

基于 PWC-Net 的多层权值和轻量化 改进光流估计算法*

胡毅轩, 吴飞[†], 熊玉洁
(上海工程技术大学, 上海 201620)

摘要: 针对现有光流估计方法实时性不够的问题, 提出轻量化的深度可分离卷积的 PWC-Net 改进模型 (depth separable pyramid, warping and cost volume, DS-PWC)。其改进是将常规二维卷积网络层解耦为深度可分离卷积层, 并且 DS-PWC 在金字塔层增加基于层数的权重系数, 从而使得网络结构在不损失精度的情况下大幅减少模型参数量。在训练过程中, 使用图像及对象感知数据随机擦除 (image and object-aware random erasing, I + ORE) 等数据增强技术, 进一步提升估计预测结果泛化能力。实验结果表明, 在数据集测试 DS-PWC 模型, 在保持质量的同时运行效率达到约 58 fps (frame per second)。同时为了验证算法有效性, 进行了模型结构和数据增强的消融实验。结果证明了 DS-PWC 模型的有效性。

关键词: 密集光流; 轻量化; 卷积神经网络; 网络解耦; 金字塔结构

中图分类号: TP391

文献标志码: A

文章编号: 1001-3695(2022)01-052-05

doi: 10.19734/j.issn.1001-3695.2021.05.0204

Multi-layer weight and lightweight improved optical flow estimation algorithm based on PWC-Net

Hu Yixuan, Wu Fei[†], Xiong Yujie

(Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: To solve the problem of insufficient real-time performance of existing optical flow estimation methods, this paper proposed a lightweight improved PWC-Net model with deep separable convolution (depth pyramid, warping and cost volume, DS-PWC). One of the improvements was decoupling the conventional two-dimensional convolutional network layer into a deep separable convolutional layer. The other one was adding a weight coefficient based on the number of layers in the pyramid layer by DS-PWC, which greatly reduced the number of model parameters in the network structure without loss of precision. In addition, in the training process, the paper applied data enhancement technologies such as I + ORE to further improve the generalization ability of estimation and prediction results. The experimental results show that the DS-PWC model was tested in the dataset and the operating efficiency reaches about 58 fps while maintaining the quality. To verify the effectiveness of the algorithm, this paper carried out the ablation experiments of model structure and data enhancement. The results verify the effectiveness of the DS-PWC model.

Key words: dense optical flow; lightweight; CNN; network decoupling; pyramid structure

0 引言

光流估计是计算机视觉中的一个经典问题。它被广泛使用在许多领域中, 如运动跟踪、动作识别、视频分割、三维重建、视频修复等。1981年, Horn等人^[1]首次提出光流基本守恒假设及相应光流估计方法, 而后他们又提出使用能量函数耦合亮度不变性和空间平滑性的能量最小化方法, 这是当时准确率最高的计算方法。这种方法的计算量十分大, 无法达到实时应用的要求。之后 Brox等人^[2]从理论上推导出基于翘曲的光流估计方法。文献[3]在 Horn等人的方法和模型上进行了改进, 提出了一个非局部项来恢复运动细节。flow fields^[4]基于局部匹配的计算方法的计算精度较高, 但是计算量仍然较大。变分方法仍然是当今比较流行的光流计算方法, 然而它需要解决复杂的优化问题, 不符合实时性程序的基本要求。

随着深度学习技术的进步与发展, 许多传统的图像问题都

采用了卷积神经网络 (convolutional neural networks, CNN)。在光流估计方面有许多相关算法使用 CNN。在有监督的学习中, Dosovitskiy等人^[5]的 FlowNet 基于有监督学习的光流估计模型, 开创性地将 U-Net^[6] 回形网络架构运用在光流估计上, 使用了 encoder-decoder 网络架构, 并且提出了计算图像对之间特征的成本量, 使两帧之间产生联系, 证明了卷积神经网络直接估计图像序列光流的可行性。为了解决 FlowNet 准确度不高以及小位移预测不精准的问题, 2017年 FlowNet 2.0^[7] 采用了 FlowNetC 和 FlowNetS 模型堆叠的方法解决了上述问题, 使光流准确度大幅上升, 但模型需要 640 MB 的存储空间, 运行速度也并不算快, 因此不适合在移动端和嵌入式设备使用。Ranjan等人^[8]将经典空间金字塔的思想与卷积神经网络结合, 提出了 SpyNet 网络模型, 显著减小了模型参数, 但是由于 SpyNet^[8] 网络结构单一, 导致其运算速度虽快但估计精度较低。2018年, Sun等人^[9]提出了 PWC-Net 网络结构, 其输入采用金字塔结构, 提高了输入特征图的置信度, 根据翘曲光流估计理

