DOI:10.19651/j.cnki.emt.2211215

深浅层特征结合的自监督立体匹配*

葛 兰 贾振堂
(上海电力大学电子与信息工程学院 上海 201306)

摘 要: 针对现有的立体匹配算法物体细节部分估计效果较差、有监督算法依赖大量真实视差图等问题,本文提出了一种深浅层特征结合的自监督立体匹配算法。该算法在特征提取网络中嵌入通道注意力机制来提取图片的浅层和更具表征能力的深层特征。基于深层特征构建代价体积预测初始视差图,并用浅层特征指导初始视差图进行优化。此外在损失函数部分在左右视差一致性损失的基础上本文提出左右特征一致性损失,加强浅层特征信息对视差的约束作用,提高算法的鲁棒性。本文在 KITTI 2015 数据集上训练评估,并应用到拍摄的实际场景中。实验结果表明,本文提出的方法与其他算法相比能获得更好的效果,特别是在视差突然变化的细节区域。

关键词: 立体匹配;自监督;通道注意力机制;浅层特征;深层特征;视差图
中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Self-supervised stereo matching combining deep features and shallow features

Ge Lan Jia Zhentang
(College of Electronical and Information Engineering, Shanghai University of Electric Power, Shanghai 201306, China)

Abstract: Aiming at the poor estimation of the existing stereo matching algorithms on the details of objects and the fact that supervised algorithms relying on a large number of groundtruth disparity maps, this paper proposes a self-supervised stereo matching algorithm combining deep and shallow features. The algorithm embeds Efficient Channel Attention in the feature extraction network to extract shallow and more expressive deep features of the picture. The cost volume predicting initial disparities are constructed based on the deep features, and the shallow features are used to guide the optimization of the initial disparities. In addition, in the loss function section, on the basis of the left and right disparity consistency loss, this paper proposes the left and right feature consistency loss, which strengthens the constraint effect of shallow feature information on disparity maps and improves the robustness of the algorithm. This article trains and evaluates on the KITTI 2015 dataset and applies it to the actual scenes taken by us. Experimental results show that the proposed method can achieve better results than other algorithms, especially in the details where the disparity changes suddenly.

Keywords: stereo matching; self-supervised; efficient channel attention mechanism; shallow features; deep features; disparity map

0 引 言

随着双目相机在手机、自动驾驶汽车和机器人中的普及,双目视觉越来越受学术界和工业界的关注。立体匹配算法通过计算双目相机拍摄的被测物体左右图像中的位置偏差,结合相机的内外参数获得被测物体的距离(深度)信息。立体匹配在获取物体深度信息方面具有成本低、灵活且易于实现等优点,它可以帮助感知环境,在现实增强、机器人导航等方面有广泛的应用。

近年来,卷积神经网络表现出强大的特征表示能力,使用深度学习方法从输入的左右图得到视差图的方法受到研究人员的广泛关注[1-3]。传统的立体匹配算法分为 4 个步骤:代价体积计算、代价聚合、视差计算和视差优化。深度学习立体匹配方法借鉴传统的 4 个步骤理论,将其统一成一个端到端的网络,进一步提升了视差估计的性能。相对于传统算法,深度学习方法具有效果好、速度快、鲁棒性强

收稿日期:2022-08-29
* 基金项目:国家自然科学基金(62105196)项目资助

和泛化能力强等优势，逐渐成为了立体匹配算法中的主流。现有的基于深度学习的立体匹配大致可以分为监督学习和自监督学习两大类。

对于监督学习的立体匹配，端到端训练的网络需要大量带有真实标签的数据进行训练，Mayer 等[4]建立了大规模的合成数据集 SceneFlow，并在此基础上提出了首个端到端的立体匹配网络 DispNetC。Kendall 等[5]提出一种新的立体匹配网络 GC-Net，取消了对于特征图的互相关操作，而是将不同视差下的特征图进行级联构成代价体积。在此基础上，Chang 等[6]提出了基于语义信息和全局信息的 PSMNet，特征提取部分的空间金字塔池化模块通过聚合不同尺度的上下文信息来构建匹配代价卷。这些有监督的方法在损失函数部分通常采用 L1 损失，通过计算预测的视差值和真实视差值间的绝对差值反馈监督训练网络，使网络训练的模型效果越来越好。虽然这些监督的方法能够预测出高质量的视差图，但是需要提供精准的真实视差图供网络学习中使用。然而针对不同的场景，想要获取精准的真实视差图代价是非常大的，这往往依赖于昂贵的专业设备和专业人员，耗费大量资源。

