

DOI:10.19651/j.cnki.emt.2208796

# 基于边缘检测与注意力机制的立体匹配算法\*

余雪飞 顾寄南 黄则栋 荆彩霞

(江苏大学机械工程学院 镇江 212000)

**摘要:** 随着深度学习理论不断进步,端到端的立体匹配网络在自动驾驶和深度传感等领域取得了显著成果。然而,最先进的立体匹配算法仍然无法精确恢复物体的边缘轮廓信息。为了提高视差预测的精度,在本研究中,提出了一种基于边缘检测与注意力机制的立体匹配算法。该算法从立体图像对中学习视差信息,并支持视差图和边缘图的端到端多任务预测。为了充分利用二维特征提取网络学习到的边缘信息,本算法提出了一种全新的边缘检测分支和多特征融合匹配代价卷。结果表明,基于本文模型的边缘检测方案有助于提高视差估计的精度,所获取的视差图在KITTI 2015测试平台上的误匹配率为1.75%,与金字塔立体匹配网络相比,视差图的精度提高了12%,且运行时间减少约20%。

**关键词:** 深度学习;立体匹配;边缘检测;注意力机制

**中图分类号:** TP751.2 **文献标识码:** A **国家标准学科分类代码:** 520.6040

## Stereo matching algorithm based on edge detection and attention mechanism

Yu Xuefei Gu Jinan Huang Zedong Jing Caixia

(School of Mechanical Engineering, Jiangsu University, Zhenjiang 212000, China)

**Abstract:** With the continuous progress of deep learning theory, end-to-end stereo matching network has achieved remarkable results in the fields of automatic driving and depth sensing. However, the most advanced stereo matching algorithm still have trouble in accurately recover the edge contour information of the object. In order to improve the accuracy of disparity prediction, in this study, we propose a stereo matching algorithm based on edge detection and attention mechanism. The algorithm learns parallax information from stereo image pairs and supports end-to-end multi task prediction of parallax map and edge map. In order to make full use of the edge information learned by the two-dimensional feature extraction network, we propose a new edge detection branch and multi feature fusion matching cost volume. The results show that the edge detection scheme based on the model helps to improve the accuracy of parallax estimation. The error matching rate of the obtained parallax map on KITTI 2015 test platform is 1.75%. Compared with pyramid stereo matching network, the accuracy of parallax map is improved by 12% and the running time is reduced by 20%.

**Keywords:** deep learning; stereo matching; edge detection; attention mechanism

## 0 引言

双目视觉作为计算机视觉最热门的研究领域之一,是一个极其重要且充满挑战的学科,其目标是从一对立体图像对中计算每个像素的视差数据。近些年来,双目视觉被广泛应用于自动驾驶,医疗影像分析,三维重建,文物保护等方面。由于双目图像普遍存在病态区域,如遮挡、弱纹理、重复纹理、透视变形、反光区域等,如何能够获取高精度的视差图到目前仍然是一项具有挑战性的研究课题。

近年来,深度学习理论表现出了强大的图像理解及推理能力,特别是在目标分类<sup>[1]</sup>、检测<sup>[2-3]</sup>和语义分割<sup>[4-5]</sup>等任务中具有良好的表现力。为了充分利用深度学习理论的优秀性能,研究者们近年来也将深度学习算法引入立体匹配领域。卷积神经网络(CNN)不仅能够从图像中提取肉眼可见的纹理特征,也可以根据提供的真实数据,学习人们无法直接辨别的特征信息,从而表现出来更加强大的学习能力和推理能力。

