计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2020.20190478
57(7): 1531-1538, 2020

适应立体匹配任务的端到端深度网络

李 瞳[1] 马 伟[1] 徐士彪[2] 张晓鹏[2]
[1](北京工业大学信息学部 北京 100124)
[2](中国科学院自动化研究所 北京 100190)
(772402345@qq.com)

Task-Adaptive End-to-End Networks for Stereo Matching

Li Tong[1], Ma Wei[1], Xu Shibiao[2], and Zhang Xiaopeng[2]
[1](*Faculty of Information Technology, Beijing University of Technology, Beijing* 100124)
[2](*Institute of Automation, Chinese Academy of Sciences, Beijing* 100190)

Abstract Estimating depth/disparity information from stereo pairs via stereo matching is a classical research topic in computer vision. Recently, along with the development of deep learning technologies, many end-to-end deep networks have been proposed for stereo matching. These networks generally borrow convolutional neural network (CNN) structures originally designed for other tasks to extract features. These structures are generally redundant for the task of stereo matching. Besides, 3D convolutions in these networks are too complex to be extended for large perception fields which are helpful for disparity estimation. In order to overcome these problems, we propose a deep network structure based on the properties of stereo matching. In the proposed network, a concise and effective feature extraction module is presented. Moreover, a separated 3D convolution is introduced to avoid parameter explosion caused by increasing the size of convolution kernels. We validate our network on the dataset of SceneFlow in aspects of both accuracy and computation costs. Results show that the proposed network obtains state-of-the-art performance. Compared with the other structures, our feature extraction module can reduce 90% parameters and 25% time cost while achieving comparable accuracy. At the same time, our separated 3D convolution, accompanied by group normalization (GN), achieves lower end-point-error (EPE) than baseline methods.

Key words stereo matching; disparity estimation; feature extraction; 3D convolution; end-to-end network

摘 要 针对现有立体匹配深度网络中特征提取模块冗余度高以及用于视差计算的3D卷积模块感受野受限问题,提出改进的端到端深度网络.相比现有网络,该网络特征提取模块遵循立体匹配特性,结构更简洁;引入分离3D卷积实现大卷积核3D卷积运算以扩充感受野.在SceneFlow数据集上,从匹配精度和计算开销等方面评估所提出网络.实验结果显示:所提出网络在准确度上达到了先进水平;相比现有同类型模块,所提出特征提取模块在保证结果精度的同时能减少90%的参数量,并减少约25%的训练时间;相比3D卷积,所提出的分离3D卷积将卷积核大小提升至覆盖整个视差维度,搭配群组归一化(group normalization, GN),其端点误差(end-point-error, EPE)较基础方法降低了12%的相对量.

收稿日期:2019-07-08;**修回日期**:2019-11-18
基金项目:国家自然科学基金项目(61771026,61671451);模式识别国家重点实验室开放课题基金
This work was supported by the National Natural Science Foundation of China (61771026, 61671451) and the Open Project Program of the National Laboratory of Pattern Recognition (NLPR).
通信作者:马伟(mawei@bjut.edu.cn)

关键词 立体匹配;视差计算;特征提取;3D 卷积;端到端网络

中图法分类号 TP391

立体匹配(stereo matching)是计算机视觉领域中的经典问题.相比激光测距等方法,立体匹配能够低成本从双目图像中便捷恢复场景深度,在自动驾驶、机器人避障、立体图像智能编辑等领域具有重要应用价值[1-5].近年来,深度学习(deep learning)受到广泛关注,并在诸多计算机视觉任务中取得显著成功[6-8].鉴于此,国内外研究者尝试设计用于立体匹配的深度网络[9-10].相比传统方法,深度学习模型尤其是端到端模型,能够在学习大量数据的基础上得到更准确的视差结果[11-15].

然而,现有深度网络尚存在诸多不足:现有网络的特征提取模块多源自于求解其他问题的网络模型,对立体匹配任务特性考虑不足,冗余度高;3D 卷积常用于立体匹配的视差计算,性能优越.然而,3D 卷积复杂度高,难以实现大卷积核运算,现有网络多采用小卷积核 3D 卷积,感受野范围受限.

针对上述问题,提出改进的端到端神经网络模型.该模型以当前性能优异的金字塔立体匹配网络(pyramid stereo matching network, PSMNet)[13]为基准方法,对其核心模块,包括特征提取模块和 3D 卷积匹配模块,进行改进.所提出特征提取模块专为立体匹配问题所设计.相比现有特征提取网络,结构更简洁,能够在保持结果精度前提下分别以 90%和 25%的幅度大幅减少参数量与计算量.在视差计算模块中,提出分离 3D 卷积,相比现有 3D 卷积复杂度低.因此,可实现大卷积核运算以扩充感受野,从而提高立体匹配准确度.在 SceneFlow 数据集上验证了所提出方法的优异性能,并通过对比实验从准确度、计算成本等方面验证 2 个模块的有效性.