自监督训练缓解了有监督训练过度依赖带有标签（真实视差图）的数据集的问题，并提高了网络的泛化能力。自监督学习算法主要从图像自身的特征结构、视差图自身的特点或者借助传统算法来构造噪声标签来推动网络的学习。对于自监督学习的立体匹配，Zhou 等[7]采用了左右一致性检查来生成一个置信图，以指导网络的训练。Li 等[8]提出了一种遮挡感知立体匹配网络，引入遮挡推理模块来进行立体匹配。但是这些有/无监督立体匹配方法通常构建 4 维代价体积（即高度×宽度×视差×通道）来表示左右图的相似性。构建的四维代价体积在后续代价聚合计算中需要使用三维卷积、反卷积操作，这将占用计算机大量内存，计算成本太高，并限制了输入图片的可能分辨率。针对这个问题，Wang 等[9-10]提出了一种视差注意力机制构建三维代价体积、降低计算复杂度的方法，但是在他们的方法中对输入图片的深浅层特征没有充分利用，计算量也可以进一步降低。此外，遮挡、扭曲、反光表面和弱纹理区域一直是立体匹配的难点问题，如何提高这些病态区域的匹配准确性也是需要深入研究的。

在上述研究工作的基础上，本文提出了一种深浅层特征结合的自监督立体匹配算法，包含的主要贡献为：1）本文提出了融合通道注意力机制的特征提取模块提取了图片的浅层和深层特征。通道注意力机制可以在不改变图像感受野的前提下，对于作用更大的图片特征区域赋予更大的权重，使提取到的深层特征更具表征能力，在构建的代价体积更高效的表示左右图像素点之间的相似关系；2）本文认为输入图片的浅层特征包含的细节信息在视差优化方面有重要作用，提出了一种新颖的基于浅层特征构建的损失函数称为左右特征一致性损失。左右特征图一致性损失有助于提高网络模型的表现力，在预测的视差图中获得更清晰的物体轮廓，缓解左右图之间的遮挡带来的边缘拖影问题，提高视差预测的准确性。

1　网络结构

立体匹配的过程通常是输入双目相机拍摄的一对左右图像，通过这对图像寻找图像中的同名点，输出稠密视差图。本文的整体网络结构如图 1 所示，主要分为 3 个部分：特征提取、构建代价体积、视差回归及优化。

