

# 基于改进卷积神经网络的稠密视差图提取方法

黄东振<sup>1,2</sup>, 赵沁<sup>1,2</sup>, 刘华巍<sup>1</sup>, 李宝清<sup>1</sup>, 袁晓兵<sup>1\*</sup>

<sup>1</sup>中国科学院上海微系统与信息技术研究所微系统技术重点实验室, 上海 201800;

<sup>2</sup>中国科学院大学, 北京 100049

**摘要** 针对现有的卷积神经网络方法所生成的视差图中细节损失严重的问题,提出了在结构上改进的新方法。将原有网络中特征提取部分的4层卷积结构提升到7层,最大化提高了精度;在网络中引入了双金字塔结构,将多尺度降采样信息和特征信息进行了融合,保持了输入图像中的原始细节信息。实验结果表明,改进后网络的错误率从3.029%降到了2.795%,生成的视差图具有更好的连通性。

**关键词** 图像处理; 视差获取; 深度学习; 卷积神经网络; 立体匹配

中图分类号 TN911.73

文献标识码 A

doi: 10.3788/LOP55.121004

## Dense Disparity Map Extraction Method Based on Improved Convolutional Neural Network

Huang Dongzhen<sup>1,2</sup>, Zhao Qin<sup>1,2</sup>, Liu Huawei<sup>1</sup>, Li Baoqing<sup>1</sup>, Yuan Xiaobing<sup>1\*</sup>

<sup>1</sup>Key Laboratory of Microsystem Technology, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 201800, China;

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract** According to the problem of the severe detail loss of the disparity map generated by the current convolutional neural network methods, a structural improvement method is proposed. The 4 layers convolutional structure of the feature extraction part from original network is added to 7 layers to maximize the accuracy. And, the proposed dual pyramid structure is introduced to the network to combine the multi-scale down-sampling information with the feature information, which keeps the details of the original input images. Experimental results show that the error rate of the improved network reduces from 3.029% to 2.795%, and the generated disparity maps have better connectivity.

**Key words** image processing; disparity acquisition; deep learning; convolutional neural network; stereo matching

**OCIS codes** 100.6890

## 1 引言

近年来,随着立体视觉算法的迅速发展,基于立体视觉算法进行三维重构的各种应用已经走进人们的视线,比如无人机勘测、无人驾驶汽车等。王琪龙等<sup>[1]</sup>提出结合双目视觉与激光测距的目标跟踪系统,解决了时滞问题,提高了算法的实时性。大多数双目立体视觉算法均是围绕视差展开。视差是指物体在左右视图中像素坐标水平方向的差值。根据简单的三角测量原理,可以由物体的视差结合像素焦

距、基线距离得到其深度,进而通过三角形相似获取该物体完整的三维空间位置信息。提取稠密准确的视差图是三维重构的关键,视差获取则需要物体在左右视图中的对应关系。

通过物体匹配获取视差的方法最早由Scharstein等<sup>[2]</sup>进行总结归纳,基本分为4个步骤:匹配代价计算、代价聚合、视差计算以及视差细化。Boykov等<sup>[3]</sup>将视差获取问题转换成全局能量最小化问题,通过建立网络图,基于图论方法求解从而得到视差图。Hirschmuller<sup>[4]</sup>基于匹配像素之间的互

收稿日期: 2018-05-04; 修回日期: 2018-06-04; 录用日期: 2018-06-08

基金项目: 微系统技术国防科技重点实验室基金(CXJJ-17S072)

\* E-mail: sinowsn@mail.sim.ac.cn

信息量提出了半全局匹配(SGM)算法,通过将目标像素以及周围区域像素纳入代价计算,在速度远快于全局匹配算法的同时取得了优异的匹配效果。Lowe<sup>[5]</sup>从局部特征提取、特征描述、特征比较的角度,提出了尺度不变特征变换(SIFT)匹配方法。Bay等<sup>[6]</sup>改进了SIFT中的特征提取和描述,提高了算法实现效率。张鑫等<sup>[7]</sup>对于SIFT中特征描述进行降维,特征匹配加入方向约束,提高了匹配准确率。祝世平等<sup>[8]</sup>基于改进Census变化,在视差选择阶段采用改进的动态规划算法,消除了扫描线效应,提高了匹配速度和正确率。

