DOI:10.19651/j.cnki.emt.2415322

基于细节信息增强的无监督双目立体匹配算法*

王晓峰[1] 孙志恒[2] 喻 骏[2] 孙贾梦[2] 丁坤岭[2] 舒 航[2]
(1.重庆科技大学数理与大数据学院 重庆 401331;2.重庆科技大学电气工程学院 重庆 401331)

摘 要:无监督立体匹配算法在自动驾驶等领域有重要的应用,然而无监督立体匹配算法在物体连续、边缘等细节信息区域的视差精度较低,本文提出了一种提高细节信息区域精度的无监督立体匹配算法。通过在特征金字塔网络中引入空间注意力机制和残差网络,设计了一种空间特征金字塔网络算法,抑制特征提取过程中边缘和小目标细节信息的丢失。构建了视差融合模块,将半全局立体匹配算法生成的原始视差和视差回归生成的初步视差进行置信度视差融合,提升连续细节信息区域的精度。对于网络损失函数,集成了原始视差监督损失和置信度遮挡损失,保留更多图像边缘和连续区域处的细节信息。实验结果表明,本文算法在 KITTI 2015 测试集中非遮挡区域和所有区域的误匹配率分别为 6.24%和 5.89%,与其他经典算法相比在细节信息区域的效果、精度方面有较大提升。

关键词:无监督;立体匹配;细节信息增强;注意力机制;视差融合
中图分类号:TP391.41 **文献标识码**:A **国家标准学科分类代码**:510.4050

Unsupervised stereo matching algorithm of binocular based on detail information enhancement

Wang Xiaofeng[1] Sun Zhiheng[2] Yu Jun[2] Sun Jiameng[2] Ding Kunling[2] Shu Hang[2]
(1. College of Mathematics, Science and Big Data, Chongqing University of Science and Technology, Chongqing 401331, China;
2. College of Electrical Engineering, Chongqing University of Science and Technology, Chongqing 401331, China)

Abstract: Unsupervised stereo matching algorithms have important applications in areas such as autonomous driving, however, unsupervised stereo matching algorithms have low disparity accuracy in the region of object continuity, edges and other detail information, this paper proposes a new unsupervised stereo matching algorithm to improve the accuracy of detail information region by combining spatial attention mechanism and parallax fusion. Specifically, the spatial feature pyramid network is designed by introducing a spatial attention mechanism and residual structure into the feature pyramid network, to suppress the loss of edge and small target details in the process of feature extraction. Further, a disparity fusion module is constructed to improve the accuracy of the continuous detail information region, where the original disparity generated by the semi-global block matching algorithm and the initial disparity generated by disparity regression are fused with confidence disparity. Moreover, For the network loss function, the original disparity supervised loss and confidence masking loss are integrated to retain more detailed information at image edges and continuous regions. Finally, the experimental results show that the mis-matching rate of the proposed algorithm in the non-occluded region and all regions in the KITTI 2015 test set are 6.24% and 5.89%, respectively, which greatly improves the effect and accuracy of the detailed information region compared with other classical algorithms.

Keywords: unsupervised; stereo matching; detail information enhancement; attention mechanism; disparity fusion

0 引 言

随着计算机视觉的不断发展,双目深度估计广泛应用于机器人导航、自动驾驶等领域。双目深度估计的核心任务是立体匹配[1],其主要是寻找校正后的图像之间的对应视差关系,进而通过三角形相似原理获得场景的深度信息。

近年来,深度学习在各项分类、检测等任务的性能方面取得了显著性地提升,在立体匹配算法中也取得一定进展。与传统算法[2-3]相比,深度学习凭借强大的学习能力,对模型性能的提升具有显著优势。早期的方法通过 CNN 计算

收稿日期:2024-01-11
*基金项目:重庆市自然科学基金面上项目(CSTB2022NSCQ-MSX0398,CSTB2022NSCQ-MSX1425)资助

立体图像匹配点的相识度之后,通过视差后处理获得视差图。由于这些方法无法进行端到端的训练,研究人员开始研究端到端的有监督立体匹配算法,例如金字塔立体匹配网络[4](pyramid stereo matching network, PSMNet)、引导聚合网络[5](guided aggregation network, GANet)等。虽然上述有监督立体匹配算法在视差估计中具有较优的性能,但需要真实视差作为标签训练网络,真实视差的采集难度大且耗时长,并且在不同数据集上的性能差异较大,增加了监督训练时微调的难度。

研究人员开始探索无监督立体匹配算法,无监督立体匹配算法[6-9]通过输入图像和输出视差自身的信息来训练立体匹配网络,缓解了有监督训练需要标签的问题,增强了模型的泛化性能。Zhou 等[10]首次提出了一种端到端的无监督立体匹配算法,主要通过一种基于左右检查的迭代过程来指导无监督训练,但是这种方法产生的视差精度较低。在此基础之上,Yin 等[11]提出的 GeoNet 模型通过对重构图像和源图像计算损失,并引入了平滑性损失提高预测视差的准确性,虽然提高了模型的精度,但是网络中含有较多的残差结构,计算量复杂。Wang 等[12]借鉴了 GeoNet 模型的重构损失,提出的视差注意力立体匹配网络(parallax attention stereo matching network, PASMNet)是目前流行的无监督立体匹配算法,其通过级联视差注意力模块解决了以往固定视差的缺陷和计算复杂的问题,但 PASMNet 在图像中的边缘、小目标、连续等细节区域会产生不精确的视差。综上,无监督立体匹配算法在小目标、边缘、连续等细节信息区域仍会产生误匹配问题,导致细节信息区域的视差精度较低,对真实场景深度信息的恢复具有重要影响。

