Computer Science and Application 计算机科学与应用, 2022, 12(8), 1964-1973
Published Online August 2022 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2022.128197

Hans汉斯

基于局部代价层级优化的立体匹配

胡　昊，刘兴林

五邑大学智能制造学部，广东　江门

收稿日期：2022年7月16日；录用日期：2022年8月16日；发布日期：2022年8月23日

摘　要

随着深度学习的出现和立体匹配数据不断丰富，立体匹配结合深度学习成为研究热点。虽然结合深度学习的立体匹配算法精度上不断地获得提升，但是精度的提升带来的计算复杂度不断加深的神经网络，导致大量的计算成本，这将导致其中立体匹配的方法并不是适用于常规的CPU或者GPU运算。因此保持较高精度并降低算法计算复杂度成为当前立体匹配适用工程上的热点问题。对此本文从降低算法复杂度出发并保持算法性能的目的，提出结合深度可分离卷积的ASPP (Atrous Spatial Pyramid Pooling)特征提取，构建局部代价体来降低计算复杂度和内存消耗，并通过层级的方式不断地对视差进行优化，从而保证性能的稳定性。研究通过实验表明，通过局部代价的方式和深度可分离卷积能够降低算法的运行时间，以及通过层级优化和ASPP方式提取特征能保证精度水平，从而在计算成本和精度水平上有一个很好的性能表现。

关键词

立体匹配，ASPP，局部代价，层级优化

Stereo Matching with Hierarchical Optimization Based on Local Cost

Hao Hu, Xinglin Liu

Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen Guangdong

Received: Jul. 16th, 2022; accepted: Aug. 16th, 2022; published: Aug. 23rd, 2022

Abstract

With the emergence of deep learning and the continuous enrichment of stereo matching datasets, stereo matching combined with deep learning has become a research hotspot. Although the accuracy of the stereo matching algorithm combined with deep learning has been continuously im-

文章引用: 胡昊, 刘兴林. 基于局部代价层级优化的立体匹配[J]. 计算机科学与应用, 2022, 12(8): 1964-1973.
DOI: 10.12677/csa.2022.128197

proved, the neural network with increasing computational complexity brought about by the improvement of accuracy leads to a large amount of computational cost, which will result in the stereo matching method not suitable for conventional CPU or GPU operation. Therefore, maintaining high accuracy and reducing the computational complexity of the algorithm has become a hot issue in the current stereo matching engineering. In order to reduce the complexity of the algorithm and maintain the performance of the algorithm, this paper proposes the ASPP feature extraction combined with the depthwise separable convolution, constructs a local cost volume to reduce the computational complexity and memory consumption, and continuously optimizes the disparity in a hierarchical manner to ensure stable performance. The research shows through experiments that the running time of the algorithm can be reduced by means of local cost and depthwise separable convolution, and the level of accuracy can be guaranteed by extracting features through hierarchical optimization and ASPP, so there are many good results in terms of computational cost and precision.

Keywords

Stereo Matching, ASPP, Local Cost, Hierarchical Optimization

1. 引言

随着计算机技术和计算能力的飞跃发展，人工智能的发展也不断突破创新，人类日常生活越来越多的智能化设备不断地涌现。立体匹配一直都是人工智能诸多任务中的重要一环，比如工业生产中的匹配系统、汽车领域的自动驾驶技术、3D 建模中的深度测量等。立体匹配也称之为视差估计，或者双目深度估计。输入的是同水平面上的两个相机同时拍摄的图像，经过极线矫正的左右图像 I_l 和 I_r，输出的是左图中的每个像素对应右图中匹配像素的视差值所构成的视差图。视差图 d 根据已知的相机参数 b 和 f 可以计算出深度图，$D=\frac{b*f}{d}$，如图 1 所示。

立体匹配工作发展了几十年之久，工作的任务都是估计立体图片之间的视差。传统立体匹配方法一般分为局部和全局两种类型[1] [2]，分别为基于支持窗口的方法和最小化能量函数。立体匹配方法根据不同算法大致可分为以下四个步骤[2]：1) 匹配代价计算；2) 代价聚合；3) 视差计算；4) 视差优化。传统的立体匹配虽然能达到一定的效果，但是对于一些简单的场景，而随着深度学习卷积网络在视觉工作中的大行其道，立体匹配工作也纷纷引入了深度学习的框架，来进一步增加立体匹配算法在现实任务中的可行性。但是深度学习下的立体匹配框架[3] [4] [5]，往往需要大量的计算资源，从而导致算法不具备工程实用性。因此有大量的工作者研究轻量化或者加快运行效率的方法，有基于可微分代价块匹配进行代价体修建的方法[6]、层级估计高分辨率来减少计算步骤的方法[7]、引入轻量化的卷积的方法[8] [9]、进行局部视差估计的方法[10]。本文通过引入深度可分离卷积到 ASPP 进行特征提取，对于边缘信息能够提供足够的感受野的同时并能够保证计算效率。对于一般方法中体量庞大的代价体部位，本文采用传统的绝对误差和的方式进行计算各个视差之间的代价，并结合稀疏代价体进一步缩减了初始化代价体的体量。在层级优化视差过程中，通过使用初始化视差来形成局部视差代价体来减少代价体的内存消耗和计算复杂度的冗余，最后进行多尺度的视差结果融合形成最终的优化视差结果。

