

基于视差信息的无参考立体图像质量评价



朱玲莹 桑庆兵 顾婷婷

江南大学物联网工程学院 江苏 无锡 214122

(zlingyingzly@163.com)

摘要 近年来,随着深度学习在图像质量评价领域的快速发展,平面图像质量评价得到了有效的改善,但是立体图像质量评价还有待提高。为此,文中结合三分支卷积神经网络,提出了基于视差信息的无参考立体图像质量评价方法,并分析了不同视差图对模型性能的影响。该方法将左右视图以及视差图小块作为输入,自动提取特征,通过训练得到回归模型,从而实现对立体图像的预测。文中使用了5种不同立体匹配算法来生成视差图,实验结果表明使用 SAD 算法得到的效果最好。在立体图像库 LIVE3D 和 MCL3D 上的实验结果表明,该方法不仅适用于评估对称失真图像,还适用于非对称失真的立体图像评价。该方法在总体失真上的结果优于其他对比算法,尤其是在 MCL3D 图像库上,所提方法的 PLCC 和 SROCC 比其他方法高出 1% 和 4%。实验数据表明,所提模型提高了立体图像质量评价的性能,与人类主观感知高度一致。

关键词: 视差信息; 立体匹配算法; 卷积神经网络; 立体图像质量评价; 非对称失真图像

中图法分类号 TP183

No-reference Stereo Image Quality Assessment Based on Disparity Information

ZHU Ling-ying, SANG Qing-bing and GU Ting-ting

School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract In recent years, with the rapid development of deep learning in the field of image quality assessment (IQA), 2D-IQA has been improved, but 3D-IQA still needs to be improved. Therefore, combining the three-branch convolutional neural network, the paper proposes a no-reference stereo image quality assessment based on disparity information and analyzes the influence of different disparity maps on the performance of the model. The algorithm takes the left/right view patches and the disparity map patches as input, automatically extracts features, and obtains the regression model through training to realize the prediction of the stereo images. In this paper, 5 different stereo matching algorithms are used to generate disparity maps. The experimental results show that the SAD algorithm is the best. The experimental results on stereo image databases LIVE3D and MCL3D show that the method is not only suitable for evaluating symmetric distortion images, but also for evaluating asymmetric distortion stereo images. The overall distortion results of this method are superior to other comparison algorithms. Especially on the MCL3D image database, the evaluation method PLCC and SROCC of the proposed method are 1% and 4% higher than other methods. The Experimental data shows that the proposed model improves the performance of stereo image quality assessment, which is highly consistent with human subjective perception.

Keywords Disparity information, Stereo matching algorithm, Convolutional neural network, Stereo image quality assessment, Asymmetric distortion image

1 引言

随着 3D 图像技术的迅速发展,3D 图像在人们的生活中变得越来越重要。3D 电影、VR 虚拟现实、医学影像等通过立体图像的展示,使人们获得接近真实的感受。为了不影响人们的视觉感知,三维技术首先要保证立体图像的质量,因此对立体图像质量评价算法进行研究就变得极其重要。与平面

图像不同,立体图像需要考虑深度信息和视差信息^[1],尤其当左右视图失真程度不同时,对图像的评价会变得更加复杂,这就使得立体图像质量评价更具挑战性。

针对立体图像质量评价,一些学者借鉴了平面图像质量评价的方法。例如,Benoit 等^[2]使用一些 2D 图像的全参考质量评价算法,分别计算参考图像、失真图像及其视差图的质量得分,并组合这些得分得到立体图像的质量分数。Zhang

收稿日期:2019-07-31 返修日期:2019-10-31 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:江苏省自然科学基金面上项目(BK20171142)

This work was supported by the Natural Science Foundation of Jiangsu Province, China(BK20171142).

通信作者:桑庆兵(sangqb@163.com)

等^[3]提出了 3DLN 算法,首先通过学习双眼对比度图的局部自然统计数据来计算质量,然后用平面图像质量评价方法来估计单眼视图质量,将单眼视图质量作为权重,组合两个质量分数计算出最终得分。

借鉴平面图像质量评价的方法虽然简单直接,但是由于立体图像和平面图像存在差异,结果并不理想。因此,很多学者提出了新的立体图像评价方法。Zhou 等^[4]研究了双目视觉原理,从图像中提取双目质量预测特征,用这些特征训练一个多元高斯模型来预测质量分数。Appina 等^[5]提出了用于亮度和视差子带系数联合分布的二元广义高斯分布模型,并基于这个模型实现了图像质量的评估。

在立体图像中,视差图以左右视图中的任一幅图像为基准,其大小为基准图像的大小,元素值即为视差值。视差图包含了场景的距离信息和深度信息,因此在立体图像质量评价中具有重要作用。部分学者基于此展开了研究^[6-10]。Li 等^[8]利用小波域的特点和三维图像的视差信息,使用深度信念网络对对称和非对称失真图像进行分类,分别提取小波域、空间的自然图像统计特征,提出了一种基于深度学习和自然场景统计的立体图像盲评价方法。Lin 等^[9]从立体图像对和视差中提取统计特征来进行质量评价,使用支持向量机的回归模型来预测质量分数。