针对上述问题,本文在 PASMNet[10]的基础上提出了一种基于细节信息增强的无监督立体匹配算法。首先,设计一种用于立体匹配特征提取过程的空间特征金字塔网络算法(spatial feature pyramid network, SFPN),以此减少特征提取过程中边缘和小目标细节信息的丢失;其次,通过融合半全局立体匹配算法[13](semi-global block matching, SGBM)产生的原始视差提升连续细节信息区域的精度;然后,为使模型在边缘和连续区域输出更多的细节信息,将原始视差监督损失函数、置信度遮挡损失函数引入基准模型的损失函数之中,有效解决了细节信息区域视差精度低的问题;最后,通过消融实验来验证本文设计的 SFPN 和视差融合的有效性,并在 KITTI 和 Scene Flow 测试集中将本算法与目前经典的无监督立体匹配算法进行对比,验证了本算法在细节信息区域产生视差的精确性。本论文提升了目前无监督立体匹配算法在细节信息区域的效果。

1 网络模型设计

本文无监督立体匹配模型的输入为双目左右图像,输出为精确的视差图。网络模型结构如图 1 所示,主要由特征提取模块、视差回归模块、视差融合模块组成。

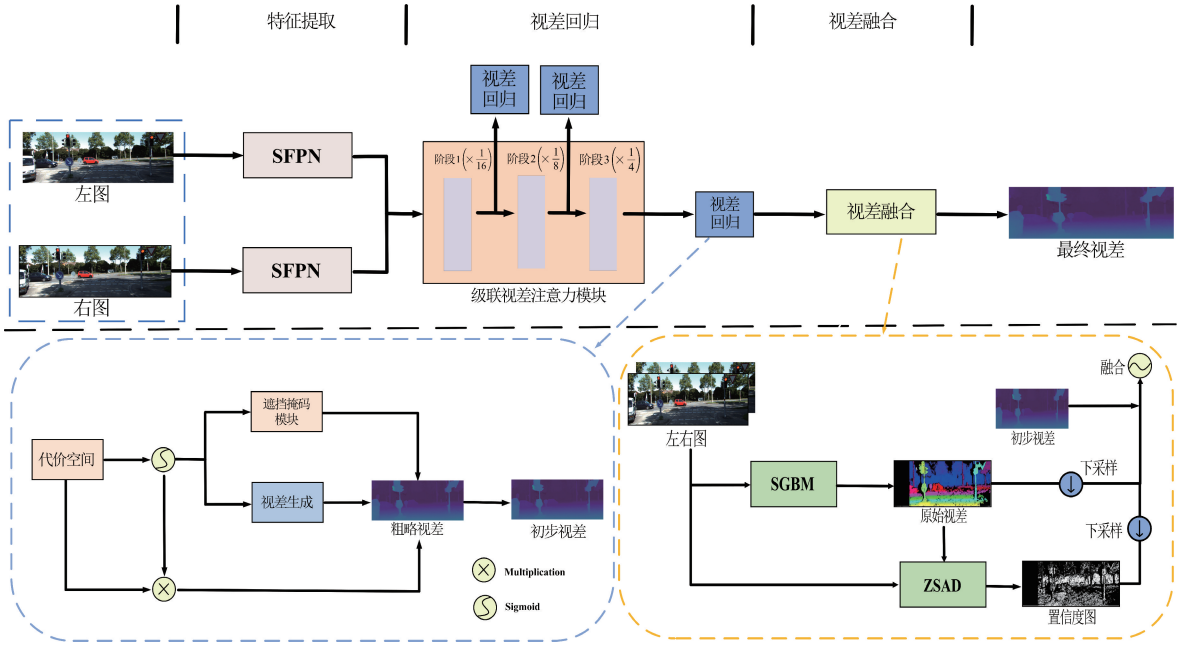


图 1 网络模型结构图

1.1 特征提取模块

特征提取是无监督立体匹配算法的关键步骤。基准模型 PASMNet 的特征提取模块在边缘和小目标细节信息区域容易产生细节信息丢失的问题。特征金字塔网络(feature pyramid network, FPN)是一种多尺度的特征提取模块,通过简单的网络连接,在基本不增加原有模型计算量的情况下,提升了小目标处细节区域的特征提取能力。

为了强化网络在边缘和小目标细节信息区域的特征提取能力，减少区域细节信息的丢失。首先，将 ResNet 中的普通卷积使用非对称卷积（asymmetric convolutional network，ACNet）进行替换，设计一种非对称卷积残差网络（asymmetric convolutional resNet，ACResNet）；其次，将 ACResNet 和空间注意力机制（spatial attention module，SAM）融入 FPN 自下而上过程，构建一种空间特征金字塔网络算法（spatial feature pyramid network，SFPN），如图 2 所示；最后，将 SFPN 算法作为立体匹配特征提取模块。SFPN 中的 SAM 将空间位置信息设置不同的权重，从而增强图像边缘和小目标处细节信息的提取；ACResNet 通过横向和纵向的卷积方式，减少模型传递过程边缘细节信息的丢失，使模型输出更多的细节信息，以此提高模型的精度。