深度学习立体匹配算法仍然无法精确恢复物体的边缘

收稿日期:2022-01-07

\* 基金项目:国家自然科学基金(51875266)项目资助

轮廓信息,原因在于研究者在特征提取阶段通常采取全局特征提取的思路。尽管引入空间金字塔池化(SPP)结构能够获取图像的上下文信息,扩大图像的感谢野,但仍然无法使网络能够对图像的边缘纹理线索进行深入的学习,从而无法恢复高精度的视差图。因此,本文所提出的算法主要针对上述问题,包含的贡献主要为:1)本文在图像二维特征提取阶段,设计了边缘检测分支网络用于学习图像的边缘纹理信息,针对边缘检测分支网络构建图像边缘信息真实数据集、边缘检测损失函数用于网络高效率学习图像的边缘线索,引导视差图的预测;2)本文提出了 3D 通道注意力机制模型,该模型在不改变图像感受野的前提下,实现对 3D 通道重点特征的提取,同时参数量相比于传统 3D 卷积减少约 25%;3)本文设计了多任务特征融合匹配代价卷,使模型能够将残差网络(Res-Net)学习的特征与边缘检测分支网络学习的特征进行融合,有助于提高模型的表现力,提高视差估计的准确度。

## 1 相关工作

传统的立体匹配方法通常包括四个步骤:匹配代价计算、代价聚合、视差预测和视差优化。其中每一步都可以用卷积神经网络(CNN)代替,用来学习匹配代价信息、代价聚合等,减少了人为设计造成的误差。一些研究人员主要关注使用 CNN 精确计算匹配代价函数,以及使用半全局匹配方法优化预测的视差图。Zbontar 等<sup>[6]</sup>首次将成对的小图像块和真实值(ground truth)作为输入,匹配代价值作为输出,并用监督学习网络(MC-CNN)训练双目图像的匹配代价。与传统方法相比上述方法在性能上取得了较大的提升。然而,将深度学习与传统算法相结合具有很大的局限性,其表现在于:为了计算不同视差下的匹配代价,网络包含多个前向通道,这将造成巨大的计算负担;遮挡区域中的像素不能用于训练,因此网络不能有效地估计遮挡区域中的视差;系统不是端到端网络,不能直接生成视差图,需要进行后处理。

端到端的立体匹配算法从整体上效果更加优异,Mayer 等<sup>[7]</sup>于 2016 年首次将端到端的 CNN 模型应用于立体匹配领域,建立了大规模人工合成数据集 Sceneflow,并对端到端神经网络进行了训练,该算法采用自动编解码结构直接输出视差图,无需后处理步骤匹配精度高、速度快。Kang 等<sup>[8]</sup>针对该算法在深度不连续区域和弱纹理区域匹配效果差的问题,提出了改进的 DispNet 神经网络学习算法,该算法利用扩展卷积建立上下文金字塔的特征提取模块,提出了一种基于图像块的匹配生成计算方法,并将视差梯度信息引入损失函数进行后处理以保持局部细节,进一步提高了视差估计的精度。

为了获得更多的图像特征信息,一些学者采用了深度特征融合和多尺度图像信息提取的卷积神经网络。Chang 等<sup>[9]</sup>提出了基于图像语义信息和全局信息的 PSM-Net

(pyramid stereo matching network, PSM)结构,该网络由两个模块组成:空间金字塔池化模块(SPP)和 3D 卷积神经网络模块。SPP 模块通过聚合不同尺度的上下文信息来构建匹配代价卷。3D 卷积神经网络模块使用多个堆叠沙漏结构(stacked hourglass),通过一系列的编码解码操作,提高全局信息利用率,来调整匹配成本量。

近年来,研究者们受到人类注意力机制的鼓舞,试图利用 CNN 设计一些网络注意力结构,以提高特征提取的性能。Hu 等<sup>[10]</sup>引入了挤压和激励模块,以充分利用网络中的信道信息。Wang 等<sup>[11-12]</sup>引入了 PASSR 网络来集成立体图像对中的超分辨率信息,并在 PASM-Net 中提出了 PAM 模块来计算沿极线的左图和相应右图的一致性得分。Ying 等<sup>[13]</sup>提出了一种通用的立体注意模块(SAM),旨在解决信息融合问题。Chen 等<sup>[14]</sup>利用交叉视差注意模块(CPAM)解决了大视差立体图像和多种极线问题。