当失真类型未知时,现有立体图像质量评价模型预测的准确性较差,尤其是在预测非对称失真图像的情况下。针对以上问题,本文提出了一个端到端的三支卷积神经网络模型,研究了 5 种不同立体匹配算法生成的视差图对该模型性能的影响。文献[6]、文献[8]和文献[9]使用的视差信息直接由左右两幅图像对应像素相减得到,而视差图的产生方法有很多,性能也各不相同,因此本文利用三支网络模型分析了视差图的好坏对立体图像质量评价的影响。其次,不同于文献[6]的模型结构,本文模型的最后一次池化采用了最大池化和均值池化联合的方法,这样既保留了纹理信息,又保留了图像的背景信息。与仅评估左右视图的方法相比,本文方法表现出了一定的优势。

2 视差图与立体匹配算法

2.1 视差图原理

视差的原理如图 1 所示, O_L 和 O_R 是左右相机的光心, Z 为场景的深度, f 为相机的焦距, b 为相机基线之间的距离, x_L 和 x_R 分别是物体 P 在像平面上投影点与最左边缘的距离,则视差公式定义为:

$$disparity = x_L - x_R \quad (1)$$

运用相似三角形的性质可知:

$$\frac{b}{Z} = \frac{b - (x_L - L/2) - (L/2 - x_R)}{Z - f} \rightarrow x_L - x_R = \frac{b \cdot f}{Z} \quad (2)$$

$$disparity = \frac{b \cdot f}{Z} \quad (3)$$

其中, b 是定值。从公式中可以看出视差与图像场景的深度成反比。在视觉研究中,视差图通过立体匹配算法得到。

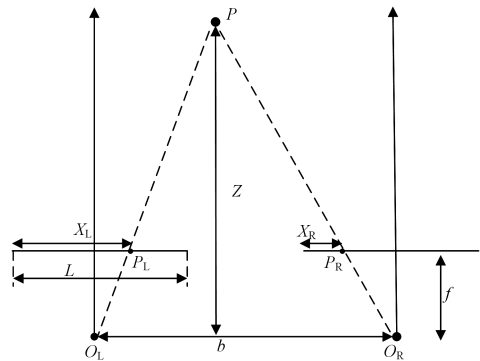


图 1 视差原理

Fig. 1 Parallax principle

2.2 立体匹配算法

在立体匹配中,匹配问题可以看成寻找两组数据相关程度的过程。根据算法运行时约束的作用范围,立体匹配算法分为全局(半全局)匹配算法和局部匹配算法。

2.2.1 全局(半全局)立体匹配算法

全局(半全局)立体匹配算法主要采用了全局的优化理论方法来估计视差,建立了一个全局能量函数,其包含一个数据项和平滑项,通过最小化全局能量函数得到最优的视差值。其中,图割(Graph Cuts, GC)、置信传播(Belief Propagation, BP)、动态规划(Dynamic Programming, DP)、遗传算法(Genetic Algorithm, GA)等优化算法,都是常用的求解能量最小化的方法。

GC 算法是一种全局优化方法,能有效去除瑕疵,其匹配效果较其他全局匹配算法更好^[11]。GC 算法的基本思想为:在一定的约束条件下建立能量函数,根据能量函数构造合适的图,由此最小化该能量函数。能量函数一般定义为:

$$E(d) = E_{data}(d) + E_{smooth}(d) \\ = \sum_{p \in R} C(p, d_p) + \sum_{q, p \in R} P(d_q - d_p) \quad (4)$$

其中, $E_{data}(d)$ 描述了匹配程度, $E_{smooth}(d)$ 为平滑项,体现了场景的约束条件; C 是匹配代价; p 点像素的视差为 d_p ; P 是惩罚项,是不同两像素 p 和 q 视差的函数。当 $d_q \neq d_p$ 时,差值越大, P 值越大;当 $d_q = d_p$ 时, $P = 0$ 。

SGM^[12] (Semi-Global Matching) 算法是对动态规划改进后形成的半全局立体匹配算法,是全局算法的近似算法。较全局算法而言,SGM 算法实现简单,计算速度快。其能量函数为:

$$E(d) = \sum_{p \in R} C(p, d_p) + \sum_{q \in N_p} P_1(|d_q - d_p| = 1) + \\ \sum_{q \in N_p} P_2(|d_q - d_p| > 1) \quad (5)$$

其中, N_p 是像素 p 的相邻像素点, P_1 和 P_2 是惩罚项。 P_1 适用于 p 和 q 视差相差 1 的像素,否则其函数值为 0; P_2 适用于 p 和 q 视差差值大于 1 的像素,否则其函数值为 0。

2.2.2 局部立体匹配算法

局部立体匹配算法又称基于窗口的方法或基于区域的方法,通常是在一个设定的窗口基础上来完成,以窗口中提取的特征向量为基础进行匹配。这种匹配可以是逐像素的匹配,也可以是成块的匹配,更进一步地有根据图像自适应调节窗

