



基于非对称空间金字塔池化的立体匹配网络

王金鹤, 苏翠丽, 孟凡云, 车志龙, 谭 浩, 张 楠

(青岛理工大学 信息与控制工程学院, 山东 青岛 266000)

摘 要: 卷积神经网络因具有强大的表征能力而被广泛用于图像处理算法, 但其在处理过程中存在耗时和信息损失等不足。为此, 提出一种基于非对称空间金字塔池化模型的卷积神经网络结构。设计非对称金字塔池化方法融入立体匹配网络, 以获取更详细的图像特征信息。分别叠加卷积核为 3×3 和 1×1 的卷积层, 用于融合多尺度信息和提升网络收敛速度, 同时将网络结构由 4 层增加至 7 层, 以提高匹配精度。在 KITTI 和 Middlebury 数据集上进行视差预测, 实验结果表明, 与基准网络相比, 该网络结构可使收敛时间缩短约 50.1%, 匹配错误率从 6.65% 降低至 4.78%, 在立体匹配中获得更平滑的视差效果。

关键词: 卷积神经网络; 非对称空间金字塔池化; 多尺度融合; 信息损失; 立体匹配

开放科学(资源服务)标志码(OSID):



中文引用格式: 王金鹤, 苏翠丽, 孟凡云, 等. 基于非对称空间金字塔池化的立体匹配网络[J]. 计算机工程, 2020, 46(7): 228-234, 242.

英文引用格式: WANG Jinhe, SU Cuili, MENG Fanyun, et al. Stereo matching network based on asymmetric spatial pyramid pooling[J]. Computer Engineering, 2020, 46(7): 228-234, 242.

Stereo Matching Network Based on Asymmetric Spatial Pyramid Pooling

WANG Jinhe, SU Cuili, MENG Fanyun, CHE Zhilong, TAN Hao, ZHANG Nan

(School of Information and Control Engineering, Qingdao University of Technology, Qingdao, Shandong 266000, China)

[Abstract] Convolutional Neural Network(CNN) is often used in image processing algorithms because of its excellent representation capabilities, but the process is time-consuming and often results in information loss. To address the problem, this paper proposes a CNN structure based on Asymmetric Spatial Pyramid Pooling(ASPP) model. An ASPP method is designed to be integrated with the stereo matching network to obtain more specific information about image features. Then convolutional layers with a 3×3 convolution kernel are superposed on those with a 1×1 convolutional kernel for multi-scale information fusion and improvement of network convergence speed. Also, the number of network layers is increased from four layers to seven layers to improve the matching accuracy. The parallax prediction is performed on the KITTI and Middlebury data sets. Experimental results show that, compared with the benchmark network, the proposed network structure shortens the convergence time by about 50.1% and reduces the matching error rate from 6.65% to 4.78%, achieving a smoother parallax effect in stereo matching.

[Key words] Convolutional Neural Network(CNN); Asymmetric Spatial Pyramid Pooling(ASPP); multi-scale fusion; information loss; stereo matching

DOI: 10.19678/j.issn.1000-3428.0055428

0 概述

三维场景信息重建是计算机视觉应用的关键, 如车辆自动驾驶和医学内窥镜成像技术的应用, 都是基于该技术对目标场景或物体的深度信息进行计算。深度信息计算目的是获取双目图像中像素点的

视差值, 而视差值取决于空间景物在左右视图中的对应关系, 这种寻找左右图像平面之间对应点的过程称为立体匹配^[1]。

传统立体匹配方法通常包含 4 个步骤, 即匹配代价计算、匹配代价聚合、视差计算和视差细化。早期的立体匹配算法分为全局匹配、局部匹配以及两

基金项目: 国家自然科学基金(31271077); 山东省高等学校科技计划项目(J17KA061)。

作者简介: 王金鹤(1963—), 男, 教授, 主研方向为图像处理、模式识别; 苏翠丽, 硕士研究生; 孟凡云, 讲师; 车志龙、谭 浩, 硕士研究生; 张 楠, 讲师。

收稿日期: 2019-07-09

修回日期: 2019-08-20

E-mail: sucuili1993@163.com

者结合的半全局匹配算法。但是这些方法均通过人工设计代价函数学习图像特征之间的线性关系,不仅计算代价昂贵,而且在遮挡、重复纹理、弱纹理等区域达不到理想的匹配效果。

深度学习技术的发展使计算机视觉研究得到重大突破。卷积神经网络(Convolutional Neural Network, CNN)能够进行复杂的非线性表示,具有强大的表征能力,近年来被广泛用于图像识别及立体匹配的应用。进而,基于卷积神经网络的立体匹配方法相继被提出。但现有方法多采用深而复杂的3D卷积网络架构进行端到端的视差图学习。尽管这些立体匹配方法在网络框架和训练方式上存在巨大差异,但是通常都采用Siamese网络作为立体图像对的特征提取器,由此可见Siamese网络结构在立体匹配中起到基础性作用。

