

主办:中国科学院空天信息创新研究院 中国图象图形学学会 北京应用物理与计算数学研究所











第27卷第2期(总第310期) 2022年2月16日

中国精品科技期刊 中国国际影响力优秀学术期刊 中国科技核心期刊 中文核心期刊

版权声明

凡向《中国图象图形学报》投稿,均视 为同意在本刊网站及CNKI等全文数据 库出版,所刊载论文已获得著作权人的 授权。本刊所有图片均为非商业目的使 用,所有内容,未经许可,不得转载或 以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use. 主管单位 中国科学院 主办单位 中国科学院空天信息创新研究院 中国图象图形学学会 北京应用物理与计算数学研究所

中国图象图形学报

主 編 吴一戎 编辑出版 《中国图象图形学报》编辑出版委员会 通信地址 北京市海淀区北四环西路19号 邮 编 100190 电子信箱 jig@aircas.ac.cn 电 话 010-58887035 网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号

总 发 行	北京报刊发行局		
订 购	全国各地邮局		
海外发行	中国国际图书贸易集团有限	灵公司	
	(邮政信箱:北京399信箱	邮编:	100048)
印刷装订	北京科信印刷有限公司		

Journal of Image and Graphics

Title inscription: Song Jian

Jian Monthly, Started in 1996

 Superintended by
 Chinese Academy of Sciences

 Sponsored by
 Aerospace Information Research Institute, CAS

 China Society of Image and Graphics
 Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

 Distributed by
 Beijing Bureau for Distribution of Newspapers and Journals

 Domestic
 All Local Post Offices in China

 Overseas
 China International Book Trading Corporation (P.O.Box 399, Beijing 100048,P.R.China))

 Printed by
 Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB ISSN 1006-8961 CODEN ZTTXFZ

国外发行代号 M1406 国内邮发代号 82-831 国内定价 60.00元

刊名题字:宋健 月刊(1996年创刊)



三维视觉与智能图形专刊简介 陈宝权,虞晶怡,周昆,郭裕兰,黄惠,刘利刚,刘烨斌,徐凯,章国锋,周晓巍 ………… 0289

1000	And in case of the local division of the loc	5	
1111		100	-
1.00	1000	-	111
		and the second se	

面向本征图像分解的高质量 渲染数据集与非局部卷积网 络(第0404页)