1 整体网络架构

本文所提出网络以 PSMNet 为基准网络,因此,先简要说明 PSMNet 网络结构,再给出本文网络设计.

PSMNet 网络结构如图 1 所示.首先,提取立体图像特征,用于构建匹配成本特征体(cost volume).之后,通过 3D 卷积运算计算视差.在立体图像特征提取过程中,PSMNet 使用类似 ResNet[16]的网络结

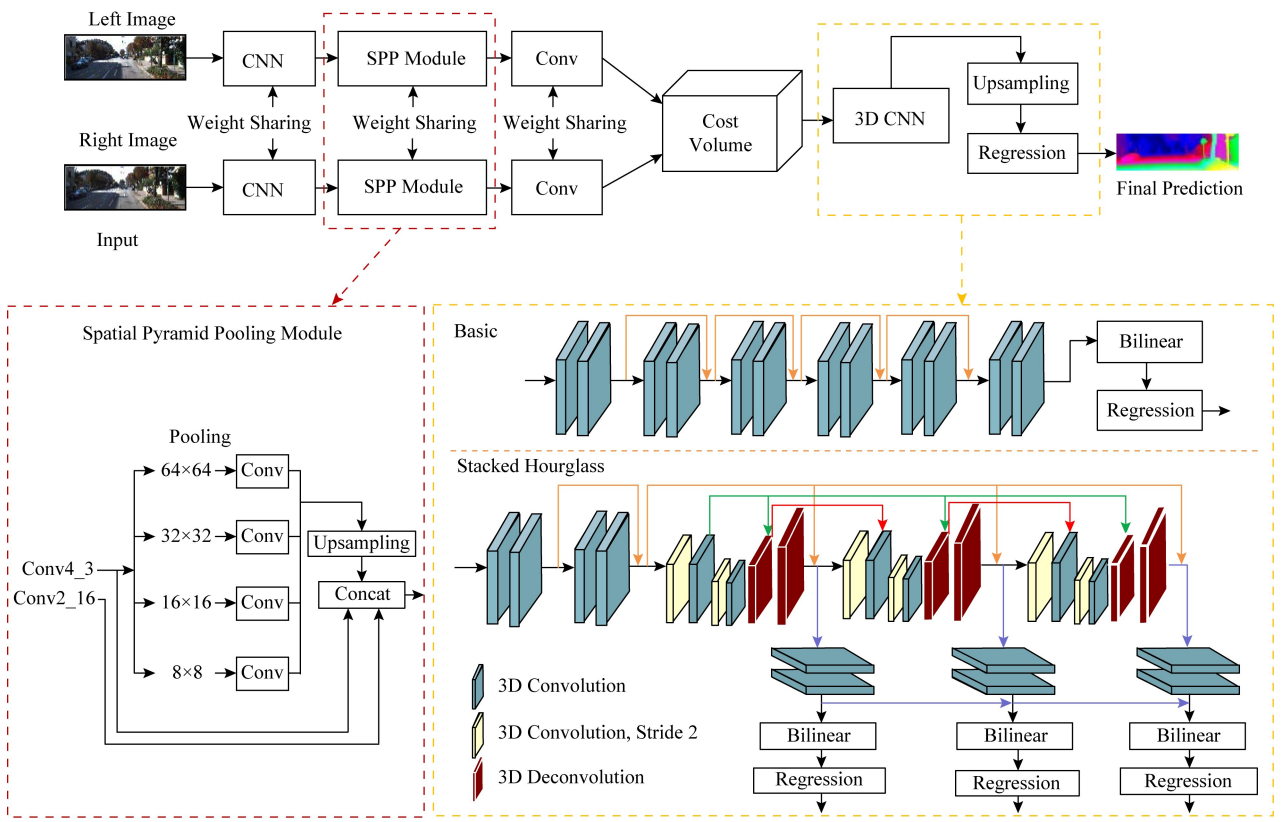

Fig. 1 Architecture of PSMNet[13]

图 1 PSMNet 网络结构[13]

构,再使用空间金字塔池化(spatial pyramid pooling,SPP)结构[13,17]融合多尺度特征增强特征表达.然后将特征连接形成匹配成本特征体.在视差计算部分,PSMNet 网络包括 2 种版本网络,分别是多次残差级联 Basic 的 3D 卷积网络和使用 Stackhourglass 结构[18]的 3D 卷积网络.