收稿日期: 2021-05-24; 修回日期: 2021-07-14 基金项目: 上海市科技学术委员会重点项目 (18511101600); 国家自然科学基金资助项目 (62006150, 61802251); 上海青年科技英才扬帆计划资助项目 (19YF1418400)

作者简介: 胡毅轩 (1996-), 男, 四川内江人, 硕士研究生, 主要研究方向为计算机视觉、无人机控制; 吴飞 (1968-), 男 (通信作者), 教授, 硕导, 博士, 主要研究方向为计算机网络 (fei_wu1@163.com); 熊玉洁 (1989-), 男, 讲师, 硕导, 博士, 主要研究方向为模式识别与智能系统。

论重新设计了图像对构成光流特征图的方法。最后采用空洞卷积估计网络使得小位移的光流估计在此网络上也有不错的估计精度。在减少耗时的同时,PWC-Net 相对 SpyNet 也提高了精度。2019年,Yang 等人^[10]的 VCN 提出了不同的图像对匹配方式,也有不错的精度。文献[11]对遮挡区域的前后帧匹配修正来解决部分遮挡区域光流估计问题。IRR-PWC^[12]是 PWC-Net 的一个改进,主要通过迭代的思想融合前后几帧信息从而提高光流估计的精度,但其估计速度相对较慢。

无人系统作为近期的热门研究方向一直备受关注,使用无人车、无人机进行自主导航、目标跟踪时若能使用光流估计数据则能大幅提高无人系统自主运行时的可靠性,现有的光流估计方法虽然就传统的计算方法已经大幅提高速度,但是依旧无法满足无人系统环境下深度学习边缘计算硬件设备的使用需求。

为了在不损失精度的情况下进一步减少耗时以及缩小模型大小,本文提出 DS-PWC 网络结构。本文算法使用了 Google^[13-15]提出的深度可分离二维卷积层代替常规卷积层构建光流预测网络的主干;同时增加了网络间权重系数,减轻了高层金字塔网络和加强低层金字塔网络的预测结果对其上一层网络的影响,使得损失能够收敛在更低的值。而使用深度可分离卷积解耦网络可大幅降低卷积所带来的计算量和参数量。图1所示是本文使用 DS-PWC 网络最终的预测效果。在预测效果不逊色于 PWC-Net 的情况下大幅提高预测速度。其中,(a)(b)为相邻帧的图像对,(c)为模型预测的光流可视化图,(d)为真实光流的可视化图。



图1 DS-PWC 在 Sintel 训练集上的结果及与真实光流的对比
Fig. 1 DS-PWC results on the Sintel training set and comparison with real optical flow

1 光流估计改进模型架构

本文估计光流模型的结构如图2所示。总共采用六层金字塔网络结构,将六层下采样后的结果输入六个相同的神经网络结构,下一层的最终输出结果输入在上一层的成本代价估计层。

a) 将输入的图像对分别进行下采样形成六层金字塔结构,然后对图像对的六层金字塔结构分别进行普通卷积操作形成输入特征图;b) 将图像对的后一张输入图片进行扭曲(warping)补偿一部分拍摄及其他硬件导致的变形所带来的部分扭曲,使输入变得更准确和平滑;c) 对金字塔输入特征图的每一层都进行成本量的计算,将两个特征输入图联系起来;d) 将上一步所得到的特征图进行深度可分离卷积后上采样生成初步的光流估计图;e) 将上一步的特征图带入上下文网络中预测光流的大位移变化;f) 将步骤d)和e)产生的光流预测进行叠加产生一个批次的训练结果与真实光流标签对比,产生的损失值反向传播,用于更新模型中参数的权重。