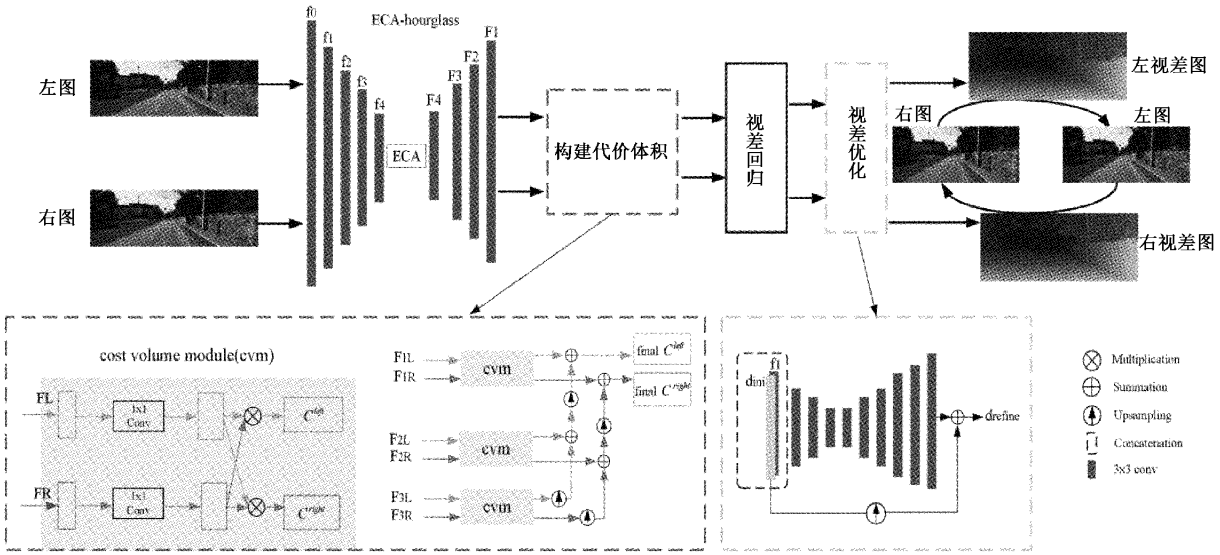


图 1　整体网络结构

1.1　ECA-hourglass 特征提取模块

1) 轻量注意力机制(ECA)

近年来,将通道注意力模块引入卷积网络引起了人们广泛关注,在性能提升方面显示出巨大潜力。ECA-Net[11]对 SE-Net[12]模块进行了一些改进,提出了一种无需降维的局部跨通道互动策略和自适应选择一维卷积核大小的方法,实现了性能上的提优。这个模块只增加了少量的参数,却实现了显著的性能提升。后根据实验证明,本文选用 ECA 模块是高效可行的。ECA 模块的结构如图 2 所示。

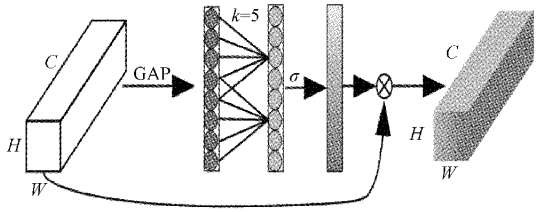


图 2　ECA 模块结构

在经过全局平均池化(图 2 中 GAP)之后,ECA 通过考虑每个通道和它的 k 个邻居来收集通道之间相互作用的局部信息。ECA 可以通过卷积核大小为 k 的快速一维卷积实现,其中卷积核大小 k 表示局部跨信道相互作用的覆盖率,为了避免通过交叉验证手动调整 k,使用与通道维度 C 正相关的 k,k 与 C 的具体关系如式(1)所示,其中 $|t|_{odd}$ 表示最接近 t 的奇数,在实验中 b 和 γ 分别设置为 2 和 1,C 为 160,经计算 $k=5$。

$$k = \left| \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{1}$$

2) ECA-hourglass

本文中使用单个 hourglass[13]作为特征提取的主干网络,第 1 部分是编码器,图像通过编码器提取特征,图 1 中的提取到的浅层特征 f0、f1 包含更多像素的位置信息,深层特征 f2、f3、f4 包含更多语义信息。将 f4 先送入嵌入的通道注意力模块,接着将这部分结果送入第 2 部分解码器,得到图片的深层特征记为 F4。连续的卷积和下采样操作会损失图片的很多像素信息,使用跳跃连接,具体如式(2)所示,降低像素信息的损失,使提取到的深层特征图 F3、F2、F1 包含更多细节信息,其中 upsample()为上采样操作。

$$F_i = f_i + \text{upsample}(F_{i+1}),\ i = 1,2,3 \tag{2}$$

1.2　构建代价体积

由极线约束可知如果已知某点在左图的映射点为 p_1,那么该点在右图的映射点 p_2 一定和 p_1 在同一条极线上,这样可以减少待匹配的点数量。对于左图中的每个像素,代价体积构建方法关注该点在右图极线方向的所有待匹配点。给定分别从左图和右图提取出的特征图 F_L 和 F_R,将特征图经过 1×1 的卷积层后调整维度,使用特征向量之间的点积作为视觉相似度的度量,得到基于左图的代价体积 C^{left}。交换左右特征图的位置,同样的计算方式得到基于右图的代价体积 C^{right}。我们通过特征提取网络提取了左右图 3 个不同尺度的特征 F_1、F_2、F_3,尺度分别为原图的 1/4,1/8 和 1/16。首先将 1/16 尺寸的左右图特征送入网络得到 1/16 尺寸的代价体积,接着将 1/8 尺寸的特征图送入代价体积构建模块得到 1/8 尺寸的代价体积,同时对 1/16 尺寸的结果上采样,与现在得到的 1/8 尺寸的结果相加得到新的 1/8 尺寸的代价体积。再将 1/4 尺寸的特征图送入代价体积构建模块得到 1/4 尺寸的代价体积,与上采样后的 1/8 尺寸的结果相加得到最终的 1/4 尺寸的代价体积。我们最终得到的 1/4 尺寸的代价体积是 3 个不同尺度的代价体积的融合,包含更多细节信息,可以更好的计算左右图的相似性。