本文在 PSMNet 网络结构基础上提出改进的立体匹配网络架构.如图 2 所示.首先,从立体匹配问题特性出发,提出简洁的特征提取网络 SimpleResNet,替换 PSMNet 网络中的特征提取部分.之后,提出用于视差计算的具备大感受野的分离 3D 卷积层,替换 PSMNet 网络的 Basic 结构中的 3D 卷积层.

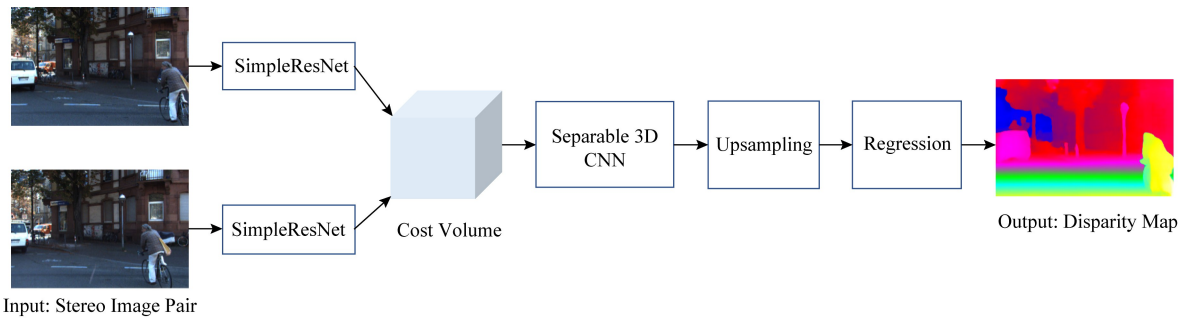

Fig. 2　Architecture of the proposed task-adaptive network for stereo matching

图 2　本文适应立体匹配任务的网络结构

2　适用立体匹配的特征提取结构

PSMNet 网络的特征提取部分使用了类似 ResNet 网络的结构.但是,ResNet 网络原是针对图像分类任务所设计.图像分类与立体匹配 2 个任务之间存在较大差别.

首先,图像分类属于平移不变任务,要求即使输入图像发生平移变换,输出分类结果不变.与图像分类不同,立体匹配需要平移敏感,即当输入图像有空间移动时对应结果也应该有相应变化.其次,图像分类需要提取高度抽象的特征用于表达语义信息,以对图像内容类型做出准确判断.但是,立体匹配求解过程并不需要语义级别特征.事实上,立体匹配需要求解 2 张图像上像素点之间的对应关系.因此,要求网络在不同空间位置提取的特征具有更多局部细节描述能力.

具体而言,在分类任务中,通常采用较深的特征提取网络,并倾向采用大的感受野.网络越深,感受野越大,可用的全局信息越多.同时,感受野足够大,才能够提取出具有平移不变性的特征.而且,随着网络层数的加深,提取到的特征更抽象,也更能够表征图像语义.但是,在立体匹配任务中,增大感受野将降低网络所得特征的差异性.特征图上相近位置的特征将因为大的感受野所导致的对应图像区域重叠度大而变得相似,特征自身位置信息也变得模糊.

综上,立体匹配任务对于高抽象程度的语义特征没有需求.采用较低抽象程度的特征更便于区别不同空间位置的像素.抽象特征不仅难以带来显著的性能提升,而且其获取需要较深层次网络结构.而较深网络意味着更多的参数量与计算量.因此,直接将原本用于处理图像分类任务的网络结构用于立体匹配不够合理.

鉴于上述事实,本文设计了专用于立体匹配任务的特征提取网络.在该网络中,通过限制卷积核大小以减少感受野;同时,简化卷积网络结构,在大幅减少参数量的同时增加网络能够提取的细节信息的比重,使提取的特征具有更强的差异性,更便于鉴别不同空间位置的像素.鉴于所提出的特征提取网络是依据立体匹配特性针对 ResNet 网络进行简化得来,将其称为 SimpleResNet 网络.其具体结构如表 1

