计算机科学
COMPUTER SCIENCE

基于改进超像素采样的立体匹配网络

徐海东, 张自力, 胡新荣, 彭涛, 张俊

引用本文

徐海东, 张自力, 胡新荣, 彭涛, 张俊. 基于改进超像素采样的立体匹配网络[J]. 计算机科学, 2024, 51(11A): 231100005-7.

XU Haidong, ZHANG Zili, HU Xinrong, PENG Tao , ZHANG Jun. Stereo Matching Network Based on Enhanced Superpixel Sampling [J]. Computer Science, 2024, 51(11A): 231100005-7.

相似文章推荐（请使用火狐或 IE 浏览器查看文章）

Similar articles recommended (Please use Firefox or IE to view the article)

基于多模态融合的动态恶意软件检测方法

Multimodal Fusion Based Dynamic Malware Detection

计算机科学, 2024, 51(11A): 240200098-7. https://doi.org/10.11896/jsjkx.240200098

基于开放集的入侵检测方法研究

Study on Open Set Based Intrusion Detection Method

计算机科学, 2024, 51(11A): 231000033-6. https://doi.org/10.11896/jsjkx.231000033

基于CNN结合BiGRU的恶意流量分类算法研究

Study on Malicious Traffic Classification Algorithm Based on CNN Combined with BiGRU

计算机科学, 2024, 51(11A): 231100106-9. https://doi.org/10.11896/jsjkx.231100106

基于深度学习智能反射面辅助通信系统的联合波束成形

Deep Learning Based Joint Beamforming in Intelligent Reflecting Surface Enhanced WirelessCommunication Systems

计算机科学, 2024, 51(11A): 231200125-5. https://doi.org/10.11896/jsjkx.231200125

MB-ATMK:融合属性权重和时序元知识的多行为序列推荐模型

MB-ATMK:Multi-behavior Sequential Recommendation Integrating Attribute Weights andTemporal Meta-knowledge

计算机科学, 2024, 51(11A): 231100047-9. https://doi.org/10.11896/jsjkx.231100047

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.231100005

基于改进超像素采样的立体匹配网络

徐海东[1,2]　张自力[1,2,3]　胡新荣[1,2,3]　彭　涛[1,2,3]　张　俊[4]
1 湖北省服装信息化工程技术研究中心　武汉 430200
2 武汉纺织大学计算机与人工智能学院　武汉 430200
3 纺织服装智能化湖北省工程研究中心　武汉 430200
4 武汉工程大学计算机科学与工程学院　武汉 430205
(2846761532@qq.com)

摘　要　针对立体匹配中细节丢失、有遮挡,以及无纹理区域匹配精度低的问题,提出了一种基于改进超像素采样的立体匹配方法。首先,利用改进的超像素采样方法对用于立体匹配的高分辨率输入图像进行下采样,随后,将下采样后的图像对输入到立体匹配网络中,利用权值共享的卷积网络进行特征提取,使用3D卷积获取特征融合后的Cost Volume并生成视差图,再将输出的视差图进行上采样还原为最终的视差图。针对超像素采样过程中容易丢失细节从而影响后续立体匹配精度的问题,引入特征金字塔注意力模块(Feature Pyramid Attention,FPA)和改进的残差结构。根据上述两个方面的创新,提出了基于超像素采样的立体匹配网络FPSMnet(Feature Pyramid Stereo Matching Network),并选取、划分图像数据集BSDS500和NYUv2作为超像素采样的训练、验证和测试的数据集。立体匹配实验结果表明,与基准方法相比,所提算法在SceneFlow和HR-VS数据集上的平均像素误差分别下降了0.25和0.52,在不影响运行时间的前提下提高了匹配精度。

关键词:深度学习;超像素;立体匹配;注意力机制
中图分类号　TP391

Stereo Matching Network Based on Enhanced Superpixel Sampling

XU Haidong[1,2],ZHANG Zili[1,2,3],HU Xinrong[1,2,3],PENG Tao[1,2,3] and ZHANG Jun[4]
1 Engineering Research Center of Hubei Province for Clothing Information,Wuhan 430200,China
2 School of Computer Science and Artificial Intelligence,Wuhan Textile University,Wuhan 430200,China
3 Hubei Provincial Engineering Research Center for Intelligent Textile and Fashion,Wuhan 430200,China
4 School of Computer Science and Engineering,Wuhan Institute of Technology,Wuhan 430205,China

Abstract　Aiming at the accuracy challenges in stereo matching related to details,occlusion,and textureless regions,a stereo matching method based on improved superpixel sampling is proposed. Initially,an enhanced superpixel sampling method is employed to downsample the high-resolution input images used for stereo matching. Subsequently,the downsampled image pairs are input into the stereo matching network,where a convolutional network with shared weights is utilized for feature extraction. Using 3D convolution,a feature-fused Cost Volume is generated,leading to the creation of a disparity map. The outputted disparity map is then upsampled to reconstruct the final disparity map. To tackle the issue of potential detail loss during the superpixel sampling process,two innovations are introduced:the feature pyramid attention module(FPA)and an improved residual structure. Based on these two innovations,a stereo matching network named FPSMnet(feature pyramid stereo matching network)is proposed. This paper selects and partitions the image datasets BSDS500 and NYUv2 for training,validation,and testing of superpixel sampling. Experimental results in stereo matching demonstrate that,compared to the baseline method,the proposed algorithm achieves a reduction of 0.25 and 0.52 in average pixel errors on the SceneFlow and HR-VS datasets,respectively. These improvements are achieved without compromising runtime efficiency.