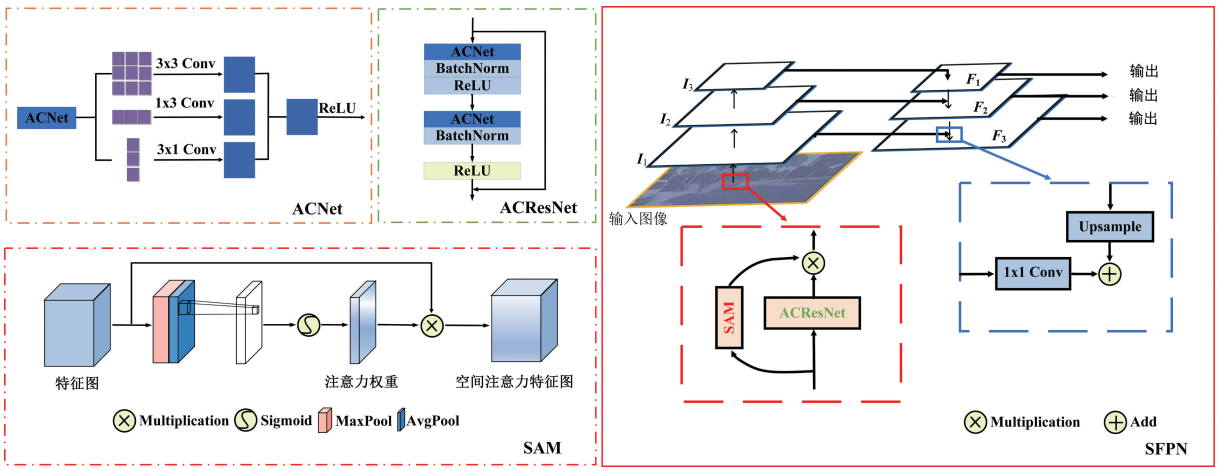


图 2　特征提取模块结构图

首先，将输入图像通过 3 次自下而上过程得到三种不同分辨率的特征图 $\{\boldsymbol{I}_1,\boldsymbol{I}_2,\boldsymbol{I}_3\}$；然后，将最低分辨率的特征图 $\boldsymbol{I}_3$ 通过 1×1 的卷积得到特征图 $\boldsymbol{F}_3$；最后，将 $\boldsymbol{F}_3$ 依次通过自上而下和横向连接得到不同分辨率的特征图 $\{\boldsymbol{F}_2,\boldsymbol{F}_1\}$。

1.2　视差回归模块

现有的无监督立体匹配算法，如 OASM-Net[6]、DispSegNet[7]，都是通过 4D 代价空间进行立体匹配任务，然而固定的最大视差限制了它们处理视差变化大的立体图像，选取不同的最大视差会使图像中连续细节信息区域产生不同的视差值。此外，4D 代价空间有较高的计算量，不满足立体匹配实时性的需求。立体图像对中左图像的像素与右图像中对应的像素在同一极线上，在这种极线约束下，级联视差注意力模块[12]通过重塑操作和矩阵乘法来计算左图像中的像素与右图像中沿极线的所有位置之间的相关性，有效地将沿极线的任意两个像素之间的特征相关性编码到视差注意图中，可以处理较大的视差变化。为了解决以往无监督立体匹配方法中需要设置最大视差的限制。故本文引用文献[12]中的级联视差注意力模块进行代价聚合来学习大视差变化下的立体匹配任务。

级联视差注意模块由 3 个阶段组成，每个阶段的过程相同。将特征提取模块得到的左右特征图 $\boldsymbol{F}_l,\boldsymbol{F}_r$ 输入到级联视差注意力模块，获取 3 个阶段的代价空间 $\boldsymbol{C}^{(1)},\boldsymbol{C}^{(2)},\boldsymbol{C}^{(3)}$，接着将 3 个阶段的代价空间输入到视差回归模块进行视差回归。由于 3 个阶段的视差回归模块相同，阶段 1、2 经过视差回归模块得到的视差只在训练中输出，因此本文只展示了阶段 3 的视差回归模块，如图 3 所示。