上述算法尽管在一定程度上提高了视差估计任务的准确度,但仍然存在一些比较明显的缺点:首先,预测视差图的边缘线索不够准确、清晰;其次,采用对细节纹理区域不敏感的全局注意力策略,导致关键区域的视差估计不准确。

## 2 网络结构

立体匹配的过程通常是通过输入双目相机拍摄的一对双目图像寻找图像中的同名点,输出稠密视差图的过程。本文算法采取端到端的网络结构,以双目图像作为输入,直接输出预测的视差图。如图 1 所示:本算法的模型结构主要包括 4 个部分,二维特征提取结构与边缘检测结构、多任务特征融合结构、3D 注意力机制模块和堆叠沙漏结构(stacked hourglass)、匹配损失计算和视差回归结构。

### 2.1 边缘检测网络结构

PSM-Net<sup>[9]</sup>网络使用 SPP 模块提取不同尺度的特征信息。但该模块不仅大大增加了模型参数量,导致推理速度变慢,同时该结构不能有效的提取图像的边缘特征信息。本文选择边缘检测模型代替 PSM-Net 网络结构中的 SPP 模块,如图 1 所示。

本文在 ResNet-50 层中添加 2 个新的输出通道,对于每个输出分支,设计了一个边缘检测结构,包括 1 个  $3 \times 3$  卷积层和 1 个  $1 \times 1$  卷积层,伴随批量规范化(Batch normalization)和 ReLU 激活函数。由于 ResNet-50 输出通道数为 64,所以本文中  $3 \times 3$  卷积层的输入通道数为 64,输出通道数为 16;  $1 \times 1$  卷积层的输入通道数为 16 输出通道数为 1。为了融合不同分支中所提取的特征信息,将所有特征映射级联起来构建边缘特征成本量。最后将 ResNet-50 网络所学习到的特征与边缘检测的信息组合起来通过多任务特征融合结构组成 3D 匹配代价卷。