口大小的区域匹配算法。局部匹配算法仅利用某一点邻域的灰度、颜色、梯度等信息来计算匹配代价,计算复杂度较低。最常见的3种匹配代价计算方法为^[13]:绝对差值和(Sum of Absolute Differences, SAD)、差值平方和(Sum of Squared Differences, SSD)、归一化互相关(Normalized Cross Correlation, NCC)。

$$SAD = \sum_{i=n} \sum_{j=m} |I_1(u+i, v+j) - I_2(u+d+i, v+j)| \quad (6)$$

$$SSD = \sum_{i=n} \sum_{j=m} (I_1(u+i, v+j) - I_2(u+d+i, v+j))^2 \quad (7)$$

$$NCC = \frac{\sum_{i=n} \sum_{j=m} I_1(u+i, v+j) I_2(u+d+i, v+j)}{\sqrt{\sum_{i=n} \sum_{j=m} I_1(u+i, v+j)^2 \sum_{i=n} \sum_{j=m} I_2(u+d+i, v+j)^2}} \quad (8)$$

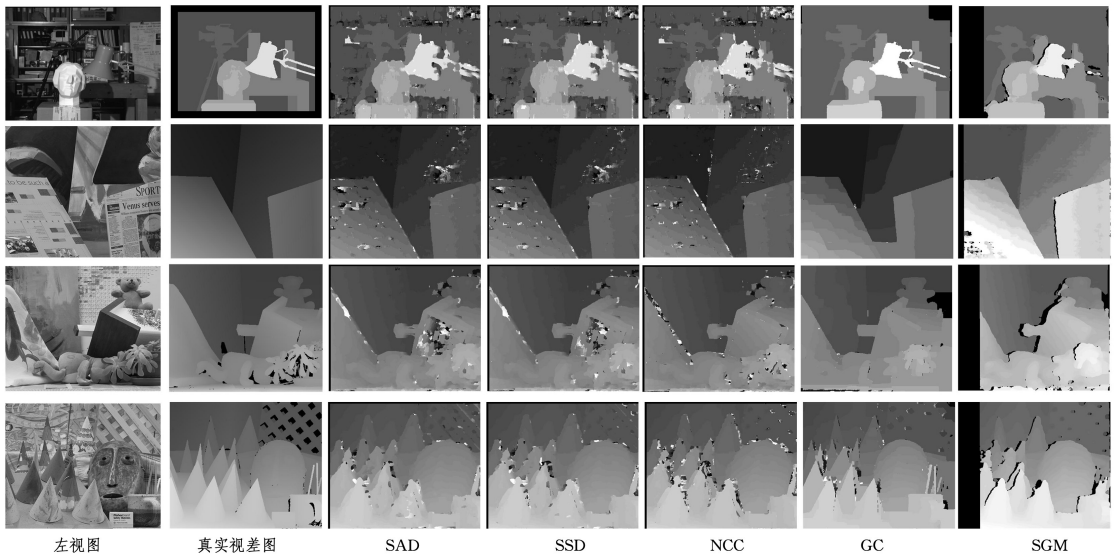


图2 5种立体匹配算法生成的视差图

Fig. 2 Disparity maps generated by five stereo matching algorithms

表1 匹配错误率

Table 1 Matching error rate

(单位: %)

algorithm		SAD	SSD	NCC	GC	SGM
tsukuba	all	9.71	10.12	10.59	16.17	13.58
	nonocc	12.99	13.43	14.10	20.68	22.93
venus	all	12.82	12.22	10.60	41.17	99.71
	nonocc	10.63	9.97	8.21	39.69	99.55
teddy	all	44.31	43.35	40.22	52.13	99.74
	nonocc	39.27	38.11	34.56	48.18	99.53
cones	all	43.34	42.37	42.36	58.33	99.69
	nonocc	36.52	35.23	35.23	54.97	99.92
Average		26.19	25.6	24.48	41.41	72.51

由表1可以看出,局部立体匹配算法(SAD, SSD, NCC)的匹配准确率优于全局(半全局)立体匹配(SGM, GC)算法,说明SAD, SSD, NCC算法生成的视差图精度更高。对于tsukuba图像来说,其图像背景简单,匹配效果好,其中SAD算法的误匹配率比SSD和NCC算法低;对于venus和teddy图像来说, NCC算法的误匹配率最低。teddy和cones图像的误匹配率较其他图像高,这是因为其存在大量的无纹理和重复纹理区域。总体来说, SAD, SSD, NCC算法的匹配准确率大体相当,相比SGM和GC算法,其处理纹理复杂图像的性能更优异。

其中, I_1 和 I_2 为左右图像对; u 和 v 为像素的坐标点; n 和 m 为窗口的长和宽,表示移动的距离。

2.3 立体匹配算法的性能比较

为了验证上文所述5种算法的性能,本文采用Middlebury测试平台提供的测试图像进行实验。匹配准确度的定量衡量指标为:整体错误率(all)、间断处错误率(disc)、非遮挡错误率(nonocc)^[14]。

图2是Middlebury第二版测试集中真实视差图及各种算法生成的视差图。该测试集共有4对图像,分别为tsukuba, venus, teddy, cones。本文使用整体错误率、非遮挡错误率以及它们的平均值来比较视差图的准确度,如表1所列。错误率越小,算法的视差图就越准确。