Keywords　Deep learning,Superpixels,Stereo matching,Attention mechanism

1　引言

双目视觉技术在计算机视觉领域具有关键地位,被广泛应用于智能驾驶[1]、姿态估计[2]、虚拟现实[3]和遥感等多个领域。通常,双目视觉任务包括4个主要步骤:相机的离线标定、图像的采集与校正、立体匹配和深度计算。其中,立体匹配是至关重要的一步,其核心目标是确定经过校正后的图像之间的像素对应关系,进而生成这两幅不同视角图像的视差图。匹配的准确性和效率直接影响着深度计算和深度图生成的性能。

基金项目:湖北省教育厅科学技术研究计划项目(B2017066)
This work was supported by the Science and Technology Research Project of Education Department of Hubei Province(B2017066).
通信作者:张自力(zlzhang@wtu.edu.cn)

传统的立体匹配算法通常分为4个阶段:代价计算、代价聚合、视差计算和视差校正[4]。然而,近年来随着深度学习的迅猛发展,深度学习方法通过深度神经网络学习到的特征通常优于传统方法手工设计的特征。在非端到端的深度学习立体匹配方法中,研究人员针对传统方法中的4个步骤,设计网络模型来取代其中一个或数个步骤,通常能够获得比传统方法更出色的效果,从而改善所生成的视差图效果。而基于深度学习的端到端立体匹配方法则将立体匹配任务中的所有步骤集成在一起,并进行联合优化,直接从双目图像对中生成视差图。

文献[5]提出了金字塔立体匹配网络PSMnet(Pyramid Stereo Matching Network),它包括空间金字塔池化和3D卷积两个核心模块。空间金字塔池化模块将全局上下文信息融入图像特征,而3D卷积则通过堆叠的沙漏网络进行正则化,以匹配代价体。相较于传统方法,这种端到端学习框架无需后处理操作,显著提高了匹配精度。然而,它在处理具有细纹理和间断小区域时仍然面临一些挑战。

此外,文献[6]中的SPSMnet(Superpixel Pyramid Stereo Matching Network)还提出了一种全卷积网络,可端到端生成超像素,并将其融入通用的超像素采样框架。传统超像素算法[7]可以分为基于图论的算法和基于聚类的算法,基于图论的算法分割出的超像素能较好地贴合边界但规则性较差;基于聚类的超像素算法能够得到形状规则的超像素,但边界不够贴合。基于深度学习的超像素方法[8]可以在边界贴合与规则性上取得平衡,但在深度学习中引入超像素需要解决标准卷积运算在不规则的超像素块上效率低下的问题。SPSMnet提出的基于全卷积网络的超像素采样方法取代了PSMnet中的传统采样策略,充分发挥了超像素在图像预处理中的优势,相较于传统采样方法,其更好地保留了图像边界和细节,提高了视差估计的准确性。但在下采样的过程中,图像细节会不可避免地随着像素的减少而丢失;上采样的过程中则会为同一超像素内的所有像素分配相同的差异值,这也会给最终差值带来误差。

文献[9]提出一种用于超像素分割的深度架构OverSegNet。该架构开发了一种NAS(Neural Architecture Search)编码器来取代其他超像素算法中的CNN编码器,改善了编码器的超像素聚类特征表示;并使用特定的前馈式聚类解码器提高解码器的计算效率。相较于SPSMnet中的全卷积采样方法,该框架在立体匹配中的应用实现了更低的终点误差,在细节区域和重复纹理区域也有更好的表现。但该框架仍然具有局限性,只能生成与初始化超像素相同级别的超像素输出。

本文在SPSMnet的基础上构建了一个基于改进超像素采样的立体匹配网络FPSMnet。在超像素编码阶段,引入了特征金字塔注意力模块[10],利用金字塔结构提取不同尺度的特征信息,同时引入了注意力机制[11],以获取全局上下文信息,并根据通道对特征进行选择。在超像素解码阶段,采用了残差结构[12],能更好地进行特征融合,从而改善了超像素采样网络可能导致图像细节丢失的问题,有助于超像素采样网络后续生成更准确的视差图。

