等[4]引入光流变分框架，将求解光流场的问题转化成求解最小能量函数的问题。BLACK等[5]构建鲁棒估计框架，对违反空间平滑和亮度恒定约束的情况进行综合处理。通过此框架，在面临噪声、遮挡以及非平稳光照等复杂条件时，光流估计运算更精确。但是该框架所采用的二次方式惩罚偏差具有明显局限性。基于此，ZACH等[6]提出采用L1数据项与总变差正则化替代原有的二次方式惩罚偏差，解除对运动连续性的束缚，同时实现对异常值的高效处理。为进一步优化该框架的性能，WEINZAEPFEL等[7]运用能量最小化方法和描述符匹配策略，融合匹配算法与变分法相，构建层次化的区域结构，引入高阶正则化项提升光流估计的准确性和鲁棒性。

近年来，深度学习在光流估计领域取得了显著进展，经过充分训练的神经网络能够直接预测帧间光流，有效避免传统算法复杂优化问题，但生成结果仍存在模糊和噪声问题。SIMONYAN等[8]通过引入变分方法，实现算法性能显著的性能提升。ILG等[9]采用多网络叠加策略提升算法性能。RANJAN等[10]将传统算法中的金字塔理念与光流估计相结合，采用由粗到细的估计方式解决光流估计过程中的大、小位移问题。SUN等[11]在此基础上引入相关体处理算法，实现了网络性能提升和端到端的训练方式。YANG等[12]通过引入4D卷积策略对相关体处理算法进行改进，显著提高了光流估计的准确性。HUI等[13]提出一种级联预测光流与特征正则化的方法，进一步优化光流估计的性能。以上深度学习的方法都采用了由粗到细的金字塔迭代优化策略。然而，该策略存在弊端，即快速移动的小物体在粗级别可能会消失。RAFT（recurrent all-pairs field transforms for optical flow）方法选择保持并更新单一的高分辨率光流场，这种方法在处理快速运动的小物体上表现出了明显的优势，为进一步提高光流估计的准确性和稳定性提供了新的思路。

RAFT方法[14]结构清晰，泛化能力强，所生成的光流图像清晰度高，在对KITTI-2015数据集[15]和MPI-Sintel数据集[16]的计算中表现出众。但RAFT方法也存在以下不足：首先，RAFT方法从所有像素的完全相关性来计算视觉相似性。但在大多数光流场景中，连续帧之间的像素位移较小，直接进行全相关计算会引入较多计算误差。其次，迭代更新依赖于卷积滤波器融合光流[17]和相关体积[18]，这使最终的光流预测图过于平滑，影响了一些关键细节和特征的呈现。为解决RAFT方法存在的不足，本文提出基于视频帧间局部相关信息的光流估计网络。该网络先对输入的连续两帧特征图进行分区处理，以强弱相关的方式计算稠密的视觉相似度，以此为基础建立更为精细的四维相关体积。在迭代更新阶段，采用残差滤波器[19]和相似卷积块的方法，尽可能地保留更多的局部小位移信息。通过评价指标，对比实验、消融实验验证该方法的有效性和优越性。

1　网络结构设计

光流估计网络结构见图1，该网络主要由特征提取模块、视觉相似度计算模块，以及迭代更新模块构成。这些模块共同协作，实现精确的光流估计。在特征提取模块中，通过特征编码器对两帧连续的图像F_1与F_2进行处理，得到特征图像对I_1和I_2。在计算视觉相似度模块中，利用光流场景中连续两帧光流图像位移相对较小的特性，在第二张特征图上预先为第一张特征图中的每个特征像素圈定几个可能的位移区域，建立“像素-区域”映射关系，并为该映射关系自适应选取相关因子。根据这种强弱相关关系，计算出两张特征图上所有像素间的视觉相似性。这种方法可以避免直接进行全像素相关计算而引起的误差。在迭代更新模块，采用残差滤波器[19]和相似卷积块构成的基准编码模块（图1中标识模块B），该模块在关注局部小位移信息的同时不会产生过于平滑的效果。