1.3　视差回归及视差优化

1) 视差回归:

在视差回归模块中,基于左图的代价体积 C^{left} 和基于右图的代价体积 C^{right} 经过 softmax 变换得到 M^{left} 和 M^{right},将点与点之间的相关性转换为在[0, 1]之间和为 1 的概率分布。在 softmax 变换中为了防止数值溢出,将每一个需要变换的代价体积减去其中最大的值,如式(3)所示。估计的初始视差是由 M^{left} 和 M^{right} 加权的所有视差候选值的总和,如式(4)所示,其中 W 为输入图像的宽度,i 为图像中的一个像素点。将公式中的 M^{left} 替为换 M^{right} 则可以得到右图的初始估计视差。

$$M^{left} = \text{softmax}(C^{left}) = \frac{e^{C^{left}-X}}{\sum e^{C^{left}-X}},\ (X = \max(C^{left})) \tag{3}$$

$$d_i^{left} = \sum_{k=0}^{W/4-1} k \times M_{(i,k)}^{left} \tag{4}$$

2) 视差优化:

由于图片的浅层特征包含更多边缘等细节信息,利用浅层特征 f0 指导视差优化,将其与初始估计视差图按通道拼接,送入 hourglass 网络输出残差优化项,将其与初始视差图相加作为优化的结果。

1.4　损失函数

本文设计的损失函数包含 4 个部分,其数学表达式如式(5)所示。

$$Loss = 2\times(L_{pleft}+L_{pright})+(L_{sleft}+L_{sright})+(L_{mleft}+L_{mright})+(L_{cleft}+L_{cright}) \tag{5}$$

式中:C_c^b 和 L_{pright} 分别表示左右图的光度损失项,L_{sleft} 和 L_{sright} 分别表示左右图的平滑损失项,L_{mleft} 和 L_{mright} 分别表示左右图特征一致项,L_{cleft} 和 L_{cright} 分别表示左右视差图的一致性损失。接下来,详细介绍每项损失左半部分的计算方式。右半部分的损失函数只需要交换每项损失中的 left 和 right 顺序,同时注意重建采样时进行反方向采样。

1) 光度损失

受文献[14-15]的启发,使用 L1 损失项和结构相似指

数(SSIM)损失项的组合作为光度损失。通过网络可以得到左视差图和右视差图,通过左视差图对右图像中的像素进行索引采样可以重建新的左图,如式(6)所示。光度损失比较原始左图像和重建得到的左图像之间的差异,监督网络学习预测准确的左视差图,具体公式如式(7)所示。

$$I_i^{\hat{left}} = I^{right}_{(i+d_i^{left})} \tag{6}$$

$$L_{pleft} = \frac{1}{N}\sum_i \alpha \frac{1-SSIM(I_i^{left}, I_i^{\hat{left}})}{2} + (1-\alpha)\| I_i^{left} - I_i^{\hat{left}} \|_1 \tag{7}$$

式中:i 表示图像的一个像素点,$\alpha = 0.85$,d^{left} 表示网络预测的左视差图,N 表示所有像素总个数,I^{left} 表示输入的左图,I^{right} 表示输入的右图,$I^{\hat{left}}$ 表示重建的左图,$\|\cdot\|_1$ 为 L1 损失。

2) 平滑损失

由于需要密集的视差图,鼓励视差局部平滑,损失项惩罚非边缘区域的视差变化。为了允许对象边缘处的视差不连续,以前的方法[15]根据图像梯度对视差进行惩罚,对应的损失为式(8),其 N 表示所有像素总个数,∂_x 和 ∂_y 分别表示计算方 R_{earth} 方向和 y 方向梯度,d^{left} 表示左视差图。

$$L_{smooth} = \frac{1}{N}\sum_i |\partial_x d_i^{left}| e^{\|\partial_x I_i^{left}\|} + |\partial_y d_i^{left}| e^{\|\partial_y I_i^{left}\|} \tag{8}$$