2 基本原理

2.1 整体网络架构

SPSMnet与PSMnet整体网络架构相似,区别之处在于其使用超像素采样模块取代了传统采样方法。SPSMnet整体网络结构如图1所示,具体流程如下:1)利用超像素模块处的超像素采样网络预测出超像素与像素之间的关联矩阵 $\boldsymbol{Q}$,然后将其应用于立体匹配高分辨率输入图像进行下采样。2)下采样后的左右图像对经主干网络PSMNet处理。在PSMNet中,图像在经过两组权重共享的特征提取模块后,会再经过一组卷积层用于特征融合。随后,左右图像对的特征被用于构建Cost Volume。3)使用相同的矩阵 $\boldsymbol{Q}$ 对PSMNet生成的低分辨率差异图进行上采样,以实现最终的视差回归。

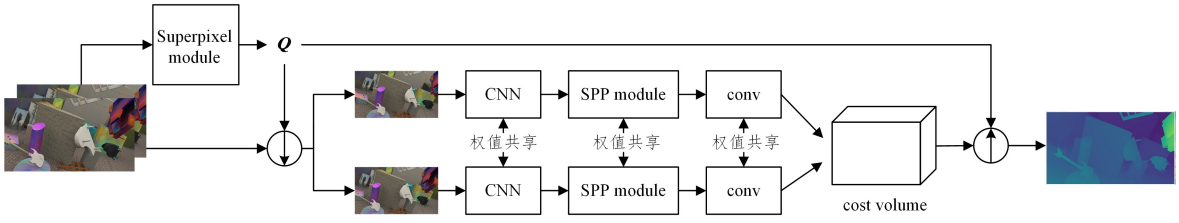


图1 SPSMnet架构

Fig. 1 SPSMnet architecture

2.2 超像素采样模块

全卷积神经网络[13]作为语义分割领域的开创性工作,将传统神经网络中的全连接层全部替换为全卷积层,网络仅包括卷积层和池化层,直接输出热力图而非类别信息。为了解决卷积和池化操作导致图像尺寸减小的问题,采用上采样方法来还原图像大小,同时结合不同深度特征的跳跃连接结构,以确保输出结果的鲁棒性和精确性。

为了解决像素空间中的位置信息丢失等问题,除了优化卷积结构外,还可以采用编码器-解码器结构[14]。编码器通过卷积层和池化层对输入图像进行特征编码,从中提取带有输入图像位置信息和语义信息的特征图。解码器则利用反卷积和反池化层来还原特征图中丢失的空间维度信息和位置信息,从而生成密集的目标预测图[15]。

本文在SPSMnet网络的基础之上对超像素模块进行改进,SPSMnet网络中的超像素采样模块采用了全卷积网络,使用了标准的编码器-解码器结构,并引入了跳跃连接来生成像素与超像素之间的关联图。本文使用联合FPA模块和残差结构的FPRE结构改进编码器-解码器结构,为了改善超像素下采样中细节丢失的情况,在编码器部分引入了FPA模块,增强采样模块的感受野,提高超像素分割边界效果。

超像素采样模块的编码器部分由5层卷积层组成,在第一层卷积层后添加FPA模块。在接收彩色图像作为输入后,第一层卷积层会对图像进行初步的特征提取并对网络通道进行扩充,经过初步处理的特征图传入FPA模块,FPA模块引

入了通道注意力机制，使得模型可以自适应地对每个通道的重要性进行建模，在每个通道上收集全局信息，帮助模型更好地理解完整图像的内容，而不仅仅是局部细节，再通过剩下的4层卷积层进行特征提取，生成高级特征图。

与此同时，在解码器部分，引入了整合后的残差结构，通过将超像素采样模块中的4层反卷积层替换为残差层，更充分地利用添加FPA模块后编码器从多个尺度提取到的特征信息，以输出更加准确的预测关联图，进一步完善超像素分割细节以改善差异估计。经过FPA模块加残差结构组合改进后的超像素采样网络FPRE能够生成更出色的超像素分割结果，为后续的立体匹配生成更高精度的视差图做出贡献。具体的网络架构如图2所示。

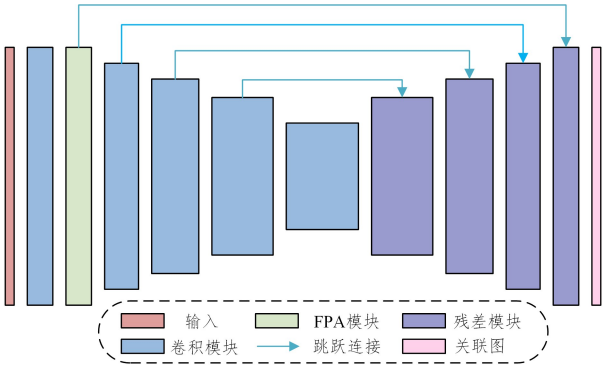


图2　超像素采样模块

Fig. 2　Superpixel sampling module

2.2.1　FPA模块

PSPNet[16]中提出的金字塔聚集模块使用金字塔结构，可以有效增加感受野，提取不同尺度的特征信息，但在不同尺度的特征图进行融合的过程中容易丢失像素定位信息。为了保证从CNN提取的高级特征中对像素的定位准确，尝试引入通道注意力，组合成FPA模块，利用全局上下文的先验注意力，根据通道来选择特征，弥补金字塔结构缺陷的同时达到更好的特征提取效果。具体网络结构如图3所示。