目前有些研究避免使用复杂的端到端CNN框架,而是针对Siamese网络结构进行改进。受此启发,本文改进基础的Siamese特征提取网络,引入空间金字塔池化(Spatial Pyramid Pooling, SPP)思想优化池化方式,以充分利用图像信息。通过构建非对称空间金字塔池化(Asymmetric SPP, ASPP)模型,在特征提取阶段对图像块进行多尺度特征提取,以期获得更准确的视差估计结果,提高匹配精度。

1 相关研究

基于卷积神经网络的立体匹配方法主要包括基于CNN的匹配代价学习、基于CNN的视差回归和基于CNN的端到端视差图获取三类。

1) 基于CNN的匹配代价学习方法主要以图像块之间的相似性作为匹配代价。文献[2]成功地将CNN应用到匹配代价计算中,通过深层Siamese网络来预测图像块之间的相似性得分,以此获取图像块的匹配代价。文献[3]在此基础上提出一种缩小型立体匹配网络,较好地解决了原模型处理过程中耗时多的问题。文献[4]研究了一系列用于图像块匹配的二分类网络,并且建立双通道网络架构进行视差估计。此后,文献[5]提出使用内积层计算图像块对应像素的相似性,该方法可实现一秒内计算准确结果,并且将匹配任务转化为多分类问题,类别是所有可能的视差。文献[6]结合残差网络^[7](ResNet)提出一种高速网络来计算匹配代价,并且设计一个全局视差网络预测视差置信度得分,进一步优化了视差图。

2) 基于CNN的视差回归方法把视差图获取作为立体匹配的核心任务,其质量的优劣对三维信息重建有直接影响,因此,目前存在较多侧重于视差图后处理方法的研究^[8-9]。文献[10]提出经典的Displets网络,使用3D卷积并通过物体识别和语义分割处理图像对间的对应关系,大幅提高了匹配精度。文献[11]提出检测-替换-优化(Detect-Replace-Refine, DRR)结构进行视差估计,通过标签检测、错误标签替换、新标签优化来改善分类效果。文献[12]提出半全局匹配网络SGM-Net,

通过训练网络学习预测SGM的惩罚项,以取代人工调整方法。

3) 基于CNN的端到端视差图获取方法不对端到端网络进行复杂后处理,而是输入左右图像对网络直接学习输出视差图。文献[13]提出端到端DispNet网络结构结合光流估计进行立体匹配任务学习的方法,并且合成一个大型立体图像数据集SceneFlow用于网络训练。此后,文献[14]在DispNet的基础上提出了级联残差学习(CRL)网络。该网络分为DispFullNet和DispResNet两部分,前者输出初始视差图,后者通过计算多尺度残差优化初始视差图,最后将两部分网络的输出结合构成最终的视差图。文献[15]提出了效果较好的GC-Net结构,其利用3D卷积获得更多的上下文信息,实现了亚像素级别的视差估计。为获得图像更多的关键信息,有学者采用深度特征融合和多尺度提取图形信息的卷积神经网络^[16-17]。文献[18]利用图像语义分割思想,同时结合全局环境信息提出了PSMNet结构,其主要由两个模块组成:金字塔池化和3D卷积神经网络。金字塔池化模块通过空间金字塔池化^[19]和空洞卷积^[20]聚合不同尺度和位置的环境信息构建匹配代价卷。3D卷积神经网络模块通过将多个堆叠的沙漏网络与中间监督相结合来调整匹配代价卷,同时提高对全局信息的利用率。

上述方法都是利用深而复杂的3D卷积网络架构进行端到端的视差图学习,通常都采用Siamese网络作为立体图像对的特征提取器。由此可见,立体匹配本质上依赖于使用Siamese网络来提取左右图像特征信息。文献[21]回避了复杂的端到端CNN框架,而是对基础的Siamese网络结构进行改进,其在文献[5]网络中引入普通池化和反卷积操作,利用池化操作扩大了网络的感受野,为后续匹配提供了更多的视觉线索,并使用反卷积操作恢复图像分辨率。

本文改进基础Siamese特征提取网络,通过引入空间金字塔池化(SPP)思想优化文献[5]方法原有的池化方式,构建非对称金字塔池化模型,在特征提取阶段对图像块进行多尺度特征提取,进而提高匹配精度。

2 非对称金字塔网络结构

如前所述,本文所提出的网络架构仍采用图像块作为输入,左右图像块的像素大小遵循先前工作,输出为图像块的相似性。本节首先介绍空间金字塔池化模型,在此基础上改进池化方式,提出非对称空间金字塔池化模型,最后详述整个网络架构及其他改进细节。

2.1 空间金字塔池化模型

空间金字塔池化^[19](SPP)对前一卷积层输出的特征图进行不同尺寸池化操作,得到不同分辨率的特征信息,从而有效提高网络对特征的识别精度。文献[19]给出了详细的池化过程:首先使用3种不同刻度的窗口对特征图像进行划分,每一种刻度代表金字塔的一层,划分后的每一个图像块大小为window_size;然后对每一