3 网络架构

3.1 三支卷积神经网络

不同于平面图像质量评价,立体图像的左右视图中包含了深度感知等信息。视差图不仅考虑了图像内容,而且考虑了立体图像的深度和距离信息,因此将视差图作为立体图像质量评价的重要信息来源。结合立体图像本身的左右视图,本文设计了一个三支卷积网络,该网络可以自适应提取失真特征,3个分支的输入分别为左视图、右视图图像块以及对应的视差图图像块。网络架构如图3所示,每个分支都有5个卷积层和3个池化层。将每个分支提取出的特征进行拼接,然后输送至全连接层,通过非线性方式拟合质量分数。受VGG-16网络^[15]的启发,每两次卷积后进行一次最大池化,子采样大小取为 2×2 ,但最后一次卷积后同时进行最大和均值池化,这样更有利于提取图像的感知特征。最大池化(Max_pool)和均值池化(Avg_pool)的计算公式分别为:

$$\mu_k = \max_{(i,j) \in \Omega} M_{i,j}^k \quad (9)$$

$$\nu_k = \text{avg}_{(i,j) \in \Omega} M_{i,j}^k \quad (10)$$

其中, k ($k=1, 2, \dots, K$) 表示卷积核的数量, M 表示特征图, $M_{i,j}^k$ 表示卷积后的第 k 个特征图, Ω 为子采样大小,即最大池化和均值池化的局部大小。

在实验中,我们采用最小的感知图像质量的尺寸 3×3 作为所有卷积层卷积核的大小。

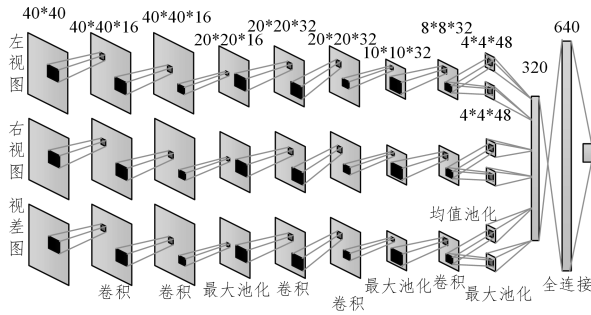


图3 三分支卷积神经网络的结构

Fig. 3 Three-branch convolutional neural network structure

3.2 损失函数及参数配置

在所有的卷积层中,使用 ReLU 函数 (Rectified Linear Units)^[16] 作为激活函数。ReLU 函数是神经网络中最常用的激活函数,能够解决梯度消失问题。在全连接层,使用 dropout 技术防止过拟合,概率设置为 0.5。

本文网络结构使用的目标函数是欧氏距离损失函数,其定义为:

$$E_{\text{loss}} = \frac{1}{2N} \sum_{n=1}^N \| Y_n - y_n \|^2 \quad (11)$$

其中, E_{loss} 代表损失值, Y_n 表示的预测值, y_n 表示图片的主观分数, N 表示图片数, C 表示图片通道数, H 表示图像的高度, W 表示图片的宽度。

3 个分支的参数配置都相同,第一分支以及全连接层的参数配置如表 2 所列。其中, Conv 表示卷积层, Pool 表示池化层, FC 表示全连接层, k 表示卷积核大小(局部窗口大小), n 为卷积核个数(通道数), s 表示步长, p 表示扩充边缘大小。

其他参数的默认设置如下:动量(momentum)为 0.9,初始学习率为 0.01,权重衰减(weightdecay)为 0.0005。学习率更新方式如下:

$$I_r = \text{base_I}_r * (1 + \text{gamma} * \text{iter})^{\wedge} (-\text{power}) \quad (12)$$

其中, base_I_r 是初始学习率; iter 为当前迭代次数; gamma 是学习速率变化因子,默认为 0.0001;指数 power 默认为 0.75。

表3 数据库图像的失真类型和数量

Table 3 Type and amount of distortion images in database

图像库	参考图像	失真图像	Gblur	FF	JP2K	WN	JPEG	AWN	Sample	Transloss
LIVE3D	Phase I	20	365	45	80	80	80	80	—	—
	Phase II	8	360	72	72	72	72	—	—	—
MCL3D		9	648	108	—	108	—	108	108	108

4.1.2 性能评价指标

本文选取皮尔森线性相关系数 (Pearson Linear Correlation Coefficient, PLCC) 和斯皮尔曼等级相关系数 (Spearman Rank Order Correlation Coefficient, SROCC) 作为分析结果的评价指标。PLCC 表示模型预测的分数与人类主观分数的线性相关关系, SROCC 是用来衡量预测的分数与人类主观分数之间秩序的相关性。PLCC 和 SROCC 的取值范围为 $[-1, 1]$, 绝对值越接近 1, 模型的性能就越好。

表2 网络的参数配置

Table 2 Network parameter configuration

操作类型	参数信息	输出数据维度	
Conv1	$k=3, s=1, p=1, n=16$	$40 * 40 * 16$	
Conv2		$40 * 40 * 16$	
Pool1 (Max_pool)	$k=2, s=3$	$20 * 20 * 16$	
Conv3	$k=3, s=1, p=1, n=32$	$20 * 20 * 32$	
Conv4		$20 * 20 * 32$	
Pool2	$k=2, s=3$	$10 * 10 * 32$	
Conv5	$k=3, s=1, p=1, n=48$	$8 * 8 * 48$	
Pool3	Max_pool	$k=2, s=2$	
	Avg_pool	$k=2, s=2$	
全连接层	FC1	ReLU	$1 * 1 * 320$
	FC2	ReLU, dropout=0.5	$1 * 1 * 640$
	FC3	$n=1, \text{sigmoid}$	1