综述 基于深度学习的单目深度估计技术综述	
米魏,朱孟飞,张明华,赵丹枫,贺琪 	0292
^{条红星,刘琅涛,谭博儿} 三维点云配准方法研究进展 李建微,占家旺	0349
多源融合SLAM的现状与挑战 王金科, 左星星, 赵祥瑞, 吕佳俊, 刘勇	0368
深度学习单目深度估计研究进展 罗会兰,周逸风 ····································	0390
数据集论文	
面同本征图像分解的局质量渲染数据集与非局部卷枳网络 王玉洁,樊庆楠,李坤,陈冬冬,杨敬钰,卢健智,Dani Lischinski,陈宝权 ····································	0404
深度估计与三维重建 高效鲁赫三维结构化重建	
高双音译二维和诗心里姓 潘珊珊,吕佳辉,方昊,黄惠····································	0421
贾迪,王子滔,李宇扬,金志楊,刘泽洋,吴思 ····································	0435
晏敏,王军政,李静	0447
苗国超,魏阳杰,侯威翰····································	0461
刘会杰,柏正尧,程威,李俊杰,许祝 ········ 单目相机轨迹的真实尺度恢复	0475
刘忠博,房立金 ····································	0486
基于显著性图的点云替换对抗攻击 刘复昌、南博 缪永伟····································	
	0500
边收缩池化的网格变分自编码器 袁字杰,来爆坤,杨洁,段琦,傅红波,高林 ····································	0500
边收缩池化的网格变分自编码器 袁宇杰,来煜坤,杨洁,段琦,傅红波,高林 雅可比矩阵引导的无翻转体映射生成 徐茂峰,刘利刚	0500 0511 0525
 边收缩池化的网格变分自编码器 袁宇杰,来煜坤,杨洁,段琦,傅红波,高林 雅可比矩阵引导的无翻转体映射生成 徐茂峰,刘利刚 嵌入Transformer结构的多尺度点云补全 刘心溥,马燕新,许可,万建伟,郭裕兰 	0500 0511 0525 0538
边收缩池化的网格变分自编码器 袁宇杰,来煜坤,杨洁,段琦,傅红波,高林 雅可比矩阵引导的无翻转体映射生成 徐茂峰,刘利刚 嵌入Transformer结构的多尺度点云补全 刘心溥,马燕新,许可,万建伟,郭裕兰 三维点云分割	0500 0511 0525 0538
边收缩池化的网格变分自编码器 袁宇杰,来煜坤,杨洁,段琦,傅红波,高林 雅可比矩阵引导的无翻转体映射生成 徐茂峰,刘利刚 嵌入Transformer结构的多尺度点云补全 刘心溥,马燕新,许可,万建伟,郭裕兰 三维点云分割 利用隐式解码器的三维模型簇协同分割 杨军,张敏敏	0500 0511 0525 0538 0550
边收缩池化的网格变分自编码器 袁宇杰,来煜坤,杨洁,段琦,傅红波,高林 雅可比矩阵引导的无翻转体映射生成 徐茂峰,刘利刚 嵌入Transformer结构的多尺度点云补全 刘心溥,马燕新,许可,万建伟,郭裕兰 三维点云分割 利用隐式解码器的三维模型簇协同分割 杨军,张敏敏 面向形状特征的多维度多层级点云分析 徐嘉利,方志军,伍世虔	0500 0511 0525 0538 0550 0562
边收缩池化的网格变分自编码器 袁宇杰,来煜坤,杨洁,段琦,傅红波,高林 雅可比矩阵引导的无翻转体映射生成 徐茂峰,刘利刚 嵌入Transformer结构的多尺度点云补全 刘心溥,马燕新,许可,万建伟,郭裕兰 三维点云分割 利用隐式解码器的三维模型簇协同分割 杨军,张敏敏 面向形状特征的多维度多层级点云分析 徐嘉利,方志军,伍世虔 多特征融合与几何卷积的机载LiDAR点云地物分类 戴莫凡,邢帅,徐青,李鹏程,陈坤	0500 0511 0525 0538 0550 0562 0574
边收缩池化的网格变分自编码器 袁宇杰,来煜坤,杨洁,段琦,傅红波,高林 雅可比矩阵引导的无翻转体映射生成 徐茂峰,刘利刚 嵌入Transformer结构的多尺度点云补全 刘心溥,马燕新,许可,万建伟,郭裕兰 三维点云分割 利用隐式解码器的三维模型簇协同分割 杨军,张敏敏 面向形状特征的多维度多层级点云分析 徐嘉利,方志军,伍世虔 多特征融合与几何卷积的机载LiDAR点云地物分类 戴莫凡,邢帅,徐青,李鹏程,陈坤 图像视频分析 聚合细粒度特征的深度注意力自动裁图	0500 0511 0525 0538 0550 0562 0574
边收缩池化的网格变分自编码器 袁宇杰,来煜坤,杨洁,段琦,傅红波,高林 雅可比矩阵引导的无翻转体映射生成 徐茂峰,刘利刚 嵌入Transformer结构的多尺度点云补全 刘心溥,马燕新,许可,万建伟,郭裕兰 三维点云分割 利用隐式解码器的三维模型簇协同分割 杨军,张敏敏 面向形状特征的多维度多层级点云分析 徐嘉利,方志军,伍世虔 多特征融合与几何卷积的机载LiDAR点云地物分类 戴莫凡,邢帅,徐青,李鹏程,陈坤 图像视频分析 聚合细粒度特征的深度注意力自动裁图 方玉明,钟裕,鄢杰斌,刘丽霞 单幅人脸图像的全景纹理图生成方法	0500 0511 0525 0538 0550 0562 0574
边收缩池化的网格变分自编码器 袁字杰,来煜坤,杨洁,段琦,傅红波,高林 雅可比矩阵引导的无翻转体映射生成 徐茂峰,刘利刚 嵌入Transformer结构的多尺度点云补全 刘心溥,马燕新,许可,万建伟,郭裕兰 三维点云分割 利用隐式解码器的三维模型簇协同分割 杨军,张敏敏 面向形状特征的多维度多层级点云分析 徐嘉利,方志军,伍世虔 多特征融合与几何卷积的机载LiDAR点云地物分类 戴莫凡,邢帅,徐青,李鵬程,陈坤 图像视频分析 聚合细粒度特征的深度注意力自动裁图 方玉明,钟裕,鄢杰斌,刘丽霞 单幅人脸图像的全景纹理图生成方法 刘洋,樊养余,郭哲,吕国云,刘诗雅 时空联合视差优化的立体视频重定向	0500 0511 0525 0538 0550 0562 0574 0586 0602
 边收缩池化的网格变分自编码器 袁字杰,来煜坤,杨洁,段琦,傅红波,高林 雅可比矩阵引导的无翻转体映射生成 徐茂峰,刘利刚 嵌入Transformer结构的多尺度点云补全 刘心溥,马燕新,许可,万建伟,郭裕兰 三维点云分割 利用隐式解码器的三维模型簇协同分割 杨军,张敏敏 面向形状特征的多维度多层级点云分析 徐嘉利,方志军,伍世虔 多特征融合与几何卷积的机载LiDAR点云地物分类 戴莫凡,邢帅,徐青,李鹏程,陈坤 图像视频分析 聚合细粒度特征的深度注意力自动裁图 方玉明,钟裕,鄢杰斌,刘丽霞 单幅人脸图像的全景纹理图生成方法 刘洋,樊养余,郭哲,吕国云,刘诗雅 时空联合视差优化的立体视频重定向 金康俊,柴雄力,邵枫 松的 表示与识别 	0500 0511 0525 0538 0550 0562 0574 0586 0602 0614
边收缩池化的网格变分自编码器 袁宇杰,来煜坤,杨洁,段琦,傅红波,高林 雅可比矩阵引导的无翻转体映射生成 徐茂峰,刘利刚 嵌入Transformer结构的多尺度点云补全 刘心溥,马燕新,许可,万建伟,郭裕兰 三维点云分割 利用隐式解码器的三维模型簇协同分割 杨军,张敏敏 面向形状特征的多维度多层级点云分析 徐嘉利,方志军,伍世虔 多特征融合与几何卷积的机载LiDAR点云地物分类 戴莫凡,邢帅,徐青,李鹏程,陈坤 图像视频分析 聚合细粒度特征的深度注意力自动裁图 方玉明,钟裕,鄢杰斌,刘丽霞 单幅人脸图像的全景纹理图生成方法 刘洋,樊养余,郭哲,吕国云,刘诗雅 时空联合视差优化的立体视频重定向 金康俊,柴雄力,邵枫 形状的全尺度可视化表示与识别 闵睿朋,李一凡,黄瑶,杨剑宇,钟宝江 结合掩码定位和漏斗网络的6D姿态估计	0500 0511 0525 0538 0550 0562 0574 0586 0602 0614 0628



高效鲁棒三维结构化重建 (第0421页)



利用隐式解码器的三维模型 簇协同分割(第0550页)

Volume 27, Number 2 Published February 16, 2022 **CONTENTS** JOURNAL OF IMAGE AND GRAPHICS



High quality rendered dataset and non-local graph convolutional network for intrinsic image decomposition(P0404)



A review of monocular depth estimation techniques based on deep learning	
Song Wei, Zhu Mengfei, Zhang Minghua, Zhao Danfeng, He Qi	0292
Review on deep learning rigid point cloud registration	
Qin Hongxing, Liu Zhentao, Tan Boyuan	0329
Review on 3D point cloud registration method	
Li Jianwei, Zhan Jiawang	0349
Review of multi-source fusion SLAM: current status and challenges	
Wang Jinke, Zuo Xingxing, Zhao Xiangrui, Lyu Jiajun, Liu Yong	0368
Review of monocular depth estimation based on deep learning	
Luo Huilan, Zhou Yifeng	0390

Dataset

High quality rendered dataset and non-local graph convolutional network for intrinsic image decomposition

Wang Yujie, Fan Qingnan, Li Kun, Chen Dongdong, Yang Jingyu, Lu Jianzhi, Dani Lischinski, Chen Baoquan 0404