为了网络能够充分学习到图像的边缘信息,本研究构建了边缘检测损失函数和真实值。对于输出通道 1 预测出的边缘图像和输出通道 2 预测出的边缘图像分别赋予不同

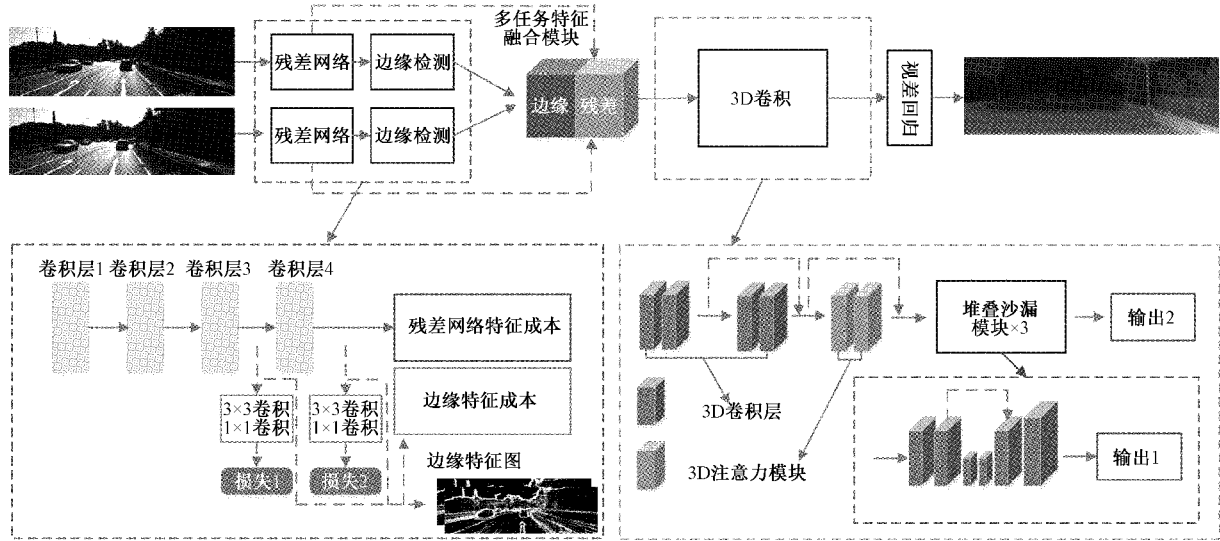


图 1 总体网络结构图

的权重(0.5 和 1),用总的损失函数来引导网络学习图像的边缘线索。

### 2.2 3D 通道注意力网络结构

过去的研究通常只关注 2D 特征信息,而忽略了 3D 通道特征的重要性。为了从 3D 卷积中充分提取图像中包含的细节纹理信息,本研究构建了 3D 通道注意力模块,该模块意在包含较多信息的通道进行重点关注,学习更多的特征信息,从而增强网络的表达能力、提高视差预测的准确性。

模型经过多任务特征融合模块之后,将会形成 3D 匹配代价卷,其中包含的特征信息为  $f \in R^{C \times D \times H \times W}$ ,  $C$  代表特征的通道数,  $D$  代表视差维度,  $H$  代表图像的高度值,  $W$  代表图像的宽度值。如图 2 所示,首先在视差维度上将该维度上的值进行元素相加操作,视差维度值变为 1,此时网络的特征信息为  $f \in R^{C \times 1 \times H \times W}$ ;对上述特征分别采用自适应全局平均池化和自适应全局最大池化操作。自适应全局池化操作可以使网络以全局的视野,获取图像的全局信息。经过上述两步操作之后特征分别为  $f_{avg} \in R^{C \times 1 \times 1 \times 1}$ ,  $f_{max} \in R^{C \times 1 \times 1 \times 1}$ ;将两个特征分别输入到两层卷积层中,其中第一层卷积的输入通道为  $C$ ,输出通道为  $C/r$ ,卷积核大小为  $1 \times 1$ (其中  $r$  代表下降率,通常取 16 可以减少网络的参数量)。第二层卷积层的输入通道为  $C/r$ ,输出通道为  $C$ ,卷积核大小为  $1 \times 1$ ;将上述特征按元素相加,再输入到 Softmax<sup>[15]</sup> 激活函数中输出特征  $f_u \in R^{C \times 1 \times 1 \times 1}$ 。为了不改变模型的输入输出参数,最后使用元素乘积操作将输出特征向量与输入特征合并,以获得最终 3D 通道注意力特征映射,其特征  $M_c \in R^{C \times D \times H \times W}$ 。

传统的 3D 卷积会对输入向量的每一个通道采取全局特征提取的机制,而现实是往往某些通道包含的特征信息相较于另外一些通道更丰富。所以,3D 通道注意力模块的作用在于对通道信息更为丰富的区域采取更多的关注,

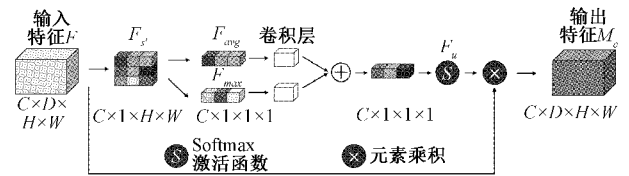


图 2 3D 通道注意力模块

提取更为复杂的内容。相比于传统的 3D 卷积,在获取相同的感受野的情况下,本算法模型生成的参数减少约 25%。

### 2.3 损失函数

本网络共输出 2 个视差预测图,为了充分利用输出特征图,在计算损失时对每个输出视差图分配不同的权重。为了使网络能够端到端的输出视差图,首先将堆叠沙漏结构输出的 3D 特征输入到 1 个  $3 \times 3$  卷积和 1 个  $1 \times 1$  卷积中获得单通道输出;此后,使用双线性插值对输出特征映射进行上采样;最后,设计了一个 Softmax 函数用于计算视差预测图。本算法总的视差平滑损失可按如下方式计算:

$$L_d(d_s, \tilde{d}_s) = \sum_{i=1}^2 \lambda_i \left[ \frac{1}{N} \sum_{j=1}^N \text{smooth}_{L_1}(d_j^i - \tilde{d}_j^i) \right] \quad (1)$$