Table 1　Architecture Parameters of SimpleResNet

表 1　SimpleResNet 网络结构参数

Layer	[Kernel Size, #Channel]	Output
*Conv*0_1	[3×3,32]	$\frac{1}{2}H\times\frac{1}{2}W\times32$
*Conv*0_2	[1×1,32]	$\frac{1}{2}H\times\frac{1}{2}W\times32$
*Conv*1_*x*	$\begin{bmatrix}1\times1,\ 32\\ 1\times1,\ 32\end{bmatrix}$	$\frac{1}{2}H\times\frac{1}{2}W\times32$
*Conv*2_*x*	$\begin{bmatrix}3\times3,\ 64\\ 1\times1,\ 32\end{bmatrix}\times4$	$\frac{1}{4}H\times\frac{1}{4}W\times64$
*Conv*3_*x*	$\begin{bmatrix}1\times1,\ 128\\ 1\times1,\ 128\end{bmatrix}$	$\frac{1}{4}H\times\frac{1}{4}W\times128$
ConvLast	[1×1,32]	$\frac{1}{4}H\times\frac{1}{4}W\times32$

所示.其中,H 和 W 分别表示输入图像高和宽.网络参数表示方法与 ResNet 网络相同[16].例如 $Conv2\_x$ 残差单元包含 4 组 2 层卷积操作. 其中,第 1 层卷积操作的卷积核尺寸为 3×3,输出通道为 64;第 2 层卷积操作的卷积核尺寸为 1×1,输出通道为 32.

为方便与 PSMNet 网络的特征提取部分进行对等比较以验证本文推论,本文在 SimpleResNet 网络中保留了 PSMNet 特征提取部分所用 ResNet 网络的整体结构,只依据推论做出 2 方面修改:限制了非下采样且具有重复结构层的卷积核大小;大幅度消减重复网络结构.如此,限制感受野和减少层数的同时保持网络整体结构不发生改变.具体地,只保留 PSMNet 中 $Conv\_1$ 与 $Conv3\_x$ 中的卷积核原尺寸,其余卷积核减少至 1×1 大小;去除 $Conv1\_x$, $Conv2\_x$, $Conv3\_x$ 模块中的重复卷积结构;将原特征提取部分的 48 层卷积层减少至 14 层.重复结构和层数减少量依据实验得来.

相比 PSMNet 网络的特征提取结构,SimpleResNet 网络复杂度和计算量减少显著:原特征提取网络约有 300 万个参数,SimpleResNet 网络参数减少至不足 20 万个,减少量超过 90%.所作简化操作均源自前述推论:立体匹配并不需要过大感受野和高度抽象的语义信息,如 ResNet 网络这类层次深、感受野覆盖范围大的结构,对于立体匹配任务而言存在极大冗余.因此,所提出方法并不会因为网络深度的大幅度减小而对立体匹配准确度有较大影响.后续通过实验验证了该观点.

3 分离 3D 卷积(separated 3D CNN)

在端到端立体匹配深度网络中,3D 卷积能够在优化匹配成本特征体的同时实现视差计算.与前段网络部分以提取特征为目的不同,此段网络的目的是利用邻域信息优化已提取的成本特征,以减少视差中错误离散值的出现.

鉴于采用更大邻域范围内的约束将更有利于优化成本特征,倾向在网络中应使用较大卷积核.然而,3D 卷积计算复杂度高,扩大 3D 卷积核将带来巨大的参数量增加.

为了克服 3D 卷积计算在扩大感受野时的参数量爆炸问题,提出分离 3D 卷积层结构.图 3(a)是正常的 3D 卷积层,图 3(b)是所提出的分离 3D 卷积层.分离 3D 卷积将 3D 卷积分解为 1 次长与宽维度的 2D 卷积与 1 次视差维度的 1D 卷积.普通 3D 卷积参数量为 $m^2\times c$(m 为卷积核宽/高,c 表示类型数),而使用分离 3D 卷积的参数量是 m^2+c.当 m 与 c 增长时,后者参数量增长远小于前者.

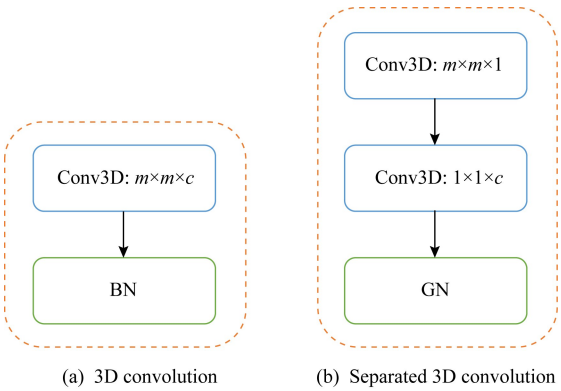

Fig. 3 3D convolution and separated 3D convolution

图 3 3D 卷积与分离 3D 卷积

此外,考虑到与分类任务不同,立体匹配任务中输入图像分辨率大,训练时每批数据数量较少,不适合在分离 3D 卷积后接批量归一化(batch normalization, BN)操作.因此,本文使用分离 3D 卷积搭配群组归一化(group normalization, GN)层替代 PSMNet 里的 3D 卷积和 BN 层结构.