Depth Estimation & 3D Reconstruction

Efficient and robust 3D structure-aware reconstruction
Pan Shanshan, Lyu Jiahui, Fang Hao, Huang Hui
Multi-stage guidance network for constructing dense depth map based on LiDAR and RGB data
Jia Di, Wang Zitao, Li Yuyang, Jin Zhiyang, Liu Zeyang, Wu Si
Reliable binocular disparity estimation based on multi-scale similarity recursive search
Yan Min, Wang Junzheng, Li Jing 0447
Scene depth extraction and application of microscopic optical blur image
Miao Guochao, Wei Yangjie, Hou Weihan 0461
Fusion attention mechanism and multilayer U-Net for multiview stereo
Liu Huijie, Bai Zhengyao, Cheng Wei, Li Junjie, Xu Zhu
Monocular camera trajectory recovery with real scale
Liu Sibo, Fang Lijin ·······0486
3D Shape Analysis
Point cloud replacement adversarial attack based on saliency map
Liu Fuchang, Nan Bo, Miao Yongwei
Mesh variational auto-encoders with edge contraction pooling
Yuan Yujie Lai Yukun Yang Jie Duan Qi Fu Hongbo Gao Lin
Generation of foldover-free volumetric mapping quided by Jacobian matrix
Xu Maofeng Liu Ligang
Multi-scale Transformer based point cloud completion network
Liu Xinnu Ma Vanvin Xu Ke Wan Jianwei Guo Vulan
3D Point Cloud Segmentation
Co-segmentation of 3D shape clusters based on implicit decoder
Vang Jun Zhang Minmin
Multi dimonoional multi lavar point algud analysis far abana fasturas
Vu lieli Eena Zhijun Wu Shigion
Au Jiali, Fariy Zhijuli, Wu Shiyani UDAD point cloud based on multi feature fusion and geometric
Semantic segmentation of arborne LIDAR point cloud based on multi-reature rusion and geometric
Convolution
Dai Moran, Xing Shuai, Xu Qing, Li Pengcheng, Chen Kun
Imaga & Video Analysia
Inage & video Analysis
Deep attention guided image cropping with fine-grained feature aggregation
Fang Yuming, Zhong Yu, Yan Jiebin, Liu Lixia
Solo face image-based panoramic texture map generation
Liu Yang, Fan Yangyu, Guo Zhe, Lyu Guoyun, Liu Shiya
Optimizing spatiotemporal disparities for stereoscopic video retargeting
Jin Kangjun, Chai Xiongli, Shao Feng
Visualized all-scale shape representation and recognition
Min Ruipeng, Li Yifan, Huang Yao, Yang Jianyu, Zhong Baojiang

6D pose estimation based on mask location and hourglass network



Efficient and robust 3D structureaware reconstruction(P0421)



Co-segmentation of 3D shape clusters based on implicit decoder(P0550)

中图法分类号:TP391 文献标识码: A 文章编号: 1006-8961(2022)02-0404-17

论文引用格式: Wang Y J, Fan Q N, Li K, Chen D D, Yang J Y, Lu J Z, Dani Lischinski and Chen B Q. 2022. High quality rendered dataset and nonlocal graph convolutional network for intrinsic image decomposition. Journal of Image and Graphics, 27(02):0404-0420(王玉洁,樊庆楠,李坤,陈冬冬, 杨敬钰,卢健智, Dani Lischinski,陈宝权. 2022. 面向本征图像分解的高质量渲染数据集与非局部卷积网络. 中国图象图形学报, 27(02):0404-0420)[DOI:10.11834/jig. 210705]

面向本征图像分解的高质量渲染数据集 与非局部卷积网络

王玉洁1,樊庆楠2,李坤3,陈冬冬4,杨敬钰3,卢健智5,

Dani Lischinski⁶,陈宝权⁷*

山东大学,青岛 266237; 2. 腾讯 AI Lab,深圳 518057; 3. 天津大学,天津 300072;
 微软云人工智能,华盛顿 98052,美国; 5. 广东三维家信息科技有限公司,广州 510000;
 耶路撒冷希伯来大学,耶路撒冷 91904,以色列; 7. 北京大学,北京 100091

摘 要:目的本征图像分解是计算视觉和图形学领域的一个基本问题,旨在将图像中场景的纹理和光照成分分离 开来。基于深度学习的本征图像分解方法受限于现有的数据集,存在分解结果过度平滑、在真实数据泛化能力较差 等问题。方法 首先设计基于图卷积的模块,显式地考虑图像中的非局部信息。同时,为了使训练的网络可以处理 更复杂的光照情况,渲染了高质量的合成数据集。此外,引入了一个基于神经网络的反照率图像优化模块,提升获得 的反照率图像的局部平滑性。结果 将不同方法在所提的数据集上训练,相比之前合成数据集 CGIntrinsics 进行训练 的结果,在 IIW (intrinsic images in the wild)测试数据集的平均 WHDR (weighted human disagreement rate)降低了 7.29%,在 SAW (shading annotations in the wild)测试集的 AP (average precision)指标上提升了2.74%。同时,所提出 的基于图卷积的神经网络,在 IIW、SAW 数据集上均取得了较好的结果,在视觉结果上显著优于此前的方法。此外, 利用本文算法得到的本征结果,在重光照、纹理编辑和光照编辑等图像编辑任务上,取得了更优的结果。结论 所提 出的数据集质量更高,有利于基于神经网络的本征分解模型的训练。同时,提出的本征分解模型由于显式地结合了 非局部先验,得到了更优的本征分解结果,并通过一系列应用任务进一步验证了结果。 关键词;图像处理;图像理解;本征图像分解;图卷积网络(GCN);合成数据集

High quality rendered dataset and non-local graph convolutional network for intrinsic image decomposition

Wang Yujie¹, Fan Qingnan², Li Kun³, Chen Dongdong⁴, Yang Jingyu³, Lu Jianzhi⁵, Dani Lischinski⁶, Chen Baoquan⁷*

1. Shandong University, Qingdao 266237, China; 2. Tencent AI Lab, Shenzhen 518057, China;

3. Tianjin Uniersity, Tianjin 300072, China; 4. Microsoft Cloud AI, Washington 98052, USA;

 Sunvega Company, Guangzhou 510000, China; 6. The Hebrew University of Jerusalem, Jerusalem 91904, Israel; 7. Peking University, Beijing 100091, China

收稿日期:2021-08-06;修回日期:2021-11-24;预印本日期:2021-12-01

^{*}通信作者:陈宝权 baoquan@pku.edu.cn

基金项目:国家自然科学基金项目(62136001)

Supported by: National Natural Science Foundation of China (62136001)