不同的是,本文在此损失的基础上,利用传统方法获得物体边缘(对应式(9)),更好地规范图片在无纹理的低图像梯度区域的视差结果,鼓励平面区域的视差平滑,允许物体边界的视差不连续,具体为式(10)所示。其中 sobel 函数是利用 sobel 算子运算得到的图像 x 方向和 y 方向的导数。

$$E^{left} = sobel(I^{left}) \tag{9}$$

$$L_{sobel} = \frac{1}{N}\sum_i |\partial_x d_i^{left}| e^{\|E_i^{left}\|} + |\partial_y d_i^{left}| e^{|E_i^{left}|} \tag{10}$$

总的平滑损失记为 L_s,如式(11),其中 $\beta = 0.1$。

$$L_{sleft} = \beta L_{smooth} + (1-\beta) L_{sobel} \tag{11}$$

3) 左右视差一致性损失

立体匹配中的重建约束是左视差图对右图中的像素进行采样,可以重建左图。同样的道理,利用左视差图对右视差图进行采样,则可以重建左视差图。将重建的左视差图与原来的左视差图计算 L1 损失,这项损失将左右视差图紧密联系起来,更好的处理左右图中的相互遮挡区域,提高视差计算的准确性。对应式(12):

$$L_{cleft} = \frac{1}{N}\sum_i \| d_i^{left} - d_i^{\hat{left}} \|_1 \tag{12}$$

其中,d^{left} 表示左视差图,$d^{\hat{left}}$ 是左视差图 d^{left} 对右视差图 d^{right} 进行采样重建出的左视差图。

4) 左右特征一致性损失

浅层网络感受野较小,感受野重叠区域也较小,因此提取到的特征包含更多纹理、边缘、棱角等细节信息。以前的无监督立体匹配在损失函数部分往往忽略了浅层特征的约束作用。本文提出的浅层特征一致性损失约束认为浅层特征同输入的左右图一样,利用左视差图对右图的浅层特征采样,可以实现对左图浅层特征的重建(参考式(6)),重建后的左图浅层特征与原来的计算 L1 损失,进一步提高视差结果的准确性。特征一致性的公式为式(13):

$$L_{mleft} = \frac{1}{N}\sum_i \| f_i^{left} - f_i^{\hat{left}} \|_1 \tag{13}$$

其中,f^{left} 表示左图的浅层特征,$f^{\hat{left}}$ 是左视差图 d^{left} 对右图的浅层特征 f^{right} 进行采样重建出的左图浅层特征。

2　实　　验

2.1　实验数据集

在 KITTI 2015[16]上训练和评估了本文的网络。KITTI 2015 数据集是一个真实道路场景下采集的街道场景数据集。使用 EIGEN[17]的拆分方式,其中 22 600 对左右图用于训练,888 对用于验证,697 对用于测试,图片尺寸为 1 242×375。

2.2　实验细节

本文实验部分使用 PyTorch 深度学习架构实现提出的网络模型,使用 Adam 优化器对网络进行训练,优化器的参数设置为 $\beta_1 = 0.9$,$\beta_2 = 0.99$。并使用两个 Nvidia RTX 2080ti GPU 对网络进行训练,网络在训练时对所有输入图像使用随机剪裁将图像大小剪裁为 960×256。使用 KITTI 2015 数据集训练,batch size 设置为 8,共训练 20 代。采用可变学习率,初始学习率为 0.001,每隔 5 代学习率降为原来的 0.1 倍。

2.3　实验结果及数据分析

为了证明提出的网络能够提高视差预测结果的准确性,与其他自监督方法进行定性和定量的比较。比较结果如表 1 和图 3 所示。表 1 是将提出的网络与使用 KITTI 2015 EIGEN 拆分数据集训练的其他现有方法进行定量比较的结果,其中红色标注的指标数值越小越好,蓝色标注的指标数值越大越好,最优结果加粗标记,次优结果用下划线标记。从表 1 可以看出,本文的网络展示了一个优越的整体性能,这表明本文的模型可以从几何约束中学习,预测出更为精准的视差图。为了证明本文预测结果的改善不仅仅是因为输入的是双目图像,我们对 Monodepth2[15]的输入进行修改,将原本单输入的网络改为双目输入。从第 3 行可以清楚的看出,虽然双目输入的测试结果优于单目输入的 Monodepth2,但仍然不如本文的预测结果好。