1.1 特征输入层

PWC 将输入图像对进行了五次下采样,每次采样的像素为上一层的一半,以此构成六层金字塔结构,第0层为原始图像。将六层金字塔同时进行两次卷积,构成六个参数互不干涉的特征图,第一次卷积步长为2,第二次为1。采用卷积特征图

作为输入层有助于提高下面网络结构计算时的置信度。

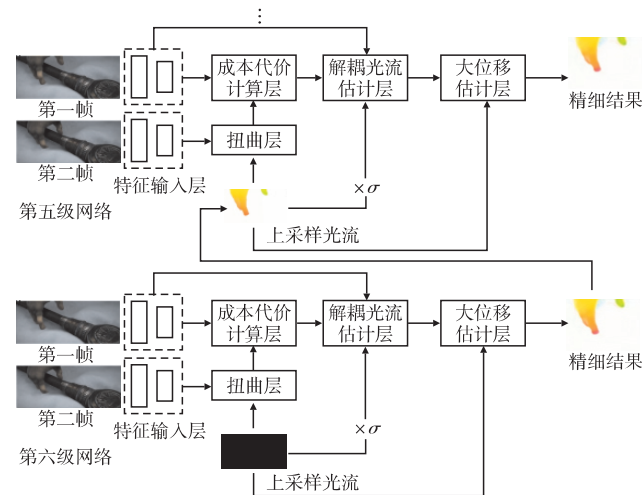


图2 DS-PWC 光流估计神经网络模型结构
Fig. 2 Structure of DS-PWC optical flow estimation neural network model

1.2 图像扭曲计算层

图像扭曲层将金字塔 $l+1$ 层预测的光流 $flow^{l+1}$ 进行 2 倍上采样后与 l 层的第二张输入特征图 c_2^l 进行一次光流扭曲得到 c_w^l ,使其与第一张特征图更接近,而最下面一层 l_0 的光流估计设置为 0。计算公式为

$$c_w^l = c_2^l(x + up_2(w^{l+1}(x))) \quad (1)$$

其中: x 是特征图像素的索引; up 是使用双线性插值进行上采样操作。使用双线性插值的方法是为了防止后续 CNN 反向传播时产生梯度消失。通过上一层光流的上采样版本将第二帧向第一帧扭曲,使下一层的输入是上一层光流的运动增量,根据文献[16],该方法使 CNN 的输入更平滑细致,且矫正小范围几何失真带来的噪声。

1.3 匹配成本代价计算层

匹配成本代价计算层将处理后的两个特征图进行相关像素的匹配。PWC-Net 提出了一种新的匹配代价计算方法,将第一幅图与扭曲过后的第二幅图的相关性定义为匹配代价。特征图对进行类卷积运算匹配代价计算公式为

$$cv^l(x_1, x_2) = \frac{1}{N} (c_1^l(x_1))^T c_2^l(x_2) \quad o \in [-k, k] * [-k, k] \quad (2)$$

其中: T 是转置操作; N 是列向量 $c_1^l(x_1)$ 的数量。为了避免计算量过于庞大,此计算方法设定限制参数 d ,使得 $|x_1 - x_2|_\infty \leq d$ 。由于顶层金字塔一个像素的移动对应了全分辨率图像 2^{l-1} 像素的移动,每层金字塔的 d 将按比例缩小。

图像扭曲层和匹配成本代价计算层均为计算层,不需要训练权重参数,减少了模型的大小和参数量。

1.4 解耦改进的光流估计层

光流估计层的输入是匹配成本代价 $corr$,第一幅图像 x_1 和上一层的光流预测图上采样结果 $flow^l$,输出是当前层的光流预测图以及上一层的光流预测的部分权重^[17]。该层的基础结构参考 PWC-Net 的子结构 PWC-Net-s,此结构减少了常规结构中的残差结构^[18],并且由五个深度可分离卷积结构代替原结构中的常规卷积,每个深度可分离卷积结构参考文献[14],由深度可分离卷积层、BN 层及 LeakyReLU 层构成,并且前置的 BN 层和 ReLU 层能在反向传播时把深层的梯度传递到任意浅层,所以不管参数再小也不会造成梯度消失。常规卷积层解耦为深度可分离卷积层可大幅度减少模型的参数量,同时保持光流估计层的表达能力。在网络中深度可分离卷积层如图3所示。