近年来,随着深度学习<sup>[9]</sup>的迅速发展,卷积神经网络在众多图像处理领域取得了优异的效果,其中也包括立体视觉领域。Zagoruyko等<sup>[10]</sup>提出了双通道输入的卷积神经网络架构,实现了宽基线条件下的特征匹配。Žbontar等<sup>[11]</sup>基于匹配代价卷积神经网络(MC-CNN)的方法,在后处理上结合了众多算法,如垂直交叉匹配代价聚类(CBCA)<sup>[12]</sup>、半全局匹配(SGM)<sup>[4]</sup>、中值滤波以及引导滤波<sup>[13]</sup>,并在实现上借鉴了绝对差统计(AD-Census)<sup>[14]</sup>方法,达到了前沿的水准。三维最小生成树(3DMST)<sup>[15]</sup>、块匹配超像素分割(PMSC)<sup>[16]</sup>等方法均是在MC-CNN的基础上进行后处理算法的改进算法。MC-CNN分为精准架构MC-CNN-acrt和快速架构MC-CNN-fst两种,其中MC-CNN-acrt的时耗为MC-CNN-fst的百倍以上。宽窗卷积神经网络(LW-CNN)<sup>[17]</sup>则是基于MC-CNN-acrt架构进行改进的,在网络中加入了池化结构以及 $1 \times 1$ 卷积层,实现了更大的匹配窗口覆盖,使得精度进一步提高。

前沿的MC-CNN方法得到的视差数据仍然较为离散,结果图中的许多物体失去了连通性。本文结合对传统特征提取匹配算法的理解,对MC-CNN中的卷积神经网络部分给予了具有实际意义的解释,并结合前沿的网络架构以及传统特征提取算法中的尺度空间理论,对其进行了结构上的改进,主要包括网络的加深以及双金字塔结构的引入。本文基于MC-CNN-fst架构进行改进,在时耗少量增加的代价下提升了精度,获得了连通性更加完整的视差图。

## 2 改进的卷积神经网络模型

介绍了所提出的双金字塔网络结构,随后引入改进网络的其他细节。本文中所有的网络模型输入为两个大小为 $9 \times 9$ 的相关图像块,输出为它们的匹维度)为64,与MC-CNN-fst相同,最终输出 $64 \times 1 \times 1$

配代价。

### 2.1 双金字塔网络结构

MC-CNN网络架构本身可以分为三个部分:卷积层部分、连接部分和全连接层部分。网络的功能实现类似特征提取匹配算法中的三个步骤:特征提取、特征描述和特征匹配。MC-CNN中的卷积层实现了特征提取和特征描述;连接部分对所得到的左右视图图像块的特征描述向量进行合并;全连接层部分实现了特征匹配,最终输出匹配代价。不同于LW-CNN<sup>[17]</sup>对全连接层部分进行改进,本文基于卷积层部分提出一种新型结构。

MC-CNN卷积层部分中采用4层卷积核为 $3 \times 3$ 、步长为1、无补零的卷积层,前三层卷积层后连接着线性整流函数(ReLU)<sup>[18]</sup>。输入图像块大小为 $9 \times 9$ ,4层卷积层依次输出像素大小分别为 $7 \times 7$ 、 $5 \times 5$ 、 $3 \times 3$ 、 $1 \times 1$ 的特征图。依据尺度空间理论,可以通过降采样得到对不同尺度下保留原始信息的图像。公路网络<sup>[19]</sup>中对于网络中的一层进行了重新定义:

$$y = H(x, \mathbf{W}_H) \cdot T(x, \mathbf{W}_T) + x \cdot C(x, \mathbf{W}_C), \quad (1)$$

式中 $x$ 与 $y$ 分别为该层网络的输入和输出, $H$ 表示该层网络的非线性变换,对应参数为 $\mathbf{W}_H$ , $T$ 与 $C$ 是公路网络<sup>[19]</sup>添加的非线性变换,相应参数为 $\mathbf{W}_T$ 与 $\mathbf{W}_C$ ,用于控制非线性变换和恒等映射之间的权重分配。