表 1　本文算法与其他现有网络定量比较结果

算法	AbsRel	SqRel	RMSE	$\delta<1.25$	$\delta<1.25^2$	$\delta<1.25^3$
Andrea[18]	0.152	1.388	6.016	0.789	0.918	0.965
Monodepth2[15]	0.109	0.873	4.960	0.864	0.948	0.975
Monodepth2(concat)	0.082	0.752	4.407	0.914	0.960	0.978
H-Net[19]	0.076	0.607	4.025	0.918	**0.966**	**0.982**
本文算法	**0.069**	**0.596**	**3.865**	**0.928**	0.962	0.979

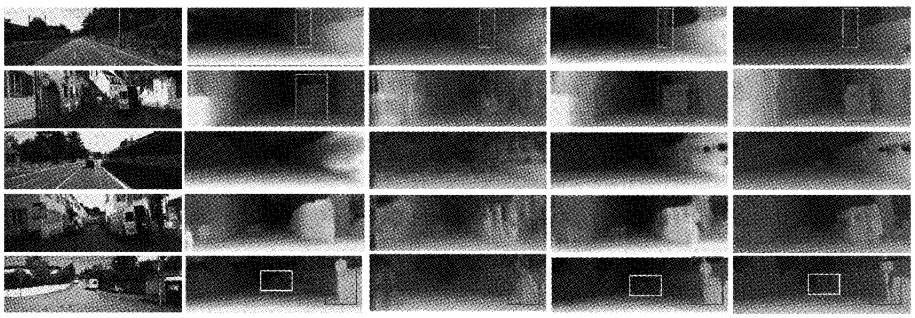


图 3　本文算法与其他现有网络定性比较结果

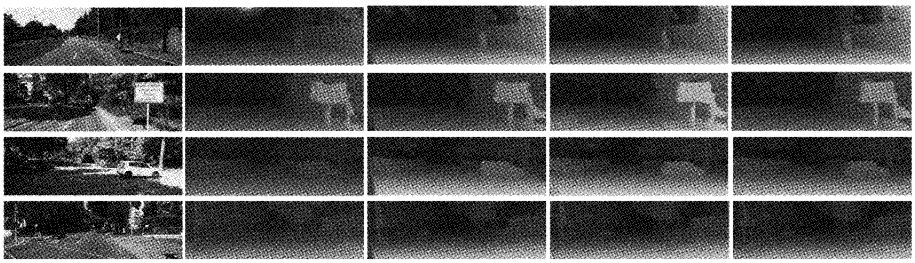
部分预测视差结果如图 3 所示,图中第一列为输入网络的左图,后面 3 列分别为修改为双目输入的 Monodepth2、文献[18]和文献[19]的预测视差图,最后一列为本文算法的视差估计结果。视差预测全部针对左输入图像,不难看出,本文算法的估计结果更光滑,对场景细节恢复更好。如第 1 行各个算法预测的视差图,对于图中方框标注的电线杆,对比可以发现本文的算法预测出的电线杆视差图轮廓是最清晰光滑的。第 2 行中车的形状,文献[18]和文献[19]的恢复效果均较比本文算法要模糊,双目的 Monodepth2 半开的车门从视差图中比较难看出,而本文的预测图可以明显看出原图中车门状态。第 3 行中,图片中间的车本文的预测结果也是最清晰的,地面区域整体的预测结果也比其他算法更平滑。第 4 行的车前有一个人,图 3(b)(c)列的结果很难看出人的存在,图 3(d)预测的人的轮廓没有本文的清晰,本文的视差结果物体边缘更加明显,界限更加清楚。第 5 行方框区域以及白色的墙面,都可以看出本文的结果准确度更高我们的算法对局部区域的细节学习更为充分,因而可以恢复出更多的细节结构。