个图像块都采取最大池化操作,提取出更高级的图像特征信息。图像划分计算公式如下:

$$\text{win_size} = \lceil a/n \rceil \quad (1)$$

$$\text{str_size} = \lfloor a/n \rfloor \quad (2)$$

其中,win_size 和 str_size 分别表示池化窗口和池化步幅的大小, a 表示金字塔池化层输入的特征图尺寸为 $a \times a$, n 表示金字塔池化层输出的特征图尺寸为 $n \times n$, $\lceil \cdot \rceil$ 表示向上取整, $\lfloor \cdot \rfloor$ 表示向下取整。

如图 1 所示,对左输入图像进行金字塔池化操作:若前一卷积层输出为 13×13 大小的特征图,池化之后的特征图尺寸为 2×2 ,则 $\text{win_size} = \lceil 13/2 \rceil$, $\text{str_size} = \lfloor 13/2 \rfloor$,结果分别为 7,6,即采用窗口为 7、步幅为 6 池化操作即可得到 2×2 尺寸的特征图。

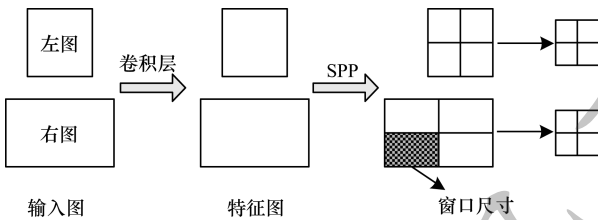


图 1 原金字塔池化过程

Fig.1 Process of primary pyramid pooling

如果对右输入图像采用相同刻度的池化操作,SPP 模型也会将右图按照刻度划分,最终输出与左图相同尺寸的特征图像。但实际右输入图像块尺寸远大于左图像块,这表示右图像块包含更多的特征信息,若采用相同尺度的池化方式,将直接导致右原始图像丢失大部分特征信息,最终降低网络视差估计的精度,影响网络的匹配效果。

2.2 非对称空间金字塔池化模型

针对上述 SPP 模型存在的图像特征信息缺失问题,本文改变对右图像块的划分方式,重新定义金字塔池化公式,如式(3)~式(5)所示:

$$k = \lceil |l - a|/n \rceil \quad (3)$$

$$\begin{cases} \text{win_h} = \lceil a/n \rceil \\ \text{win_w} = \lceil l/(n+k) \rceil \end{cases} \quad (4)$$

$$\begin{cases} \text{str_h} = \lfloor a/n \rfloor \\ \text{str_w} = \lfloor l/(n+k) \rfloor \end{cases} \quad (5)$$

$$\begin{cases} \text{str_h} = \lfloor a/n \rfloor \\ \text{str_w} = \lfloor l/(n+k) \rfloor \end{cases}$$

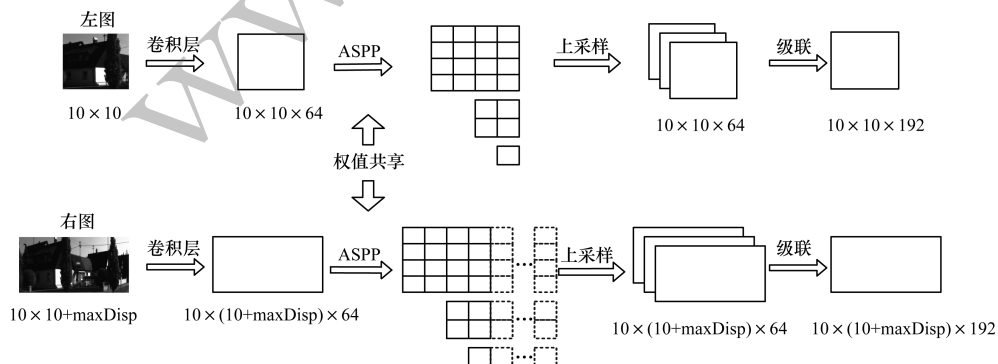


图 3 非对称空间金字塔池化模型

Fig.3 Asymmetric spatial pyramid pooling model

其中, a 、 l 表示金字塔池化层的输入特征图尺寸, n 表示金字塔池化层输出的特征图尺寸,win_h 和 win_w 表示池化窗口高度和宽度,str_h 和 str_w 表示两个维度上池化步幅的大小。

改进后的金字塔池化过程如下:首先按照金字塔池化原理对右图多于左图的部分进行初步划分,得到初始划分图像块大小 k ;然后以整幅图像为池化层的输入,在图像高度的维度上仍采用原金字塔划分刻度,大小为 n ,但在图像宽度的维度上采用新的划分刻度 $(n+k)$ 。改进后的池化过程如图 2 所示。