令 $T(x, \mathbf{W}_T) = 1$ , $H(x, \mathbf{W}_H)$ 替换为MC-CNN卷积层中一层的非线性变换 $H^*(x, \mathbf{W}_H^*)$ ;同时将 $x \cdot C(x, \mathbf{W}_C)$ 替换为降采样 $d(x)$ 。为了不引入另外的参数,这里降采样 $d(x)$ 通过窗口大小为 $3 \times 3$ 、步长为1、无补零的平均池化实现,从而得到保留有原始输入信息的降采样图像。基于(1)式,每一层网络可以表示为

$$y^* = H^*(x, \mathbf{W}_H^*) + d(x). \quad (2)$$

改进后的卷积层部分网络架构如图1所示,其中AP代表平均池化,conv代表卷积层,在前三层的卷积层后省略了ReLU<sup>[18]</sup>。输入为 $9 \times 9$ 的图像块,经过平均池化和卷积层分别得到 $7 \times 7$ 的降采样图像和 $7 \times 7$ 的特征图,然后将降采样图像和特征图按像素强度逐像素相加,得到的结果作为该层的输出。随后三层操作相同,最后得到大小为 $1 \times 1$ 的特征描述。

本文所提出的结构选择卷积核通道数(特征图的特征描述。从结构上可以看出,图像尺度的变化类

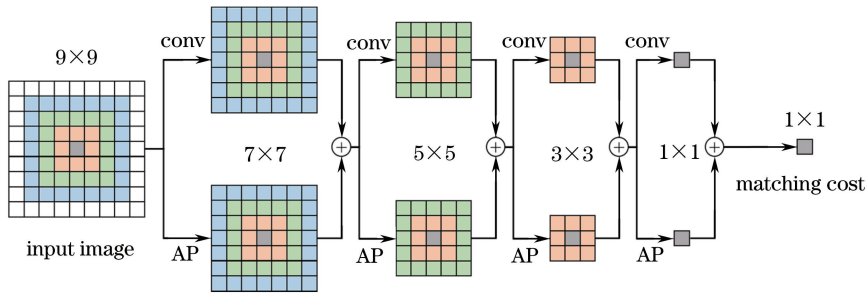


图1 双金字塔网络结构

Fig. 1 Dual pyramid network structure

似于图像金字塔,本文将这种结构称为双金字塔结构。

## 2.2 改进网络结构

基于双金字塔网络结构,提出了如图2所示的网络结构,其中 conv、ReLU、identity、average pooling、normalize 分别代表卷积、线性整流、恒等映射、平均池化以及归一化, stereo join 对左右视图归一化之后的特征描述向量求点积,得到的结果作为网络输出的匹配代价。MC-CNN-fst 架构中 stereo join 取代了全连接层的作用。方括号中为卷积核通道数,也是特征图数量;  $3 \times 3$  和  $5 \times 5$  分别指卷积核大小以及池化窗口大小;注明 pad 的网络层为补零的网络结构,此外均为无补零网络结构。所提网络中所有卷积层和池化结构中的步长均为1,卷积层中所有卷积核通道数为64。

本文所提出的网络结构相对于 MC-CNN-fst 架构主要具有两处改进:1) 将网络深度由4层(卷积层)提到7层高,提高了精度;2) 引入了双金字塔结构,保持了图像的连通性。当直接将  $9 \times 9$  的图像块输入双金字塔结构时,平均池化会在  $7 \times 7$  的尺度空间产生大量相同的降采样图像,使得网络中的信息大量冗余而降低效率。因此,在输入图像和双金字塔结构之间引入64通道卷积核的卷积层进行特征图维度的提升。对于额外增加卷积层,在  $9 \times 9$  的图像块特征提取过程中:采用  $1 \times 1$  卷积核则无法提取卷积中心与周围像素的相关信息;采用  $5 \times 5$  卷积核则将输入图像块补零至  $13 \times 13$ ,在原始81 pixel的基础上额外引入了88 pixel的零值。而  $3 \times 3$  卷积核在考虑到与周围像素的相关性的同时尽量少地引入零值。将其与  $1 \times 1$  和  $5 \times 5$  卷积核进行对比,结果如图3(a)所示,实线表示预测误差,虚线表示训练误差。额外卷积层的层数选择方式为添加至过拟合,如图3(b)所示。考虑到网络加深会增加时耗,本文最终引入层数为三层、卷积核大小为  $3 \times 3$  的卷积层进