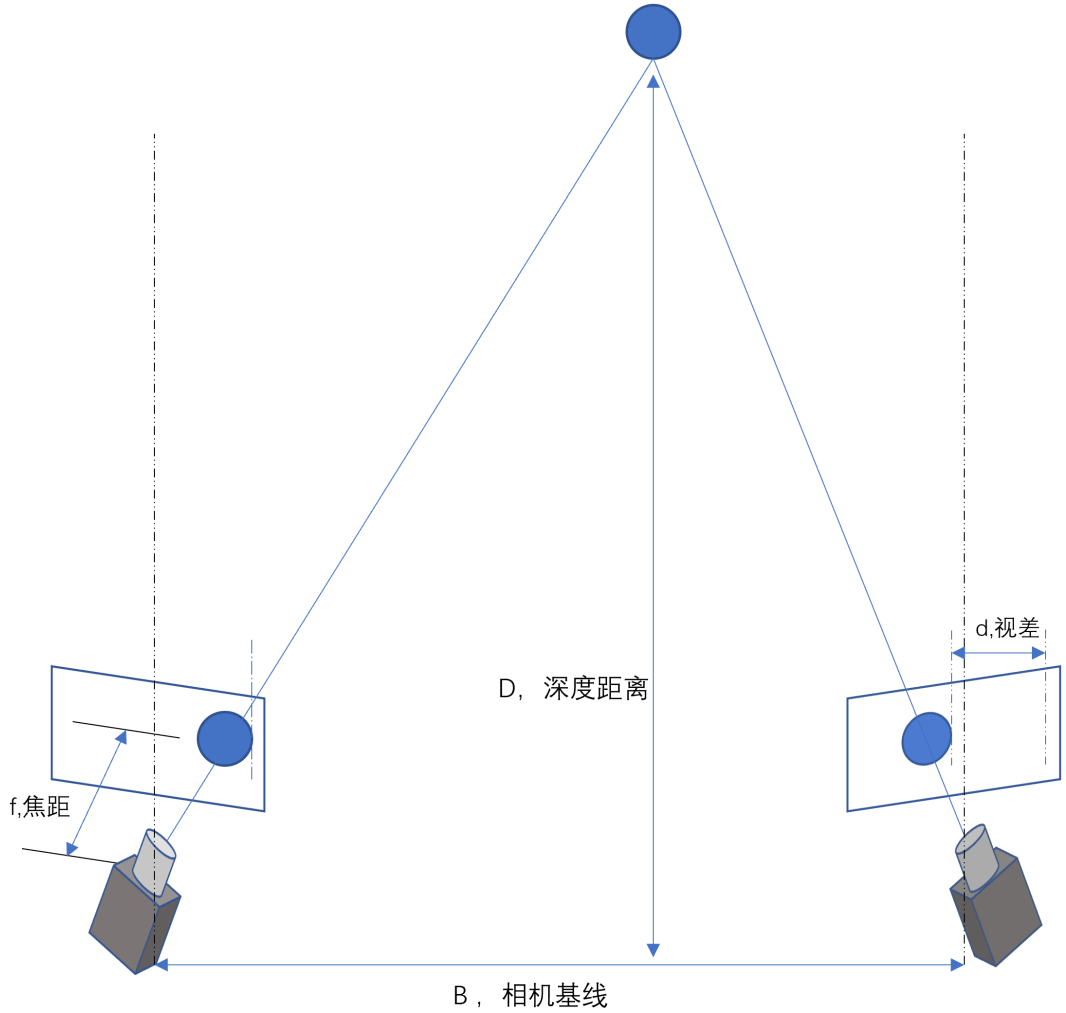

Figure 1. Basic structure of stereo matching
图 1. 立体匹配基本结构

2. 相关工作

2.1. 代价体

在立体匹配过程中，代价算是非常重要的一部分，承载着立体匹配结果的好坏。代价计算本质上是计算左右匹配特征之间的相似度，传统算法上通常使用简单的计算，比如绝对差、海明距离或者相关性，虽然计算简单，但是性能不差。在深度学习的框架下，可分为两种代价体构建的方式。第一种就是计算单元特征相关性构建 3D 代价体，而后使用 2D 卷积进行代价聚合，在效率和准确度上更倾向于效率；第二种直接拼接单元特征来构建 4D 代价体，或者进行变体同样构建 4D 代价体来增强估计精确度，比如基于方差的方式[11]、分组计算相关性的方式[12]、金字塔代价体[13]。通过 3D 卷积对 4D 代价体进行代价聚合所需要的计算资源是庞大的，也有方法[14]采用多尺度融合代价体，低分辨率特征形成 4D 代价体，高分辨率形成 3D 代价体，来权衡 3D 代价体和 4D 代价体的优缺。为了符合传统立体匹配的过程，使用深度学习端到端训练学习的方式，把整个流程放入 GPU 进行计算能大大提升计算效率。但是由于大尺度的代价体和代价聚合步骤，需要庞大的内存和计算资源，这使得方法的实际应用价值不高。