式中:  $\lambda_i$  为第  $i$  个输出视差图;  $N$  代表单个视差图中像素的总个数;  $d_j^i$  为图像中的第  $j$  个元素,其对应的真实值为  $\tilde{d}_j^i$ 。平滑损失函数计算公式为:

$$\text{Smooth}_{L_1}(x) = \begin{cases} x^2, & |x| < 1 \\ |x| - 0.5, & \text{其他} \end{cases} \quad (2)$$

图像中物体边缘轮廓信息的准确预测有助于视差估计任务。为了边缘检测网络能够端到端的学习和输出准确的边缘预测图像,本网络使用了一种用于端到端边缘检

测交叉熵损失函数,其计算公式为:

$$L_c(x) = \frac{1}{N} \sum_{j=1}^N \begin{cases} \log(1 - P(x^j)), & y^j = 0 \\ \log P(x^j), & \text{其他} \end{cases} \quad (3)$$

其中,  $x^j$  和  $y^j$  分别表示图像中像素  $j$  处的预测值和真实值;  $P(x)$  表示 Sigmoid 激活函数,  $N$  表示图像中像素总数。

### 3 实验

#### 3.1 实验细节

本研究使用 PyTorch 工具实现了本文的架构,所有方法均使用 Adam 优化器进行网络的训练任务 ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.99$ )。由于实验条件的限制,本研究将网络的批量大小设置为 4,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  $256 \times 512$ ;最大视差值搜索范围设置为 192;对 2 个输出的预测视差图分别将其权重系数赋值为 0.7 和 1。本文使用 Scene Flow<sup>[7]</sup>作为网络训练的数据集并且迭代次数设置为 16,其中学习率(learning rate)为 0.001。当迭代数为 10、12 和 14 时,模型的学习率下降一半。对于 KITTI<sup>[16-17]</sup>数据集,使用预先用 Scene Flow 数据集训练后的模型权重来重新训练另外 300 个迭代。将初始学习率设置为 0.001,当模型的训练次数超过 200 个迭代时学习率下降 10 倍。

#### 3.2 数据集

本研究使用了两个公开的数据集进行网络的训练和测试,分别为 Scene Flow<sup>[7]</sup>和 KITTI<sup>[16-17]</sup>。

##### 1) Scene Flow

人工合成双目数据集。数据集包含 3 个子集: Driving、Flyingthings 3D 和 Monkaa,其中包括 35 454 张用于训练的图像和 4 370 张用于测试的图像;图像的高度为 540、宽度为 960,以及包含对应的真实值。由于 Scene Flow 数据集中的图片数量足够大训练后的网络模型具有很强的泛化能力,因此常常被用来作为立体匹配任务中的预训练数据集。

##### 2) KITTI 2012 和 KITTI 2015

使用激光雷达扫描获取空间点三维坐标的真实驾驶场景数据集。KITTI 2012 包括 194 个训练立体图像对和 195 个测试图像对。KITTI 2015 包含 200 个分别用于训练和测试的双目图像对。KITTI 数据集源自现实中真实的街景深度数据,因此 KITTI 数据集官网中的立体匹配算法排名往往具有较强的说服力,被众多研究人员使用。

#### 3.3 边缘检测模型实验

在本节中,与 PSM-Net 算法相比,本文对特征提取网络进行了一些关键性的修改。具体而言,本文舍弃了 PSM-Net 算法中的 SPP 结构,取而代之的是本算法提出的边缘检测模型用于提取特征信息;同时为了融合不同的特征向量,本算法设计了一个多任务特征融合匹配代价卷,