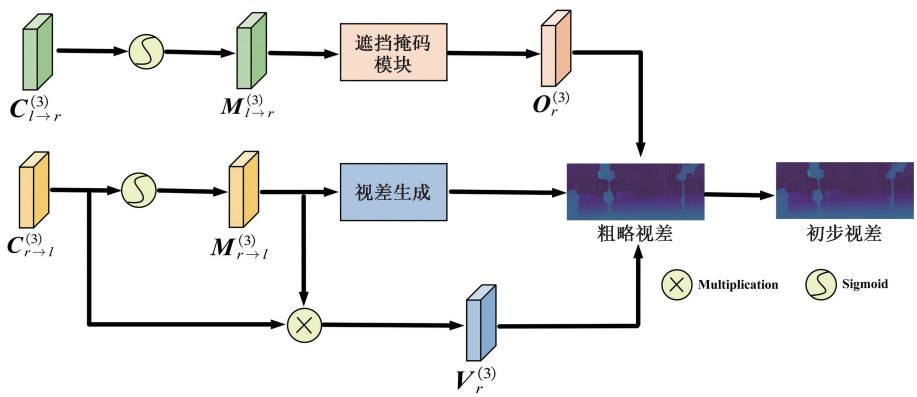


图 3　阶段 3 的视差回归模块结构图

首先,将3阶段 $\boldsymbol{C}^{(3)}$ 中右到左的代价空间 $\boldsymbol{C}_{r\to l}^{(3)}$ 和左到右的代价空间 $\boldsymbol{C}_{l\to r}^{(3)}$ 分别经过 Sigmoid 激活函数得到右到左、左到右的视差注意力图 $\boldsymbol{M}_{r\to l}^{(3)}$、$\boldsymbol{M}_{l\to r}^{(3)}$;然后,对 $\boldsymbol{M}_{l\to r}^{(3)}$ 进行遮挡掩码生成得到右图的遮挡特征图 $\boldsymbol{O}_r^{(3)}$,如式(1)[12]所示。

$$\boldsymbol{O}_r^{(3)}(i,k)=\begin{cases}1, & \sum_{j\in[1,W]}\boldsymbol{M}_{l\to r}^{(3)}(i,j,k)>0.8\\ 0, & \text{其他}\end{cases} \tag{1}$$

然后,对 $\boldsymbol{M}_{l\to r}^{(3)}$ 进行视差生成得到粗略视差 $\boldsymbol{D}'$,如式(2)[12]所示。

$$\boldsymbol{D}'(i,j)=\sum_{k=0}^{w/4-1}(j-k)\boldsymbol{M}_{r\to l}^{(3)}(:,:,k) \tag{2}$$

式中:$\boldsymbol{M}_{l\to r}^{(3)}$ 表示阶段3中右到左的视差注意力图,w 表示特征维度,i,j 表示粗略视差中的像素,k 表示左右特征图中对应像素的相关性。

最后,使用 $\boldsymbol{O}_r^{(3)}$ 过滤粗略视差中的遮挡视差,将 $\boldsymbol{C}_{r\to l}^{(3)}$ 与 $\boldsymbol{M}_{r\to l}^{(3)}$ 进行矩阵相乘得到有效特征 $\boldsymbol{V}_r^{(3)}$,通过 $\boldsymbol{V}_r^{(3)}$ 对 $\boldsymbol{D}'$ 中过滤掉的遮挡视差进行填充,进而生成初步视差 $\hat{\boldsymbol{D}}$。

1.3 视差融合模块

基准模型 PASMNet 在连续细节信息区域容易产生不准确的视差,融合其他立体匹配算法在连续细节信息区域生成的精确视差可以缓解此问题,从而提高立体匹配算法在连续细节信息区域的视差精度。传统的立体匹配 SGBM 算法在连续细节信息区域具有视差效果好、速度快的特点。

本文首先引用 SGBM 算法生成原始视差 $\tilde{\boldsymbol{D}}$;然后,为了充分融合原始视差中精确的视差信息,过滤掉不正确的视差信息,提高立体匹配算法在连续细节信息区域的精度,利用绝对差的零均值和[14](zero-mean sum of absolute difference, ZSAD)算法生成左右图像的置信度图 $\boldsymbol{A}$;再后,为了确保视差分辨率一致,通过下采样将 $\tilde{\boldsymbol{D}}$ 和 $\boldsymbol{A}$ 调整到与 $\hat{\boldsymbol{D}}$ 相同的分辨率;最后,通过置信度图将初步视差 $\hat{\boldsymbol{D}}$ 和原始视差 $\tilde{\boldsymbol{D}}$ 进行置信度视差融合生成最终视差 $\boldsymbol{D}$,融合过程如下:

$$\boldsymbol{D}=\tilde{\boldsymbol{D}}+(1-\boldsymbol{A})\hat{\boldsymbol{D}} \tag{3}$$

式中:$\tilde{\boldsymbol{D}}$ 表示原始视差,$\hat{\boldsymbol{D}}$ 表示初步视差,$\boldsymbol{A}$ 表示置信度图,$\boldsymbol{D}$ 表示最终视差。

1.4 损失函数

由于基准模型 PASMNet 在边缘和连续等细节信息区域的视差精度较低,本文对 PASMNet 的损失函数进行了改进,原始的损失函数 L 如式(4)[12]所示。

$$L=L_p+\lambda_s L_s+\lambda_{PAM}(0.2L_{PAM}^1+0.3L_{PAM}^2+0.5L_{PAM}^3) \tag{4}$$

式中:L_p 为光度损失,L_s 为视差平滑性损失,L_{PAM}^i 为不同尺度 i 的视差注意力损失,其中 $i\in\{1,2,3\}$,λ_s、λ_{PAM} 分别表示光度损失和视差注意力损失的权重。

为了保留更多图像边缘和连续区域处的细节信息,提升这些细节信息区域的视差精度。本文在基准模型损失函数 L 之上引入原始视差监督损失 L_d、遮挡置信度损失 L_{ocf},以无监督方式训练网络,改进的损失函数 L' 如下:

$$L'=\lambda_{PAM}(0.2L_{PAM}^1+0.3L_{PAM}^2+0.5L_{PAM}^3)+\lambda_d L_d+\lambda_{ocf}L_{ocf}+L_p+\lambda_s L_s \tag{5}$$