行升维。

He等<sup>[20]</sup>提出残差结构,在50层以下网络中将(1)式中的  $H(x, \mathbf{W}_H)$  替换为  $F = \mathbf{W}_2 \sigma(\mathbf{W}_1 x)$ ,其中  $\mathbf{W}_1$ 、 $\mathbf{W}_2$  为两层卷积层的权值参数,简洁起见,省略了偏置参数,  $\sigma$  表示激活函数 ReLU<sup>[18]</sup>,并在每个激活函数前添加批量归一化层<sup>[21]</sup>。受该结构启发,本文将双金字塔结构进一步简化为每两个尺度空间取一次降采样,即采用  $5 \times 5$  的平均池化。网络被增加至7层后,考虑到第一层卷积层用于特征图维度的变化,改进网络在第二层和第三层前后引入恒等映射,达到保留原始输入信息的目的。整体网络最终如图2所示。 $3 \times 3$  平均池化构造双金字塔和简化后  $5 \times 5$  平均池化构造双金字塔的网络运算时间和测试误差如表1所示。

表1 复杂度与精度对比

Table 1 Comparison of complexity and accuracy

Item	MC-CNN-fst	Proposed network	
		$3 \times 3$	$5 \times 5$
Time consumption / s	<b>0.1902</b>	0.2561	0.2438
Test error / %	3.029	<b>2.784</b>	2.795

由表1可以看出,改进的网络经简化后精度略微下降,而运行速度明显上升。本文后续如无特殊说明,默认以简化后  $5 \times 5$  平均池化构造双金字塔的网络作为所提出的网络。

## 3 实验与结果分析

本文进行卷积神经网络训练、预测和性能测试的服务器的CPU和GPU分别为 Intel Xeon E5-2620 v4 与 Nvidia GeForce GTX 1080Ti,操作系统为 Ubuntu 14.04,采用 Torch 7 作为深度学习框架。

### 3.1 实验细节

为了验证改进网络以及改进模块的效果,实验与 MC-CNN-fst 方法进行对比。实验所采用的图像训练集、预测集来源于 KITTI Stereo 2012 数据集。该数据集分为两部分,其中194对图像提