4 实验与结果

本节通过实验验证所提出方法和模块.首先,介绍所用数据集、评价标准与网络训练过程.然后,通过消融实验验证本文方法和模块的有效性.之后,通过与当前优秀方法进行性能比较与分析,证明本文方法的先进性.最后,通过对比基准网络 PSMNet,证实所提出网络在时间、显存占用方面的优势.

4.1 数据集与训练过程

使用 SceneFlow 数据集完成网络的训练与对比实验.SceneFlow 数据集[11]属于大型合成数据集,共计包含 35 454 对训练用立体图像数据与 4 370 对测试用立体图像数据.每对数据都提供了稠密真值信息与相机参数信息.数据集中所有图像分辨率均为 960×540.SceneFlow 数据集按照场景内容分为 3 个子数据集,分别是 FlyingThings3D, Driving, Monkaa.其中,Flying-Things3D 子数据集的场景包含大量漂浮的随机类型物体,图像内物体较多,细节内容较为丰富;Driving 子数据集是在模拟汽车驾驶过程中所抓取,其内容是开阔的街道场景;Monkaa 子数据集的场景内容是深林环境中的猴子,其中有较多近距离物体,即存在较多视差值较大的区域.

本文实验使用端点误差(end-point-error, EPE)作为量化指标评.端点误差是真值与估计视差值之间的平均欧氏距离.其计算为

$$e_{\mathrm{EPE}}=\frac{1}{N}\sum_{n=1}^{N}|d_n-gt_n|, \tag{1}$$

其中,d 与 gt 分别表示视差的估计值和真值,N 是参与计算的像素总数,n 是指示变量.EPE 评价指标分值越低表示方法性能越好.

在 Pytorch 框架下构建网络.在 Ubuntu 操作系统下完成网络训练.训练所使用 GPU 为单个 NVIDIA GeForce TITAN X Pascal.

网络使用 Smooth L1 距离作为监督训练的损失函数.需要注意的是,所提出网络和 PSMNet 网络相同,均设置视差上限为 192.该值适用于大部分实用情形.但是,SceneFlow 数据集中存在部分数据视差真值超过该上限.为避免此部分数据的干扰,训练过程只取用视差真值小于等于 192 数据参与计算损失函数回传.训练损失函数:

$$l_{\mathrm{Loss}}=\frac{1}{N'}\sum_{n=1}^{N}l_{\mathrm{smoothL1}}(d_n,gt_n), \tag{2}$$

其中:

$$l_{\mathrm{smoothL1}}(d_n,gt_n)=\begin{cases}0.5|d_n-gt_n|^2, & |d_n-gt_n|<1,\\ |d_n-gt_n|-0.5, & \text{otherwise},\end{cases}$$

其中,d_n 与 gt_n 分别表示第 n 点的网络估计视差值与视差真值,N'表示视差真值小于等于 192 的像素点总数.

网络每次训练数据批量大小设置为 3.训练过程是端到端训练,无需其他后处理方法.针对数据集中每对图像以原尺寸输入网络,训练学习率设置为 0.001,在 SceneFlow 数据集上使用 Adam 优化方法[19]训练 10 轮.Adam 方法参数为 $\beta_1=0.9$,$\beta_2=0.999$.训练时间持续约 40 h.

4.2　消融实验与分析

在 SceneFlow 测试集上通过消融实验测试所提出网络的核心模块,包括特征提取模块 SimpleResNet 和分离 3D 卷积模块 Globalconv.结果如表 2 所示.

用 ResCNN+Basic+BN 表示 PSMNet 网络的原始基础结构.SimpleResNet+Basic+BN 表示使用本文提出的特征提取模块 SimpleResNet 代替 PSMNet 中的特征提取模块 ResCNN.从替换前后的 EPE 数值可看出:相比 ResCNN,所提出特征提取网络结构 SimpleResNet 网络参数降低超过 90%(见第 3 部分分析)的情况下,依然能够保持较低的 EPE.

Table 2　EPEs of PSMNet and Our Networks

表 2　本文网络与原网络 EPE 对比

Methods	e_{EPE}
ResCNN+Basic+BN	1.56
SimpleResNet+Basic+BN	1.59
SimpleResNet+Globalconv+BN	1.53
SimpleResNet+Globalconv+GN	1.40
ResCNN+SPP+Stackhourglass+BN	1.09
SimpleResNet+Stackhourglass+BN	1.03

进一步地,以 PSMNet 网络性能最佳的组成结构 ResCNN+SPP+Stackhourglass+BN 为基准方法验证所提出的特征提取网络.表 2 中 SimpleResNet+Stack hourglass+BN 是用所提出特征提取网络替换原特征提取网络.从表 2 实验数据看:在此框架下,所提出特征提取网络在参数量减少情况下,EPE 更低.