图 3 深度可分离卷积结构

Fig. 3 Depth separable convolutional structure

深度可分离卷积^[15]的原理如图 4 所示,先分别对不同通道的卷积核进行卷积,再用 C_{out} 个大小为 $1 \times 1 \times C_{in}$ 的卷积核进行卷积,所输出的特征图仍然符合常规卷积的卷积输出。 K_h 、 K_w 分别为卷积核的长和宽; C_{in} 、 C_{out} 分别为输入输出的通道数; F_h 则为特征图的长度减去卷积核的长度; F_w 为特征图的宽度减去卷积核的宽度。

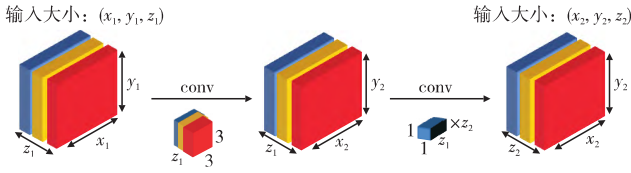


图 4 深度可分离原理示意图

Fig. 4 Principle of depth-separable

常规卷积的参数量:

$$P_{conv} = K_h \times K_w \times C_{in} \times C_{out} \quad (3)$$

深度可分离卷积的参数量:

$$P_{depth} = K_h \times K_w \times C_{in} \quad P_{point} = 1 \times 1 \times C_{in} \times C_{out} \\ P_{dconv} = P_{depth} + P_{point} \quad (4)$$

深度可分离卷积的参数量由逐深度卷积和逐点卷积的参数量线性叠加,其参数量明显小于常规卷积。

常规卷积的计算量:

$$C_{conv} = K_h \times K_w \times C_{in} \times C_{out} \times F_h \times F_w \quad (5)$$

深度可分离卷积的计算量:

$$C_{depth} = K_h \times K_w \times C_{in} \times F_h \times F_w \quad C_{point} = C_{in} \times C_{out} \times F_h \times F_w \\ C_{dconv} = C_{depth} + C_{point} \quad (6)$$

通过比较可以得出将常规卷积替换为深度可分离卷积后计算量的减少比例:

$$\frac{C_{dconv}}{C_{conv}} = \frac{1}{C_{out}} + \frac{1}{K_h \times K_w} \quad (7)$$

该结构的网络表达能力在文献[15]中被证明与常规卷积基本相似。在 MobileNet^[14] 中,采用深度可分离卷积层进行网络解耦使其计算量减小到原来的九分之一,而识别精准度仅下降 1.7%。

1.5 金字塔结构改进的光流估计层

PWC-Net-s 中低层数金字塔的网络结构由于其下采样所导致的特征及参数量的数据量较少,误差始终无法收敛至一个较低的水平。若将这种光流预测数据于上层网络中用于训练输入,并且所占信息的权重过高,则会在网络的收敛后期成为干扰项,成为阻碍网络收敛于更低值的原因。所以本文网络对于上层网络训练的输入,增加了一个基于金字塔的层数 l 的权重系数 σ 。

$$input^l = [x^l | corr^l | l \times \sigma \times flow^{l-1}] \\ output^l = net(input^l) + l \times \sigma \times flow^{l-1} \quad (8)$$

如图 5 所示,输入输出的上一层光流上采样后结果的权重随着金字塔层数的提升而提升。权重系数 σ 的计算公式如下:

$$\sigma = k \times U \quad (9)$$

U 是经过上一个训练数据集训练后第 0 层光流预测数据的平均端点误差值 k 是常数,本文 k 是经验值 1.1。第一个数据集光流估计层的输入输出与文献[9]中的相同。