式中:L_d 为原始视差监督损失,L_{ocf} 为遮挡置信度损失,λ_d、λ_{ocf} 分别表示 L_d 和 L_{ocf} 的权重。本文将融入的原始视差作为一种监督信号进行无监督训练,形成原始视差监督损失 L_d,来提升连续细节信息区域的视差精度。如式(6)所示。

$$L_d=\sum_p \boldsymbol{A}(p)smooth_{L1}(\boldsymbol{D}(p)-\tilde{\boldsymbol{D}}(p)) \tag{6}$$

式中:p 表示图像中的像素信息,$\tilde{\boldsymbol{D}}(p)$ 表示原始视差的像素,$\boldsymbol{A}(p)$ 表示置信度图的像素,$\boldsymbol{D}(p)$ 表示最终视差的像素。L_d 采用 $smooth_{L1}$ 损失函数。为了精准地去除原始视差中不准确的视差和粗略视差中的遮挡视差,提升边缘和连续细节信息区域的视差精度,本文将遮挡特征图和置信度图作为一种监督信号进行无监督训练,形成遮挡置信度损失 L_{ocf},如式(7)所示。

$$L_{ocf}=\sum_{p^f}-ln(\boldsymbol{O}_r(p)\boldsymbol{A}(p)) \tag{7}$$

式中:$\boldsymbol{O}_r(p)$ 表示遮挡特征图的像素,本文将二进制交叉熵损失[15]应用于遮挡置信度损失,防止 $\boldsymbol{O}_r$、$\boldsymbol{A}$ 收敛到0。

2 实验结果和分析

2.1 数据集与实验细节

本文在3个数据集对模型进行训练和测试:

1)Scene Flow 数据集是1个合成的数据集,含有 Flyingthings3D、Driving、Monkaa 3个子数据集,共有35 454张训练图像和4 370张测试图像,图像分辨率为960×540。

2)KITTI 2015 数据集是一个含有汽车、街道、建筑物等真实场景的数据集,它有200对具有真实视差的训练集立体图像和200对不具有真实视差的测试集图像,图像分辨率为376×1 240。本文将训练数据中的160对图像作为训练集,40对图像作为验证集。

3)KITTI 2012 数据集是一个真实场景的数据集,并且有灰色图像和彩色图像两种,它含有194对具有真实视差的训练集立体图像对和195对不具有真实视差的测试集图像对,图像分辨率为376×1 240。本文将训练数据中的160对彩色图像作为训练集,34对彩色图像作为验证集。

本文的实验环境为 GTX1080Ti 显卡,开发环境为 Python 3.6.6、Pycharm 2019,深度学习框架为 Pytorch 1.7。为了与基准模型进行对比,本文将超参数设置与基准模型 PASMNet 相同,批次设置为14,采用 Adam 优化器对模型进行优化,$\beta_1=0.9$,$\beta_2=0.999$。首先,在 Scene

Flow 数据集上对模型进行预训练，将图像随机裁剪为 256×512，迭代周期为 10，前 5 个周期学习率为 1×10^{-3}，后 5 个周期为 1×10^{-4}；最后，分别在 KITTI 2015 和 KITTI 2012 数据集上对模型进行微调，获得最优模型，迭代周期为 80，前 60 周期学习率为 1×10^{-4}，后 20 个周期学习率为 1×10^{-5}。

2.2　消融实验

1）验证设计模型的有效性

本文在 Scene Flow 和 KITTI 2015 数据集上对 SFPN 特征提取模块、视差融合模块以及改进的损失函数进行消融实验，所选的基础网络为 PASMNet[12]，通过与表 1 中的第一行数据进行对比，来评估本文设计的特征提取模块 SFPN，视差融合模块以及改进的损失函数对模型性能的影响。为了验证设计模型的有效性，采用端点误差[16]（end-point-error，EPE）和大于 t 像素（>t-pixels）误差[17]作为模型的评价指标，EPE 表示视差图中的真实视差和预测视差的欧氏距离，>t-pixels 误差表示预测和真实视差的误差大于 t 像素在所有像素中的比例。表 1 中“√”表示网络中运用该结构。

由表 1 的实验结果可知，本文在 PASMNet 基础之上，设计的 SFPN 特征提取模块相比于沙漏网络[10]的 EPE 降低了 0.45，提升了模型的准确率。进行视差融合之后各项误差均有降低，融合原始视差与未融合之前相比 EPE 降低了 0.75，3 像素误差（>3－pixels）降低了 0.8%。增加的 L_d 和 L_{ocf} 均降低了 3 像素误差和 EPE，尤其增加 L_d 的效果较为显著。本文模型最佳设置在 Scene Flow 数据集上的 EPE 降低到 2.58，3 像素误差降低到 13.23%，与基准模型相比，模型精度有较大提升。

表 1　模型在不同设置下的性能评估

特征提取		视差融合	损失函数			KITTI 2015	Scene Flow
沙漏网络[12]	SFPN		L	L_d	L_{ocf}	>3-pixels/%	EPE/pixel
√			√			15.91	4.54
	√		√			14.64	4.09
	√	√	√			13.83	3.34
	√	√	√	√		13.69	2.98
	√	√	√	√	√	**13.23**	**2.58**