有大量的关于代价体缩减或者进行稀疏的研究，Deeppruner [6]中沿用块匹配(Patch Match)的思想进行并结合控制每个部位的最大最小视差来形成修建的代价体，来减少大体量代价体所带来的庞大内存占用。SCV-Net [15]中使用再视差维度以步长的跳跃形成稀疏的代价体，来形成小体量/稀疏的代价体。本文中使用局部视差的代价体并结合置信度和视差梯度，在多尺度层级的不断优化视差。此过程中的局部代价体所计算的视差范围在±1 之间，所以所占用的计算和内存需要是较少的，并且引入置信和视差的梯度融合弥补精度上的不足。

2.2. 可分离卷积

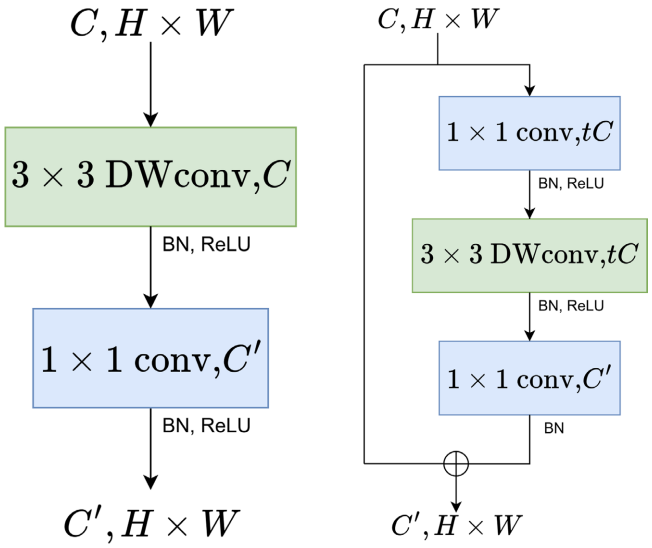

Figure 2. Left: MobileNet-V1 block, Right: MobileNet-V2 block [9]
图 2. 左：MobileNet-V1 块，右：MobileNet-V2 块[9]

深度可分离卷积[16] [17]能够实现常规卷积相似的性能，并实现显著减少计算复杂度和参数量从而来提升整体网络的效率。MobileNets [18]应用深度可分离卷积到移动设备的视觉任务上，获得绝佳的效果，而后又提出 MobileNets-V2 [19]使用模块内的残差方式来增强了 MobileNets 性能并提升输入通道兼容性，可见图 2。MobileStereoNet [9]结合两种分离卷积模块融合立体匹配中进行研究，分别对特征提取和代价体优化步骤中的卷积进行替换，研究结果表示其模型在参数量上得到显著的优化，并同时保证了立体匹配的性能。

3. 框架方法

框架的设计遵守立体匹配中的基本步骤[2]。首先通过可分离卷积构成的 ASPP 特征提取，再构建局部视差代价体逐级优化视差结果，最后获取最佳的结果。

3.1. 特征提取

特征提取网络中要求生成可靠的点特征，立体匹配过程中通过一系列可靠的特征能实现优质的匹配结果。根据相关研究[7] [14]表明，在特征提取的过程中使用 SPP [20]能够显著地增加提取过程中的感受野，而在 SPP 基础上再添加空洞卷积可在不丢失分辨率(不进行下采样)的情况下扩大卷积核的感受野。ASPP 虽然在特征提取以及边缘信息提取效果比较好，但是其的计算效率却有所增大，所以为了保证性能并提升计算效率，DeepLabv3+ [21]提出的融合深度可分离卷积和 ASPP，再融合更多的语义信息和详细

边缘目标信息的基础上来降低运行复杂度，从而来保证特征提取过程中的性能和效率。本文的特征提取便由[21]的 DeepLabv3+改进而来，在普通卷积上面替换成 MobileNets-V2 提出的深度可分离卷积，对于 ASPP 中的最大池化层，使用步长为 2 的深度可分离卷积来替代，如图 3 所示，这能保证空洞分离卷积能保证在任意分辨率下都能进行，其他部分的卷积同样使用 MobileNets-V2 块进行替换来减少参数里量。通过可分离卷积融合的 ASPP 特征提取方式，能够实现较低计算复杂度的情况下保证匹配结果。而后通过提取的多尺度特征，构建局部视差代价来不断优化视差估计结果。对左右两图分别进行特征提取，获取两个多尺度的特征表示为 f^L 和 f^R。