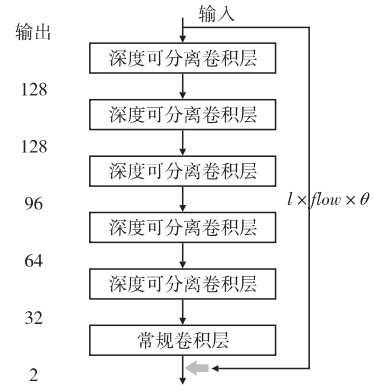


图 5 光流估计层网络结构

Fig. 5 Optical flow estimation layer network architecture

1.6 大位移光流估计层

使用大位移估计层作为此网络模型的后处理网络层,进一步改善大位移光流的估计精度。大位移估计层由膨胀率为 1、2、4、8 的空洞卷积构成,这样能够提升像素信息获取范围并且减少空洞卷积所带来的网格效应。使用空洞卷积构成大位移光流估计网络能够有效地在不增加参数量的基础上增加网络的感受野,提高卷积特征图远距离像素之间的相关性。大位移估计层不使用残差网络结构,防止空洞卷积所带来的网格效应产生的高频信号向下继续传播。

1.7 训练损失

定义 Θ 为神经网络中可训练参数的集合,包括特征金字塔层、光流估计、大位移光流估计层。其中扭曲层和成本代价层不包含可训练的参数,仅为计算层。定义 W_{θ}^l 为第 l 层金字塔所预测的光流,定义 W_{GT}^l 为对应的真实光流值。由文献[7]中提出多尺度端点误差损失公式式(10)计算损失值。

$$\mathcal{L}(\Theta) = \sum_{l=0}^L \alpha_l \sum_x |W_{\theta}^l(x) - W_{GT}^l|_2 + \gamma |\Theta|_2 \quad (10)$$

其中: α 的值是 $\alpha_0 = 0.32$ $\alpha_1 = 0.08$ $\alpha_2 = 0.02$ $\alpha_3 = 0.01$ $\alpha_4 = 0.005$; $\gamma |\Theta|_2$ 用于对网络的参数进行正则化; γ 的权重设置为 0.0004。

2 数据增强(data augmentation)

现实世界中的真实光流是非常难获取的,而且无法通过人工手动标注。Baker 等人^[19]通过游戏引擎自动生产了图像对以及相关光流图,但对于光流的学习仍然数据量太小,由于光流数据集是图像数据集,仅 2 万份数据就占用 75 GB 容量,但 2 万份数据不足以训练出一个好的模型。为了弥补数据集 FlyingChairs 及 FlyingThings 和 MPI Sintel 非真实数据的缺陷,对图像对其光流数据进行相同的随机裁剪、随机旋转,对图片镜像色彩增强、噪声叠加。这就需要使用数据增强的方法扩展数据集的样本,目的是提高模型鲁棒性,降低过拟合的风险。

2.1 随机裁切

随机裁切可以使原来的一张图变为多张图,防止训练网络过度拟合,同时也能增加训练速度。随机裁剪的方法是:取两个随机数的集合 X, Y 服从式(11) $(x, y) \in (X, Y)$ 作为左上角顶点进行裁剪,裁剪大小根据数据集不同选择裁剪为 348×448 或 348×768 。

$$X \sim U(0, image_size, h - crop_size, h)$$

$$Y \sim U(0, image_size, w - crop_size, w) \quad (11)$$

2.2 色彩增强

色彩增强也是丰富数据集数据的一种有效方式,能有效地提高神经网络模型的鲁棒性。此处与文献[20]提到的色彩增

强方法基本一致,如表 1 所示。

表 1 数据集色彩增强方法

Tab. 1 Dataset color enhancement methods

数据增强种类	取值范围
对比度增强	$[-0.8, 0.4]$ 的对比度增强
RGB 增强	通道乘数因子为 $[0.5, 2]$ 的增强
Gamma 增强	$[0.7, 1.5]$ 的非线性变换方差
亮度增强	为 0.2 的高斯噪声增强

2.3 图像和对象感知随机擦除

随机擦除增强策略是在训练集中随机选择图像的矩形区域,并使用 0~255 的随机值擦除其像素。生成具有遮挡级别的训练图像,同样会降低过拟合风险并使得模型对遮挡具有一定的鲁棒性。该增强方法是根据文献[21]使用图像和对象感知随机擦除(I+ORE)的方法。如图 6 所示,(a)和(b)是连续两帧图像根据(c)对象检测框的位置生成相同的随机擦除区域,(c)是根据真实光流检测出的对象框。读取裁剪后的标签光流数据,检测光流图中对象的边界。在边界内选取图像随机擦除块的区域位置,其大小按照图像分辨率的 0.02~0.15 倍取随机数,长宽比在 0.33~3.33 均匀分布取随机数,使用随机的像素值作为掩码填充擦除区域的图像。此方法能够更有效地遮挡部分关键信息,提高网络的联想学习能力。