为验证所提出分离卷积的效果,在 SimpleResNet+Basic+BN 基础上,使用所提出的分离 3D 卷积替换 Basic 中所用到的普通 3D 卷积,记作 SimpleResNet+Globalconv+BN.从表 2 可以看出:相比普通 3D 卷积,分离 3D 卷积能够有效降低误差.表 2 中 SimpleResNet+Globalconv+GN 方法是用分离 3D 卷积加 GN 层替换了 SimpleResNet+Basic+BN 中原属于 PSMNet 的后 2 部分.从结果可看出:该方法的 EPE 相比 SimpleResNet+Basic+BN 的 EPE,降低约 12%的相对量.该实验说明:所提出分离 3D 卷积搭配 GN 操作相比普通 3D 卷积搭配 BN 操作,性能显著更佳.

需要注意的是:以上对比方法(除 PSMNet 网络性能最佳的组成结构外),以及本文方法均未加入 SPP 结构.既是为了方便与 PSMNet 中基础方法(不含 SPP 结构)对比,也是因为 SPP 结构将破坏 SimpleResNet 网络局部性特征提取能力.

4.3　横向对比实验与分析

所提出方法与其他现有深度学习立体匹配方法在 SceneFlow 测试集上的 EPE 对比如表 3 和图 4

Table 3　EPEs of Our Networks and Other Methods

表 3　本文网络与其他网络方法 EPE 对比

Methods	e_{EPE}
MC-CNN[10]	3.79
DispNet[11]	1.84
DispFulNet[11]	1.75
CRL[12]	1.32
PSMNet[13]	1.09
Ours	1.03

所示.其中 Ours 为本文性能最佳方法,对应表 2 中的 SimpleResNet+Stackhourglass+BN.从表 3 和图 4 可见:相比其他 5 种方法,PSMNet 网络效果最佳;而相比 PSMNet 网络,本文方法在大幅度降低特征提取模块参数的同时,EPE 更低.

图 5 是采用 SimpleResNet+Stackhourglass+BN 方法与 PSMNet 方法在 ScenenFlow 数据集上所得视差结果.如图 5 所示,本文方法与 PSMNet 网络在大部分区域表现相近,均能够得到与真值相近的结果.但是在方框标示的物体内部的连续区域,PSMNet 网络所得结果出现明显错误,而所提出方法结果基本无误.