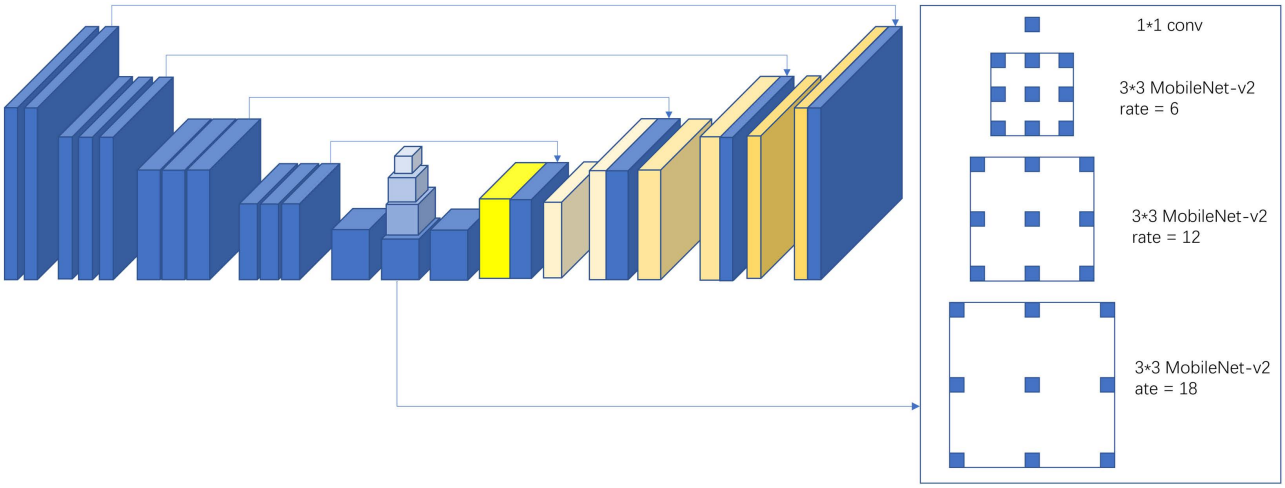

Figure 3. ASPP feature extraction with separable convolution fusion
图 3. 可分离卷积融合的 ASPP 特征提取

3.2. 局部视差代价优化

3.2.1. 初始化

视差初始化的步骤与传统立体匹配中的代价匹配计算类似，但是为了压缩代价体的体积，所以再沿着视差维度上构建代价体的形式采用了与 SCV-Net [15]相似的稀疏代价体构建方式，所以通道上系数对应的视差值是两倍的关系。构建代价体过程中的代价计算采用较为简单的绝对误差和算法(SAD)，这种计算方式在运算过程中还可以优化成更小的内存占用，所以对于匹配代价 c 表示在坐标 (x,y)，分辨率 l 和视差 d 下可以定义为：

$$c(l,x,y,d)=\sum_{h\in f_c}\left|f^L_{h,l,x,y}-f^R_{h,l,x,y}\right| \tag{1}$$

其中 $h\in f_c$，表示 h 的值取自于特征向量的通道维度。

初始视差根据视差维度对应的代价值取最小值表示初始化的视差，然后初始视差队对应的代价值保留用作为置信作为后面视差优化的引导。

$$d^{Init}_{l,x,y}=2*\arg\min_{d\in(0,D)}c(l,x,y,d) \tag{2}$$

考虑当前维度下的最大视差范围 0 到 D，取最小值也即匹配度最高的视差作为初始化视差。由于通过稀疏代价体的构建方式，所以视差的值需要进行比例的放大。由此得出的视差是十分粗糙的，根据层级优化的思想，对于粗糙的视差下一个尺度需要不断地优化。为了更好的在下一个尺度中进行优化，沿用[22]的思想在初始化的视差中引入视差在 x,y 方向上的梯度变化，并初始化 dx,dy 都为 0。拼接初始化

视差 $d_{l,x,y}^{Init}$、dx 和 dy 形成初始化视差描绘体 $d_v=\left[d_{l,x,y}^{Init},0,0\right]$。在优化的过程仍然需要初始化视差中的引导信息，这里使用初始化视差的代价值与左特征作为引导信息，并通过一个感知器 P 对引导信息进行整合和激活。

$$F_{l,x,y}^{init}=P\left(\delta\left(d_{l,x,y}^{Init}\right),f_{l,x,y}^{L}\right) \tag{3}$$