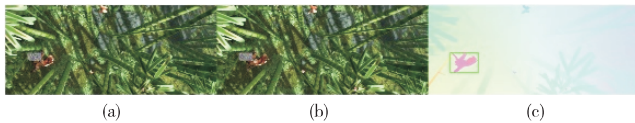


图 6 图像和对象感知随机擦除

Fig. 6 Image and object-aware random erasure

3 训练策略

文献[21]的实验证明训练策略对模型最后的结果十分关键,本模型的学习率训练策略与文献[22]类似,如图 7 所示,按顺序在三个数据集上训练。

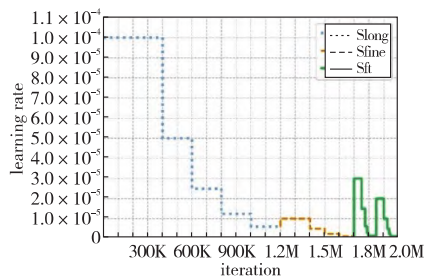


图 7 DS-PWC 训练时的学习率变化

Fig. 7 Change of learning rate during DS-PWC training

a) 在 FlyingChairs 上按照 FlowNet 2.0 中所提到的 Slong 对学习率进行调整,节点在迭代次数为 0.4 M、0.6 M、0.8 M 和 1 M 时调整为之前的一半。batch-size 为 8、裁切图像大小为 488×348 ,在简单的 FlyingChairs 上可以帮助网络学习色彩匹配的一般概念,而不会让 3D 运动和照明产生先验混淆。

b) 在 FlyingThings3D 上继续微调(finetune),FlyingThings3D 的数据集采用 FlowNet 2.0 团队提出高难度图像对(运动幅度超过 1 000 像素)后的子数据集,按照 Sfine 的调整路线调整学习率, batch-size 为 4、裁切图像大小为 768×348 ,在 FlyingThings3D 上学习物体的三维移动模式。FlyingThings3D 数据集能够让网络模型继续学习物体的三维运动与光流估计的联系。

c) 在 MPI Sintel 上按照 Sft 路线微调模型参数,使模型能够学习现实中存在的运动模糊和照明、环境对光流估计的干扰, batch-size 为 4、裁切图像大小为 768×348 。Sintel 的 Final 数据集能够使网络模型学习光照、运动模糊对光流估计的影

响,使其更接近现实相机拍摄的情况。最后得到本文最终提出的改进的光流预测网络。

4 实验结果与分析

4.1 评价标准

EPE(end-point error)^[19]是指同一个像素点的预测光流矢量(u_1, v_1)与真实光流矢量(u_0, v_0)之间的距离,用来度量预测结果的精度。

$$EPE = \sqrt{(U_1 - U_0)^2 + (V_1 - V_0)^2} \quad (12)$$

AEPE 是光流图像中每个对应像素端点误差的平均值。

$$AEPE = \overline{EPE} \quad (13)$$

FPS 是指每秒能够预测的光流帧数。帧数越高,计算速度越快,预测间距越小,越能保证算法的实时性。

4.2 消融实验

4.2.1 数据增强

本文设置两组对照实验,DS-PWC 在 Sintel 数据训练集上进行训练,使用相同的训练策略 Sft。一组为进行数据增强学习,一组使用原数据集学习。训练后的结果分别在 Sintel 训练集和 FlyingChairs 数据集上进行测试。测试结果如图 8、9 所示。