2）验证不同损失函数权重对模型的影响

由于本文在基准模型之上改进了损失函数，增加的两种损失函数的权重对模型具有一定影响。将损失权重设置为 0～1 的组合进行实验，通过计算 KITTI 2015 和 KITTI 2012 验证集的 3 像素误差来分析增加的两种损失权重 λ_d、λ_{ocf} 对模型的影响，不同损失函数的权重在两个验证集上产生的三像素误差如图 4 所示。

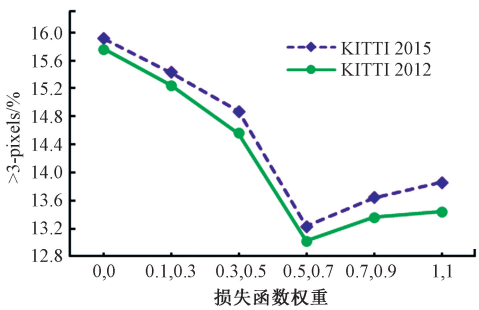


图 4　损失函数权重对模型的影响

从图 4 可以看出，当损失函数的权重 λ_d、λ_{ocf} 分别为 0.5、0.7 时模型达到了最优性能，在 KITTI 2015、KITTI 2012 的验证集上 3 像素误差分别为 13.23%、13.02%。

2.3　实验结果

1）KITTI 2015 上的实验结果

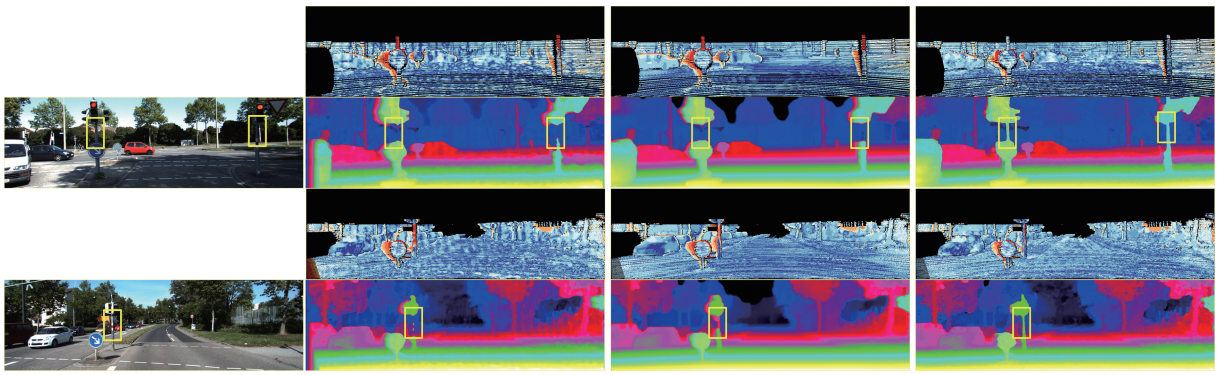
本文利用最优模型对 KITTI 2015 中 200 对测试集图像进行测试，将测试得到的视差图提交到 KITTI 2015 在线评估网络上，与相关算法进行对比，其结果如表 2 所示。表中“All”表示进行误差估计图像中的所有像素，“Noc.”表示图像中非遮挡区域的像素，“D1-bg”、“D1-fg”和“D1-all”分别表示在误差估计中背景像素、前景像素和所有区域像素[18]。

表 2　KITTI 2015 测试集排行榜结果

算法	All/%			Noc./%		
	D1-bg	D1-fg	D1-all	D1-bg	D1-fg	D1-all
Zhou[10]	—	—	10.23	—	—	9.91
OASM-Net[6]	6.89	19.42	8.98	5.44	17.30	7.39
AAFS[8]	6.27	13.95	7.54	5.96	**13.01**	7.12
Flow2Stereo[9]	5.01	14.62	6.61	4.77	14.03	6.29
PASMNet[12]	5.41	16.36	7.23	5.02	15.16	6.69
本文算法	**4.57**	**14.57**	**6.24**	**4.32**	13.84	**5.89**

由表 2 可以看出，本文算法除无遮挡区域的背景误差之外，其余各项误差均为最低，取得了最佳性能，与基准模型 PASMNet[12]相比，模型的各项误差均有所降低，提升了模型的精度，其中所有区域的误差由原来的 7.23%降低到 6.24%，非遮挡区域的误差从 5.02%降低到 4.32%，并且本文算法与其他无监督算法相比也具有一定优势。实验结果表明，基于细节信息增强的无监督立体匹配算法在视差预测中具有可行性。在 KITTI 2015 评估排行榜中本文

算法、基准模型 PASMNet、OASM-Net[6]模型估计的一些视差图和相应的误差图如图 5 所示,与左图对齐的为对应算法产生的视差图,视差图上方为对应算法产生的误差图,黄色矩形框表示左图部分细节信息区域。



(a) 左图 (b) OASM-Net (c) PASMNet (d) 本文算法

图 5 KITTI 2015 测试集的视差评估结果对比

由图 5 中矩形框的视差图可以得到,OASM-Net 和 PASMNet 在交通指示牌、路边细杆的边缘和连续细节信息区域无法产生精确的视差,然而,本文算法可以在这些细节信息区域产生较准确的视差。