4 实验结果及分析

4.1 图像数据库及性能评价指标

4.1.1 图像数据库

本文采用的是美国德州大学 LIVE 实验室提供的对称和非对称失真 3D 图像数据库, 它分为测试库 I (Phase I) 和 II (Phase II)。Phase I 中左右图像是对称失真, 共有 365 张失真图像; Phase II 包含了对称和非对称失真图像, 共有 360 张失真图像。在 LIVE 3D 图像库中, 一共有 5 种失真情况, 分别为 JPEG 压缩 (JPEG)、JPEG2000 压缩 (JP2K)、加性高斯白噪声 (Additive white Gaussian noise, WN)、高斯模糊 (Gaussian blur, Gblur) 和快速衰落 (Fast-fading, FF) 失真。LIVE 3D 图像库的图像分数由平均主观得分差 (Differential Mean Opinion Score, DMOS) 来表示, 分数越高, 图像质量就越差。

为了验证算法的通用性, 采用 MCL3D 数据库中的立体图像进行验证, 其包含的图像都为对称失真图像, 共有 9 个场景, 每个场景有 6 种不同类型的失真, 分别为 JPEG、JP2K、Gblur、加性白噪声 (Additive White noise, AWN)、采样模糊 (Sampling blur, Sample) 和传输损耗 (Transmission loss, Transloss) 失真。MCL3D 图像库中的图像分数由平均主观得分 (Mean Opinion Score, MOS) 来表示, 分数越高, 图像质量就越好。LIVE3D 和 MCL3D 数据库中图像的失真类型与各自的数量如表 3 所列。

$$\text{SROCC} = 1 - \frac{6 \sum_{i=1}^N (x_i - y_i)^2}{N(N^2 - 1)} \quad (13)$$

$$\text{PLCC} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (14)$$

其中, x_i 和 \bar{x} 分别表示第 i 张图片的客观预测值和所有图片客观预测值的平均值, y_i 和 \bar{y} 分别表示第 i 张图片的人类主观评分和所有图片的人类主观评分的平均值, N 表示图片数量。

4.2 实验结果分析

LIVE 3D 图像库的图像标签需要进行归一化处理来减小数据分布的差异,使 DMOS 取值为 $[0, 1]$ 。整个网络模型的训练需要大量的实验样本,但是 LIVE 3D 图像库 Phase I 和 Phase II 都只有 300 多对图片,因此将每幅图像切成大小为 40×40 的无重叠小块后增加了数据量,满足网络训练的需要。将图像库数据的 80% 作为本实验的训练集,10% 作为测试集,10% 作为验证集,3 个数据集的数据相互独立、无交叉。验证集的设置有助于调整网络参数,也可以防止网络的过拟合。为了平衡数据在随机选择时对结果的影响,将进行 5 次训练实验,每次都按上述要求随机选择数据,计算 5 次结果的平均值作为最后的实验结果。

为了确定迭代次数,观察 Phase I 上某次实验总体失真的 SROCC 值和 PLCC 值随着迭代次数改变而变化的情况,结果如图 4 所示。综合实验结果和时间成本的因素,选取迭代次数为 200000 时的实验结果。

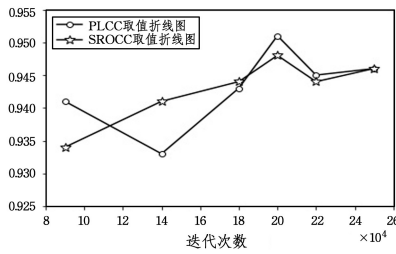


图 4 SROCC 值和 PLCC 值的变化折线图

Fig. 4 Line chart of SROCC and PLCC

为了验证视差图在立体图像质量评价中的效果,本文将三分支模型与仅评估左右视图的两分支卷积神经网络模型(两分支模型)进行对比。该模型只有两个分支,两个分支的输入分别为左视图和右视图图像块,每个分支的结构与三分支模型的结构相同。

为了验证算法的有效性,本文将本文算法与以下 5 种方法进行比较:2 种有效的全参考方法(PSNR^[17], SSIM^[18])和 4 种无参考方法(Zhang 等^[6]提出的方法、Lin 等^[9]提出的方法和 QRSIVS^[19])。

表 4 和表 5 列出了在 LIVE 3D Phase I 库上进行实验获得的评价指标值,表 6 和表 7 列出了在 LIVE 3D Phase II 库上进行实验获得的评价指标值。总体来看,当三分支模型使用局部立体匹配算法(SAD, SSD, NCC)生成的视差图作为输入时,三分支模型性能优于其他对比算法,但是当使用全局(半全局)立体匹配算法生成的(SGM 和 GC)视差图时,三分支模型的性能不佳;使用 SAD 视差图的三分支模型的性能优于两分支模型,但是使用 SGM 视差图和 GC 视差图的三分支模型的性能远不如两分支模型。因此,网络是否提取视差图的特征会影响模型的性能。2.2 节中已证明 SAD, SSD 和 NCC 算法生成的视差图的准确度优于 SGM 和 GC 算法,因此三分支模型的性能与视差图的准确度相关,提供的视差图越精确,网络越能提取到有益的视差特征,从而提升模型性能;反之,视差图的误匹配率越高,越会降低模型的性能。