Abstract: Objective Intrinsic decomposition is a key problem in computer vision and graphics applications. It aims at separating lighting effects and material-oriented characteristics of object surfaces of the depicted scene within the image. Intrinsic decomposition from a single input image is highly ill-posed since the amount of unknowns is twice of the known values. Most classical approaches model intrinsic decomposition task with handcrafted priors to generate reasonable decomposition results. But they perform poorly in complicated scenarios as the prior knowledge is too limited to model complicated lightmaterial interactions in real-world scenes. Deep neural network based methods can automatically learn the knowledge from data to avoid using handcrafted priors to model the task. However, due to the dependency on training datasets, the performance of current deep learning based methods is still limited because of various constraints in the current intrinsic datasets. Moreover, the learned networks tend to suffer from poor generalization once there is a large difference between the training and target domain. Another issue of deep neural network based methods is that the limited receptive field probably constrains the ability of the models to exploit the non-local information in the intrinsic component prediction process. Method A graph convolution based module is designed to fully utilize the non-local cues within the input feature space. The module takes a feature map as input and outputs a feature map with same resolution as the input feature map. For producing the output feature vector for each position, the module uses information that includes the feature of itself, the information extracted from the local neighborhood and the information aggregated from the non-local neighbors that are likely to be very distant. The full intrinsic decomposition framework is constructed by integrating the devised non-local feature learning module into a U-Net network. In addition, to improve the piece-wise smoothness of the produced albedo results, we incorporate a neural network based image refinement module into the full pipeline, which is able to adaptively remove unnecessary artifacts while preserving structural information within the scenes depicted in input images. Simultaneously, there are noticeable limitations in existing intrinsic image datasets including limited sample amount, unrealistic scene and achromatic lighting in shading and sparse annotations, which will cause generalization issues for deep learning models and limit the decomposition performance as well. A new photorealistic rendered dataset for intrinsic image decomposition is proposed, which is rendered by leveraging large-scale 3D indoor scene models, along with high-quality textures and lighting to simulate the real-world environment. The chromatic shading components are first implemented. Result To validate the effectiveness of the proposed dataset, several state-of-the-art methods are trained on both the proposed dataset and CGIntrinsics dataset, a previously proposed dataset, and tested on intrinsic image evaluation benchmarks, i. e., intrinsie images in the wild (IIW) /shading annotations in the wild (SAW) test sets. Compared to the variants trained on CGIntrinsics dataset, the variants trained on the proposed dataset demonstrate a 7.29% improvement in averaging weighted human disagreement rate (WHDR) on IIW test set and a 2.74% gain for average precision (AP) on SAW test set. Simultaneously, the proposed graph convolution based network achieves comparable quantitative results on both IIW and SAW test sets and gets significantly better qualitative results. To further investigate the intrinsic decomposition quality for different methods, a number of application tasks including re-lighting and texture/lighting editing are conducted utilizing the generated intrinsic components. The proposed method demonstrates more promising application effects comparing with two state-of-the-art methods, further highlighting its superiority and application potential. Conclusion Based on the non-local priors in classical methods for intrinsic image decomposition, a graph convolutional network for intrinsic decomposition is proposed, in which non-local cues are utilized. To mitigate the issues existed in current intrinsic image datasets, a new high quality photorealistic dataset is rendered, which provides dense labels for albedo and shading. The depicted scenes in the images of the proposed dataset have complicated textures and illuminations that closely approximate general indoor scenes in reality, which helps to mitigate the domain gap issues. The shading labels in this dataset first consider chromatic lighting, which allows the neural networks to better separate material properties and lighting effects, especially for the effects introduced by inter-reflections between diffuse surfaces. The decomposition results of both the proposed method and two current state-of-the-art methods are applied to a range of application scenarios, visually demonstrating the superior decomposition quality and application potentials of the proposed method.

Key words: image processing; image understanding; intrinsic image decomposition; graph convolutional neural network (GCN); synthetic dataset

0 引 言

406

图像的物理成分,如反照率和阴影,对于许多计 算机视觉和图形应用来说十分关键。提取这些关键 成分是一个重要的中级视觉问题,称为本征图像分 解,该问题由 Barrow 和 Tenebaum(1978)首次定义。 在理想的漫反射环境中,一幅输入图像的每个像素 可以分解为反照率和亮度的乘积。

由于未知量的数目是已知量的两倍,从单个输入图像恢复场景的反照率成分和亮度成分是高度不适定的。但由于其巨大的应用潜力,这项任务一直受到学者们的广泛关注。之前的工作提出了各种基于先验和统计的模型,包括 Retinex 模型(Land 和 McCann,1971)、非局部纹理线索(Zhao 等,2012)和 全局稀疏先验(Gehler 等,2011;Shen 和 Yeo,2011)。

与许多其他具有挑战性的问题一样,基于深度 学习的方法逐渐被用以克服本征图像分解任务的不 适定性。在基于深度学习的本征图像分解方法中, 通常使用一个编码器—解码器的深度学习网络结构 从输入的图像中恢复场景的本征成分(Fan 等, 2018; Li 和 Snavely, 2018a, b; Narihira 和 Gehler, 2015b; Shi 等, 2017)。为了实现反照率图像的局部 平滑性质,这些方法通常引入后处理的滤波模块 (Fan 等, 2018; Li 和 Snavely, 2018a; Nestmeyer 和 Gehler, 2017)或在目标函数中加入正则化项(Li 和 Snavely, 2018a, b)来实现平滑先验。然而, 一般的 卷积神经网络结构较为擅长保持滤波器的空间局部 性, 从而由于其感受野大小的限制, 可能会影响本征 图像分解任务的性能, 因为从图像中分解本征成分 需要利用整幅图像的信息。

受到图卷积神经网络在形状和语义理解任务中 的成功应用,以及经典的本征图像分解算法(Bi等, 2015;Chen 和 Koltun, 2013;Shen 等, 2008;Shen 和 Yeo, 2011;Sinha 和 Adelson, 1993;Zhao 等, 2012)中 所采用的非局部稀疏性先验的启发,本文提出了一 种适合于本征图像分解任务的非局部图卷积网络。 提出的非局部图卷积运算的设计借鉴了用于形状分 类的图卷积网络(graph convolutional neural network, GCN)(Simonovsky 和 Komodakis, 2017),但针对 2D 图像结构进行了修改。提出的图卷积网络将特征图 中的点视为图中的一个顶点,对特征图像中的每个 点,与在整幅特征图上定义的非局部邻域点之间建 立连接,进行特征融合,从而使得图卷积层学习更多 的全局知识。

另一方面,数据集对基于深度学习网络的方法 非常重要(沙浩和刘越,2021)。一旦训练数据和测 试数据之间存在较大的差别,基于学习的方法往往 泛化能力较差。并且,目前现有的本征图像数据集 均存在缺陷,例如数据量不足(MIT (Grosse 等, 2009))、场景不够逼真(MPI-Sintel (Butler 等, 2012))、仅包含单个物体的简单场景(ShapeNet (Shi 等,2017)),或只用于弱监督的稀疏注释(IIW) (intrinsic images in the wild) (Bell 等, 2014), SAW (shading annotations in the wild) (Kovacs 等, 2017))。这些缺点使得深度学习技术在本征图像 分解这个任务上无法充分发挥其性能。Li 和 Snavely(2018a)提出了一个基于 SUNCG 室内场景数据集 (Song 等,2017) 渲染的逼真的本征图像数据集。尽 管该方法提出的数据集在场景的逼真度和渲染图片 的质量上均有明显提升,但是该数据集中的图片中 仍然具有明显的噪声,以及该数据集中并未提供逐 像素的亮度图标签,并在使用时默认亮度图中光照 是白光。