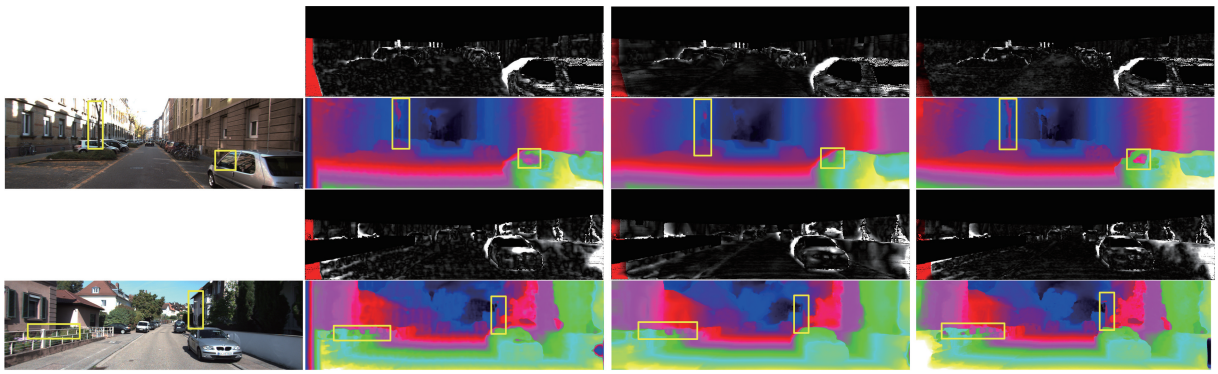
2)KITTI 2012 上的实验结果

与 KITTI 2015 评估方法类似,利用最优模型对 KITTI 2012 中的测试集图像进行预测,并将测试的视差图提交在 KITTI 2012 在线评估网络,和其他无监督算法进行对比,其对比结果如表 3 所示。

表 3 KITTI 2012 测试集排行榜结果

算法	>2-pixels/%		>3-pixels/%		>5-pixels/%	
	Noc	All	Noc	All	Noc	All
Zhou[10]	—	14.32	—	9.86	—	7.88
OASM-Net[6]	9.01	11.17	6.39	8.60	4.32	6.50
AAFS[8]	10.64	11.69	6.10	6.94	3.28	3.81
DispSegNet[7]	7.05	8.28	4.68	5.66	**2.76**	**3.43**
PASMNet[12]	8.57	10.08	5.71	6.96	3.62	4.47
本文算法	**6.20**	**7.67**	**4.33**	**5.60**	2.83	3.82

由表 3 可得,本文算法与相关算法对比,在 2 像素误差(>2-pixels)中非遮挡区域和所有区域均为最低,分别为 6.20%和 7.67%,3 像素误差(>3-pixels)中的非遮挡区域和所有区域也为最低,分别为 4.33%和 5.60%,虽然本文算法在 5 像素误差(>5-pixels)中没能达到最优,但是与具有最好性能的 DispSegNet[7]模型相差不大。>2-pixels、>3-pixels 最低表明本文算法更加适于小目标和边缘细节信息区域。综合来看,所提算法模型具有较好的优势,与基准模型 PASMNet[12]相比,各项误差均有降低。本文算法与 PASMNet、OASM-Net[6]在 KITTI 2012 测试集上视差估计的结果和相应的误差图如图 6 所示。从图 6 矩形框中细节信息区域的视差图可以看出,PASMNet、OASM-Net 模型在围栏、树干、车窗等小目标和连续细节信息区域具有较多噪声,无法产生精确的视差信息,本文算法在这些细节信息区域可以产生精确的视差信息。实验结果表明,本文提出的融合原始视差和 SFPN 特征提取模块改善了边缘和小目标细节信息区域的视差效果,提高了模型的精度。



(a) 左图 (b) OASM-Net (c) PASMNet (d) 本文算法

图 6 KITTI 2012 测试集的视差评估结果对比

3)Scene Flow 上的实验结果

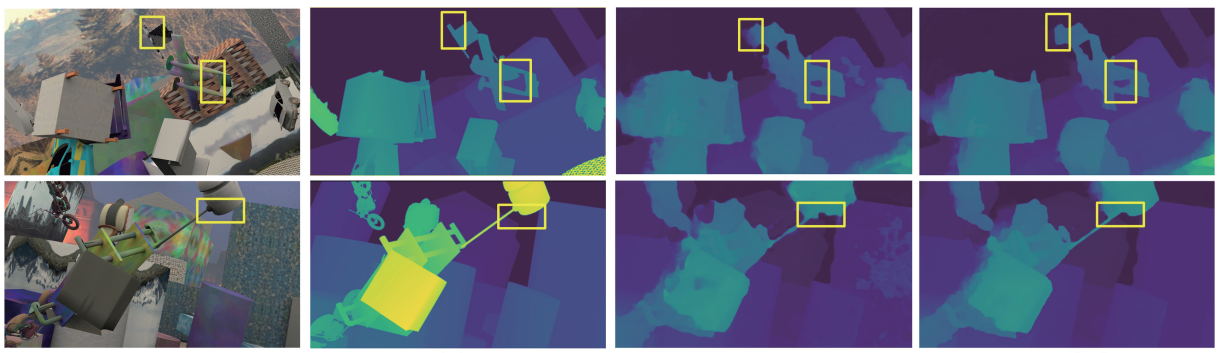
本文还利用 Scene Flow 测试集中数据与基准模型 PASMNet[12] 和其他无监督模型 OASM-Net[6]、DispSegNet[7]、AAFS[8]进行对比,来展示本文模型的优越性。对比结果如表 4 所示。