融合二维特征提取网络学习的图像边缘轮廓信息与残差网络提取的信息。

对于本文提出的多任务特征融合匹配代价卷,其包括两个组成部分:首先对于 Resnet-50 残差网络输出的特征图,将最终的输出通道设置为 32;另一部分为边缘检测输出结构,为了寻找最优的边缘检测网络输出通道值,进行了一系列消融实验如表 1 所示;当边缘检测网络的输出通道为  $16 \times 2$  时,模型在 KITTI 2015 与 KITTI 2012 的 EPE 值均为最小。同时本文所提出的算法相比于 PSM-Net,在 KITTI 2015 数据集中 EPE 值下降了约 10%;在 KITTI 2012 数据集中 EPE 值下降了约 14%。同时,本文模型在运算速度上也有所提高。

表 1 边缘检测模块通道数对比结果

模型	边缘通道数	KITTI2015 EPE/%	KITTI2012 EPE/%	Time/ms
本模型	$10 \times 2$	0.612	0.633	215.7
	$12 \times 2$	0.586	0.621	218.5
	$14 \times 2$	0.579	0.617	222.4
	$16 \times 2$	<b>0.575</b>	<b>0.616</b>	225.9
	$18 \times 2$	0.580	0.619	229.4
PSM-Net	无	0.639	0.713	246.1

#### 3.4 3D 通道注意力模型实验

在本节中,为了验证 3D 通道注意力模型的性能,本研究在 PSM-Net 网络上将其本身的 3D 卷积更换为本文所提出的 3D 通道注意力模型,以观察本文结构是否对网络表现力有所提高,其结果如表 2 所示。

表 2 3D 通道注意力模型实验

模型	注意力模块	EPE/px	>1px/%	>3px/%	Time/ms
PSM-Net	无	0.98	10.60	5.02	246.1
	II	0.89	10.10	4.80	249.4
本模型	II	<b>0.77</b>	<b>8.49</b>	<b>3.84</b>	<b>222.4</b>

本文所提出的 3D 通道注意力模块可以直接用于三维卷积层,因为其模型不会改变输入特征向量的通道数和图像大小。实验结果如表 2 所示,在计算时间略有增加的前提下,提出的 3D 通道注意力模块能够明显提高视差预测的准确度,提升网络的性能。在 Scene Flow 数据集中,PSM-Net 网络的 EPE 误差值为 0.98,在添加 3D 通道注意力模块后,其值降为 0.89,下降了约 10%;同时,本模型对视差预测的 3 像素误差值也具有明显的降低。将 SPP 模型由边缘检测模型替代后可以看出,与基准算法相比,所提算法不仅提高了算法运行效率,而且在各方面评测中都具有一定的优势。



### 3.5 几种典型算法的性能对比

将训练好的视差预测网络得到的结果上传至 KITTI 评测网站与其他优秀的算法进行比较。表 3 中展示了 KITTI 2015 数据集上部分算法以及本文算法的表现,其中“All”表示在输出的视差图评测时包含所有像素,“Noc”表示在评测时只考虑非遮挡区域内的像素。

表 3 不同算法的误差对比 (KITTI 2015)

模型	All/%			Noc/%			Time/s
	D1-bg	D1-fg	D1-all	D1-bg	D1-fg	D1-all	
GC-Net <sup>[18]</sup>	2.21	6.16	2.87	2.02	5.58	2.61	0.9
CRI <sup>[19]</sup>	2.48	3.59	2.67	2.32	3.12	2.45	0.47
PSM-Net <sup>[9]</sup>	1.98	4.87	2.32	1.71	4.51	2.24	0.41
本模型	<b>1.75</b>	<b>4.15</b>	<b>2.05</b>	<b>1.63</b>	<b>3.59</b>	<b>1.86</b>	<b>0.33</b>