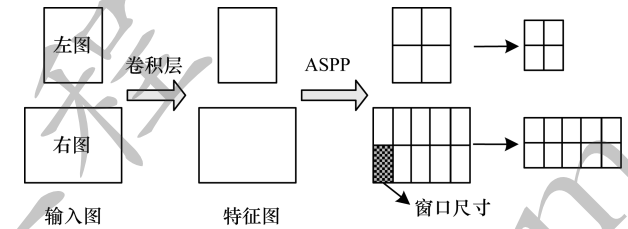


图 2 改进的金字塔池化过程

Fig.2 Process of improved pyramid pooling

改进后的池化模型对左右特征图像采用不同刻度的划分方式,针对右图实现了两个维度上不同尺寸的池化,右图被划分为更均匀的图像块,进行后续最大池化操作时能够得到更详细的特征信息。由于从整个池化结构上看,左右分支呈现非对称性,因此将本文改进后的池化操作称为非对称空间金字塔池化(ASPP)。

如图 3 所示,ASPP 网络的左右分支分别以不同尺寸的图像块作为输入,左、右输入分别为 10×10 、 $10 \times (10 + \text{maxDisp})$ 的图像块(maxDisp 表示最大视差值)。输入图像经卷积层输出通道数为 64 的特征图像,图像尺寸不变。然后经过非对称金字塔池化,池化尺寸分别为 4×4 、 2×2 、 1×1 ,对应生成 3 种不同尺度和精度的特征图像,通道数均为 64。之后,通过上采样将 3 种特征图恢复为原始图像尺寸,最后采用级联操作将其连接得到通道数为 $192(64 \times 3)$ 的特征图像。本文将池化及后续上采样、级联步骤归一化称为非对称池化模块。此外,整个网络模型左右分支共享权重。

2.3 改进的网络结构

为充分利用图像特征信息, 本文在 LuoNet 网络基础上引入金字塔池化层, 扩大目标像素位置的感受野, 获取更多的图像信息, LuoNet 网络结构如图 4(a) 所示。由于原有的金字塔池化操作会造成部分图像信息缺失问题, 因此本文改进了网络结构, 如图 4(b) 所示, 使 ASPP 模型可以提取到更多的特征信息, 达到更精确的匹配效果。

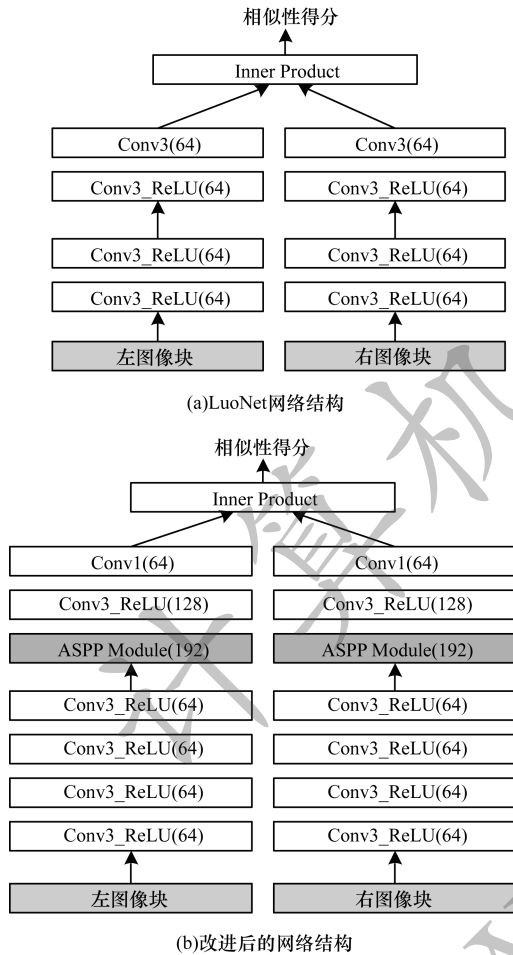


图 4 改进前后网络结构对比

Fig.4 Comparison of network structure before and after improvement

本文网络结构设计流程如下:

- 1) 图像先经过 4 个卷积层 (Conv)、批标准化和线性整流函数 (ReLU)。
- 2) 非对称池化模块 (ASPP) 对特征图像进行不同刻度的池化操作。
- 3) 再次经过两层卷积操作 (Conv) 融合特征信息。
- 4) Inner Product 层利用内积计算每个视差下左右特征的相似性得分。
- 5) Softmax 函数计算每个视差的概率值, 形成多分类网络模型。数字 3 和 1 分别代表卷积核的大小为 3×3 、 1×1 , 64、128、192 代表卷积核通道数。

此外说明, 本文网络结构中所有卷积层和池化层中的步长均为 1。

本文主要有两处结构上的改进: 基于改进的金字塔池化模型提出非对称金字塔网络结构, 网络通过多尺度提取图像特征信息改善了匹配效果; 将网络深度由 4 层加深至 7 层, 提升了网络匹配精度, 并且额外的卷积层使用较小卷积核, 减少了特征维度, 加快了网络收敛速度。