表 4 实验在 LIVE 3D Phase I 库上获得的 SROCC 值

Table 4 SROCC for different methods on LIVE 3D Phase I database

Model	JP2K	JPEG	WN	Gblur	FF	ALL
PSNR	0.799	0.121	0.932	0.902	0.587	0.834
SSIM	0.803	0.569	0.886	0.907	0.803	0.854
Zhang	0.931	0.690	0.946	0.909	0.834	0.943
Lin	0.902	0.772	0.929	0.903	0.820	0.937
QRSIVS	0.881	0.513	0.944	0.930	0.686	0.916
two-branch	0.786	0.619	0.881	0.900	0.881	0.925
ours (GC)	0.548	0.095	0.952	0.300	0.738	0.699
ours(SGM)	0.833	0.643	0.905	0.900	0.857	0.554
ours(NCC)	0.905	0.524	0.809	0.300	0.952	0.907
ours(SSD)	0.929	0.595	0.952	0.700	0.952	0.938
ours(SAD)	0.905	0.691	0.667	0.900	0.929	0.948

表 5 实验在 LIVE 3D Phase I 库上获得的 PLCC 值

Table 5 PLCC for different methods on LIVE 3D Phase I database

Model	JP2K	JPEG	WN	Gblur	FF	ALL
PSNR	0.785	0.219	0.935	0.916	0.703	0.834
SSIM	0.817	0.582	0.869	0.902	0.797	0.851
Zhang	0.926	0.740	0.944	0.930	0.883	0.947
Lin	0.934	0.761	0.933	0.932	0.861	0.942
QRSIVS	0.909	0.617	0.949	0.951	0.778	0.924
two-branch	0.929	0.438	0.979	0.981	0.913	0.934
ours (GC)	0.514	0.071	0.968	0.395	0.834	0.769
ours(SGM)	0.739	0.605	0.960	0.711	0.843	0.725
ours(NCC)	0.951	0.734	0.960	0.705	0.923	0.919
ours(SSD)	0.952	0.610	0.958	0.983	0.944	0.939
ours(SAD)	0.934	0.747	0.910	0.981	0.955	0.951

表 6 实验在 LIVE 3D Phase II 库上获得的 SROCC 值

Table 6 SROCC for different methods on LIVE 3D Phase II database

Model	JP2K	JPEG	WN	Gblur	FF	ALL
PSNR	0.597	0.491	0.919	0.690	0.730	0.665
SSIM	0.735	0.709	0.886	0.820	0.818	0.799
Zhang	0.897	0.942	0.941	0.489	—	0.914
Lin	0.897	0.774	0.903	0.922	0.881	0.906
QRSIVS	0.953	0.773	0.826	0.770	0.831	0.714
two-branch	0.678	0.714	0.714	0.714	0.286	0.881
ours (GC)	0.500	0.286	0.429	0.393	0.393	0.075
ours(SGM)	0.464	0.393	0.857	0.536	0.357	0.187
ours(NCC)	0.857	0.643	0.964	0.643	0.750	0.931
ours(SSD)	0.857	0.536	1.000	0.679	0.536	0.938
ours(SAD)	0.750	0.857	0.964	0.893	0.857	0.952

表 7 实验在 LIVE 3D Phase II 库上获得的 PLCC 值

Table 7 PLCC for different methods on LIVE 3D Phase II database

Model	JP2K	JPEG	WN	Gblur	FF	ALL
PSNR	0.597	0.491	0.919	0.690	0.730	0.665
SSIM	0.747	0.715	0.892	0.831	0.829	0.797
Zhang	0.921	0.927	0.957	0.899	—	0.912
Lin	0.901	0.812	0.895	0.979	0.896	0.913
QRSIVS	0.957	0.789	0.829	0.784	0.868	0.745
two-branch	0.756	0.836	0.765	0.976	0.590	0.901
ours (GC)	0.391	0.481	0.587	0.236	0.537	0.153
ours(SGM)	0.468	0.309	0.848	0.554	0.062	0.191
ours(NCC)	0.871	0.743	0.978	0.941	0.821	0.920
ours(SSD)	0.936	0.573	0.921	0.936	0.811	0.933
ours(SAD)	0.900	0.884	0.824	0.976	0.956	0.933

Phase II 图像库中包含非对称失真图像,结合表 6 和表 7 中的数据来看,三分支模型能够很好地预测非对称失真图像的质量。本文提出的方法与文献[6]和文献[9]提出的方法具有竞争性,虽然在单失真 JPEG 和 JP2K 上的效果不如文献[6],但在总体失真上具有极大的优越性。

散点图可以直观地反映预测得分是否与主观得分一致,能可视化所提出模型的性能。散点分布越集中于直线两侧,模型性能就越好。图 5(a)~图 5(c)、图 5(e)、图 5(f)为应用

不同种类视差图的三分支模型在 Phase I 库上预测得分的散点图,图 5(d)为两分支模型预测得分的散点图。图 5 中,每个圆点代表一对左右图像的预测分数(纵轴)和主观分数(横轴),线条为拟合曲线,圆点越靠近拟合曲线,表示性能越好。从图中可以看到,图 5(a)的散点分布集中,离曲线更近,明显优于两分支模型(图 5(d))和使用全局(半全局)立体匹配算