图 3 所示为 GC-Net<sup>[18]</sup>、PSM-Net 和本研究提出的基于边缘检测与 3D 注意力机制的立体匹配算法生成的部分视差图和 KITTI 测评网站测试出的误差图(在误差图中蓝色的区域展示的是正确匹配的点,黄色区域表示匹配存在错误的点,红色区域展示的是匹配错误较多的点,黑色区域表示忽略的

点)。由表 3 可知,本研究所提出的算法,与网络[9,18-19]相比在视差预测的精度上有了明显的提高,同时其推理时间也具有一定的优势;与参考网络 PSM-Net 相比,整体的误差率由 2.32%下降至 2.05%,并且推理时间也缩短了将近 20%(从 0.41 s 下降至 0.33 s);在非遮挡区域中,本研究的总体误差相较与对比算法下降了约 17%(由 2.24%下降到 1.86%)。

图中白色的方框标记出了 GC-Net 和 PSM-Net 匹配效果较差的区域,通常出现在车辆边缘、栏杆等包含细节纹理的位置,而本研究设计的网络通过对物体的边缘信息进行有监督的学习,同时使用 3D 通道注意力机制获取图像的纹理信息,有效的提高了在困难区域视差预测的精度。

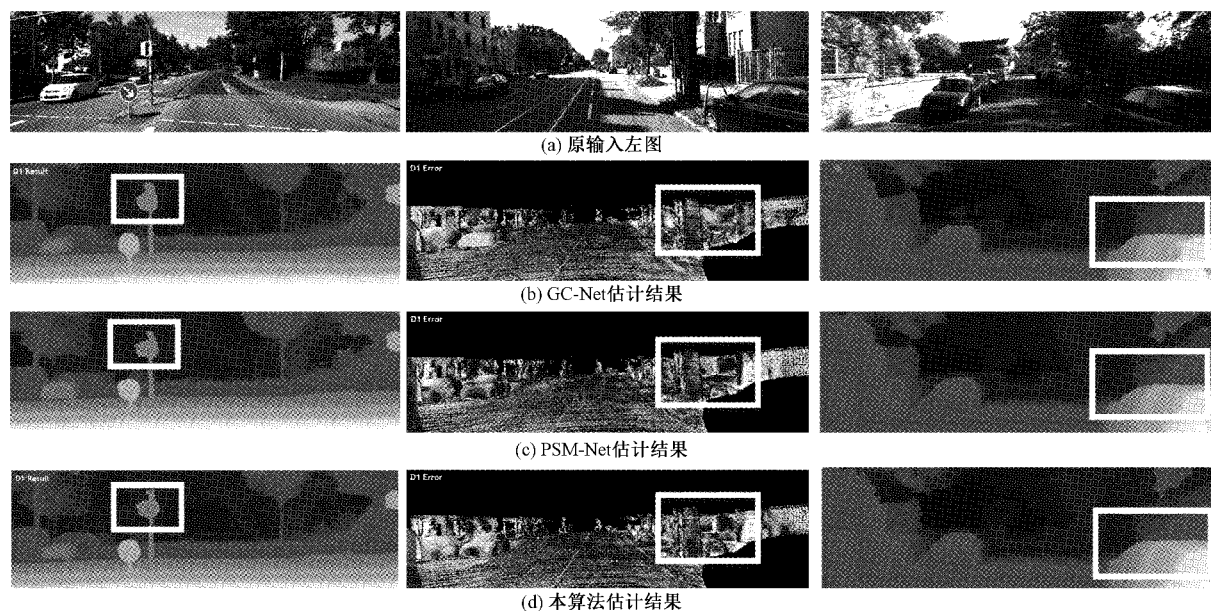


图 3 KITTI 数据集视差估计结果

## 4 结 论

本研究提出一种基于边缘检测与 3D 注意力机制的立体匹配算法,在 KITTI 2012、KITTI 2015、Scene Flow 数据集上,与其他几种典型的基于深度学习的算法相比,所提算法在利用边缘检测网络学习图像的边缘信息的同时,在 3D 特征提取阶段,使用 3D 通道注意力机制模型,获取图像的细节纹理信息。本文算法不仅从整体上提高了视差预测的精度,同时也缩短了推理的时间。但是本文算法