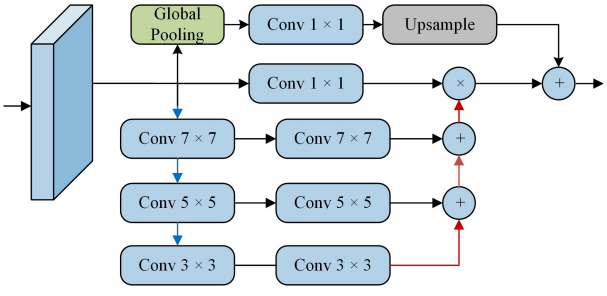


图3　FPA模块

Fig. 3　FPA module

FPA模块采用了类似于特征金字塔网络的U型结构，融合了3个不同金字塔尺度下的特征。为了更有效地捕获不同金字塔尺度的上下文信息，分别使用了3×3、5×5和7×7这3种不同尺寸的卷积核进行多层次的特征提取。在分层对特征信息进行不同程度的提取之后，将这3种尺寸的特征图逐层相加进行融合，以获得对当前任务更有意义的特征表示，同时抑制不相关的特征。金字塔提取到的特征与经过1×1卷积的原始特征相乘，最后将其与从全局平均池化分支提取到的全局特征相加，以生成最终的特征金字塔注意力特征。由于FPA模块的输入高层特征图具有较低分辨率，因此使用较大的卷积核也不会显著增加计算负担。

2.2.2　残差结构

在全卷积网络中，为了充分利用编码部分提取到的特征，在解码部分采用了跳跃连接，将编码部分的特征与反卷积得到的特征进行融合。这一策略改善了解码过程中感受野逐渐扩大导致细节丢失的问题。为了更好地保留细节信息，在解码上采样的过程中引入了残差结构，以改进超像素的分割边界，提高生成的超像素的泛化能力。残差结构如图4所示。

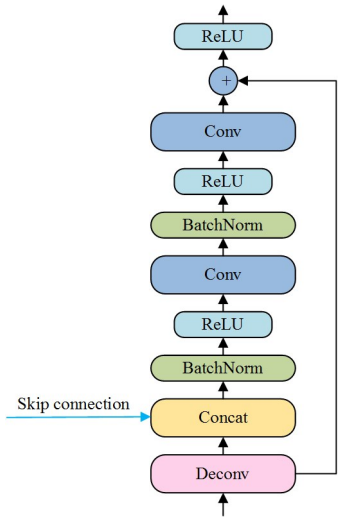


图4　残差结构

Fig. 4　Residual structure

SPSMnet中的全卷积网络解码过程中，会进行4次与编码部分相对应的反卷积操作，每一次反卷积后都通过跳跃连接与编码部分的特征图进行Concat处理。在此处引入改进后的残差结构，经过Concat后，特征会再经过两层BatchNorm、ReLU和卷积处理，然后与反卷积残差分支相加，最后再经过一次ReLU激活函数。上述操作在保留逐层获取的深层信息的同时，更好地整合了跳跃连接传递过来的原始特征信息，提高了模型的性能。

2.2.3　超像素采样和损失函数

受传统超像素算法的初始化策略[17]启发，本文采用超像素分割方法，把$H\times W$图像划分为大小为$h\times w$的规则网格，让每一个网格关联一个超像素，作为初始超像素的种子，将超像素分割任务转化为了预测像素与超像素之间的关联图。对于图像中的每一个像素$\boldsymbol{p}=(i,j)$和初始超像素种子$\boldsymbol{s}=(x,y)$，如果两者的关联$g_s(\boldsymbol{p})=g_{x,y}(i,j)$为1，则该像素属于当前超像素，每一个规则网格中的像素仅需从周围9个初始超像素种子中寻找映射$G\in\mathbb{Z}^{H\times W\times 9}$，使用超像素采样网络直接学习映射，找出像素与超像素种子之间的最大关联。

本文采用的损失函数包含两个部分，第一项鼓励模型将像素按相似的属性进行分组，第二项希望超像素在空间上保持紧凑：

$$L_{\mathrm{SLIC}}(\boldsymbol{Q})=\sum_{\boldsymbol{p}}\left\|\boldsymbol{f}_{\mathrm{col}}(\boldsymbol{p})-\boldsymbol{f}'_{\mathrm{col}}(\boldsymbol{p})\right\|_2+\frac{m}{S}\left\|\boldsymbol{p}-\boldsymbol{p}'\right\|_2 \quad (1)$$

其中，超像素的属性$\boldsymbol{f}_{\mathrm{col}}(\boldsymbol{p})$使用CIELAB颜色向量表示，并使用$\ell_2$准则进行度量，$S$是超像素采样间隔，$m$是平衡两个项的权重。