网络将不同尺寸的左右图像块作为输入, 首先利用 4 层卷积神经网络层进行特征提取, 每个卷积层对图像做卷积核为 3×3 、通道数均为 64 的空间卷积; 然后经过空间批标准化和 ReLU 层; 之后非对称池化模块将特征图像压缩到 3 个不同尺度中, 得到多尺度特征图像; 最后通过双线性插值将特征图像上采样至原始图像分辨率, 并利用级联操作进行连接输出通道数为 192 的特征图。由于经过非对称金字塔池化后的特征图像含有不同精度的特征信息, 因此本文叠加了两层额外的卷积层进行特征信息融合。为加快网络的收敛速度, 额外的两个卷积层的卷积核尺寸分别设置为 3×3 和 1×1 , 其通道数分别为 128 和 64。为保留以负值编码的特征信息, 同样删除了最后卷积层的 ReLU 线性激活函数。内积层以一种简单的方式计算特征之间的内积, 并将其作为不同像素间的匹配代价。最后通过 Softmax 函数进行视差分类预测, 类别是所有可能的视差值。本文网络结构中所有卷积层和池化层中的步长均为 1。

3 实验与评估

在本文实验中, 所有模型的训练、验证及预测均采用 Tensorflow 深度学习框架, 操作系统为 ubuntu18.04, 网络训练及性能测试的 GPU 服务器为 NVIDIA GeForce GTX 1060Ti。实验使用 KITTI 2012^[22]、KITTI 2015^[23] 数据集及 Middlebury 测试集上进行模型训练和视差评估。

3.1 实验过程

KITTI2012 数据集包含 194 对带有视差真值的左右图像, 随机选取 160 对图像作为模型训练集, 剩余 34 对图像作为模型验证集。KITTI2015 数据集包含 200 对左右图像, 同样从中随机选取 160 对图像训练模型, 剩余 40 对图像用于模型验证。每对图像大小为 375 像素 \times 1 241 像素。使用 Middlebury 测试集提供的 Cones 测试图像对、Teddy 测试图像对和 Tsukuba 测试图像对来评估视差图, 每对图像都含有视差真值, 大小为 375 像素 \times 450 像素。在模型训练之前, 对所有立体图像均进行预处理操作, 分别通过减去平均值、除以像素强度标准差的方式将图像归一化为零均值和单位标准差的图像。

实验步骤与文献[5]中一致,首先从左右立体图像数据集中随机提取图像块,将提取出的右图像块以最大视差值进行拓展。如前所述,分别使用尺寸为 4×4 、 2×2 、 1×1 的池化操作。网络训练时图像块尺寸设置为 10×10 ,最大视差值设置为128,验证时KITTI数据集上的图像块均随机截取275像素 \times 640像素,Middlebury测试集的图像对大小仍为375像素 \times 450像素。

网络所有参数使用He初始化^[24]随机初始化,后续采用随机梯度下降(SGD)算法进行参数更新,网络损失函数采用softmax交叉熵损失函数,并采用AdamOptimizer算法^[25]优化损失函数。本文以初始学习率 $1e^{-3}$ 进行网络训练,每迭代400次对学习率进行一次更新,逐步减小学习率直至模型训练稳定。本文设计批处理尺寸为128,网络总迭代次数为40k次,学习率由 $1e^{-3}$ 降至 $1e^{-4}$ 。

3.2 超参数分析

对算法性能产生影响的超参数分析实验,第1组是对2.1节中的 k 值进行影响分析,第2组是额外卷积层参数对比分析,参数为卷积核的尺寸。两组实验均使用KITTI2015数据集。

1) k 值影响分析

本文通过引入 k 值对文献[19]金字塔池化方式进行改进,因此,对于 k 值的分析以文献[19]为基准算法进行直接对比。当网络池化刻度为 $\{4, 2, 1\}$ 时,由式(1)、式(2)计算出经金字塔池化后的输出尺寸分别为 $\{(4 \times 4), (2 \times 2), (1 \times 1)\}$;采用非对称金字塔池化,则由式(3)~式(5)进行相关计算,首先是对应参数 k 为 $\{32, 64, 128\}$,其对应输出的特征图尺寸分别为 $\{(4 \times 46), (2 \times 28), (1 \times 14)\}$,无疑包含更多的图像信息。

本文实验采用图4(a)所示的网络结构(在最初的两个卷积层之后引入不同的池化结构),两组实验不同之处仅在于池化方式不同,验证数据均使用KITTI2012验证集。以网络匹配率为衡量标准,比较结果如表1所示。表中数据指标为视差值与基准视差差距分别大于2个、3个和5个像素的像素点比例,其中加粗数据表示最优数据。实验结果表明,在同等网络结构设置下,本文方法较对比方法在精度上均有提升,并且取得了最低误匹配率,这证明了改进网络的匹配效果优于其他模型。

表1 k 值对匹配误差的影响

Table 1 Influence of k value on matching error

方法	视差值与基准视差的差距范围/%			运行时间/s
	大于2个 像素	大于3个 像素	大于5个 像素	
LuoNet	10.87	8.61	7.00	0.14
SPPNet	7.65	6.06	4.46	1.06
本文方法	7.22	5.63	3.65	1.47