法的三分支模型(图 5(e)、图 5(f))。图 6(a)、图 6(c)为应用 SAD 和 SGM 视差图的三分支模型在 Phase II 库上预测得分的散点图。图 6(b)显示了两分支模型预测得分的散点图。图 6(a)的散点紧密地分布在拟合线两边,而图 6(b)、图 6(c)的散点分散,说明提取积极的视差特征有利于提高模型性能,消极的视差特征反而降低模型性能。

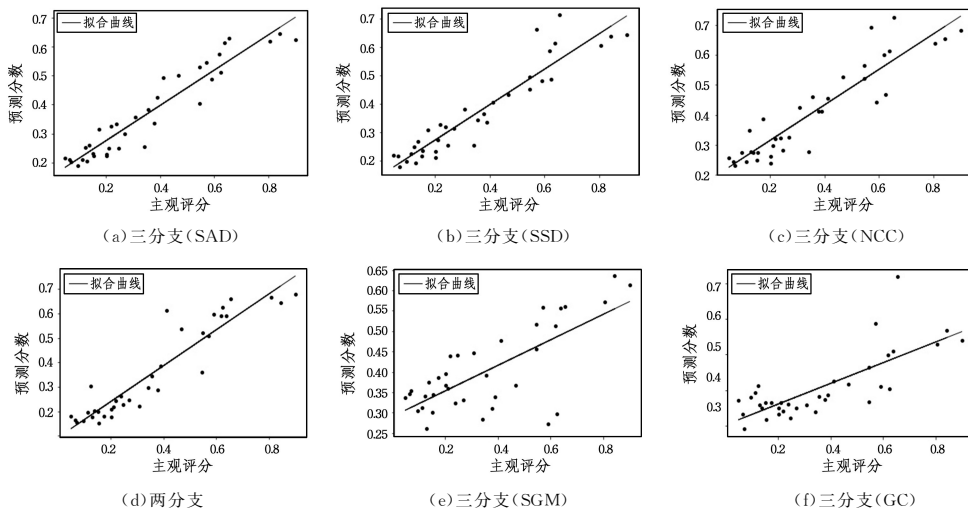


图 5 Phase I DMOS 值和预测得分的散点图

Fig. 5 Scatter plot of DMOS values and predicted scores in Phase I

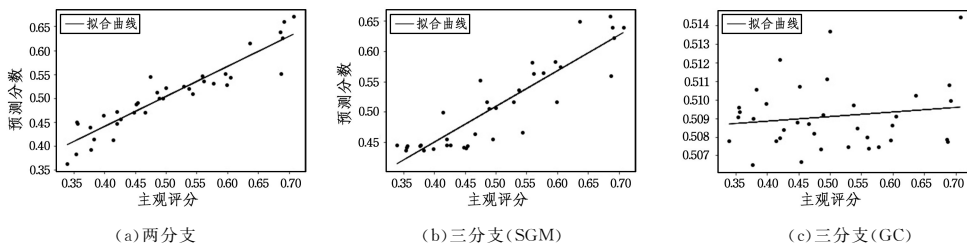


图 6 Phase II DMOS 值和预测得分的散点图

Fig. 6 Scatter plot of DMOS values and predicted scores in Phase II

4.3 算法通用性

本文采用 MCL3D 图像库验证该模型是否具有通用性,实验模型以及训练方法都不变,依旧使用 SROCC 和 PLCC 作为评价指标。本文将本文算法与 5 种方法进行了比较:2 种全参考方法(Benoit 等^[20]提出的方法和 Ryu 等^[21]提出的方法)和 3 种无参考方法(NIQSV 等^[22]、SIQM 等^[23]和 Yang 等^[24]提出的方法)。

表 8 各方法在 MCL3D 图像库上获得的 PLCC 值和 SROCC 值

Table 8 PLCC and SROCC for different methods on

MCL3D database			
Model	Type	SROCC	PLCC
Benoit	FR	0.767	0.755
Ryu	FR	0.850	0.842
NIQSV	NR	0.621	0.678
SIQM	NR	0.776	0.774
Yang	NR	0.905	0.912
Ours(GC)	NR	0.830	0.791
Ours(SGM)	NR	0.871	0.788
Ours(NCC)	NR	0.928	0.912
Ours(SSD)	NR	0.938	0.919
Ours(SAD)	NR	0.945	0.923

支模型性能最好,其结果相比其他方法提高了 1% 以上。文献[21]所采用的方法(NIQSV)不适用于非对称失真图像,而本文方法适用于对称和非对称失真,说明该模型与人类主观感知一致,有较强的泛化能力。

结束语 本文提出的深度卷积神经网络,将左右视图的图像小块以及视差图作为 3 个分支输入,自适应学习特征,最后采用最大、均值池化联合的方法,将 3 个分支学习到的特征图线性拼接一起输入全连接层,最终通过训练得到的回归模型预测立体图像质量。本文采用立体匹配算法对图像进行视差计算,比较了 5 种经典算法生成的视差图的准确性,研究了视差图对立体图像质量评价算法性能的影响。在 LIVE 3D 和 MCL3D 图像库上的对比实验说明,视差图的精确度越高,越能提高三分支模型的性能,且本文方法比仅评估左右视图的方法更优异。本文采用 SAD 算法处理的视差图得出的结果具有优势,且能准确地评价对称和非对称失真的立体图像质量,与人类主观感知高度一致。