从表 4 中数据可知,本文算法准确率优于基准模型和其他无监督方法,在 Scene Flow 测试集上可以取得较好的效果。本文算法和 PASMNet[12] 在 Scene Flow 测试集上的对比结果如图 7 所示。

由图 7 矩形框中的视差图可以看出,本文算法在物体的边缘和连续等细节信息产生的视差图明显优于 PASMNet[10],改善了细节信息区域的视差不精确的问题,并在多个数据集上相比于其他模型均展示了较优的性能,因此本文算法具有较好的泛化性能。

表 4　Scene Flow 测试集中模型对比

算法	EPE/pixel
OASM-Net[6]	3.86
AAFS[8]	2.88
DispSegNet[7]	3.14
PASMNet[12]	3.54
本文算法	**2.59**



(a) 左图　(b) 真实视差　(c) PASMNet　(d) 本文算法

图 7　Scene Flow 测试集中视差估计结果对比

3　结　　论

本文提出了一种基于图像细节信息增强的无监督立体匹配算法,设计一种集成的 SFPN 模块,通过 SFPN 模块中横向和纵向两种卷积方式以及 SAM 设置的不同权重,抑制特征提取过程中边缘和小目标细节信息的丢失;置信度视差融合进一步提高了连续细节信息区域的精度,引入的原始视差监督损失和置信度遮挡损失,使模型获取更多边缘和连续区域的细节信息。在 KITTI 2015、KITTI 2012、Scene Flow 3 个公开数据集上验证可得,与基准模型 PASMNet 相比,提高了视差估计的精度,尤其在一些细节信息区域,与其他典型的无监督立体匹配算法相比也有较高的准确率。随着轻量化模型的不断优化,网络的实时性也具有重要意义,模型的轻量化是本文的下一步研究目标。

参考文献

[1]　葛兰,贾振堂.深浅层特征结合的自监督立体匹配[J].电子测量技术,2023,46(12):143-149.

[2]　LI P X, LIU P F, CAO F D, et al. Weight adaptive cross-scale algorithm for stereo matching[J]. Acta Optica Sinica, 2018, 38(12): 248-253.

[3]　HAN X J, LIU Y L. A stereo matching algorithm guided by multiple linear regression[J]. Journal of Computer-Aided Design and Computer Graphics, 2019, 31(1): 84-93.

[4]　CHANG J R, CHEN Y S. Pyramid stereo matching network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5410-5418.

[5]　ZHANG F, PRISACARIU V, YANG R, et al. GA-net: Guided aggregation net for end-to-end stereo matching[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 185-194.

[6]　LI A, YUAN Z. Occlusion aware stereo matching via cooperative unsupervised learning[C]. Proceedings of the IEEE Asian Conference on Computer Vision, 2018: 197-213.

[7]　JUNMING Z, SKINNER K A, VASUDEVAN R, et al. DispSegNet: Leveraging semantics for end-to-end learning of disparity estimation from stereo imagery[J]. IEEE Robotics and Automation Letters, 2019, 4(2): 1162-1169.

[8]　CHANG J R, CHANG P C, CHEN Y S. Attention aware feature aggregation for real-time stereo matching on edge devices[C]. Proceedings of the IEEE Asian Conference on Computer Vision, 2020: 365-380.

[9] LIU P P, KING I, LYU M, et al. Flow2Stereo: Effective self-supervised learning of optical flow and stereo matching [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 6647-6656.

[10] ZHOU C, ZHANG H, SHEN X Y, et al. Unsupervised learning of stereo matching [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 1576-1584.

[11] YIN Z C, SHI J P. GeoNet: Unsupervised learning of dense depth, optical flow and camera pose [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:1438-1452.

[12] WANG L G, GUO Y L, WANG Y Q, et al. Parallax attention for unsupervised stereo correspondence learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 2108-2125.

[13] 顾梦娇,朱宇锋,郭迎庆. 基于半全局立体匹配算法的改进研究[J]. 电子测量技术, 2020, 43(19): 89-93.

[14] YE X Y, ZHOU X C, WANG W Y, et al. Research on facial fatigue detection of drivers with multi-feature fusion[J]. Instrumentation, 2023, 10(1): 23-31.

[15] BADKI A, TROCCOLI A, KIM K. Bi3D: Stereo depth estimation via binary classifications [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 1597-1605.

[16] 余雪飞,顾寄南,黄则栋. 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.

[17] WANG H, SANG X, CHEN D, et al. Self-supervised stereo depth estimation based on bi-directional pixel-movement learning[J]. Applied Optics, 2022, 61(7): 7-14.

作者简介

王晓峰(通信作者),博士,教授,硕士生导师,主要研究方向为计算机视觉、图像处理和智能控制等。

E-mail: xfwang828@126.com

孙志恒,硕士研究生,主要研究方向为计算机视觉。