3 实验结果与分析

3.1 实验环境

本文实验环境配置如表 1 所列。

表 1 实验环境配置

Table 1 Experimental environment configuration

实验环境	配置
操作系统	Ubuntu 16.04
处理器	Inter(R)Xeon(R)Gold 5218
GPU	Tesla V100(32G)
CUDA 版本	CUDA 10.1
Python 版本	Python 3.7
深度学习框架	Pytorch
Torch 版本	Torch 1.6.0

3.2 超像素实验结果

3.2.1 超像素分割数据集

超像素采样实验部分选择了 BSDS500[18] 和 NYUv2[19] 数据集作为超像素实验数据集。BSDS500 数据集广泛用于图像分割和物体边缘检测，是标准的轮廓检测基准数据集。该数据集旨在评估自然边缘检测，包括物体轮廓、物体内部边界和背景边界等多个方面。BSDS500 包含 500 张自然图像，其中有 200 张用于训练，100 张用于验证，其余 200 张用于测试，所有图像都经过精细注释。本文所提算法及对比算法均遵循数据集所提供的方式进行划分。

NYUv2 数据集则是一个针对室内场景理解任务的 RGB-D 数据集。它包括了 1 449 张带标注的 RGB 和深度图像，这些图像采集自 3 个不同城市的 464 个场景，还包括 407 024 张未标注的图片。每个目标都附带一个类别和标识号，文献[20]在这个数据集上开发了一个用于超像素评估的基准方法，通过移除图像边界附近未标记的区域，选择了一个包含 400 张测试图像的子集，这些图像的大小为 608×448 像素。

在 BSDS500 数据集上对超像素采样模型进行训练和评估，并将本文算法 FPRE 与 SLIC[21]，ERS[22] 和 FCN 进行了比较。SLIC 作为传统聚类算法中的代表，使用 K 均值迭代聚类的方法来生成超像素，能够在保持图像结构的同时减少计算开销，生成的超像素也较为匀称。ERS 则使用信息熵率(Entropy Rate)来度量图像中的像素分布，根据信息熵率合并那些具有相似颜色和强度分布的相邻像素类别，减少不同类别像素的数量。ERS 生成的超像素对图像的纹理变化和颜色变化有较好的适应性。为了测试本文提出的方法在不同数据集上的泛化效果，直接将在 BSDS500 上训练的模型应用于 NYUV2 数据集，而不进行任何微调。图 5 展示了这两个数据集上的分割效果。通过局部放大图对比可以看出，与两个代表性算法以及基准模型 FCN 相比，本文算法 FPRE 在保持超像素的规则性的同时与边界贴合得更好。

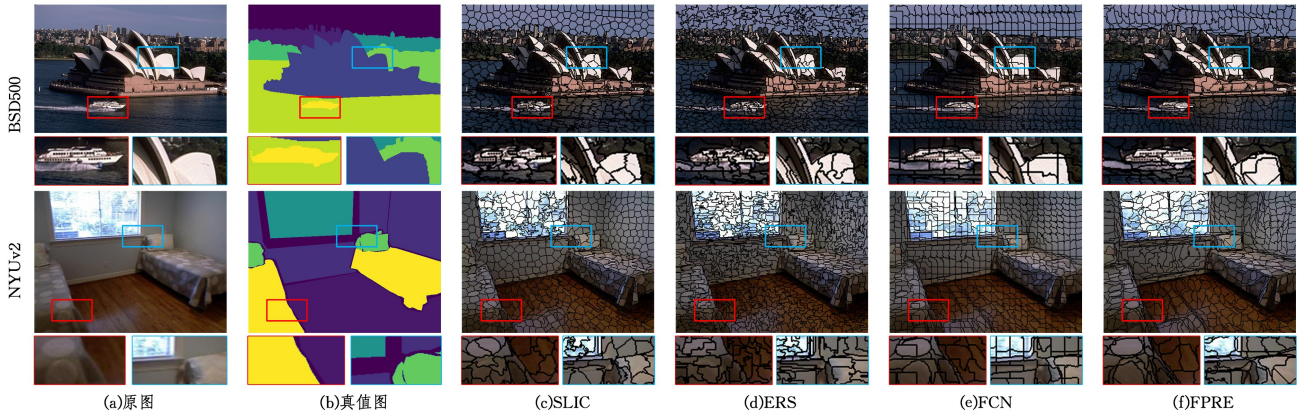


图 5 超像素分割结果可视化

Fig. 5 Visualization of superpixel segmentation results

3.2.2 超像素评估指标

实验选取可达分割精度(ASA)、边界召回率和边界精度(BR-BP)以及紧凑度(CO)作为超像素分割的评价指标。ASA 用于评估超像素作为预处理步骤可以达到的分割精度，如式(2)所示。

$$ASA(G,S)=\frac{1}{N}\sum_{S_j}\max_{G_i}\{|S_j\cap G_i|\} \tag{2}$$

其中，S 代表超像素分割，G 代表真实分割；S_j 表示超像素分割部分，G_i 表示真实分割部分，两者交集越大说明超像素分割效果越好。BR-BP 用于衡量超像素和真实值的边界一致性，如式(3)和式(4)所示：