2.4　消融实验结果分析

为了验证本文算法中各个模块在视差估计中的性能,本节在 KITTI 2015 数据集上进行消融实验,主要分析通道注意力机制、左右视差一致性损失、特征一致性损失 3 个模块作用。设置 3 个消融方式:1)编码器端是否嵌入通道注意力;2)损失中是否包含左右视差一致性损失和加强的平滑损失;3)损失是否包含特征一致性项。3 种消融方式对应图 4 的(b)～(d),可视化结果如图 4 所示。观察图 4(b)的去除通道注意力模块后的结果,与图 4(e)对比可以发现第 1 张图片天空这样的无纹理、无限远区域,在去除通道注意力模块后会的视差预测结果会出现了一个大的光圈,误差很大。第 2 张图片的标识牌、第 3 张图片的汽车、第 4 张图片的树,这些地方在去除通道注意力后的预测准确度都大打折扣,边缘处出现斑驳模糊的情况。观察对比图 4(c)与图 4(e)的结果,可以得出结论本文提出的左右一致性损失可以在左右图中遮挡区域填补适当的视差值。图 4(c)所有结果在最左边都出现一片未知视差值,这是由于左图中的最左边缘部分在右图中是不可见的,构建代价体积后时右图寻找不到与之最佳匹配的点,最终导致结果图出现这样现象。同时对比可以看出图 4(e)物体边缘更加平滑,说明我们的加强的平滑约束是有积极作用的。图 4(d)所在结果是去除特征一致性损失训练得到的,特征一致是为了丰富细节信息,第二张图的标识牌杆的部分尤为明显,图 4(d)所在位置杆不完整,缺失了一块,对比看图 4(e)在细节处有明显优势。

2.5　应　　用

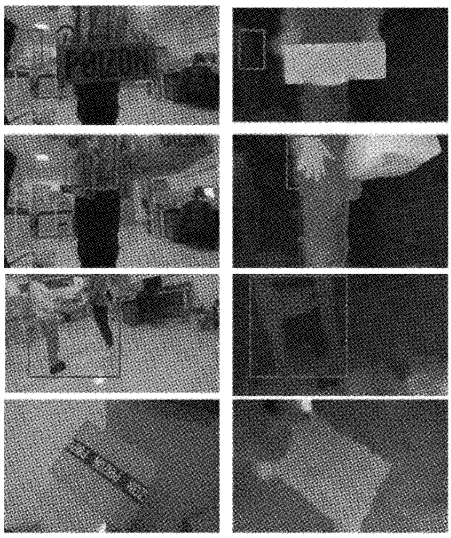
为了测试本文方法的实际应用情况,在实验室采集了 160 对左右图像,在经过上述 KITTI 2015 训练的模型上直接测试,具体结果图 5 所示。图 5 中第 1 行的视差结果图可以清楚看出人双手举盒子的姿势,人和盒子的边缘轮廓都很清晰,红色方框处还可以看出桌子变化的细节。第 2 行视差结果中红色框框出部分可以清晰的看出手部动作,只看视差图就可以明显看出五指张开的动作。第 3 行视差图也可以一眼看出是两人抬着一个盒子,两个人的身形轮廓都很清楚,视差值也合理,但是第 3 行的图片中有大



(a) 左图　(b) 去除ECA　(c) 去除L_c和L_s　(d) 去除L_m　(e) 本文算法

图 4　消融实验结果可视化

量无纹理的地板存在，并且在地板上有灯光的影子，这导致它地面部分的视差估计结果有大块斑驳，估计结果准确性较差。第 4 行图片左侧是无纹理的地板，右侧是无纹理的纸板，这大片的无纹理区域导致视差图整体的准确性没有前几张图片高，尤其是纸盒的边缘没有 1、2 行中的那么清晰，地面也不平滑，左上角的地面还有一块黑点。尽管拍摄数据集的相机参数和场景特征与 KITTI 2015 存在一定的差异，但是仍然可以在视觉上产生令人信服的视差估计结果，因此本文的网络模型具有一定的实用价值。