为了克服上述局限性,提出了一个新的基于真 实感绘制的本征图像数据集,该数据集中的渲染数 据来自大规模设计精细的3维室内场景模型,并结 合高质量的纹理和光照来模拟真实环境。提出的本 征图像数据集提供了逐像素的亮度图标签,且亮度 图中没有使用白光光照。实验结果表明,所提出的 数据集比现有的本征分解数据集具有更好的图像质 量,有效地缓解了基于学习的方法在真实图像上的 泛化能力。

此外,对所提出的方法与前沿的本征图像分解 方法进行了综合比较,不仅在主流的本征图像评价 基准(IIW/SAW测试集)上进行了比较,而且在各种 图像编辑应用场景中更加直观地对比了不同方法产 生的本征分解结果的质量。

本文的贡献可总结如下:

1)提出了第一个针对本征图像分解问题的图 卷积网络,在网络设计中显示利用了非局部图像 信息。

2)通过大规模真实感渲染,提出了一个新的场 景级本征图像数据集,并在亮度图像标签中提供了 非白光光照,使亮度图像分量更加逼真,从而使得在 本文数据集上训练的深度学习网络具有更好的泛化 能力。

3)对提出的方法及数据集在基准评价指标上 进行了测试和对比,并在一系列应用任务上对比分 析了本文方法所产生的分解结果的质量。

1 相关工作

20世纪70年代,从图像中分解场景的物理属 性就开始引起学术界的关注(Barrow和Tenenbaum, 1978;Land和McCann,1971)。Barrow和Tenenbaum (1978)介绍了本征图像的定义,此后众多的本征图 像分解算法相继提出,包括一系列经典算法(Chen 和 Koltum, 2013;Shen和Yeo,2011;Shi等,2015; Zhao等,2012)和最近的深度学习方法(Fan等, 2018;Li等,2018a;Li等,2018b;Ma等,2018)。由 于对单幅图像进行本征分解的不适定性,一部分方 法探索使用额外的信息来缓解这种困难,例如使用 同一场景下的图像序列、深度信息(Chen和Koltum, 2013;Wang等2017)和用户涂鸦等。本文主要研究 基于单幅图像的本征图像估计,适用于最常见的没 有提供额外信息的真实场景。

1.1 方法

1.1.1 传统方法

在传统方法中, Retinex 理论(Land 和 McCann, 1971)影响深远,它假设反照率图像是分块平滑的, 亮度图像内整体变化比较平缓,成为之后许多基于 先验的本征图像分解模型的基石。为了进一步减少 问题的不适定性,此前的方法引入了许多其他先验, 其中最常用的是非局部先验(Chen 和 Koltum, 2013; Shen 等,2008; Zhao 等,2012) 和稀疏性约束(Gehler 等,2011; Shen 和 Yeo, 2011)。例如, Shen 等人 (2008)结合了非局部相关性——不相邻的两点,如 果它们具有相似的纹理,那么它们仍然极有可能具 有相同的反照率值。Shen 和 Yeo(2011)基于自然 图像中通常只包含几种颜色的假设,提出了一种全 局稀疏性约束。一些方法通过聚类的算法实现稀疏 先验: Garces 等人 (2012) 使用 K-means 算法在 CIELab 色彩空间内将像素聚类为一些像素组, Meka 等人(2016)在构建本征图像分解模型时使用了较 为简单的直方图聚类。此外,一些方法(Bell 等, 2014;Chen 和 Koltum,2013)在整个图像内构建点对 之间的连接,以利用非局部先验。综上,非局部先验 在传统的本征图像分解模型中得到广泛使用,并被 验证是有效的。

1.1.2 基于深度学习的方法

如 Barron 和 Malik (2015)所述,对真实世界中 的光照和几何分布的统计,对于解决本征图像分解 中的歧义性是有效的。2009 年以来,随着本征图像 数据集,包括 MIT、MPI-Sintel、IIW、SAW 等的发布, 以及深度学习技术在计算机视觉任务中取得的进 展,越来越多的工作使用深度学习技术来构建本征 图像分解模型(Fan 等,2018; Ma 等,2018; Narihira 等,2015a,b; Shi 等,2017; Zhou 等,2015)。为了提 高反照率分量的分块平滑性,一些方法探索将传统 方法与深度学习方法结合,如后处理滤波操作(Li 和 Snavely,2018a; Nestmeyer 和 Gehler,2017)、联合 学习的引导滤波器(Fan 等,2018)和平滑度损失函 数(Li 和 Snavely,2018a,b)。这些方法大多只考虑 反照率图像的局部平滑性,而没有显式地考虑反照 率图像的全局信息。

1.1.3 图卷积网络

图卷积网络(GCN)在需要处理非规则结构数 据的任务中,取得了明显的提升,如点云上的任务 (Yi 等, 2017; Wang 等, 2018)。基于谱图论, Bruna 等人(2013)设计图卷积的变体,该工作是 GCN 领 域的先驱工作。之后,GCN 被其他工作提升或扩 展,包括基于谱图卷积的工作(Henaff 等, 2015; Li 等,2018a),以及一些基于空域的图卷积工作 (Hamilton 等, 2017; Monti 等, 2017)。Simonovsky 和 Komodakis(2017)提出了基于边的图卷积,它在每 个点周围的邻域上进行加权聚合,并根据边上标 签确定连接权重。对于本征图像分解任务,本文 方法将特征图中的每一个点视为一个节点,通过 将空间距离相近的点以及距离较远但局部特征近 似的点选取为邻域,在节点(特征)间建立非规则 的图。然后,本文方法将图卷积扩展到建立的图 结构上。

1.2 数据集

当给定足够多样的训练数据时,深度学习方法 能够从数据中学习到比手工先验更复杂的先验知 识。然而,由于现有的本征分解数据集均存在不足, 使得深度学习技术无法充分发挥出优势。例如,

MIT 数据集(Grosse 等,2009)是通过拍摄几百个物 体级的场景构建的.获得的图像没有背景,而且样例 较少。Bell 等人(2014) 在亚马逊平台上让用户在真 实图像的稀疏点对上标注相对的反照率关系.从而 构建了 IIW 本征数据集。Kovacs 等人(2017)通过 众包形式收集了多种形式的亮度图标注,该数据集 命名为 SAW。最近,一些大规模渲染的本征图像数 据集相继提出,包括从开源的 3D 动画电影中提取 的数据集(Butler 等, 2012)、室外环境的渲染数据集 (Baslamisli 等.2018)和在一些 3D 形状上渲染的数 据集(Shi等人,2017)。与本文所提数据集最相近 的本征分解数据集是 CGIntrinsics (Li 和 Snavely, 2018a), 它是基于 3D 室内模型数据集 SUNCG(Song 等,2017)进行渲染的,包含约20K幅图像。此外, Li 等人(2018b)提出了一个大规模渲染数据集 InteriorNet, 它包含了更多(20 M)的图像。但是, 在该数 据集中,为每个场景渲染了约1000幅图像,因此该 数据集在图像多样性上和 CGIntrinsics 数据集、本文 数据集接近。同时,以上两个数据集中都不提供渲染的亮度标签图像,而是通过使用输入图像与反照 率图像计算而来,从而造成亮度图像中的伪影,且图 像中有明显的噪声。