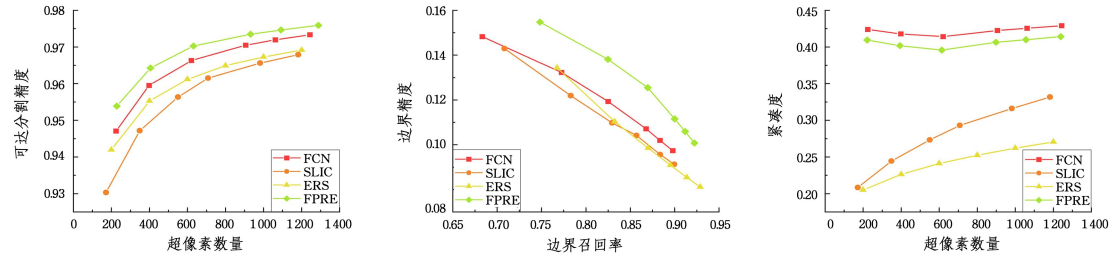
$$Rec(G,S)=\frac{TP(G,S)}{TP(G,S)+FN(G,S)} \tag{3}$$

$$Pre(G,S)=\frac{TP(G,S)}{TP(G,S)+FP(G,S)} \tag{4}$$

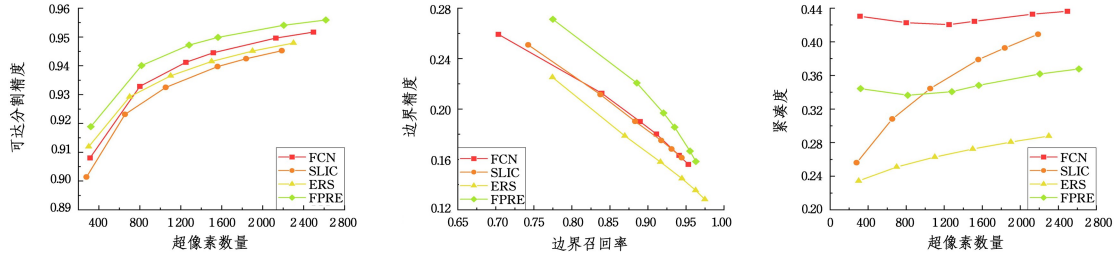
其中，$TP(G,S)$ 表示 S 中的边界像素在规定的邻域中存在 G 中的边界像素，$FN(G,S)$ 表示 S 中的边界像素在规定的邻域中不存在 G 中的边界像素，Rec 值越高表示边界一致性越高，Pre 值越高表示边界精度越高；CO 则用于评估超像素的紧凑性，如式(5)所示，$A(S_j)$ 表示超像素 S 的面积，$P(S_j)$ 表示与超像素 S 周长相同的圆的面积，两者比值越高，超像素越紧凑。

$$CO(G,S)=\frac{1}{N}\sum_{S_j}|S_j|\frac{4\pi A(S_j)}{P(S_j)} \tag{5}$$

本文算法与其他算法对比结果如图 6 所示。在 BSDS500 数据集上本文所用方法 FPRE 在 ASA，BR-BP 上均优于其他算法，这意味着本文算法能获取边界分割精度更高、边界贴合更好的超像素，这有利于后续的立体匹配任务。然而如文献[23]中所说，边界一致性与紧凑性之间需要权衡。本文所用的方法在紧凑性上略逊于 FCN 方法但仍优于其他算法；基于 NYUv2 数据集的评估，未对模型进行任何微调，主要是为了研究本文方法的泛化性。虽然数据集不同，针对测试的超像素数量不同，但本文算法仍然取得了优于其他算法的效果，这验证了本文算法的泛化性。



(a)BSDS500



(b)NYUv2

图 6　超像素评估

Fig. 6　Superpixel evaluation

3.3　立体匹配实验结果

3.3.1　立体匹配数据集

立体匹配实验在 SceneFlow[23] 和 HR-VS[24] 两个公开的数据集上进行评估。SceneFlow 是一个具有密集实况差异的大规模合成数据集。该数据集提供 35454 对立体图像用于训练，4370 对立体图像用于测试。HR-VS 是一个具有城市驾驶视图的合成数据集。包含 780 幅图像，分辨率为 2056×2464。有效差异范围为[9.66，768]。由于没有发布测试集，参考文献[6]随机选择了 680 幅用于训练，其余用于测试。模型的可视化结果如图 7 所示。

从可视化结果与局部放大图可以看出，在木头纹理区域和车顶的细节区域，本文方法取得了更好的视差效果，超像素采样上的改进不仅改善了超像素分割的边界，同样改善了视差图的边界。