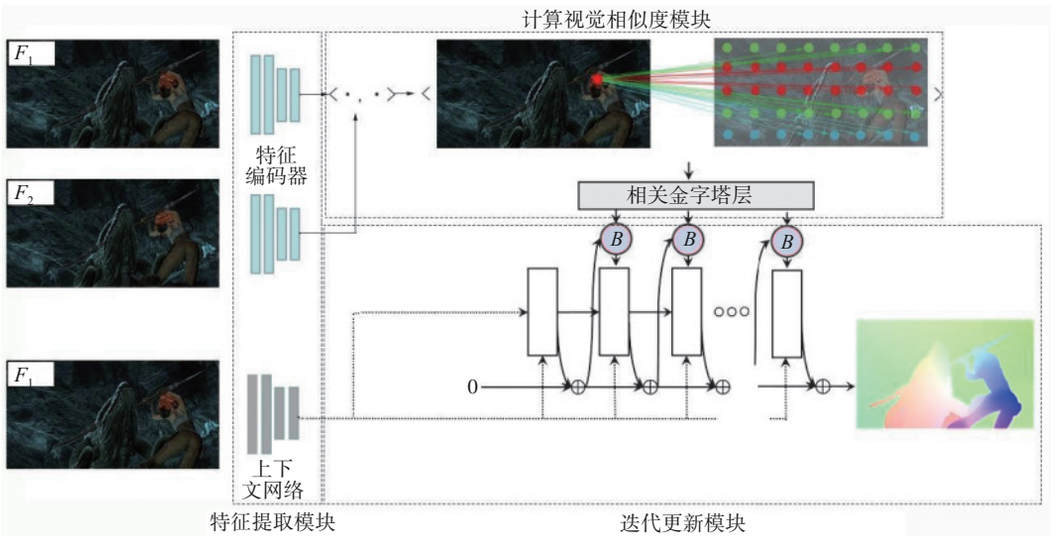


图1 光流估计网络结构

Fig.1 optical flow estimation network structure

1.1 特征提取模块

特征提取模块包括特征编码器和上下文网络两个部分。特征编码器$g(\theta)$的实现主要依托残差结构块，残差结构块见图2。

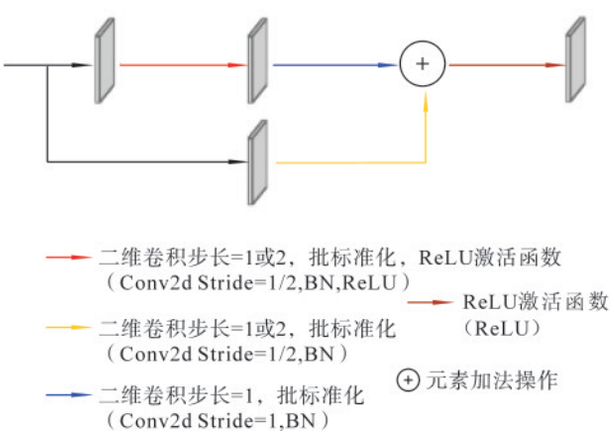


图2 残差结构块

Fig.2 residual block

特征编码器$g(\theta)$以权重共享的方式对连续两帧图像F_1、F_2进行处理，并以1/8分辨率输出编码后的特征图像对I_1，$I_2\in\mathbf{R}^{(H/8)\times(W/8)\times D}$，其中，$H$为特征图的高度，$W$为特征图的宽度，$D$为维度。

上下文网络与特征编码器的结构一致，但仅用于提取F_1图像特征信息，最终输出编码后的特征$I_3\in\mathbf{R}^{(H/8)\times(W/8)\times D}$。

1.2 视觉相似度计算模块

光流估计的核心在于计算视觉相似度，在整体架构中不可或缺。然而，RAFT方法采用特征向量内积的计算方式，忽视了光流场景中像素位移相对较小的特性，因此引入了较多的误差信息。为解决这一问题，提出一种基于分区思想的强弱相关计算方法。

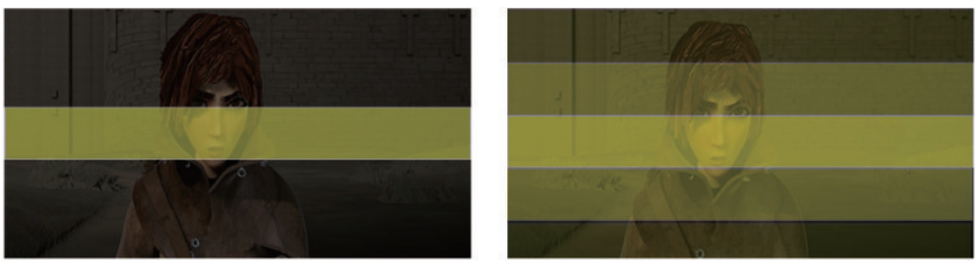
视觉相似度计算模块将图像划分为多个区域，以减少误匹配。设定像素位移的阈值，圈定两帧图像中的相关区域。在两个特征图像对中，相应区域的像素值表现出高度相关性，因此将相关因子设为1。对于非对应区域的像素，将根据像素间的距离自适应地确定相关因子。特征图相关性强弱见图3。图3中，颜色深浅代表相关性强弱，颜色越深相关性越强；颜色越浅相关性越弱。该方法不仅获得了所有像素对的相关性，而且显著减少了区域外错误匹配造成的误差。相关体积为

$$C_{\mathrm{v}}=\bigcup_{\substack{A=A_1A_2\ldots A_n,\\ B=B_1B_2\ldots B_n}} C_{AB}, \tag{1}$$

$$\begin{cases} I_1=\bigcup\limits_{A\in(1,2n)}\mathrm{region}(A) \\ I_2=\bigcup\limits_{B\in(1,2n)}\mathrm{region}(B) \end{cases}, \tag{2}$$

$$\begin{cases} C_{AB}=\lambda_k\cdot\mathrm{region}(A)\otimes\mathrm{region}(B) \\ B=J(A) \end{cases}。 \tag{3}$$