2 本文方法

首先描述如何在图像中构建和进行图卷积、提 出的整个本征分解网络框架以及本文算法中设计的 损失函数。

2.1 图卷积层

如图 1 所示,图卷积层接收的输入是特征图 $F^{in} \in \mathbb{R}^{e \times h \times w}$,其通道数为c,尺寸为 $h \times w_{\circ}$ 该图卷 积层具有 3 个分支,包括 $g_s \setminus g_l 和 g_{nl}$ 。其中, g_s 将 每个点自身的特征转换为 F^s_i , g_l 将每个点周围 3 × 3 邻域的特征融合为 $F^l_i \circ g_s$ 和 g_l 分别实现 1 × 1 和 3 × 3 的卷积,旨在提取每个点周围局部区 域内的信息。





Fig. 1 Non-local graph convolutional layer

对于图中的每个点 p_i , g_{nl} 分支在整幅特征图 内随机选择 k 个点作为邻域, 然后将 F_i^{in} 与每个邻域 点之间建立连接。在每个点与其邻域点建立起连接 后, 每个点的特征向量视做图的顶点, 所有的连接视 做图中的边, 则在整幅特征图上, 图 G 被建立, 表 示为

$$\begin{cases} \boldsymbol{G} = \{ \boldsymbol{V}, \boldsymbol{E} \} \\ \boldsymbol{V} = \{ p_i \mid \forall p_i \in \boldsymbol{F}^{\text{in}}, i \in [1, M] \} \\ \boldsymbol{E} = \{ e_{i,j} \mid e_{i,j} = \langle p_i, p_j \rangle, p_j \in \boldsymbol{N}(p_i), p_i \in \boldsymbol{V} \} \end{cases}$$

$$(1)$$

式中, $M = h \times w$, 是特征图中点的总数, $V 表示 F^{in}$ 中所有点的集合, $N(p_i) \ge p_i$ 的邻域点的集合, E表示建立的所有连接的集合, $e_{i,j}$ 表示点 p_i 与点 p_j 建立的连接。 g_{nl} 分支计算为

$$\boldsymbol{F}_{i}^{\mathrm{nl}} = \frac{1}{k} \sum_{p_{j} \in N(p_{i})} \boldsymbol{g}_{nl}(\boldsymbol{F}_{i}^{\mathrm{in}}, \boldsymbol{F}_{j}^{\mathrm{in}})$$
(2)

式中, F_i^{nl} 为 g_{nl} 分支的输出, g_{nl} 以加权的方式对 $N(p_i)$ 中的点的特征 F_i^{in} 进行组合, 计算为

$$_{nl}(\boldsymbol{F}_{i}^{\mathrm{in}},\boldsymbol{F}_{j}^{\mathrm{in}}) = \boldsymbol{w}_{i,j} \odot \boldsymbol{F}_{j}^{\mathrm{in}}$$
 (3)

式中, \odot 表示将两个向量的对应元素进行相乘, 权 重向量 $w_{i,j}$ 是根据 $p_i = p_j$ 之间的特征差异和位置上的距离进行计算, 即

$$\boldsymbol{w}_{i,j} = g_{nl}((\boldsymbol{F}_i^{\text{in}} - \boldsymbol{F}_j^{\text{in}}) \oplus d_{i,j})$$
(4)
式中, ① 表示将两个向量连接, g_{nl} 是通过一个多层
感知机实现的, $d_{i,j}$ 是 p_i 与 p_j 之间的距离, 计算为

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$
(5)

式中, (x_i, y_i) 和 (x_j, y_j) 是 p_i 与 p_j 在特征图 F^{in} 中的2 维坐标。得到3 个分支的输出后,图卷积层的最终输出计算为

第27卷/第2期/2022年2月

$$\boldsymbol{F}_{i}^{\mathrm{o}} = \boldsymbol{F}_{i}^{\mathrm{s}} + \boldsymbol{F}_{i}^{\mathrm{l}} + \boldsymbol{F}_{i}^{\mathrm{nl}}$$
(6)

2.2 本征图像分解网络

本文提出的本征图像分解网络采用了一个编码器一解码器的网络结构。如图 2 所示,整个网络的输入是一幅自然图像 I,首先经过一个由 6 个步长为 2 、卷积核大小为 4 × 4 的卷积层构成的编码器 h_e ,以提取不同尺度的特征。在不引入本文提出的非局部模块 h_{alm} 的情况下,编码器提取到的特征将被输入到 3 个解码器中, d_s 为亮度图解码器, d_a 为反照率图像解码器, d_a 为反照率图像解码器, d_a 为反照率图像解码器,

它们的结构与编码器的结构是对称的。每个解码器 具有6个连续的步长为2、卷积核大小为4×4的反 卷积层,分别用于产生反照率图像 Ã、亮度图像 Š 和反照率边缘图像 Ã。。

在编码器和解码器中,每个卷积层或反卷积层 之后均使用批标准化层(batch normalization)和 Re-LU 激活层。并且,由于编码器和每个解码器之间的 结构是对称的,在编码器和解码器的对应特征尺度 间,分别建立对应的跳过连接(skip connections)。



图 2 本文所提本征分解网络结构图 Fig. 2 Structure of the proposed intrinsic decomposition network

从图像中推理场景中的本征属性需要对整个图像的整体理解。例如,图3中左侧的例子中,地砖上呈现的粉色(图3中蓝色圈中的区域)是由较远的粉色橱柜间接反射而来的(图3中绿色圈中的区域)。为了正确地恢复地板区域的本征信息,深度神经网络需要利用整幅图的全局信息。对于图3中右面的例子,蓝色圈中的区域和绿色圈中的区域都属于地板区域,是由相同的材质制成的,因此它们具有相同的反照率信息。然而,由于阴影的存在,使得它们在输入图中的像素信息相差较大。因此,为了正确地估计绿色圈中区域的反照率信息,需要利用图片中相隔较远的区域的信息,如蓝色圈中的区域的信息。



图 3 全局理解对本征属性感知的必要性 Fig. 3 Requirement for global understanding of intrinsic property perception