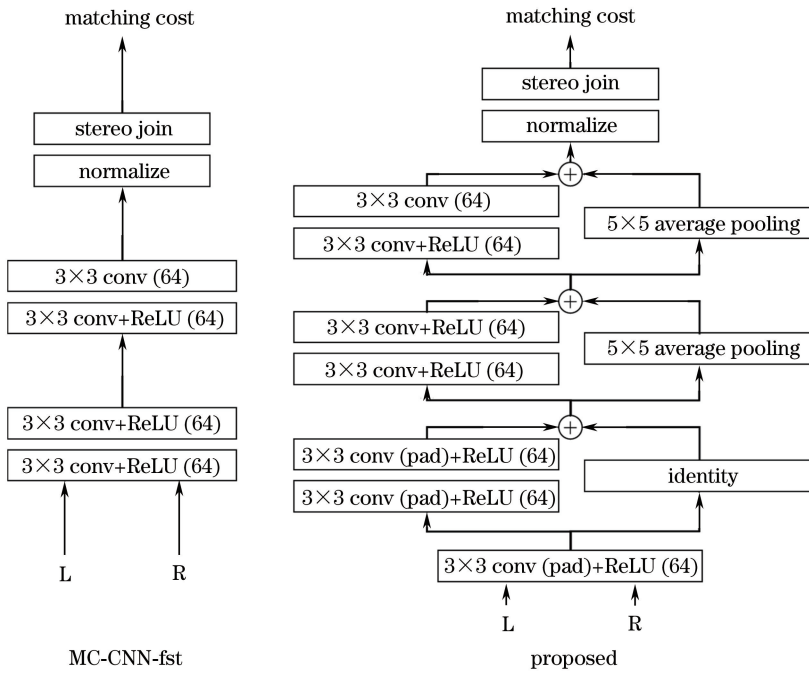


图 2 网络结构对比

Fig. 2 Comparison of network structures

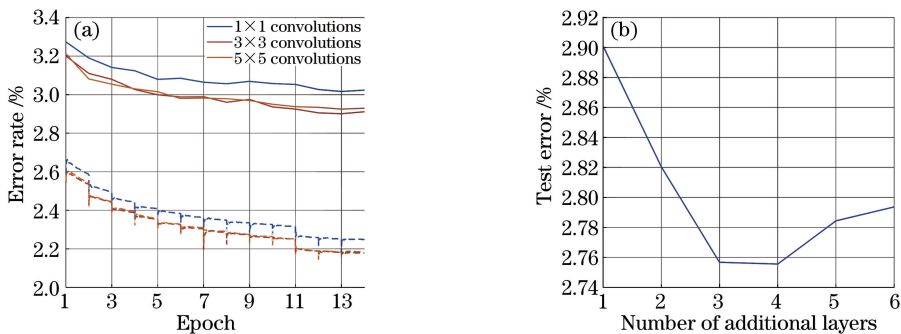


图 3 错误率与额外卷积层参数的关系。(a)不同卷积核大小下的错误率;(b)不同卷积层层数下的测试误差

Fig. 3 Relationship between error rate and parameters of additional convolutional layers. (a) Error rate with different convolutional kernel sizes; (b) test error with different number of convolutional layers

供真实视差值,另外 195 对图像作为测试使用。为了便于计算出所得视差图的错误率,本文沿用 MC-CNN<sup>[11]</sup>中的方法,在 194 对提供真实视差值的图像中随机选取 40 对图像作为测试集,剩余图像对作为训练集,训练和预测均只考虑未遮挡区域。一个训练样本包含截取自左、右图像的两个图像块,即  $\langle P_{9 \times 9}^L(p), P_{9 \times 9}^R(q) \rangle$ , 其中  $P_{9 \times 9}^L(p)$  指截取自左视图的  $9 \times 9$  图像块,中心坐标为  $p = (x, y)$ ,  $P_{9 \times 9}^R(q)$  指截取自右视图的  $9 \times 9$  图像块,中心坐标为  $q$ 。该位置的真实视差值  $d$  已知,对应该位置的正样本为  $q = (x - d + o_{\text{pos}}, y)$ , 负样本为  $q = (x - d + o_{\text{neg}}, y)$ , 其中  $o_{\text{pos}}$  从集合  $\{-P_{\text{hi}}, P_{\text{hi}}\}$  中随机选取,  $o_{\text{neg}}$  从集合

$\{-N_{\text{hi}}, \dots, -N_{\text{lo}}, N_{\text{lo}}, \dots, N_{\text{hi}}\}$  中随机选取, 实验中  $P_{\text{hi}} = 1, N_{\text{lo}} = 4, N_{\text{hi}} = 8$ 。错误率由视差图中视差值不同于真实值的像素占总像素的百分比表示,  $P_{\text{hi}} = 1$  对应错误阈值为 3 pixel。网络的损失函数采用二分类铰链损失, 即

$$L(\mathbf{o}, \mathbf{t}) = \frac{1}{N} \sum_{i=1}^N \max(0, C_{\text{margin}} - \mathbf{o}[i] \cdot \mathbf{t}[i]), \quad (3)$$

网络采用随机梯度下降 (SGD) 算法进行参数更新。(3) 式中  $N$  为网络一次迭代的输入训练样本数, 实验中取 128,  $\mathbf{o}[i]$  为其中一个输入样本经过网络得到的输出,  $\mathbf{t}[i]$  为该样本对应的期望输出。  $C_{\text{margin}}$  为常量, 实验中取 1。初始学习率为 0.003, 在第 12 个训练周期及以后缩小为 0.0003。

### 3.2 主观效果

将4对左右视图经过MC-CNN-fst以及所提出的改进网络,得到的结果作为主观效果。图4(a)~(f)分别为6组对比图片,每组图片由上往下依次为原始输入左视图、MC-CNN-fst网络生成的视差图、本文所提出网络生成的视差图。

本文所提出的网络生成视差图与MC-CNN-fst

整体上相似,在一些细节上的轮廓要明显优于MC-CNN-fst。图5为图4中红色方框区域的放大图。所提出的网络生成的视差图在图5(a)和(b)中树、围栏轮廓更加完整,而在图5(c)中车辆部位的噪点明显减小,图5(d)中对于路标形状呈现得更加清晰,图5(e)和(f)对于路标以及路标杆的显示更为完整。