(a) 左图　(b) 视差图

图 5　在本文拍摄的实际场景测试结果

3　结　　论

本文提出了一种深浅层特征结合的自监督视差估计算法，在特征提取网络中融入通道注意力机制，通过学习的方法重点关注对视差估计性能贡献较大的通道，提取图片的浅层特征和深层特征。提取到的深层特征对图片有更强的表征能力，用来构建级联的代价体积进行图像的视差估计，浅层特征主要用于视差优化和左右特征一致性的损失约束。在 KITTI 2015 数据集上的测试结果与之前的方法相比，在视差突然变化的区域中，获得的物体边缘更加准确，并且包含更多细节信息。同时应用到我们拍摄的实际场景中，网络能够准确的得到场景中物体的边缘，进一步推动了双目视差估计的实用化进程。但是在像地面这样反光无纹理的区域，视差图会出现斑驳情况，误差比较大，所以下一步的研究目标为提高无纹理、镜面反光、透明等病态区域的视差预测准确度，以进一步提升算法性能。

参考文献

[1] 余雪飞，顾寄南，黄则栋，等. 基于边缘检测与注意力机制的立体匹配算法[J]. 电子测量技术，2022，45(11)：167-172.

[2] 王正家，陈长乐，徐研彦，等. 基于跨尺度 PatchMatch 的立体匹配算法[J]. 电子测量技术，2022，45(12)：114-119.

[3] 赵倩. 基于 3D 卷积模块和视差分割的立体匹配方法[J]. 电子测量技术，2021，44(18)：72-77.

[4] MAYER N, ILG E, HAUSSER P, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016, DOI:10.1109/CVPR.2016.438.

[5] KENDALL A, MARTIROSYAN H, DASGUPTA S, et al. End-to-end learning of geometry and context for deep stereo regression[C]. IEEE International Conferenceon Computer Vision, 2017, DOI:10.1109/ICCV.2017.17.

[6] CHANG J, CHEN Y. Pyramid stereo matching network[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, DOI:10.1109/CVPR.2018.00567.

[7] ZHOU C, ZHANG H, SHEN X Y, et al. Unsupervised learning of stereo matching[C]. IEEE International Conference on Computer Vision, 2017, DOI:10.1109/ICCV.2017.174.

[8] LI A, YUAN Z J. Occlusion aware stereo matchingvia cooperative unsupervised learning[C]. Asian Conference on Computer Vision, 2019, DOI:10.1007/978-3-030-20876-9_13.

[9] WANG L G, WANG Y Q, LIANG Z F, et al.

Learning parallax attention for stereo image super-resolution[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, DOI:10.1109/CVPR.2019.01253.

[10] WANG L G, GUO Y L, WANG Y Q, et al. Parallax attention for unsupervised stereo correspondencelearning[J]. IEEE Transactions on Pattern Analysis andMachine Intelligence, 2020, 44(4):2108-2125.

[11] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. IEEE Conference on Computer Visionand Pattern Recognition, 2020, DOI:10.1109/CVPR42600.2020.01155.

[12] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8):2011-2023.

[13] NEWELL A, YANG K Y, DENG J, et al. Stacked hourglass networks for human pose estimation[C]. European Conference on Computer Vision, 2016, DOI:10.48550/arXiv.1603.06937.

[14] CLEMENT G, OISIN M, GABRIEL J. Unsupervised monocular depth estimation with left-right consistency[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017, DOI:10.1109/CVPR.2017.699.

[15] CLEMENT G, OISIN M, MICHAEL F, et al. Digging into self-supervised monocular depth estimation[C]. IEEE International Conference on Computer Vision, 2019, DOI: 10.1109/ICCV.2019.00393.

[16] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]. IEEE Conference on Computer Vision andPatternRecognition, 2012, DOI: 10.1109/CVPR.2012.6248074.

[17] EIGEN D, FERGUS R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture [C]. IEEE InternationalConference on Computer Vision, 2015, DOI:10.1109/ICCV.2015.304.

[18] ANDREA P, DAN X, MIHAI P, et al. Unsupervised adversarial depth estimation using cycled generative networks[J]. International Conference on 3D Vision, 2018, DOI:10.1109/3DV.2018.00073.

[19] HUANG B, ZHENG J Q, GIANNAROU S, et al. H-Net: Unsupervised attention-based stereo depth estimation leveraging epipolar geometry[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2021, DOI:10.48550/arXiv.2104.11288.

作者简介

葛兰,硕士研究生,主要研究方向为图像处理。

贾振堂,副教授,硕士生导师,主要研究方向为智能视频监控(涉及多模态深度学习、目标识别、人体姿态分析、立体视觉等)。

E-mail:462458081@qq.com