在进行视差预测时图像的分辨率往往较小(KITTI 数据集的图像分辨率为  $1\ 242 \times 375$ , Scene Flow 数据集图像的分辨率为  $960 \times 540$ ),在计算较大分辨率的图像时会出现显存爆炸的情况,所以下一步的研究目标为对高分辨率图像进行端到端视差预测。

### 参考文献

- [1] 毛向向,王红军,韩凤霞,等.基于深度卷积神经网络的机电系统故障分类识别方[J].电子测量与仪器学报,2021,35(2):87-93.

- [2] QIN Z, LI Z, ZHANG Z, et al. ThunderNet: Towards real-time generic object detection on mobile devices [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6718-6727.
- [3] 王文胜, 李继旺, 吴波, 等. 基于 YOLOv5 交通标志识别的智能车设计[J]. 国外电子测量技术, 2021, 40(10): 158-164.
- [4] LIU X, JI W, YOU J, et al. Severity-aware semantic segmentation with reinforced wasserstein training[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 12566-12575.
- [5] 陈彦彤, 李雨阳, 陈伟楠, 等. 基于深度语义分割的遥感图像海面舰船检测研[J]. 仪器仪表学报, 2020, 41(1): 233-240.
- [6] ZBONTAR J, LECUN Y. Stereo matching by training a convolutional neural network to compare image patches[J]. The Journal of Machine Learning Research, 2016, 17(1): 2287-2318.
- [7] MAYER N, ILG E, HAUSSER P, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:4040-4048.
- [8] KANG J, CHEN L, DENG F, et al. Context pyramidal network for stereo matching regularized by disparity gradients[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 157: 201-215.
- [9] 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.
- [10] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [11] WANG L, WANG Y, LIANG Z, et al. Learning parallax attention for stereo image super-resolution [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12250-12259.
- [12] WANG L, GUO Y, WANG Y, et al. Parallax attention for unsupervised stereo correspondence learning[J]. Ieee Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(4):2108-2125.
- [13] YING X, WANG Y, WANG L, et al. A stereo attention module for stereo image super-resolution[J]. IEEE Signal Processing Letters, 2020, 27: 496-500.
- [14] CHEN C, QING C, XU X, et al. Cross parallax attention network for stereo image super-resolution [J]. IEEE Transactions on Multimedia, 2021,24:202-216.
- [15] JOULIN A, CISSÉ M, GRANGIER D, et al. Efficient softmax approximation for gpus [C]. International Conference on Machine Learning, 2017: 1302-1310.
- [16] GEIGER A, LENZ P, URTASUN R. Arc we ready for autonomous driving? the kitti vision benchmark suite[C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition,2012:3354-3361.
- [17] MENZE M, GEIGER A. Object scene flow for autonomous vehicles [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3061-3070.
- [18] KENDALL A, MARTIROSYAN H, DASGUPTA S, et al. End-to-end learning of geometry and context for deep stereo regression[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 66-75.
- [19] PANG J, SUN W, REN J S, et al. Cascade residual learning: A two-stage convolutional neural network for stereo matching [C]. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017: 887-895.

### 作者简介

余雪飞, 硕士研究生, 主要研究方向为基于深度学习的三维重建技术、目标检测技术、语义分割技术等。

顾寄南, 工学博士, 教授, 博士生导师, 主要研究方向为智能制造技术及系统、智能机器人及机器视觉、复杂产品建模与创新设计理论、基于网络的设计制造及远程监控、机械CAD/CAE及计算机图像处理技术等。

E-mail: gujinan@tsinghua.org.cn

黄则栋, 博士研究生, 主要研究方向为基于深度学习的三维重建技术、点云分割技术等。

荆彩霞, 硕士研究生, 主要研究方向为基于深度学习的目标检测技术、图像分类、图像分割等。