然而,通常的卷积网络所产生的感受野大小是

有限的,因此只能学习较为局部的特征,并且没有显 式地考虑图像中距离较远的点之间的联系。为了克 服这个限制,在编码器—解码器结构中引入了由 4个连续的图卷积层构成的非局部注意力模块 h_{alm}, 从而使得网络框架可以显式地利用图中的非局部信 息。因此,3个编码器接收到的是经过 h_{alm} 处理的 特征,由于它们所预测成分的不同,其输出层的通道 数分别被设置为 3,3,1。

2.3 反照率优化模块

此外,所提的本征分解框架还包括一个反照率 图像优化模块 h, 该网络由 20 层卷积核为3×3、步 长为1 的卷积层构成。每一层卷积层之后使用实例 规范化层(instance normalization)和 ReLU 激活层。 第1 层的输入通道数为4(由输入图像和引导边缘 图像构成),最后一层的输出通道数为3,中间层的 输入/输出通道数均为64。此前,一些基于学习的 方法均利用了后处理的滤波模块对神经网络预测得 到的反照率图像 albedo 进行优化(Fan 等,2018;Li 和 Snavely,2018a;Nestmeyer 和 Gehler,2017);另一 些工作在损失函数中引入了稀疏先验以生成干净平 整的反照率图像。受这些工作的启发,所提的网络 框架引入一个优化模块对解码器预测的反照率图像 进行优化。与之前工作不同的是,所提框架中引入

的优化模块是基于神经网络的优化模块。

给定一幅输入图像以及对应的结构图,该基于 深度网络的优化模块可以在结构图像的引导下,保 持图像中的重要结构,同时去除图像中一些不重要 的细节。在训练阶段,该优化模块接受的输入是自 然图像和一幅边缘图像,并使用经过 Bi 等人(2015) 算法处理之后的输入图像作为训练的标签。其中输 入中使用的边缘图像是根据 Fan 等人(2018)的做 法从标签图像中提取的。经过 Bi 等人(2018)的做 法从标签图像中提取的。经过 Bi 等人(2015)算法 处理之后的图像,在图像颜色相近的区域内是分块 平滑的,这个特点符合反照率图像的特性。经过训 练后,该优化模块学习到如何在输入的边缘图像的 引导下,将输入图像映射到目标图像上。在测试阶 段,该优化模块接收网络预测的反照率图像和反照 率边缘图,输出优化之后的反照率图像*Ã*'。

由于本文所提网络框架中的反照率图像优化模 块是由一个深度学习网络实现的,相比于一个固定 的传统滤波算法,它支持更加复杂的图像操作,从而 实现了更好的优化效果,其作用在本文的消融实验 中得到了验证。

2.4 损失函数

对本文网络框架中所设计的损失函数进行描述。为了训练所提出的网络框架,建立一个具有照 片真实感的本征分解数据集。在该数据集中,为了 减小合成图像与自然图像之间的差异,除了逐像素 的反照率标签(A)和亮度标签(S),也渲染了4 个其他的光照成分,包括:高光成分(SP)、反射成 分(RE)、折射成分(RA)和自发光成分(SI)。 这些其他成分的渲染使得输入图像更接近自然图 像,但是在训练本文网络框架时,由于这些其他成分 不包含在反照率图像和亮度图像中,在损失函数中 需要将这些区域排除。首先计算一个蒙版图像

$$M_i = \begin{cases} 0 & (SP_i + RE_i + RA_i + SI_i) > 0\\ 1 & \ddagger \psi \end{cases}$$
(7)

式中,*i*表示像素索引,*M_i*是该像素的蒙版值。对于蒙版值为1的区域,损失函数定义为

$$\mathcal{L} = \mathcal{L}_{\rm IE} + \mathcal{L}_{\rm IR} \tag{8}$$

式中, \mathcal{L}_{IE} 为本征分解估计损失, 以训练本征分解网络, \mathcal{L}_{IE} 为监督反照率优化网络的损失。 \mathcal{L}_{IE} 用来监督本征分解网络预测得到的反照率图像 \tilde{A} 、亮度图像 \tilde{S} 和反照率边缘图像 \tilde{A}_{e} , 定义为

Vol. 27, No. 2, Feb. 2022

 $\mathcal{L}_{\rm IE} = \mathcal{L}_{\rm MSE} + \lambda_{\rm g} \mathcal{L}_{\rm g} + \lambda_{\rm r} \mathcal{L}_{\rm r}$ (9)

式中, λ_{g} 和 λ_{r} 是平衡系数,用来平衡不同损失项的 监督力度。 \mathcal{L}_{MSE} 和 \mathcal{L}_{g} 分别为预测图像与标签图像 之间的均方误差、图像梯度损失,即

$$\mathcal{L}_{\text{MSE}} = \|\widetilde{\boldsymbol{A}} - \boldsymbol{A}\|_2^2 + \|\widetilde{\boldsymbol{S}} - \boldsymbol{S}\|_2^2 +$$

$$\|\boldsymbol{A}_{e} - \boldsymbol{A}_{e}\|_{2}^{2} \tag{10}$$

 $\mathcal{L}_{g} = \|\nabla \tilde{A} - \nabla A\|_{2}^{2} + \|\nabla \tilde{S} - \nabla S\|_{2}^{2}$ (11) 式中, \tilde{A} 和 \tilde{S} 是网络预测得到的反照率图像和亮度 图像, \tilde{A}_{e} 是预测得到的反照率边缘图像, ∇ 表示沿 着 x, y 方向求取图像的梯度。 A, S, A_{e} 是对应的标 签图像。 A_{e} 是通过边缘提取算法从 A 中计算得到 的, \tilde{A}_{e} 将用于对 A 的优化模块 h_{r} 中。 \mathcal{L}_{r} 是一项用于 自监督的重建损失, 即假设物体表面的反射属性为 漫反射时, 输入图像可以由与其对应的反照率图像 和亮度图像重建, 该损失项定义为

$$\mathcal{L}_{r} = \|\tilde{\boldsymbol{A}} \odot \tilde{\boldsymbol{S}} - \boldsymbol{I}\|_{2}^{2}$$
(12)

式中, ⊙ 表示两个图像的对应像素相乘。并借鉴 Fan 等人(2018)的损失函数定义方式,将用于训练 优化模块的损失函数 *L*_R 定义为

$$\mathcal{L}_{\rm IR} = \|\widetilde{\boldsymbol{R}} - \boldsymbol{R}\|_2^2 \tag{13}$$

式中, **R** 是优化网络的输出结果, **R** 是输入图像 **I** 经 过 Bi 等人(2015)算法处理后的结果。在实现中,本 征分解网络和反照率优化网络是分开训练的,只在 测试阶段一起使用。