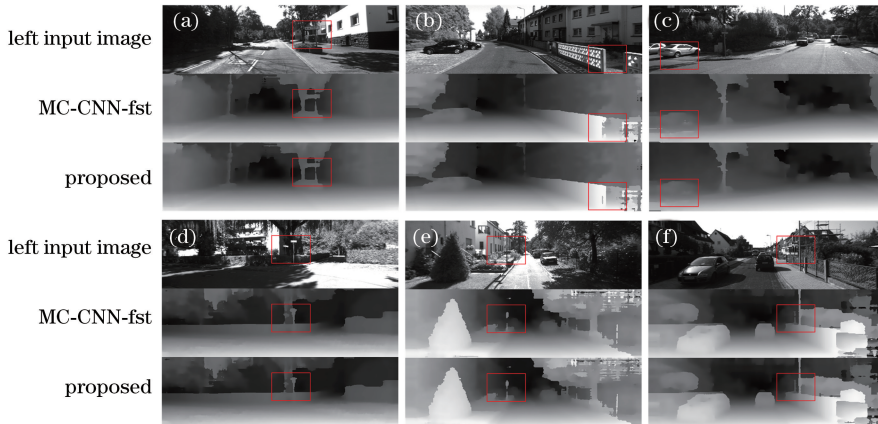


图4 视差图对比。(a)路标1;(b)围栏;(c)汽车;(d)路标2;(e)交通指示牌;(f)路标3

Fig. 4 Comparison of disparity maps. (a) Signpost 1; (b) fence; (c) car; (d) signpost 2; (e) traffic sign; (f) signpost 3

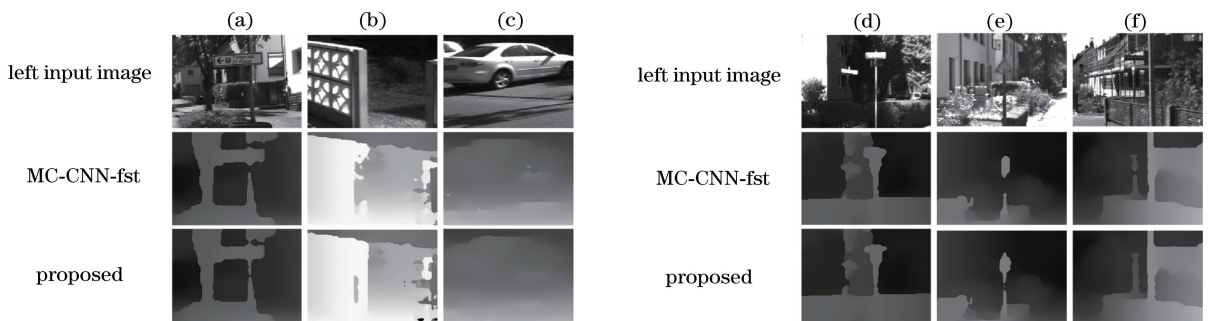


图5 图4中方框区域放大图。(a)路标1;(b)围栏;(c)汽车;(d)路标2;(e)交通指示牌;(f)路标3

Fig. 5 Zoomed parts over region marked by box in Fig. 4. (a) Signpost 1; (b) fence; (c) car; (d) signpost 2; (e) traffic sign; (f) signpost 3

### 3.3 客观指标

本文采用3 pixel阈值的百分比错误率作为评价视差图效果的客观指标,实验中采用40幅测试图的平均结果作为衡量标准,平均结果如表1所示。前10幅测试图的详细测试结果,如表2所示。

如表2所示,在10幅测试图中,本文提出的网络与MC-CNN-fst相比在客观指标上,9幅具有更高的精度,1幅精度近似相同。由表2可知,改进网络整体平均错误率为2.795%,而MC-CNN-fst的整体平均错误率为3.029%,因此,本文所提出的网络具有更高的精度。

### 3.4 复杂度

对比了MC-CNN-fst与本文所提网络的复杂度

表2 10幅测试图的客观指标

Table 2 Objective index of 10 test images

Image	MC-CNN-fst	Proposed (5×5)
1	0.807	0.807
2	4.180	<b>3.892</b>
3	1.210	<b>1.147</b>
4	1.207	<b>1.117</b>
5	8.960	<b>8.006</b>
6	6.228	<b>5.866</b>
7	1.588	<b>1.475</b>
8	0.293	<b>0.213</b>
9	2.934	<b>2.658</b>
10	3.020	<b>2.877</b>
Average	3.029	<b>2.795</b>