根据式（1）~式（3）可知，输入的特征图I_1、I_2被分为$2n$个区域，A和B分别为I_1和I_2分区的区域索引，$J(\cdot)$为帧间各个区域间的映射关系；λ_k为可学习的自适应相关因子，表示各区域间的相关性，采用ReLU激活函数将λ_k取值范围约束在0到1之间；region$(\cdot)$为分区后的图像区域；$\otimes$表示特征图之间对应区域像素的点积；C_{AB}表示区域A、B建立强弱关系后，它们之间的相关体积。



（a）特征图I_1　　（b）特征图I_2

图3　分区映射的相关性强弱

Fig.3　strength of partition mapping

由式（1）~式（3）可计算出I_1中每个特征像素与I_2中所有特征像素的相关关系，即相关体积C_v，维度为$w\times h\times w\times h$，其中$(w\times h)=(W/8,H/8)$。随后运用尺寸为1、2、4、8的四个不同卷积核，对相关体的后两个维度实施降采样处理。经过此步骤可获得四层金字塔结构$\{C_v^1,C_v^2,C_v^3,C_v^4\}$，即四维相关体。这种做法能够保存高分辨率的信息，更好地计算快速移动的小物体的运动。金字塔层标号k与其维度关系为

$$C_v^k\to h\times w\times(h/2^k)\times(w/2^k)。\tag{4}$$

由式（4）可知，相关金字塔的每一层均保持前两个维度不变，对后面的两个维度进行降采样，这种操作既可以保存图片高分辨率信息又可以完成小位移运动的追踪。

基于$\{C_v^1,C_v^2,C_v^3,C_v^4\}$定义查询操作，用于光流的迭代更新。设上一次迭代计算得到的在x和y两个方向上的光流分别为$(\boldsymbol{f}^1,\boldsymbol{f}^2)$，其中$\boldsymbol{f}$为包含所有像素点的光流信息矩阵，通过$(\boldsymbol{f}^1,\boldsymbol{f}^2)$可得$I_1$图上像素点$x=(u,v)$在$I_2$上对应位置$x'$，$x'=(u+f^1(u),v+f^2(v))$，其中$u$和$v$分别为每个像素点在$x$和$y$两个方向上的坐标。$x'$邻域点集$L(x')_r$为

$$L(x')_r=\{x'+\mathrm{d}x|\mathrm{d}x\in\mathbf{Z}^2,\|\mathrm{d}x\|\leqslant r\},\tag{5}$$

式中：$\mathrm{d}x$为整数；r为4像素的搜索半径。

将该邻域内的所有点视为I_1中像素x在I_2上潜在位置。在进行光流的后续迭代计算时，若需获取像素点x与其潜在位置的相关特征信息，可直接通过插值查询方式从相关金字塔$\{C_v^1,C_v^2,C_v^3,C_v^4\}$中提取相关数据，将来自不同金字塔层的查询结果在特征维度上进行整合，得到最终的相关体C_{final}。

1.3　迭代更新模块

基于视频帧间局部相关信息的光流估计网络，通过迭代更新模块中的光流序列$\{\boldsymbol{f}_1,\cdots,\boldsymbol{f}_N\}$完成光流估计，其迭代更新过程可以描述为

$$\boldsymbol{f}_{k+1}=\boldsymbol{f}_k+\Delta\boldsymbol{f}_k,\tag{6}$$

式中：$\Delta\boldsymbol{f}_k$为每次迭代后的更新量；$\boldsymbol{f}_k$为当前光流；$\boldsymbol{f}_{k+1}$为更新后的光流；k为迭代更新次数，最终输出光流序列$\{\boldsymbol{f}_1,\cdots,\boldsymbol{f}_N\}$完成光流估计。