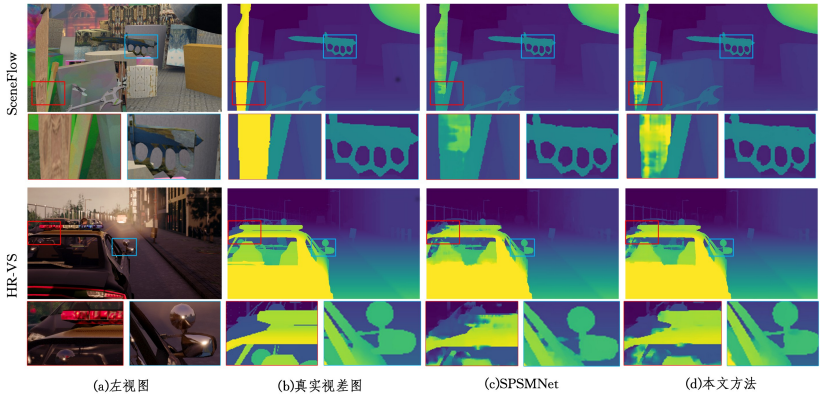


图 7　立体匹配视差图结果可视化

Fig. 7　Visualization of stereo matching parallax map results

3.3.2　消融实验

为了验证金字塔注意力模块和残差模块设计的合理性和有效性，在 SceneFlow 数据集上进行消融实验，首先对单独添加 FPA 模块的模型进行评估，然后对单独添加 RES 残差模块的模型进行评估，最后将这两种方法组合在一起与基准网络 SPSMNet 进行对比。其中打“√”表示使用了对应模块，未打“√”表示未添加对应模块。实验结果如表 2 所列。

表 2　消融实验

Table 2　Ablation experiment

算法	FPA	RES	EPE/px
本文	√		0.720
		√	0.723
	√	√	0.684
SPSMNet			0.930

如式(6)所示，EPE 表示平均误差像素，是模型估计的视差与真实视差之间的误差平均值。N表示像素点的总数，d_i 表示真实视差值，$\hat{d}_i$表示预测视差值。EPE 越低，视差精度越高，模型立体匹配的性能越好。同基准模型 SPSMnet 相比，两个模块均在不同程度上降低了平均像素误差，而两个模块结合之后，得到了最佳实验结果。消融实验表明了金字塔注意力模块和残差模块的设计是合理且有效的。

$$EPE=\frac{1}{N}\sum_{i=1}^{N}(|d_i-\hat{d}_i|) \tag{6}$$

3.3.3　对比实验

在 SceneFlow 上将本文算法与其他算法进行性能对比，对比结果如表 3 所列。结果显示，本文所采用的算法与 OverSegNet 相比，EPE 降低了 0.17；与基准方法 SPSMNet

相比，EPE 降低了 0.25。虽然计算速度有所下降，但与其他算法相比仍具有竞争力。参考文献[5]的建议，在训练和测试时都排除了色差大于 192 的像素。为了最大程度发挥超像素采样框架的效果，在 SceneFlow 数据集上使用 SLIC 损失函数同时训练超像素采样网络和立体匹配网络。将损失函数中的平衡权重 m 设置为 30，并将输入图像随机裁剪为 512×256。模型进行了总计 13 次的耗时训练。初始学习率为 1×10^{-3}，经过 11 次和 12 次后，学习率分别降至 5×10^{-4} 和 1×10^{-4}。

表 3 SceneFlow 数据集上的网络模型效果对比

Table 3 Comparison of network model effects on SceneFlow dataset

算法	EPE/px	Runtime/ms
CRL[25]	1.32	470
GCNe[26]	2.51	900
PSMNet[5]	1.09	410
SPSMNet[6]	0.93	323
OverSegNet[9]	0.85	—
本文方法	0.68	328

在 HR-VS 数据集上，对之前在 SceneFlow 上训练的模型进行微调。对比结果如表 4 所列，由于高分辨率和大差距，原始 PSMNet 无法直接应用于全尺寸图像。通常的做法是，将输入图像和差异图下采样到 1/4 大小进行训练，然后将结果上采样到全分辨率进行评估。在本文的方法中，预测网格单元大小为 16×16 的超像素，以执行 16 倍的下采样或上采样。在训练过程中，同样设置 m 为 30，并将图像随机裁剪为 2048×1024。对本文方法进行了 200 次训练，批量大小为 4。初始学习率为 1×10^{-3}，150 次后降低到 1×10^{-4}。3 像素误差(3px-Error)表示预测视差值与真实视差值之间差值的绝对值超过 3 的像素点的数量占整个图像像素数量的比例。同基准模型 SPSMnet 相比，本文方法 EPE 明显降低，3px-Error 略有下降。与 SceneFlow 相比，本文方法在 HR-VS 这个高分辨率数据集上获得了更大的性能提升。