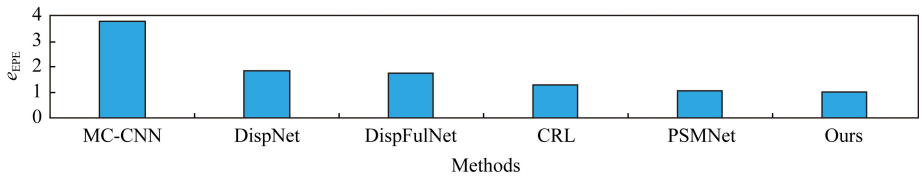

Fig. 4 Visualized EPEs of our network and other methods

图 4 本文网络与其他网络方法 EPE 可视化对比

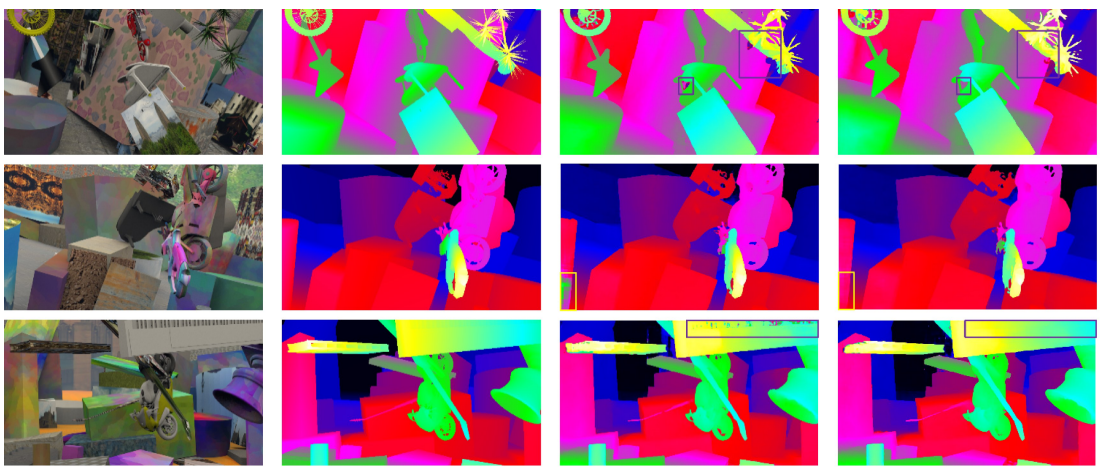

Fig. 5 Disparity maps obtained by our network and PSMNet

图 5 本文网络与 PSMNet 网络视差计算结果

4.4 计算成本分析

表 4 和图 6 给出了所提出方法和基准方法的时间与显存开销的实验数据.在此实验中,所有方法在每批次训练过程中统一采用 1 张输入图像.从表 4 和图 6 中可以看出:本文方法 SimpleResNet+Basic 比基准方法 ResCNN+Basic 减少了约 0.4 GB 显存

Table 4 Computation Time and Memory Costs of Different Networks

表 4 各网络计算时间开销与显存占用

Methods	Computation Time/s	Memory Size/GB
SimpleResNet+Basic+BN	0.3	2.4
SimpleResNet+Globalconv+BN	0.9	3.1
ResCNN+Basic+BN	0.4	2.8
ResCNN+SPP+Stackhourglass+BN	0.6	4.2

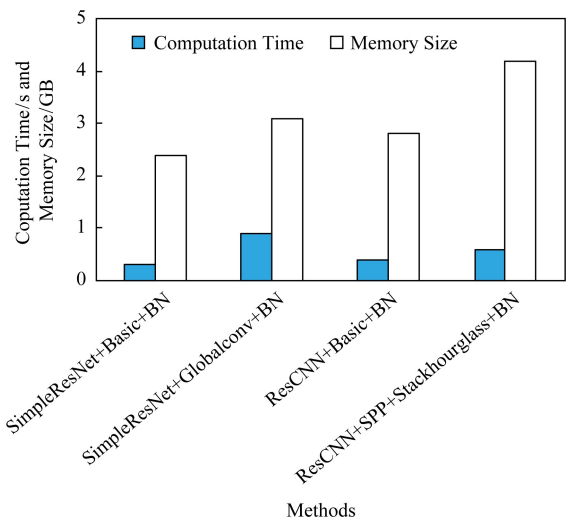

Fig. 6 Visualized time and memory costs of different networks

图 6 各网络计算时间开销与显存占用可视化对比

占用.同时,单次训练速度加快约 0.1 s,提升量为 25%.相较 Basic 方法,使用 Globalconv 模块显存增量约为 0.7 GB.显存占用增长是由于 Globalconv 模块所用更大的卷积核以及所增加的中间缓存所导致.其中,由中间缓存带来的显存占用增长占比更多.在实际使用中可以通过及时释放中间缓存以释放更多的显存.Basic 结构使用相同大小的卷积核会因为显存不足而无法运行.

改进后 Globalconv 模块相比 Basic 方法时间增长较为明显.这是因为分离卷积后,在视差维度上将卷积核扩大至覆盖所有视差的同时,需要对原特征填补较大区域,该操作带来较多额外耗时.

5 结　论

在分析现有端到端立体匹配网络不足之处的基础上,提出了改进的立体匹配深度网络.首先,根据立体匹配任务特性设计了新的特征提取结构.该结构在减少 90%参数量的同时,能够保持稳定的网络性能.之后,提出了分离 3D 卷积运算.相比传统 3D 卷积,该运算能够在增大卷积核尺寸扩大感受野的同时抑制参数量的增加,从而提高视差计算的精度.

所提出网络也存在不足之处.例如分离 3D 卷积需要额外的填补操作、增加了较多额外耗时.本文将在未来工作中着手解决该问题.

参　考　文　献

[1] Ma Wei, Yang Luwei, Zhang Yu, et al. Fast interactive stereo image segmentation [J]. Multimedia Tools and Applications, 2016, 75(18): 10935-10948

[2] Ma Wei, Qin Yue, Yang Luwei, et al. Interactive stereo image segmentation with RGB-D hybrid constraints [J]. IEEE Signal Processing Letters, 2016, 23(11): 1533-1537

[3] Ma Wei, Qin Yue, Xu Shibiao, et al. Interactive stereo image segmentation via adaptive prior selection [J]. Multimedia Tools and Applications, 2018, 77(21): 28709-28724

[4] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite [C] //Proc of Computer Vision & Pattern Recognition. Piscataway, NJ: IEEE, 2012: 3354-3361