基准编码模块结构设计见图4，其输入由两部分组成：一是基于当前光流位移在相关金字塔中查询出的相关体C_{final}；二是当前的光流位移值。

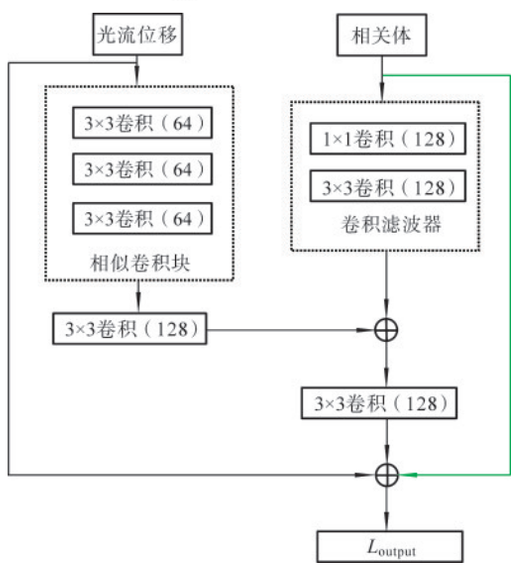


图4　基准编码模块结构设计

Fig.4　basic encoder module structure design

当前光流位移$\boldsymbol{f}_k$的处理，采用细粒度模块进行运算。此模块由3个并行的3×3卷积核组成。这3个并行的小卷积核不仅成功地解决了感受野过小的问题，而且在提高局部小运动特征关注度方面也取得了显著效果。具体处理方法为

$$F_{\text{flow}}^i=\mathrm{ReLU}\left(\mathrm{Conv}_{3\times3}\left(\boldsymbol{f}_k\right)\right),i=1,2,3\tag{7}$$

$$F_{\text{export}}=\mathrm{cat}\left(F_{\text{flow}}^1,F_{\text{flow}}^2,F_{\text{flow}}^3\right),\tag{8}$$

式中：F_{flow}^i（i=1,2,3）为光流位移$\boldsymbol{f}_k$通过3个并行卷积核的不同输出通道处理得到的光流特征，通过拼接操作cat(·)进行合并，生成光流特征拼接后的结果F_{export}。最后，通过ReLU激活函数进行处理。

经过卷积滤波器和一个3×3卷积操作，相关体C_{final}已成功整合了不同相关金字塔层的关键信息。然而，基于卷积的基准编码模块在生成输出结果时过于平滑，影响了一些关键细节和特征的呈现。为了解决这个问题，引入残差连接的设计（如图4中绿色线所示），缓解小位移运动在卷积堆叠运算过程中难以维持局部精细特征的问题，并纠正局部小位移的运动。经过引入残差连接，输出的光流预测图在局部细节上得到了显著提升。改进后的卷积滤波器输出为

$$C_{\text{export}} = C_{\text{final}} + \text{Conv}_{3\times3}\left(\text{Conv}_{3\times3}\left(F_{\text{export}}\right) + \text{Conv}_{3\times3}\left(\text{Conv}_{1\times1}\left(C_{\text{final}}\right)\right)\right), \tag{9}$$

式中：$\text{Conv}_{3\times3}$为3×3卷积核；$\text{Conv}_{1\times1}$为1×1卷积核。

卷积滤波器的输出C_{export}和当前光流位移$\boldsymbol{f}_k$在特征维度上进行拼接成为基准编码模块的输出L_{output}。将L_{output}编码后的特征I_3在特征维度合并作为ConvGRU的输入，完成光流的迭代更新。该方法能够精准捕捉到更细致的特征信息，使预测结果更加贴近实际光流情况，提升输出光流场的精确性与细致度。此外，鉴于特征提取阶段输出特征图的分辨率仅为原图的1/8，因而在迭代更新过程中生成的初始光流预测图亦维持这一较低分辨率。为获得与原图相匹配的高分辨率光流场景，采用上采样操作进行必要的处理。

1.4 算法伪代码描述

输入：连续两帧图像（image1, image2）

输出：光流场（flow_predictions）

步骤1 对每一对图像帧进行处理。

```
for frame1, frame2 in image_pairs:
# 使用特征提取模块提取图像帧的特征
features1 = feature_extraction(frame1)
features2 = feature_extraction(frame2)
# 对特征进行归一化处理 normalize_features(features1)
normalize_features(features2)
```

步骤2 对图像进行分区处理（四分条为例）。

```
fmap1, fmap2, region1, region2, region3, region4 = self.fnet([image1, image2])
fmap1 = fmap1.float()
fmap2 = fmap2.float()
region1 = region1.float()
region2 = region2.float()
region3 = region3.float()
region4 = region4.float()
#计算视觉相似度构建4D相关体
corr_fn = CorrBlock(fmap1, fmap2, region1, region2, region3, region4,radius=self.args.corr_radius)
corr = CorrBlock. corr(fmap1, fmap2, region1, region2, region3, region4)
batch, h1, w1, dim, h2, w2 = corr.shape
corr = corr.reshape(batch*h1*w1, dim, h2, w2)
```

步骤3 更新光流。

```
# 初始化光流的坐标信息
coords0, coords1 = self.initialize_flow(image1)
if flow_init is not None:
coords1 = coords1 + flow_init
#进行光流更新迭代
flow_predictions = []
net, up_mask, delta_flow = self. update_block(net, inp, corr, flow)
#基准编码模块结构
self.encoder = BasicMotionEncoder(args)
self. gru = SepConvGRU(hidden_dim=hidden_dim, input_dim=128+hidden_dim)
self. flow_head = FlowHead(hidden_dim, hidden_dim =256)
coords1 = coords1 + delta_flow # 更新光流
#上采样
flow_up = self. upsample_flow(coords1 - coords0, up_mask)
#输出光流
return flow_predictions
```