$$3px\text{-}Error=\frac{1}{N}\sum_{i=1}^{N}(|d_i-\hat{d}_i|>3) \tag{7}$$

表 4 HR-VS 数据集网络模型效果对比

Table 4 Comparison of network model effects on HR-VS dataset

算法	EPE/px	Runtime/ms	3px-Error/%
SPSMnet	2.77	375	2.20
本文文献	2.25	371	2.17

结束语 本文构建了一个基于超像素采样的立体匹配网络，利用特征金字塔注意力模块提取更多的特征信息，通过引入残差模块对提取的特征进行优化，以此提升超像素采样效果；使用超像素采样替换传统采样方法改善立体匹配网络。实验结果表明，本文算法不仅改善了超像素分割效果，还提高了视差预测的精度，而且对计算效率的影响较小。但本文算法所使用的超像素采样网络需要预先进行训练，后续将研究如何将其更好地运用到立体匹配及其他网络中去。

参 考 文 献

[1] YAO A Q, XU J M. Electric vehicle charging port recognition and positioning system based on binocular vision[J]. Sensors and Microsystems, 2021, 40(7): 81-84.

[2] QI Y F, MA Z Y. Multi-loss head pose estimation based on deep residual networks[J]. Computer Engineering, 2020, 46(12): 247-253.

[3] ZENATI N, ZERHOUNI N. Dense stereo matching with application to augmented reality[C]// 2007 IEEE International Conference on Signal Processing and Communications. IEEE, 2007: 1503-1506.

[4] SCHARSTEIN D, SZELISKI R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[J]. International Journal of Computer Vision, 2002, 47(1/2/3): 7-42.

[5] CHANG J R, CHEN Y S. Pyramid Stereo Matching Network[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition: [Volume 8 of 13]. IEEE, 2018: 5410-5418.

[6] YANG F, SUN Q, JIN H, et al. Superpixel segmentation with fully convolutional networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 13964-13973.

[7] SONG X Y, ZHOU L L, LI Z G, et al. A comprehensive survey of superpixel methods in image segmentation[J]. Journal of Image and Graphics, China, 2015, 20(5): 599-608.

[8] JAMPANI V, SUN D Q, LIU M Y, et al. Superpixel Sampling Networks[C]// Computer Vision－ECCV 2018: 15th European Conference. Springer, 2018: 363-380.

[9] LI P, MA W. OverSegNet: A convolutional encoder-decoder network for image over-segmentation[J]. Computers and Electrical Engineering, 2023, 107: 108610.

[10] LI H, XIONG P, AN J, et al. Pyramid attention network for semantic segmentation[J]. arXiv: 1805.10180, 2018.

[11] 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.

[12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[13] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[14] TIAN X, WANG L, DING Q. A review of image semantic segmentation methods based on deep learning[J]. Journal of Software, 2019, 30(2): 440-468.

[15] WANG Y R, CHEN Q L, WU J J. A review of image semantic segmentation methods for complex environments[J]. Computer Science, 2019, 46(9): 36-46.

[16] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2881-2890.

[17] VAN DEN BERGH M, BOIX X, ROIG G, et al. Seeds: Superpixels extracted via energy-driven sampling[C]// Computer Vision－ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, Part VII 12. Springer Berlin Heidelberg, 2012: 13-26.

[18] ARBELAEZ P, MAIRE M, FOWLKES C, et al. Contour detection and hierarchical image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(5): 898-916.

[19] SILBERMAN N，HOIEM D，KOHLI P，et al. Indoor segmentation and support inference from rgbd images[C]// Computer Vision—ECCV 2012：12th European Conference on Computer Vision，Florence，Italy. Springer Berlin Heidelberg，2012：746-760.

[20] STUTZ D，HERMANS A，LEIBE B. Superpixels：An evaluation of the state-of-the-art[J]. Computer Vision and Image Understanding，2018，166：1-27.

[21] ACHANTA R，SHAJI A，SMITH K，et al. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2012，34(11)：2274-2282.

[22] LIU M Y，TUZEL O，RAMALINGAM S，et al. Entropy rate superpixel segmentation[C]// CVPR 2011. IEEE，2011：2097-2104.

[23] 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.

[24] YANG G，MANELA J，HAPPOLD M，et al. Hierarchical deep stereo matching on high-resolution images[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019：5515-5524.

[25] PANG J H，SUN W X，JIMMY S J R，et al. Cascade Residual Learning：A Two-stage Convolutional Neural Network for Stereo Matching[C]// 2017 IEEE International Conference on Computer Vision Workshops(ICCVW 2017). Venice，Italy，2[v. 2]. ：Institute of Electrical and Electronics Engineers，2017：878-886.

[26] KENDALL A，MARTIROSYAN H，DASGUPTA S，et al. End-to-End Learning of Geometry and Context for Deep Stereo Regression[C]// ICCV 2017. IEEE，2017.

XU Haidong，born in 1999，postgraduate，is a member of CCF(No. Q6975G). His main research interests include machine learning and image processing.

ZHANG Zili，born in 1981，Ph.D，lecturer，is a member of CCF(No. 99006M). His main research interests include machine learning and image processing.