$\delta\left(d_{l,x,y}^{Init}\right)$表示取视差下的代价值作为匹配的置信，感知器通过 1*1 的卷积和 leaky ReLU 构成。最后整合初始化中的所有信息构成下一尺度所需要的视差优化信息。

$$I_{l,x,y}^{init}=\left(d_{l,x,y}^{Init},0,0,F_{l,x,y}^{init}\right) \tag{4}$$

3.2.2. 构建局部代价体

局部代价体是建立在右特征进行已知视差的变换与左特征进行的代价计算，把给定的视差作为右特征的偏移量，对右特征进行平移变换再与左特征进行 SAD 算法计算的到当前的视差下的相似度，也即代价值。

$$C_{warp}=\sum\left|f^{L}-f_{warp}^{R}\right| \tag{5}$$

构建局部代价体的目的是对视差进行优化，所以已求出得视差进行邻近视差范围进行寻找匹配从而构建局部代价体，这种方法能有效地优化视差。根据尺度之间的比例，选用邻近视差范围为 1 个视差范围，也即 $d'\in\left\{-1+d,0,1+d\right\}$，对三个视差进行变换代价计算从而构建局部代价体 CV_l。

对于由上一尺度产生的视差，因为数据尺度上存在比例差异，所以需要对视差进行上采样处理。在这里的上采样处理不使用常用的线性或者邻近插值方式，因为引入了视差的梯度来表示视差的变化，所以采用视差的梯度变化来上采样视差到指定的尺度。

$$d'_{i,j}=d+\left(i-0.5\right)dx+\left(j-0.5\right)dy \tag{6}$$

上采样过后的是视差同样进行局部代价体的构建，并于当前尺度下形成的局部代价体组合成局部代价体集 S_{cv}。

3.2.3. 视差优化

对于获取到的局部代价体集，是需要与同尺度的视差优化信息融合作为整合信息，并通过处理获取原视差所需要优化信息。当前尺度中的视差优化信息直接与对应的局部代价体进行拼接形成整合信息。对于上一尺度中的视差优化信息，由于尺度上与局部代价体存在比例差异，所以需要对其中优化信息进行上次采样到相同尺度下在进行拼接成整合信息。所有尺度形成的整合信息再拼接融合成优化信息体。

$$I_{l-1}=concat\left\{d',dx_{l-1}^{up},dy_{l-1}^{up},F_{l-1}^{up}\right\} \tag{7}$$

$$OI_v=concat\left\{I_l,CV_l,I_{l-1},CV_{l-1},\cdots\right\} \tag{8}$$

其中带有上标 up 的表示经过上采样得来。

为了更好地利用优化的信息体，所以再聚合的过程中使用了残差网络进行聚合，同时为了保证提取的感受野足够，同样采用了空洞卷积进行聚合的步骤。聚合得出的优化结果为对每个视差优化信息进行优化的值，所以得出的结果与原视差优化信息进行整合。对于 $l=1$ 的情况，因为原信息和结果都为单个所以采用直接相加的方式进行。对于多个，聚合过程会多计算值分别对应其中的置信度，最后通过置信度择机的选择其中的优化结果对相应的视差优化信息进行整合，也即相加。对于最后尺度优化得出的结果，使用之前所有尺度上的视差和梯度计算局部代价体，进行最后一次提炼结果，最终输出通道为 1 的最终视差。

3.3. 损失函数

整体使用深度学习的端到端的方式进行训练，并使用的真实视差标签 d_{gt} 进行监督学习。其中的损失函数采用了多尺度上的多种损失和进行计算 $Loss_{total}=\sum_l L_l^{Init}+L_l^d+L_l^{disp}$。其中的多尺度上的初始化视差的损失，通过真实视差标签进行线性下采样进行尺度缩放形成多尺度上的标签数据。对于初始化视差和最终视差损失都使用标准的 $smooth-\mathcal{L}_1$ 损失。对于梯度的标签，采用 sobel 算子使用卷积加速计算标签数据中的视差梯度，引入梯度值便是为了更好的考虑到边缘信息，所以对于梯度的损失采用有阈值的损失函数。

$$L^d\left(dx,dy\right)=\begin{cases}sum\left(\left|d_x^{gt}-d_x\right|,\left|d_y^{gt}-d_y\right|\right), & \left|d^{gt}-\hat{d}\right|<1\\ 0, & \text{else}\end{cases} \tag{9}$$

4. 实验与结果分析

4.1. 实验数据

采用的数据集一共有两个数据集 SceneFlow [23]数据集和 KITTI2015 [24]数据集，其中 SceneFlow 数据集是合成数据集有三个子集，并提供完整的真实场景流(包括前后方向上的视差变化)。KITTI 包含市区、乡村和高速公路等场景采集的真实图像数据，KITTI 15 从 KITTI 原始数据集中收集了 400 个高度动态的场景，并使用半密集的场景流地面真实度进行增强。

4.2. 实验结果与分析

根据实验训练的策略[25]，本文实验过程同样采样预训练和数据增强的步骤，来增强模型的泛化性和鲁棒性。考虑到 SceneFlow 是大体量数据集，KITTI2015 为小体量的数据集，所以实验先采样 SceneFlow 预训练模型，在使用 KITTI2015 来优化模型。