与计算耗时,其中复杂度的计算公式为  $O\{\sum_l n_{l-1} f_l^2 n_l\}$ ,即网络中卷积层参数数量的近似值,其中  $n_l$  为网络卷积层第  $l$  层输出的特征图数量,  $f_l$  为网络卷积层第  $l$  层的卷积核大小。计算耗时的测试结果表示为 30 对左右视图经过卷积神经网络方法生成视差图的平均耗时,单位为 s。测试结果如表 3 所示。

表 3 复杂度对比

Table 3 Comparison of complexity

Item	MC-CNN-fst	Proposed(5×5)
Complexity of convolution layers	<b>111168</b>	221760
Time consumption / s	<b>0.1902</b>	0.2438
Whole time consumption / s	<b>0.4040</b>	0.4770

表 4 不同网络结构的对比

Table 4 Comparison of different network structures

Item	MC-CNN-fst	MC-CNN-fst (1×1)	Without additional convolutional layers	Without dual pyramid structure	Default
3 pixel / %	3.029	3.017	3.045	<b>2.757</b>	2.795
Time consumption / s	<b>0.1902</b>	0.2081	0.2251	0.2252	0.2438

从表 3 可以看出,3 pixel 阈值标准下错误率的下降主要得益于网络的加深,双金字塔结构在一定程度上反而抑制了错误率的下降。本文认为双金字塔结构在网络的特征提取过程中引入了未经过学习的原始图像信息,因此抑制了网络的学习,从而导致了错误率的上升,这也是为了保留原始信息所造成的负面影响。

实验结果如图 6 所示,分别去除额外卷积层和双金字塔结构的网络生成的视差图在连通性上表现较差。本文认为去除了额外卷积层使得网络精度大幅

由表 3 可知,本文提出的网络相比 MC-CNN-fst,在卷积层部分复杂度上升了约 99%,网络部分前向传递耗时提高了约 28%,结合后处理算法整体耗时提高了约 18%。

### 3.5 网络结构的影响

通过对比实验探究 2.2 节中两个结构上的改进对网络性能的影响。网络层加深的对比实验通过移除改进网络中额外卷积层实现。由于双金字塔结构需要对输入图像进行升维,这里保留一层影响最小的  $1\times 1$  卷积层进行维度扩展,同时提供 MC-CNN-fst 的基础上添加一层  $1\times 1$  卷积层的网络作为对比。双金字塔结构的对比实验通过移除双金字塔结构实现。

客观指标的实验结果如表 4 所示。

下降,使得网络未能获取稠密准确的视差图而导致主观效果下降;而去除了双金字塔结构使得网络在特征提取的卷积层部分未能保留输入图像的原始信息,导致了生成的原始视差图虽然准确但是较为稀疏离散,经过后处理以后未能保持图像中物体信息的完整性,使得主观效果较差。实验结果证明了两个结构在图像连通性的保持上都是必不可少的。去除了双金字塔结构后,精度会更高,但是在视差图的细节上并没有直观的影响,因此本文认为双金字塔结构对于特征提取过程中原始信息的保护起着不可替代的作用。

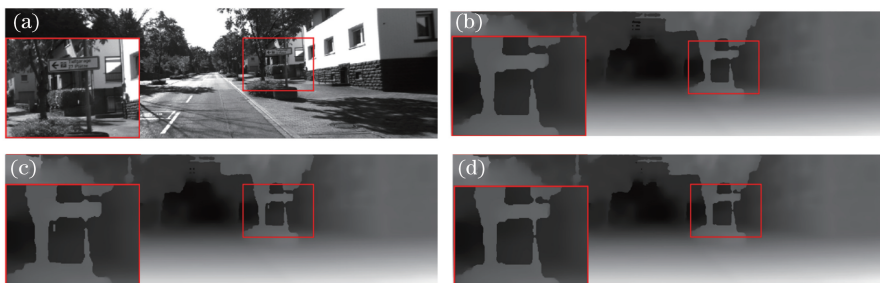


图 6 结构对于主观效果的影响。(a)输入左视图;(b)去除额外卷积层;(c)去除双金字塔结构;(d)视差图

Fig. 6 Effect of structure on subjective quality. (a) Left input image; (b) without additional convolutional layers; (c) without dual pyramid structure; (d) default