2) 卷积层参数影响分析

在CNN中卷积核的尺寸对模型训练至关重要,较大的卷积核通常可以带来较大感受野,从而获取到更多的图像信息,但同时较大卷积核会导致昂贵的计算成本,降低模型计算性能。所以,使用更小的卷积核是当前保证模型精度的情况下,提升网络训练速度的一种方式。因此,实验对于图4(b)中额外的卷积层进行分析对比。实验分为3组,分别将卷积核尺寸设置为 $\{1 \times 1, 1 \times 1\}$ 、 $\{3 \times 3, 3 \times 3\}$ 、 $\{3 \times 3, 1 \times 1\}$,通道数不变,均在KITTI2012训练集上进行实验,实验结果包括网络错误率和网络损失函数值,如图5和图6所示。可以看出:卷积核 $\{1 \times 1, 1 \times 1\}$ 的卷积层因为含有最少的参数,其损失函数收敛速度最快,但错误率最高; $\{3 \times 3, 3 \times 3\}$ 的卷积层在网络训练初期误匹配率较低,但随着训练时间的增加,其误匹配率降低缓慢,并且最终的损失函数值最高;卷积核为 $\{3 \times 3, 1 \times 1\}$ 的卷积层错误率最低,损失函数值收敛到最小,并且随着训练时间和迭代次数的增加,其错误率和损失函数值持续减小。因此,改进算法采用 $\{3 \times 3, 1 \times 1\}$ 的卷积核。

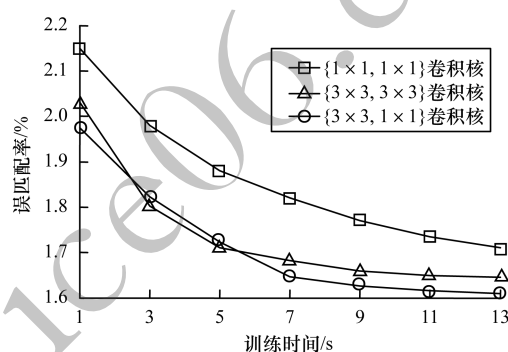


图5 不同卷积核尺寸下的误匹配率曲线

Fig. 5 Error matching rate curves under different convolutional kernel sizes

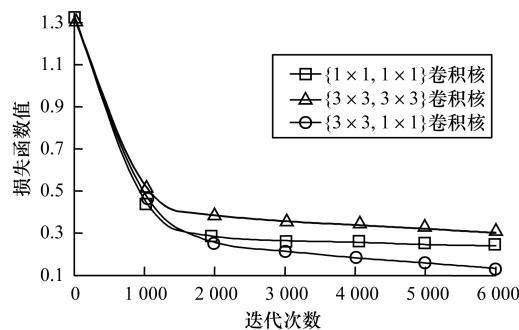


图6 不同卷积核尺寸下的损失函数曲线

Fig. 6 Loss function curves under different convolutional kernel sizes

3.3 实验结果评估

使用本文方法在KITTI2012和KITTI2015验证集上进行视差估计,并分别与文献[5,21]方法进行误差对比分析,阈值依次设置为2个、3个和5个像素,实验结

果如表 2 和表 3 所示, 其中加粗数据表示最优数据。可以看出, 本文方法的错误率在各项指标上均小于对比方法, 尤其是像素阈值为 3、5 时的网络匹配错误率分别从 5.84%、4.80% 降低至 3.88%、2.94%。虽然相比原始方法 LuoNet 较为耗时, 但比 SPPNet 在时间上提升了约 50%。相比于数据集 KITTI2012, 其在数据集 KITTI2015 上错误率略高, 但是仍低于两种对比方法。由此可以证明, 本文方法在精度和速度上都具有一定优势。

表 2 在 KITTI2012 数据集上的测试结果对比

Table 2 Comparison of test results in KITTI2012 data set

方法	视差值与基准视差的差距范围/%			运行时间/s
	大于 2 个 像素	大于 3 个 像素	大于 5 个 像素	
LuoNet	10.87	8.61	7.00	0.14
SPPNet	6.65	5.84	4.80	5.27
本文方法	4.78	3.88	2.94	2.63

表 3 在 KITTI2015 数据集上的测试结果对比

Table 3 Comparison of test results in KITTI2015 data set

方法	视差值与基准视差的差距范围/%			运行时间/s
	大于 2 个 像素	大于 3 个 像素	大于 5 个 像素	
LuoNet	9.96	7.23	5.04	0.34
SPPNet	6.79	5.92	4.92	5.27
本文方法	4.90	4.31	3.12	3.50

本文方法与经典 MC-CNN^[2] 匹配方法的比较如表 4 所示, 以 2 个像素的像素点比例作为阈值, 表中加粗数据表示最优数据, 结果表明, 本文方法较 MC-CNN-acrt^[2] 误差降低了约 14%, 较 MC-CNN-fast^[2] 方法降低了约 11%, 本文网络结构在 KITTI2012 和 KITTI2015 两个数据集上均实现了较低的错误率。