经过实验结果(图 4，图 5)可以看出在 SceneFlow 和 KITTI 数据集中模型的表现，对合成数据集 SceneFlow 的表现视觉上要优于 KITTI2015 的数据集，这种结果可能来源于数据集大小的问题或者两种数据集的数据域存在一定的差距，导致由合成数据泛化到真实拍摄数据存在一定的难度。通过实验预测

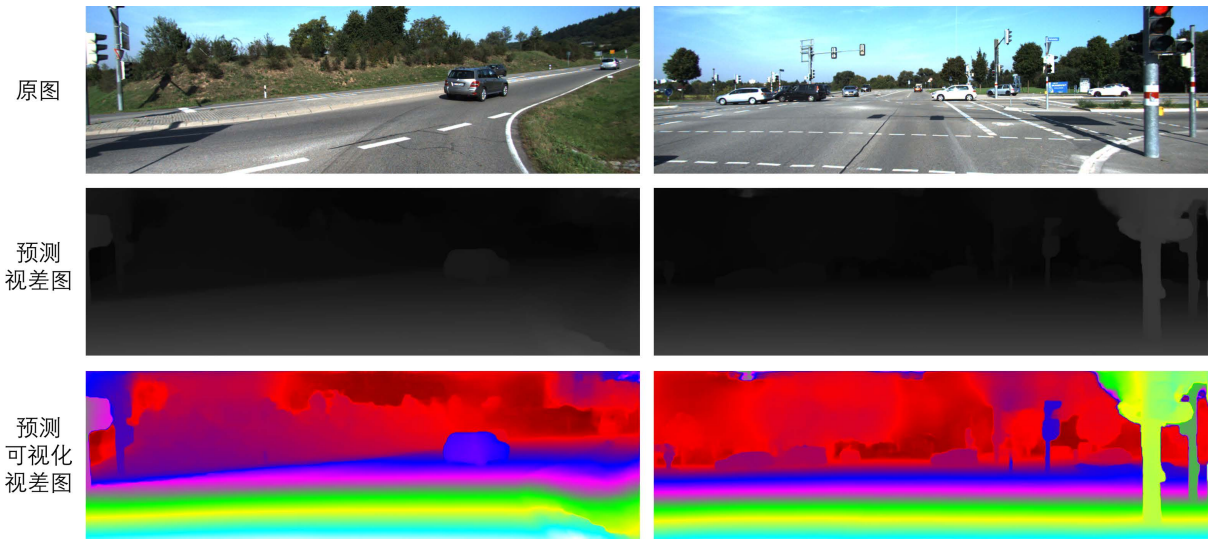

Figure 4. Results on KITTI2015 test
图 4. KITTI 的测试结果

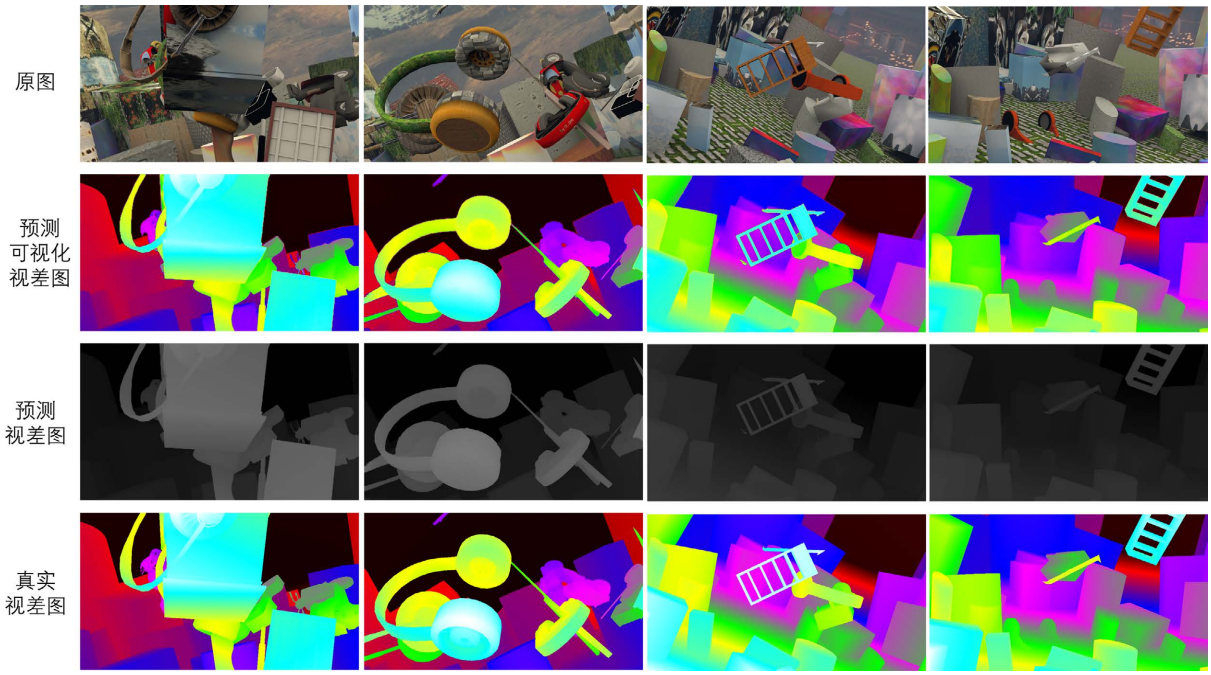

Figure 5. Results on SceneFlow test
图 5. SceneFlow 的测试结果

结果和真实数据之间的对比，可以比较直观地看出在物体边缘部分拥有比较好的提取，这是因为引入了视差梯度，能够更好地对同平面的视差做出正确的估计，从而来更好地区分出物体的边缘变化，同时在特征提取中引入了 ASPP 增大感受野，从而能够更多地考虑到细节部分，因此在对一些比较细微的部分能够保证良好的结果。

在两个数据集上，本文对比了一些算法的立体匹配效果。如表 1 所示，在 SceneFlow 数据集上经过深度可分离卷积的引入和局部代价体的轻度复杂度，本文的算法在运行时间有一定优势，并通过 ASPP 和视差梯度的引入从而不至于损失过多的精度，所以结果相比于 StereoNet 拥有比较好的精度，而对于 LEAStereo 的搜索网络和 EdgeStereo 的边缘网络拥有更好的运行速度。在 KITTI 数据上，如表 2 所示，综合性评估结果与 SceneFlow 大体一致，对于 LEAStereo 在精度上具有更佳的表现，因为 LEAStereo 是神经搜索网络，其主要在寻找最合适的神经网络，所以在数据迁移和泛化性上具备一定的优势。对于 EdgeStereo 其主要根据边缘信息和感知来增强算法对边缘的优化，所以在转移到复杂的域中就缺失一定的泛化性，所以在 KITTI 的数据集上表现欠佳。综合来看，本文中的算法综合运行时间和精度的各优势，在运行时间和精度上拥有比较好的平衡，也取得比较良好的性能。

Table 1. Performance evaluation of different methods on the SceneFlow
表 1. 不同算法在 SceneFlow 测试集的性能评价

Method	EPE	Runtime
StereoNet [26]	1.1	0.015 s
GC-Net [5]	2.51	0.9 s
LEAStereo [27]	0.78	0.3 s
EdgeStereo [28]	0.74	0.32 s
本文	0.77	0.215 s

Table 2. Performance evaluation of different methods on the KITTI2015
表 2. 不同算法在 KITTI2015 测试集的性能评价

Method	D1-bg	D1-fg	D1-all	Runtime
StereoNet [26]	4.3	7.45	4.83	0.015 s
GC-Net [5]	2.21	6.16	2.87	0.9 s
LEAStereo [27]	1.4	2.91	1.65	0.3 s
EdgeStereo [28]	1.88	4.07	2.25	0.6 s
本文	1.8	3.67	2.4	0.2 s