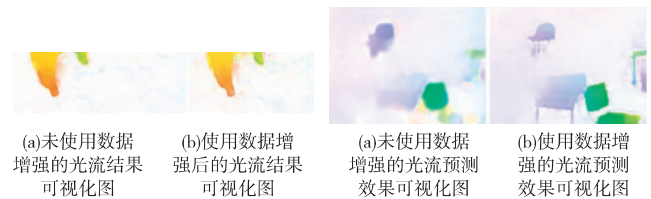


图 8 训练集上的预测效果对比
Fig. 8 Comparison of prediction effects on the training set

图 9 测试集上预测效果对比
Fig. 9 Comparison of prediction effects on test set

未经过图像增强的预测模型在训练集与经过图像增强的预测模型准确率相似。而后者在测试集上的预测准确度要高于前者。本文的图像增强方法能有效地提高模型的泛化能力。

4.2.2 金字塔权值改进

本文设置了两组对照实验,使用 PWC 原始金字塔结构和加权金字塔结构模型进行相同训练,使用相同的学习率 0.000 05,在相同的 Sintel 训练集上训练。由于最顶层第六层不会有来自更上一层的光流输入,所以采用第五层损失值曲线体现收敛效果。如图 10 所示,多层权值结构能够使损失值收敛在更低的值。

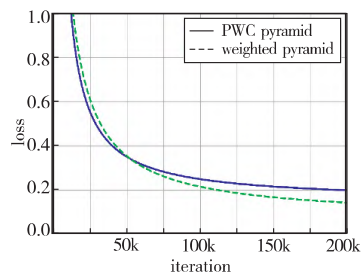


图 10 多层权值结构和普通金字塔结构对比

Fig. 10 Comparison of multi-layer weight structure and ordinary pyramid structure

4.3 实验结果与分析

本文方法和 DS-PWC 模型在最新的公共基准数据集进行了测试和比较。经过了训练策略中的两种数据集的基础训练后,在 Sintel 中进行微调训练,将图片和真实光流进行同步的随机裁剪为 768×384 ,并且对数据集进行了色彩增强、上下翻

转、图像和对象感知随机擦除的数据增强操作。最后将训练后的卷积神经网络模型作为本文结果,预测结果如图 1 所示。

在撰写本文时,DS-PWC 在 MPI-Sintel 的基准测试中是 PWC-Net 等方法中速度最快的,光流估计效果如图 1 所示。DS-PWC 使用深度可分离卷积结构重构光流估计网络,能保持平均端点误差为 4.47 的情况下拥有每次预测 0.017 s 的计算速度,达到了 58.8 fps 的计算帧率。使得模型参数从原来的 8.63 MB 降到了 1.96 MB,是 PWC-Net 参数量的 0.22 倍。网络体积仅为 8.5 MB。与 SpyNet 相比,参数量虽略高于此网络,但精准度却明显高于此网络。与 VCN 相比,DS-PWC 的平均端点误差略微逊色,但是其拥有更快的预测速度、更小的模型容量,最重要的是 DS-PWC 的光流估计速度比 PWC-Net 提高了 76%。

DS-PWC 的模型参数减少量和预测速度减少量未成线性关系,其原因与 GPU 特性相关。GPU 在训练时以每一层结构为单位导入参数,解耦网络结构在减少网络参数的同时也会增加网络的层数,也就是增加 GPU 读取缓存数据的次数,所以网络层数的增加会为 GPU 预测光流带来一定的负担,带来的减益效果会抵消部分减少参数所带来的时间缩短。性能效果对比如表 2 所示。

表 2 主流光流估计方法性能比较(Sintel final test)
Tab. 2 Performance comparison of mainstream optical flow estimation method(Sintel final test)

模型	时间/s	参数/MB	EPE
FlowFields	28	—	5.81
FlowNet2	0.12	638.5	6.02
SpyNet ^[8]	0.069	1.2	8.36
PWC-Net ^[9]	0.03	8.75	4.95
VCN ^[10]	0.03	6.2	4.40
DS-PWC	0.017	1.98	4.47