2 实验分析

2.1 实验环境及参数设置

网络架构采取端到端的处理方式，对数据集实施标准的数据增强技术，包括随机噪声添加和随机翻转等操作，并通过单次训练流程完成样本集的构建。在实施环境方面，选择PyTorch框架与Adamw优化器相结合，设定Adamw优化器的参数wdecay为10^{-5}。同时，采用NVIDIA 3090显卡作为计算资源，设置批量大小为5。在训练过程中，针对KITTI-2015数据集进行5×10^4轮的训练迭代，对MPI Sintel数据集则进行1.2×10^6轮的训练迭代。初始学习率设定为10^{-4}，并每经过5 000轮训练后，学习率递减10^{-5}。

采用地面真实流与预测值之间的距离作为监督网络的依据，损失函数为

$$F_{\text{Loss}}=\sum_{i=1}^{k}\varphi^{i-k}\left\|\boldsymbol{F}_g-\boldsymbol{f}_i\right\|_1, \tag{10}$$

式中：初始化光流$\boldsymbol{f}_1=0$；$\boldsymbol{f}_i$为光流序列，即$\{\boldsymbol{f}_1,\cdots,\boldsymbol{f}_N\}$；$\boldsymbol{F}_g$为地面真实流；$\varphi$为初始权重，取0.8；$k$为迭代次数，取12。

2.2　数据集

KITTI-2015数据集包含394组训练数据集和395组测试数据集。该数据集为真实的交通环境下的街景数据集。

MPI-Sintel光流数据集包含1 041组训练数据和552组测试数据，内容全面且丰富。该数据库被划分为Clean类和Final类。Clean类数据集针对大位移、弱纹理以及非刚性大形变等挑战性场景进行设计，主要目标在于测试算法在复杂多变环境下的性能表现。Final类数据集在Clean类数据集的基础上，通过融入运动模糊、雾化效果，以及图像噪声等多种元素，使数据集更加贴近真实世界的场景，从而更加客观地评估算法在实际应用中的表现。

2.3　评价指标

针对KITTI-2015数据集，采用端点误差（end-point-error，EPE）和光流异常值百分比（Fl）来评估光流估计的精准度。EPE通过计算所有像素点真实标签值与预测光流之间的平均欧氏距离，反映预测光流与真实光流之间的偏差程度。Fl为图像整体区域中光流误差大于3像素或超过5%的误差比率。

EPE可表示为

$$\text{EPE}=\sqrt{\sum_{i=1}^{n}\left(F_i-F_{gi}\right)^2}, \tag{11}$$

式中：F_i代表预测光流值；F_{gi}代表地面真实值。

在MPI-Sintel数据集上，采用EPE以及1px、3px、5px作为评估指标。其中，1px指标统计EPE值小于1的像素占比，体现高精度层面光流估计的准确性；3px指标统计EPE值小于3的像素占比，从稍宽但合理的误差区间评估整体效果；5px指标统计EPE值小于5的像素占比，从更大误差容忍区间考查像素情况，帮助把握不同精度层次的表现。这些指标从不同误差范围维度考量MPI-Sintel数据集上的相关表现，为光流估计网络提供全面精确的性能评估依据。

2.4　实验结果对比

不同方法在KITTI-2015数据集、MPI-Sintel测试集上光流估计性能见表1、表2，其中Ours(4)、Ours(6)、Ours(8)代表4分条区域、6分条区域以及8分条区域。

表1　不同方法在KITTI-2015数据集上光流估计性能

Tab.1　optical flow estimation performance of the different methods on the KITTI-2015 dataset

方法	EPE	Fl
VCN[12]	1.409 878	2.660 123
DICL[17]	1.318 453	2.645 445
RAFT[14]	0.770 121	2.149 829
Ours(4)	0.719 823	1.849 874
Ours(6)	0.709 878	1.843 491
Ours(8)	0.707 865	1.843 123

表2　不同方法在MPI-Sintel测试集上的光流估计性能

Tab.2　performance of optical flow estimation of the different methods on the MPI-Sintel test set

方法	Clean类				Final类			
	EPE	1px	3px	5px	EPE	1px	3px	5px
VCN[12]	1.288 978	0.819 898	0.877 893	0.878 943	1.763 459	0.813 234	0.883 565	0.903 231
DICL[17]	0.978 794	0.884 623	0.939 874	0.939 847	1.383 832	0.858 965	0.924 783	0.944 232
RAFT[14]	0.893 562	0.903 938	0.949 875	0.968 985	1.279 886	0.863 457	0.938 973	0.953 453
Ours（4）	0.853 432	0.911 298	0.957 985	0.969 438	1.244 548	0.873 354	0.941 097	0.960 023
Ours（6）	0.851 023	0.911 879	0.958 933	0.971 232	1.243 341	0.872 449	0.941 123	0.960 220
Ours（8）	0.842 398	0.912 012	0.959 012	0.972 334	1.239 996	0.873 453	0.942 022	0.960 522