5. 结论

本文提出了一个基于局部代价层级优化的网络，以轻量的代价体不断优化视差的网络。其中为了保证网络的性能，同时引入了 ASPP 模块和视差梯度，来提高特征提取和视差优化的能力。另外为了保证特征提取中的轻量，同时在 ASPP 模块中引入了深度可分离卷积降低模块的参数量。此网络在精度和效率上有一个比较好的体现，并通过实验表明在 SceneFlow 和 KITTI 上数据集上拥有优异的结果。

参考文献

[1] Hamzah, R.A. and Ibrahim, H. (2016) Literature Survey on Stereo Vision Disparity Map Algorithms. *Journal of Sensors*, **2016**, Article ID: 8742920. https://doi.org/10.1155/2016/8742920

[2] Scharstein, D. and Szeliski, R. (2002) A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, **47**, 7-42. https://doi.org/10.1023/A:1014573219977

[3] Chang, J.-R. and Chen, Y.-S. (2018) Pyramid Stereo Matching Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 5410-5418. https://doi.org/10.1109/CVPR.2018.00567

[4] Zhang, F., Prisacariu, V., Yang, R., and Torr, P.H.S. (2019) Ga-Net: Guided Aggregation Net for End-to-End Stereo Matching. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 185-194. https://doi.org/10.1109/CVPR.2019.00027

[5] Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., *et al*. (2017) End-to-End Learning of Geometry and Context for Deep Stereo Regression. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 66-75. https://doi.org/10.1109/ICCV.2017.17

[6] Duggal, S., Wang, S., Ma, W.-C., Hu, R. and Urtasun, R. (2019) Deeppruner: Learning Efficient Stereo Matching via Differentiable Patchmatch. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October 2019-2 November 2019, 4383-4392. https://doi.org/10.1109/ICCV.2019.00448