5 结束语

为了提高光流估计在嵌入式及微型电脑平台上的可用性,本文提出了一种基于 PWC-Net 的光流估计改进网络结构。一是将常规二维卷积结构解耦为深度可分离卷积;二是构造光流估计层和大位移估计层之间的改进残差网络结构。两种改进方法结合能够大幅降低模型参数和存储空间,提高模型的预测速度。使用卷积特征图代替原始输入以及使用扭曲估计方法生成光流特征图和改进的金字塔输入输出结构能够提高模型的估计精度和训练时的收敛速度。此外,为了提高模型的泛化能力、鲁棒性、真实性,对数据和标签进行了同步的随机裁切、随机翻转、色彩增强,图像和对象感知随机擦除,使改进后的网络能够在网络精度较为优秀的同时还能大幅减小网络参数及其存储空间,提高了大位移移动和遮挡下的光流估计能力,可用于无人机的光流悬停和视觉避障的物体识别上。接下来,将 DS-PWC 应用在机器人导航等应用场景,在实际的复杂工程问题上进行优化,如训练时间较长、迭代次数多、小位移估计精度优化问题是下一步的研究重点。

参考文献:

- [1] Horn B, Schunck B G. Determining optical flow[J]. *Artificial Intelligence*, 1981, 17(1-3): 185-203.
- [2] Brox T, Bruhn A, Papenberg N, et al. High accuracy optical flow estimation based on a theory for warping[C]//Proc of European Conference on Computer Vision. 2004.
- [3] Black M J, Anandan P. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields[J]. *Computer Vision and Image Understanding*, 1996, 63(1): 75-104.
- [4] Bailer C, Taetz B, Stricker D. Flow fields: dense correspondence fields for highly accurate large displacement optical flow estimation[C]//Proc of International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2015: 4015-4023.
- [5] Dosovitskiy A, Fischer P, Ilg E, et al. Flow net: learning optical flow with convolutional networks[C]//Proc of International Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2015: 2758-2766.
- [6] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//Proc of International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2015: 234-241.
- [7] Ilg E, Mayer N, Saikia T, et al. FlowNet 2.0: evolution of optical flow estimation with deep networks[C]//Proc of International Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017: 1647-1655.
- [8] Ranjan A, Black M J. Optical flow estimation using a spatial pyramid network[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017: 4161-4170.
- [9] Sun Deqing, Yang Xiaodong, Liu Mingyu, et al. PWC-Net: CNNs for optical flow using pyramid warping and cost volume[C]//Proc of International Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 8934-8943.
- [10] Yang Gengshan, Ramanan D. Volumetric correspondence networks for optical flow[C]//Proc of the 33rd Conference on Neural Information Processing Systems. 2019.
- [11] Ilg E, Saikia T, Keuper M, et al. Occlusions, motion and depth boundaries with a generic network for disparity, optical flow or scene flow estimation[C]//Proc of European Conference on Computer Vision. Berlin: Springer, 2018: 614-630.
- [12] Hur J, Roth S. Iterative residual refinement for joint optical flow and occlusion estimation[C]//Proc of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2019: 5754-5763.
- [13] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017.
- [14] Howard A G, Zhu Menglong, Chen Bo, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[C]//Proc of Computer Vision and Pattern Recognition. 2017.
- [15] Guo Jianbo, Li Yuxi, Lin Weiyao, et al. Network decoupling: from regular to depthwise separable convolutions[C]//Proc of British Machine Vision Conference. 2018.
- [16] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks[C]//Advances in Neural Information Processing Systems. 2015.
- [17] Mayer N, Ilg E, Hausser P, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 4040-4048.
- [18] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Identity mappings in deep residual networks[C]//Proc of European Conference on Computer Vision. 2016: 630-645.
- [19] Baker S, Scharstein D, Lewis J P, et al. A database and evaluation methodology for optical flow[J]. *International Journal of Computer Vision*, 2011, 92(1): 1-31.
- [20] 张聪炫,周仲凯,陈震,等.深度学习光流计算技术研究进展[J]. *电子学报*, 2020, 48(9): 1841-1849. (Zhang Congxuan, Zhou Zhongkai, Chen Zhen, et al. Research progress of deep learning optical flow computing technology[J]. *Acta Electronica Sinica*, 2020, 48(9): 1841-1849.)
- [21] Zhong Zhun, Zheng Liang, Kang Guoliang, et al. Random rising data augmentation[C]//Proc of Association for the Advancement of Artificial Intelligence. 2020.
- [22] Sun Deqing, Yang Xiaodong, Liu Mingyu, et al. Models matter so does training: an empirical study of CNNs for optical flow estimation[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2020, 42(6): 1408-1423.