表 4 4 种方法的视差结果误差对比

Table 4 Comparison of disparity results errors of four methods %

方法	KITTI2012 数据集	KITTI2015 数据集
MC-CNN-fast	17.71	18.47
MC-CNN-acrt	15.02	15.20
LuoNet	9.96	7.23
本文方法	4.90	4.31

在 KITTI2012 和 KITTI2015 数据集及 Middlebury 测试集上进行视差图预测, 输出结果如图 7~图 9 所示。图 7 和图 8 中列出 2 组视差效果对比图, 图 9 列出 3 组对比结果。特别说明, 在本文采用的文献[5]方法中, 视差图未经过视差后处理, 本文采取同样的处理方式, 因此, 输出视差图中含有一些噪音, 但并不影响实验结果分析。

从图 7~图 9 可以看出, 本文提出的非对称金字塔池化网络获得的视差效果更平滑, 尤其在 Middlebury 测试集中, 所得视差图不仅含有较少的噪声, 而且能够预测到图像物体及图像边缘的视差。例如图 7 中汽车边缘等细节区域、图 9 中 Cones 测试图像(左列)物体的边缘和 Tsukuba 测试图像(右列)台灯的边缘, 本文方法输出的视差图效果均优于 LuoNet。图 8 的匹配效果虽然在某些细节方面不够理想, 目标物轮廓较为模糊, 但是在没有进行任何后处理的情况下, 完整保留了目标物的整体信息, 即使视差图中有些区域含有较多的噪声, 但本文方法仍能完整地保留不同场景的目标像素信息, 并且取得更好的视差效果。

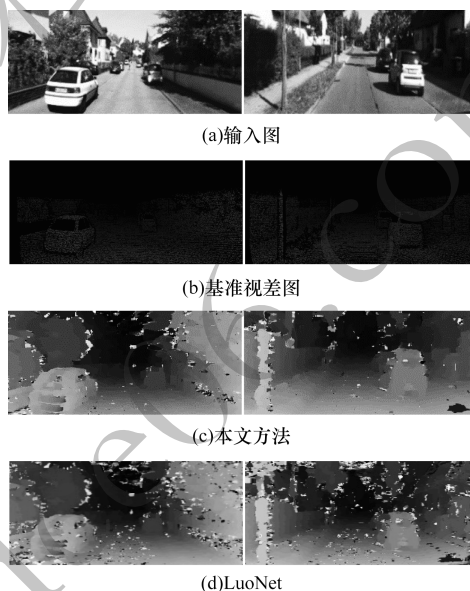


图 7 KITTI2012 数据集视差图

Fig. 7 Disparity maps of KITTI2012 data set

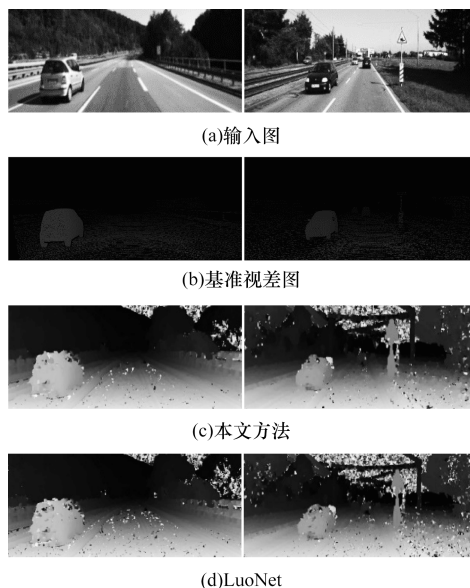


图 8 KITTI2015 数据集视差图

Fig. 8 Disparity maps of KITTI2015 data set

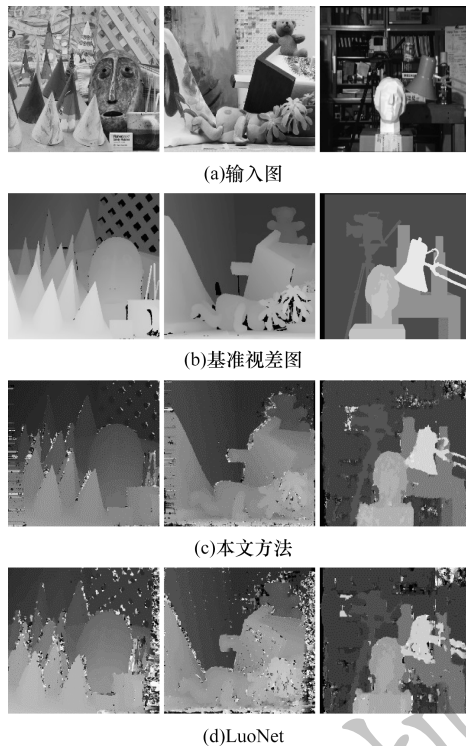


图 9 Middlebury 数据集视差图

Fig.9 Disparity maps of Middlebury data set

4 结束语

针对卷积神经网络用于立体匹配时的耗时和信息损失问题,本文结合现有的卷积神经网络架构及池化结构,对金字塔池化方法进行改进,提出非对称金字塔池化模型,并且使用较小的卷积核增加网络深度。实验结果表明,与 LuoNet 网络结构相比,改进后的网络结构不仅能够加快网络训练收敛速度,而且能够进行多级特征信息提取,提升了网络匹配精度,改善了最终视差效果。但本文设计未使用视差后处理操作,网络处理对象仍是小尺寸图像块。下一步将研究如何平衡输入图像块尺寸与视差效果,并利用残差网络多级信息融合技术对视差图进行优化。