参考文献

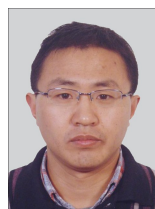
[1] CHEN M J, LAWRENCE K, CORMACK, et al. No-Refere-

根据表 8 的实验数据可以看出,使用 SAD 视差图的三分

- rence Quality Assessment of Natural Stereopairs [J]. IEEE Trans on Image Processing, 2013, 22(9): 3379-3391.
- [2] BENOIT A, CALLET P L, CAMPISI P, et al. Quality Assessment of Stereoscopic Images [J]. Eurasip Journal on Image & Video Processing, 2008, 2008(1): 1-13.
- [3] ZHANG Y, CHANDLER D M. Learning Natural Statistics of Binocular Contrast for No Reference Quality Assessment of Stereoscopic Images [C] // IEEE International Conference on Image Processing. Washington, USA: IEEE, 2018: 186-190.
- [4] ZHOU W, YU L, QIU W, et al. Utilizing Binocular Vision to Facilitate Completely Blind 3D Image Quality Measurement [J]. Signal Processing, 2016, 129(C): 130-136.
- [5] APPINA B, KHAN S, CHANNAPPAYYA S. No-reference Stereoscopic Image Quality Assessment Using Natural Scene Statistics [J]. Signal Processing Image Communication, 2016, 43: 1-14.
- [6] ZHANG W, QU C, MA L, et al. Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network [J]. Pattern Recognition, 2016, 59 (C): 176-187.
- [7] SHEN L, LEI J, HOU C. No-reference stereoscopic 3D image quality assessment via combined model [J]. Multimedia Tools and Applications, 2017, 77(7): 8195-8212.
- [8] LI Y, HU X. No-Reference Stereoscopic Image Quality Assessment Using Natural Scene Statistics [C] // International Conference on Multimedia and Image Processing. Washington, USA: IEEE Computer Society, 2017: 123-127.
- [9] LIN C T, LIU T J, LIU K H. Visual Quality Prediction on Distorted Stereoscopic Images [C] // IEEE International Conference on Image Processing. Washington, USA: IEEE, 2018: 3480-3484.
- [10] DING Y, DENG R, XIE X, et al. No-Reference Stereoscopic Image Quality Assessment Using Convolutional Neural Network for Adaptive Feature Extraction [J]. IEEE Access, 2018, 6: 37595-37603.
- [11] 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.
- [12] HIRSCHMULLER H. Stereo Processing by Semiglobal Matching and Mutual Information [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(2): 328-341.
- [13] WEIS D, LAI S H. Fast Template Matching Based on Normalized Cross Correlation with Adaptive Multilevel Winner Update [J]. IEEE Transactions on Image Processing, 2008, 17(11): 2227-2235.
- [14] LI JIAN J Q, LI Y C. Research Progress on Calculation Method of Disparity Map in Stereo Matching [J]. Remote Sensing Information, 2017, 32(2): 7-14.
- [15] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Science, 2014: 1409-1556.
- [16] HARA K, SAITO D, SHOUNO H. Analysis of Function of Rectified Linear Unit Used in Deep Learning [C] // International Joint Conference on Neural Networks. Washington, USA: IEEE, 2015: 1-8.
- [17] ALAIN H, ZIOU D. Image Quality Metrics: PSNR vs. SSIM [C] // 2010 International Conference on Pattern Recognition. Washington, USA: IEEE Computer Society, 2010: 2366-2369.
- [18] WANG Z, BOVIK A C, SHEIKH H R, et al. Image Quality Assessment: From Error Visibility to Structural Similarity [J]. IEEE Trans on Image Processing, 2004, 13(4): 600-612.
- [19] WANG X, MA L, KWONG S, et al. Quaternion representation based visual saliency for stereoscopic image quality assessment [J]. Signal Processing, 2018, 145: 202-213.
- [20] BENOITA. LE CALLET P, CAMPISI P, et al. Quality Assessment of Stereoscopic Images [J]. EURASIP Journal on Image and Video Processing, 2007, 2008(1): 1-13.
- [21] RYU S, KIM D H, SOHN K. Stereoscopic Image Quality Metric Based on Binocular Perception Model [C] // IEEE International Conference on Image Processing. Washington, USA: IEEE, 2013: 609-612.
- [22] TIAN S, ZHANG L, MORIN L, et al. NIQSV: A No Reference Image Quality Assessment Metric for 3D Synthesized Views [C] // IEEE International Conference on Acoustics. Washington, USA: IEEE, 2017: 1248-1252.
- [23] FARID M S, LUCENTEFORTE M, GRANGETTO M. Evaluating Virtual Image Quality using the Side-Views Information Fusion and Depth Maps [J]. Information Fusion, 2018, 43: 47-56.
- [24] YANG J C, SIM K, JIANG B, et al. No-reference stereoscopic image quality assessment based on hue summation-difference mapping image and binocular joint mutual filtering [J]. Applied Optics, 2018, 57(14): 3915-3926.



ZHU Ling-ying, master student. Her main research interests include image quality assessment and so on.



SANG Qing-bing, Ph.D, professor. His main research interests include image processing, quality assessment and machine learning.