[5] Xu Sheng, Ye Ning, Zhu Fa, et al. Solving the stereo matching in disparity image space using the Hopfield neural network [J]. Journal of Computer Research and Development, 2013, 50(5): 1021-1029
(徐昇, 业宁, 朱发, 等. Hopfield 网络在视差空间上的立体匹配求解[J]. 计算机研究与发展,2013,50(5): 1021-1029)

[6] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149

[7] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4): 640-651

[8] Lin Guosheng, Milan A, Shen Chunhua, et al. RefineNet: Multi-path refinement networks for high-resolution semantic segmentation [C] //Proc of Computer Vision & Pattern Recognition. Piscataway, NJ: IEEE, 2017: 5168-5177

[9] Luo Wenjie, Schwing A G, Urtasun R. Efficient deep learning for stereo matching [C] //Proc of Computer Vision & Pattern Recognition. Piscataway, NJ: IEEE, 2016: 5695-5703

[10] Bontar J, LeCun Y. Computing the stereo matching cost with a convolutional neural network [C] //Proc of Computer Vision & Pattern Recognition. Piscataway, NJ: IEEE, 2014: 1592-1599

[11] Mayer N, Ilg E, Häusser P, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation [C] //Proc of Computer Vision & Pattern Recognition. Piscataway, NJ: IEEE, 2016: 4040-4048

[12] Pang Jiahao, Sun Wenxiu, Ren J SJ, et al. Cascade residual learning: A two-stage convolutional neural network for stereo matching [C] //Proc of Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 878-886

[13] Chang Jiaren, Chen Yongsheng. Pyramid stereo matching network [C] //Proc of Computer Vision & Pattern Recognition. Piscataway, NJ: IEEE, 2018: 5410-5418

[14] Yang Guorun, Zhao Hengshuang, Shi Jianping, et al. SegStereo: Exploiting semantic information for disparity estimation [C] //Proc of European Conf on Computer Vision. Piscataway, NJ: IEEE, 2018: 660-676

[15] Song Xiao, Zhao Xu, Hu Hanwen, et al. EdgeStereo: A context integrated residual pyramid network for stereo matching [J]. arXiv preprint, arXiv:1803.05196, 2018

[16] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition [C] //Proc of Computer Vision & Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770-778

[17] Zhao Hengshuang, Shi Jianping, Qi Xiaojuan, et al. Pyramid scene parsing network [C] //Proc of Computer Vision & Pattern Recognition. Piscataway, NJ: IEEE, 2017: 6230-6239

[18] Newell A, Yang Kaiyu, Deng Jia. Stacked hourglass networks for human pose estimation [J]. arXiv preprint, arXiv: 1603.06937, 2016

[19] Kingma D P, Ba J. Adam: A method for stochastic optimization [J]. arXiv preprint, arXiv: 1412.6980, 2014

Li Tong, born in 1993. Master. His main research includes stereo vision and deep learning.

Xu Shibiao, born in 1986. PhD, associate professor. Member of CCF. His main research interests include image based three-dimensional scene reconstruction and scene semantic understanding.

Ma Wei, born in 1980. PhD, associate professor. Member of CCF. Her main research interests include image/video repairing, image/video semantic understanding and 3D vision.

Zhang Xiaopeng, born in 1963. PhD, professor. Member of CCF. His main research interests include computer graphics and computer vision.

《计算机研究与发展》征订启事

《计算机研究与发展》(Journal of Computer Research and Development)是中国科学院计算技术研究所和中国计算机学会联合主办、科学出版社出版的学术性刊物,中国计算机学会会刊.主要刊登计算机科学技术领域高水平的学术论文、最新科研成果和重大应用成果.读者对象为从事计算机研究与开发的研究人员、工程技术人员、各大专院校计算机相关专业的师生以及高新企业研发人员等.

《计算机研究与发展》于 1958 年创刊,是我国第一个计算机刊物,现已成为我国计算机领域权威性的学术期刊之一.并历次被评为我国计算机类核心期刊,多次被评为“中国百种杰出学术期刊”.此外,还被《中国学术期刊文摘》、《中国科学引文索引》、“中国科学引文数据库”、“中国科技论文统计源数据库”、美国工程索引(Ei)检索系统、日本《科学技术文献速报》、俄罗斯《文摘杂志》、英国《科学文摘》(SA)等国内外重要检索机构收录.

国内邮发代号:2-654;国外发行代号:M603

国内统一连续出版物号:CN11-1777/TP

国际标准连续出版物号:ISSN1000-1239

联系方式:

100190 北京中关村科学院南路 6 号《计算机研究与发展》编辑部

电话: +86(10)62620696(兼传真);+86(10)62600350

Email:crad@ict.ac.cn

http://crad.ict.ac.cn