以RAFT[16]为例，EPE的降低比率为

$$\eta_{\text{EPE}}=\frac{\left(\text{EPE}_{\text{RAFT}}-\text{EPE}_{\text{Our}(n)}\right)}{\text{EPE}_{\text{RAFT}}}\times 100\%, \tag{13}$$

Fl的降低百分点为

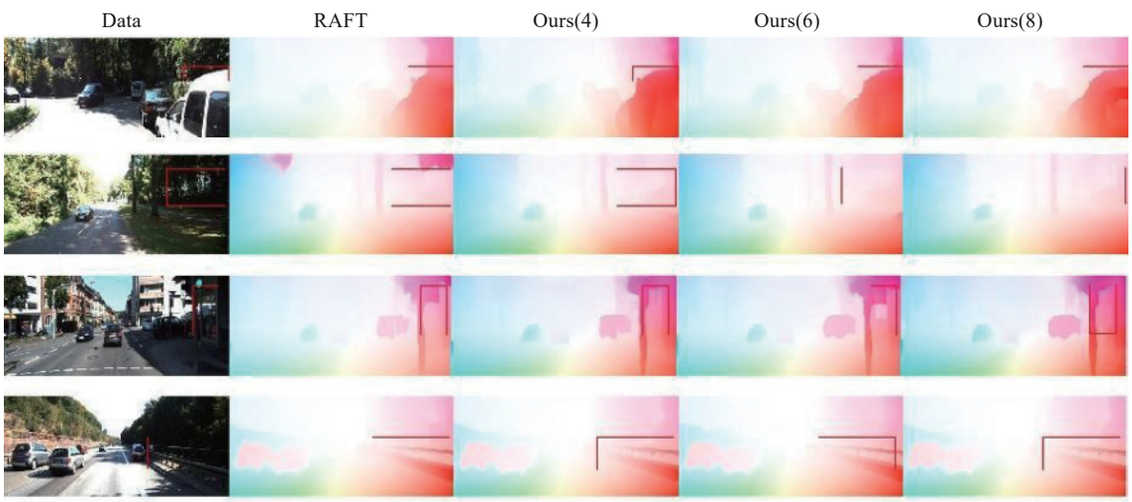
$$\text{PCT}_{\text{Fl}}=\frac{(\text{Fl}_{\text{RAFT}}-\text{Fl}_{\text{Our}(n)})}{\text{Fl}_{\text{RAFT}}}, \tag{14}$$

式中：EPE_{RAFT}为RAFT的EPE值；$\text{EPE}_{\text{Our}(n)}$为$n$分条时EPE数值；$\text{Fl}_{\text{Our}(n)}$为$n$分条的Fl值，$n$取4，6，8；$\text{Fl}_{\text{RAFT}}$表示RAFT的Fl数值。

由表 1 可见，在 KITTI-2015 测试集上，与 VCN[12]方法相比，基于视频帧间局部相关信息的光流估计网络的 EPE 和 F1 分别降低了 49.8% 和 0.31 个百分点；与 DICL[20]方法相比，基于视频帧间局部相关信息的光流估计网络的 EPE 和 F1 分别降低了 46.3% 和 0.30 个百分点；与 RAFT[14]相比，EPE 最多降低了 8.0%，F1 最多降低了 0.15 个百分点。

由表 2 可见，在 MPI-Sintel 测试集上，通过式（13）和式（14）计算可得基于视频帧间局部相关信息的光流估计网络。与 VCN[12]相比，EPE 最多降低了 67.1% 和 29.7%；与 DICL[20]相比 EPE 最多降低了 13.9% 和 10.4%；与 RAFT[14]比，EPE 最多降低了 5.7% 和 3.1%，1px 最多提高了 0.008 074 和 0.009 996，3px 最多提高了 0.009 137 和 0.003 029，5px 最多提高了 0.003 349 和 0.007 069。

基于视频帧间局部相关信息的光流估计网络在 KITTI 数据集和 MPI-Sintel 数据集上光流估计结果见图 5。



（a）K1TT1-2015 数据集实验

（b）MPI-Sintel 数据集

图 5 KITTI 和 MPI-Sintel 数据集上光流估计结果

Fig.5 results of optical flow estimation on the KITTI and MPI-Sintel datasets

图 5（a）展示了在 KITTI-2015 数据集上的光流预测结果，图中编号为 4、6、8 的区域分别代表的 4 分条区域、6 分条区域以及 8 分条区域。由图 5 中的第一行图像可见，RAFT 车身轮廓和背景分割不清晰，而基于视频帧间局部相关信息的光流估计网络将车身轮廓和背景较好地区分开，如图 5 中红框所示。此外，由图 5（a）中的最后一行图像可见，RAFT 构建的光流图中仅构建出栏杆的大致轮廓，而基于视频帧间局部相关信息的光流估计网络将栏杆的细节也大致展现出来。