## 4 结 论

针对现有的卷积神经网络方法所获取的视差图细节损失严重的问题,通过解析卷积神经网络方法

与传统经典特征提取算法之间的相似之处,结合前沿的卷积神经网络架构与尺度空间理论对现有的卷积神经网络方法进行了改进,主要特点包括:1) 对于网络的加深;2) 引入了所提出的双金字塔结构。

实验结果表明,与 MC-CNN-fst 架构相比较,所提出的网络在整体时间消耗增加 18% 的代价下生成了细节更加准确的视差图,同时将 3 pixel 阈值错误率从 3.029% 降到了 2.795%。通过对比实验,证明了两处改进都对结果具有正面的影响,网络的加深提高了精度,双金字塔结构的引入保持了图像连通性。

从实验结果中可以看出,所提出的双金字塔结构在浅层卷积神经网络中对精度产生了负面的影响,这种影响在一定程度上可以被网络的加深所削减。因此,下一步研究内容应该考虑减少所提出结构产生的负面影响,从而提高精度。

### 参 考 文 献

- [1] Wang Q L, Li J Y, Shen H K. Target tracking system of binocular vision and laser range sensor[J]. *Acta Optica Sinica*, 2016, 36(9): 0912002.  
王琪龙, 李建勇, 沈海阔. 双目视觉-激光测距传感器目标跟踪系统[J]. *光学学报*, 2016, 36(9): 0912002.
- [2] Scharstein D, Szeliski R, Zabih R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[C] // *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision*, 2001: 131-140.
- [3] Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(9): 1124-1137.
- [4] Hirschmuller H. Stereo processing by semiglobal matching and mutual information [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(2): 328-341.
- [5] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [6] Bay H, Ess A, Tuytelaars T, *et al.* Speeded-up robust features (SURF) [J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346-359.
- [7] Zhang X, Jin Y X, Xue D. Image matching algorithm based on SICA-SIFT and particle swarm optimization [J]. *Laser & Optoelectronics Progress*, 2017, 54(9): 091002.  
张鑫, 靳雁霞, 薛丹. SICA-SIFT 和粒子群优化的图像匹配算法[J]. *激光与光电子学进展*, 2017, 54(9): 091002.
- [8] Zhu S P, Yan L N, Li Z. Stereo matching algorithm based on improved Census transform and dynamic programming[J]. *Acta Optica Sinica*, 2016, 36(4): 0415001.  
祝世平, 闫利那, 李政. 基于改进 Census 变换和动态规划的立体匹配算法[J]. *光学学报*, 2016, 36(4): 0415001.
- [9] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [10] Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks[C] // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015: 4353-4361.
- [11] Žbontar J, LeCun Y. Computing the stereo matching cost with a convolutional neural network[C] // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015: 1592-1599.
- [12] Zhang K, Lu JB, Lafruit G. Cross-based local stereo matching using orthogonal integral images[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2009, 19(7): 1073-1079.
- [13] He KM, Sun J, Tang X O. Guided image filtering [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(6): 1397-1409.
- [14] Mei X, Sun X, Zhou M C, *et al.* On building an accurate stereo matching system on graphics hardware [C] // *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011: 467-474.
- [15] Li LC, Yu X, Zhang S L, *et al.* 3D cost aggregation with multiple minimum spanning trees for stereo matching[J]. *Applied Optics*, 2017, 56(12): 3411-3420.
- [16] Li L C, Zhang S L, Yu X, *et al.* PMSC: PatchMatch-based superpixel cut for accurate stereo matching [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, 28(3): 679-692.
- [17] Park H, Lee K M. Look wider to match image patches with convolutional neural networks[J]. *IEEE Signal Processing Letters*, 2017, 24(12): 1788-1792.
- [18] Fan B, Niu J C, Zhao J. Three-phase full-controlled rectifier circuit fault diagnosis based on optimized neural networks[C] // *2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, 2011: 6048-6051.
- [19] Srivastava R K, Greff K, Schmidhuber J. Training very deep networks [C] // *Neural Information Processing Systems*, 2015: 2377-2385.

- [20] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.
- [21] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning, 2015: 448-456.