参考文献

- [1] 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.
- [2] ŽBONTAR J, LECUN Y. Stereo matching by training a convolutional neural network to compare image patches [J]. Journal of Machine Learning Research, 2016, 17(1): 1-32.
- [3] XIAO Jinsheng, TIAN Hong, ZOU Wentao, et al. Stereo matching based on convolutional neural network [J]. Acta Optica Sinica, 2018, 38(8): 179-185. (in Chinese) 肖进胜, 田红, 邹文涛, 等. 基于深度卷积神经网络的双目立体视觉匹配算法 [J]. 光学学报, 2018, 38(8): 179-185.
- [4] ZAGORUYKO S, KOMODAKIS N. Learning to compare image patches via convolutional neural networks [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 4353-4361.
- [5] LUO W, SCHWING A G, URTASUN R. Efficient deep learning for stereo matching [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 5695-5703.
- [6] SHAKED A, WOLF L. Improved stereo matching with constant highway networks and reflective confidence learning [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 1-5.
- [7] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.
- [8] WANG Yufeng, WANG Hongwei, YU Guang, et al. Stereo matching algorithm based on three-dimensional convolutional neural network [J]. Acta Optica Sinica, 2019, 39(11): 1-8. (in Chinese) 王玉峰, 王宏伟, 于光, 等. 基于三维卷积神经网络的立体匹配算法 [J]. 光学学报, 2019, 39(11): 1-8.
- [9] WANG An, WANG Fangrong, GUO Baicang, et al. Disparity map optimization based on edge detection [J]. Computer Applications and Software, 2019, 36(7): 236-241. (in Chinese) 王安, 王芳荣, 郭柏苍, 等. 基于边缘检测的视差图效果优化 [J]. 计算机应用与软件, 2019, 36(7): 236-241.
- [10] GUNAY F, GEIGER A. Displets: resolving stereo ambiguities using object knowledge [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 1-5.
- [11] GIDARIS S, KOMODAKIS N. Detect, replace, refine: deep structured prediction for pixel wise labeling [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 1-5.
- [12] SEKI A, POLLEFEYS M. SGM-nets: semi-global matching with neural networks [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 1-5.
- [13] MAYER N, ILG E, HÄUSSER P, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 4040-4048.
- [14] PANG J H, SUN W X, REN J S J, et al. Cascade residual learning: a two-stage convolutional neural network for stereo matching [C]// Proceedings of IEEE International Conference on Computer Vision-Workshop on Geometry Meets Deep Learning. Washington D. C., USA: IEEE Press, 2017: 887-895.
- [15] KENDALL A, MARTIROSYAN H, DASGUPTA S, et al. End-to-end learning of geometry and context for deep stereo regression [C]// Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 66-75.

(上接第 234 页)

- [16] YUN Weiguo, SHI Qiqi, WANG Min. Multi-feature fusion gesture recognition based on deep convolutional neural network[J]. Chinese Journal of Liquid Crystals and Displays, 2019, 34(4): 417-422. (in Chinese)
俞卫国, 史其琦, 王民. 基于深度卷积神经网络的多特征融合的手势识别[J]. 液晶与显示, 2019, 34(4): 417-422.
- [17] XI Lu, LU Jixiang, TU Ting. Stereo matching method based on multi-scale CNN[J]. Computer Engineering and Design, 2018, 39(9): 2918-2922. (in Chinese)
习路, 陆济湘, 涂婷. 基于多尺度卷积神经网络的立体匹配方法[J]. 计算机工程与设计, 2018, 39(9): 2918-2922.
- [18] CHANG J R, CHEN Y S. Pyramid stereo matching network[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 1-6.
- [19] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [20] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [21] BRANDAO P, MAZOMENOS E, STOYANOV D. Widening Siamese architectures for stereo matching [EB/OL]. [2019-05-25]. <https://arxiv.org/pdf/1711.00499.pdf>.
- [22] URTASUN R, LENZ P, GEIGER A. Are we ready for autonomous driving? the kitti vision benchmark suite[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D.C., USA: IEEE Press, 2012: 1-5.
- [23] MENZE M, GEIGER A. Object scene flow for autonomous vehicles[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 1-5.
- [24] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification [C]// Proceedings of 2015 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 1-5.
- [25] KINGMA D P, BA J. Adam: a method for stochastic optimization[C]// Proceedings of International Conference on Learning Representations. Banff, Canada: [s. n.], 2015: 1-15.

编辑 金胡考