图 5（b）展示了在 MPI-Sintel 数据集上的光流预测结果。对于前两行展示的 Clean 类数据，可观察

到基于视频帧间局部相关信息的光流估计网络在物体轮廓和边缘细节的处理上，显著超越了RAFT方法。特别是在第一行中，RAFT未能成功捕捉蝴蝶的光流信息，而基于视频帧间局部相关信息的光流估计网络则能准确、清晰地展现蝴蝶的光流动态。

图5（b）后两行所展示的Final类数据预测结果，即使面对含有运动模糊的挑战性场景，依然展现出了强大的预测能力。腿部边缘的光流信息得到了完整且准确的预测，进一步验证了基于视频帧间局部相关信息的光流估计网络在复杂场景下的稳定性和有效性。

2.5　消融实验

为了验证分区强弱相关计算、相似卷积块以及残差滤波器的有效性，在MPI-Sintel数据集和KITTI-2015数据集上进行消融实验。在实验过程中，确保所有不同的方法组合都采用了相同的训练参数设置和训练轮数，以保证实验结果的公正性和准确性。在分区强弱相关计算的实验中，尝试了4、6、8三种不同的分区策略，并仅在消融实验中展示了表现最优的策略。其中：A代表相似卷积块，B代表残差滤波器，C代表分区强弱相关计算。消融实验结果见表3。

由表3可知，相似卷积块与残差滤波器的结合，在光流预测任务中展现出卓越的性能，特别是在处理局部小物体时表现尤为出色。在此基础上，通过引入强弱相关计算，不仅显著降低了端点误差，还进一步提升了各项评估指标。综上所述，只有将3种方法有机结合，才能充分发挥网络性能，验证了基于视频帧间局部相关信息的光流估计网络的有效性和优越性。

表3　消融实验

Tab.3　ablation experiment

实验组合	KITTI-2015数据集		MPI-Sintel数据集							
	EPE	Fl	Clean(EPE)	Clean(1px)	Clean(3px)	Clean(5px)	Final(EPE)	Final(1px)	Final(3px)	Final(5px)
AB	0.723 10	1.837 82	0.834 48	0.904 84	0.920 99	0.962 37	1.217 45	0.865 43	0.944 45	0.963 45
AC	0.737 82	1.989 27	0.840 82	0.903 81	0.963 24	0.974 23	1.283 45	0.898 43	0.944 33	0.959 58
BC	0.728 37	2.183 21	0.870 99	0.916 35	0.960 19	0.974 54	1.246 73	0.874 34	0.941 34	0.963 45
ABC	0.707 86	1.843 12	0.842 39	0.912 01	0.959 01	0.972 33	1.239 99	0.873 45	0.942 02	0.960 52

注:A代表相似卷积块，B代表残差滤波器，C代表分区强弱相关计算。

消融实验对比结果见图6。如图6可见，使用相似卷积块和残差滤波器的组合（AB）在识别车轮局部方面效果最佳；相似卷积块和分区强弱相关计算的组合（AC）能够准确识别车身的大体轮廓，但在车轮部分表现略显不足；而残差滤波器和分区强弱相关的组合（BC）在车轮局部细节方面表现较好，但对车尾部分的识别较粗糙。只有在本文提出分区思想的强弱相关计算方法、残差滤波器和相似卷积块共同作用下，才能更好地完成光流图的构建。

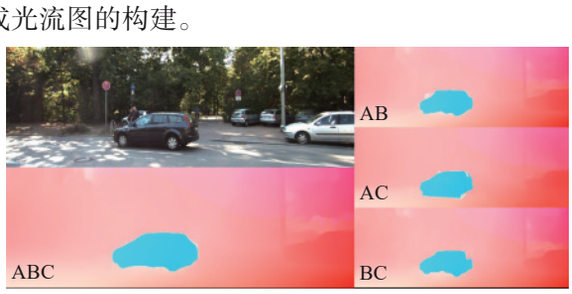


图6　消融实验对比

Fig.6　ablation experiment comparison

3　结论

（1）本文基于视频帧间局部相关信息提出一种光流估计网络，该网络通过分区计算视觉相似度方法，剔除大量误差信息，利用残差滤波器和相似卷积块保留局部微小位移信息，解决光流结果缺乏局部细粒度的问题。

（2）与VCN、DICL、RAFT等传统方法相比，基于视频帧间局部相关信息的光流估计网络在KITTI-2015和MPI-Sintel数据集上表现出显著优势，提高了光流估计网络的准确率。

（3）后续研究可着重从如何降低网络运行时间成本和参数量，提升复杂场景下的光流信息提取准确率，拓展其在自动驾驶、智能安防等领域的应用前景等方面展开。

参考文献（References）：

[1] 焦红虹,周浩,方淇. 基于光流场的时间分段网络行为识别[J]. 云南大学学报(自然科学版),2019,41(1):36-45.
JIAO Honghong, ZHOU Hao, FANG Qi. The Temporal Segment

Network based optical flow for action recognition[J]. Journal of Yunnan University(Natural Sciences Edition),2019,41(1):36-45.