[7] Yang, G., Manela, J., Happold, M. and Ramanan, D. (2019) Hierarchical Deep Stereo Matching on High-Resolution Images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 5510-5519. https://doi.org/10.1109/CVPR.2019.00566

[8] Xiao, J., Ma, D. and Yamane, S. (2021) Optimizing 3D Convolution Kernels on Stereo Matching for Resource Efficient Computations. *Sensors*, **21**, Article No. 6808. https://doi.org/10.3390/s21206808

[9] Shamsafar, F., Woerz, S., Rahim, R. and Zell, A. (2022) MobileStereoNet: Towards Lightweight Deep Networks for Stereo Matching. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, 3-8 January 2022, 677-686. https://doi.org/10.1109/WACV51458.2022.00075

[10] Badki, A., Troccoli, A., Kim, K., Kautz, J., Sen, P. and Gallo, O. (2020) Bi3d: Stereo Depth Estimation via Binary Classifications. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13-19 June 2020, Seattle, 1597-1605. https://doi.org/10.1109/CVPR42600.2020.00167

[11] Rao, Z., He, M., Dai, Y., Zhu, Z., Li, B. and He, R. (2020) NLCA-Net: A Non-Local Context Attention Network for Stereo Matching. *Apsipa Transactions on Signal and Information Processing*, **9**, Article No. e18. https://doi.org/10.1017/ATSIP.2020.16

[12] Guo, X., Yang, K., Yang, W. and Li, H. (2019) Group-Wise Correlation Stereo Network. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 3268-3277.

https://doi.org/10.1109/CVPR.2019.00339

[13] Wu, Z., Wu, X., Zhang, X., Wang, S. and Ju, L. (2019) Semantic Stereo Matching with Pyramid Cost Volumes. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 7483-7492. https://doi.org/10.1109/ICCV.2019.00758

[14] Shen, Z., Dai, Y. and Rao, Z. (2021) Cfnet: Cascade and Fused Cost Volume for Robust Stereo Matching. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 13901-13910. https://doi.org/10.1109/CVPR46437.2021.01369

[15] Lu, C., Uchiyama, H., Thomas, D., Shimada, A. and Taniguchi, R.-I. (2018) Sparse Cost Volume for Efficient Stereo Matching. *Remote Sensing*, **10**, Article No. 1844. https://doi.org/10.3390/rs10111844

[16] Sifre, L. and Mallat, S. (2014) Rigid-Motion Scattering for Image Classification. Ph.D. Thesis, Ecole Polytechnique, Palaiseau.

[17] Vanhoucke, V. (2014) Learning Visual Representations at Scale. *ICLR Invited Talk*, 1.

[18] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., *et al*. (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. Arxiv Preprint arxiv:170404861.

[19] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.-C. (2018) Mobilenetv2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4510-4520. https://doi.org/10.1109/CVPR.2018.00474

[20] He, K., Zhang, X., Ren, S. and Sun, J. (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**, 1904-1916. https://doi.org/10.1109/TPAMI.2015.2389824

[21] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. and Adamm H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, 8-14 September 2018, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49

[22] Tankovich, V., Hane, C., Zhang, Y., Kowdle, A., Fanello, S. and Bouaziz, S. (2021) Hitnet: Hierarchical Iterative Tile Refinement Network for Real-Time Stereo Matching. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 14357-14367. https://doi.org/10.1109/CVPR46437.2021.01413

[23] Menze, M. and Geiger, A. (2015) Object Scene Flow for Autonomous Vehicles. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3061-3070. https://doi.org/10.1109/CVPR.2015.7298925

[24] Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A., *et al*. (2016) A large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 4040-4048. https://doi.org/10.1109/CVPR.2016.438

[25] Rao, Z., Dai, Y., Shen, Z. and He, R. (2022) Rethinking Training Strategy in Stereo Matching. *IEEE Transactions on Neural Networks and Learning Systems*, 1-14. https://doi.org/10.1109/TNNLS.2022.3146306

[26] Khamis, S., Fanello, S., Rhemann, C., Kowdle, A., Valentin, J. and Izadi, S. (2018) Stereonet: Guided Hierarchical Refinement for Real-Time Edge-Aware Depth Prediction. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, 8-14 September 2018, 596-613. https://doi.org/10.1007/978-3-030-01267-0_35

[27] Cheng, X., Zhong, Y., Harandi, M., *et al*. (2020) Hierarchical Neural Architecture Search for Deep Stereo Matching. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Vancouver, 6-12 December 2020, Article No. 1858.

[28] Song, X., Zhao, X., Hu, H. and Fang, L. (2018) Edgestereo: A Context Integrated Residual Pyramid Network for Stereo Matching. *Proceedings of the Asian Conference on Computer Vision*, Perth, 2-6 December 2018, 20-35. https://doi.org/10.1007/978-3-030-20873-8_2