[2] 杨华,王姣,张维君等. 基于光流估计的轻量级视频插帧算法[J]. 沈阳航空航天大学学报,2022,39(6):57−64.
YANG Hua, WANG Jiao, ZHANG Weijun, et al. Lightweight video frame interpolation algorithm based on optical flow estimation[J]. Journal of Shenyang Aerospace University,2022,39(6):57-64.

[3] 李志慧,胡永利,赵永华,等.基于车载的运动行人区域估计方法[J].吉林大学学报(工学版),2018,48(3):694-703.
LI Zhihui, HU Yongli, ZHAO Yonghua, et al. Locating moving pedestrian from running vehicle[J]. Journal of Jilin University (Engineering and Technology Edition),2018,48(3):694-703.

[4] HORN B K P, SCHUNCK B G. Determining optical flow[J]. Artificial Intelligence,1981,17(1/2/3):185-203.

[5] BLACK M J, ANANDAN P. A framework for the robust estimation of optical flow[C]//1993(4th) International Conference on Computer Vision. May 11-14, 1993, Berlin, Germany. IEEE, 1993: 231-236.

[6] ZACH C, POCK T, BISCHOF H. A duality based approach for realtime TV-L1 optical flow[C]//Pattern Recognition. Berlin, Heidelberg:Springer Berlin Heidelberg,2007:214-223.

[7] WEINZAEPFEL P, REVAUD J, HARCHAOUI Z, et al. DeepFlow: large displacement optical flow with deep matching[C]//2013 IEEE International Conference on Computer Vision. December 1-8, 2013, Sydney,NSW,Australia.IEEE,2013:1385-1392.

[8] SIMONYAN K, ZISSERMAN A, SIMONYAN K, et al. Two-stream convolutional networks for action recognition in videos[C]// Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1. December 8-13, 2014, Montreal,Canada.ACM,2014: 568-576.

[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. July 21-26, 2017, Honolulu,HI,USA. IEEE,2017:1647-1655.

[10] RANJAN A, BLACK M J. Optical flow estimation using a spatial pyramid network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. July 21-26,2017,Honolulu,HI,USA.IEEE, 2017: 2720-2729.

[11] SUN D Q, YANG X D, LIU M Y, 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. June 18-23,2018,Salt Lake City,UT,USA.IEEE,2018:8934-8943.

[12] YANG G, RAMANAN D. Volumetric Correspondence Networks for Optical Flow[C]//Annual Conference on Neural Information Processing Systems, December 2019. Vancouver, BC, Canada: NeurIPS,2019:793-803.

[13] HUI T W, TANG X O, LOY C C. LiteFlowNet: a lightweight convolutional neural network for optical flow estimation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23,2018,Salt Lake City,UT,USA.IEEE,2018: 8981-8989.

[14] TEED Z, DENG J. RAFT: recurrent all-pairs field transforms for optical flow[C]//Computer Vision-ECCV 2020. Cham: Springer International Publishing,2020:402-419.

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

[16] BUTLER D J, WULFF J, STANLEY G B, et al. A naturalistic open source movie for optical flow evaluation[C]//Computer Vision-ECCV 2012. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 611-625.

[17] 张水发,张文生,丁欢,等.融合光流速度与背景建模的目标检测方法[J].中国图象图形学报,2011,16(2):236-243.
ZHANG Shuifa,ZHANG Wensheng,DING Huan,et al.Background modeling and object detecting based on optical flow velocity field [J].Journal of Image and Graphics,2011,16(2):236-243.

[18] 许广富,曾继超,刘锡祥.融合光流法和特征匹配的视觉里程计[J].激光与光电子学进展,2020,57(20):270-278.
XU Guangfu, ZENG Jichao, LIU Xixiang. Visual odometer based on optical flow method and feature matching[J]. Laser & Optoelectronics Progress,2020,57(20):270-278.

[19] 安峰,戴军,韩振,等.引入注意力机制的自监督光流计算[J].图学学报,2022,43(5):841-848.
AN Feng, DAI Jun, HAN Zhen, et al. Self-supervised optical flow estimation with attention module[J]. Journal of Graphics, 2022, 43 (5):841-848.

[20] WANG J Y, ZHONG Y R, DAI Y C, et al. Displacement-invariant matching cost learning for accurate optical flow estimation[EB/OL]. (2020-10-28)[2023-11-17].https://arxiv.org/abs/2010.14851v1.