3 合成数据集

为了实现更高质量的本征分解效果,提出一个 具有照片真实感的渲染数据集,提供逐像素的反照 率和亮度标签。该数据集中包含 21 478 组由输入 图像、反照率图像和亮度图像组成的数据样例。在 训练所提出的本征分解网络时,数据集中 18 256 组 数据用于训练,其他数据用于测试。如图 4 和图 5 所示,提出的数据集具有更高的逼真度,主要来源于 以下因素:

1)场景布局。为了建立该数据集,从其室内设 计平台收集了5730个合成的3维室内场景模型, 类型包括客厅、卧室、厨房和浴室等。它们由数百名 专业设计师/艺术家设计,场景内物体的摆放与真实 室内场景高度吻合。

2)光照设置。在渲染数据集中的图像时,本文

王玉洁,樊庆楠,李坤,陈冬冬,杨敬钰,卢健智,Dani Lischinski,陈宝权 面向本征图像分解的高质量渲染数据集与非局部卷积网络

数据集中使用了来自多种光源的光照,以模拟真实 世界的视觉效果。除了常见的全局光照和相互反射 之外,本文数据集中的图像还包含由透明物体、镜子 和自反光物体引起的折射、镜面反射和灯光等效果。 如图 4 所示,这些其他的光照成分,使得渲染后的图 像更接近日常室内场景拍摄的图像。



图 4 其他光照成分对图像渲染的影响

Fig. 4 Comparison between rendered images with and without the extra illumination effects ((a) specular; (b) self-illumination;

(c) refraction; (d) reflection; (e) input image; (f) input image (without other illuminations); (g) enlarged regions)



(a) 本文数据集

(b) CGIntrinsics数据集

 $\label{eq:Fig.5} Fig. 5 \quad Comparison \ between \ the \ proposed \ dataset \ and \ CGIntrinsics \ dataset((a) \ the \ proposed \ dataset;(b) \ CGIntrinsics))$

3) 纹理多样性。为了使得渲染得到的数据包 含丰富的纹理,数据集中所包含的场景中的物体表 面的纹理是从约80万张材质贴图中随机采样得到 的。如图4所示,本文数据集所渲染的图像具有较 高的纹理多样性。

4) 渲染设置。本文数据集中的图像是采用 Embree4 渲染引擎,使用确定性蒙特卡罗(deterministic Monte carlo, DMC)算法渲染得到的。图像的分辨 率是1 280×960 像素,每个像素的采样数为3 228。 使用普通台式机渲染一张具有照片级真实感的图像 需要数小时的计算,十分耗时。为了加速渲染过程, 本文通过 32 台服务器进行分布式渲染,平均每张图 的渲染时间为 90 s。

对此前提出的本征图像分解数据集和本文数据

集进行了对比,比较结果总结于表1中。

如表1所示, MIT(Grosse 等, 2009)和 ShapeNet intrinsics (Shi 等, 2017)这两个数据集, 仅提供了物 体级别的图像及对应的逐像素标签, 与真实环境中 的场景有较大的差距。MIT 数据集受限于数据集的 大小, Shi 等人(2017)数据集中的图像与真实图像 差别较大。IIW(Bell 等, 2014)数据集和 SAW 数据 集(Kovacs 等, 2017)是建立在真实场景拍摄得到的 照片上的, 但是两者均只提供基于人类判断的稀疏 标注。此外, MPI-Sintel(Butler 等, 2012)数据集提 供了场景级的图像及逐像素的本征属性标签, 但是 其中的场景均来自于与现实环境差别很大的动画场 景。与本文数据集最相关的数据集是 CGIntrinsics 数据集(Li和Snavely, 2018a)。与该数据集相比, 本

Table 1 Comparison between different intrinsic datasets							
数据集	MIT	ShapeNet	MPI-Sintel	IIW	SAW	CGIntrinsics	本文数据集
年份	2010	2017	2012	2014	2017	2018	2021
场景类型	单个物体	单个物体	场景	场景	场景	场景	场景
图像获取方式	拍摄	渲染	渲染	拍摄	拍摄	渲染	渲染
图像数量	220	2 443 336	890	5 000	5 000	20 000	21 000
反照率标注形式	逐像素	逐像素	逐像素	稀疏点对标注	无	逐像素	逐像素
亮度图标注形式	逐像素	逐像素	逐像素	无	稀疏标注	逐像素	逐像素
亮度图光照	白光	白光	非白光	无	无	白光	非白光
图像尺寸/像素	300×400 等	256×256	1 024 ×436	640×480	640×480	640×480	$1\ 280 \times 960$

表 1 不同本征分解数据集的对比 able 1 Comparison between different intrinsic dataset

文数据集提供了更高分辨率和带有非白光光照的亮 度图像标签,而 CGIntrinsics 数据集只提供反照率标 签图像,亮度图像通过使用输入图像除以反照率图 像计算得到。由于该公式只对漫反射的区域满足, 那么在非漫反射的区域,计算得到的亮度图像会引 起误差。为了直观地体现本文数据集与 CGIntrinsics 数据集的差别,图5(a)中提供了来自本文数 据集的4组样例,图5(b)中展示了来自 CGIntrinsics 数据集的3组样例,从上到下依次展示了输入 图、反照率标签图像、亮度标签图像和放大的局部区 域。如图5所示,相较于 CGIntrinsics 数据集,本文 数据集展现了更逼真的视觉效果,具有更丰富的纹 理,在场景布局上更加复杂,并且图像中噪声更少。 本文数据集中的亮度图像中的光照是非白光的,更 加符合真实室内场景的光照。

4 实验结果

IIW/SAW 数据集中只提供了人工标注的一些 稀疏标签,并未提供逐像素的反照率图、亮度图标 签,因而无法计算如 MSE(mean squared error)、 PSNR(peak signal-to-noise ratio)、SSIM(structure similarity index measure)等图像质量评价指标上的 数值结果。因此,本文使用了基于 IIW、SAW 数据 集的标签类型而设计的评价指标 WHDR(weighted human disagreement rate),该指标越低越好,AP(average precision),该指标越高越好。

4.1 实现细节

本文本征图像分解网络模型使用 PyTorch 框架

实现,批大小为6,使用 Adam 算法进行优化。在训练过程中,初始学习率设置为0.01,在 IIW/SAW 数据集上进行微调时更改为0.0005。优化网络模块单独训练,学习率设置为0.01,批大小设置为4。

4.2 数据集的有效性

通过将所提数据集与该数据集进行详细比较, 来验证本文数据集对提升本征分解结果的有效性。 首先,在所提出的数据集和 CGIntrinsics 数据集上分 别训练了3种目前效果领先的本征分解深度学习网 络,然后在 IIW/SAW 数据集的测试集上对训练之 后的网络的分解结果进行测试,得到的测试结果汇 总在表2中。

为了公平比较不同数据集的质量,在测试时,对 所有方法均去掉了其中的后处理优化模块,在训练 时均使用作者提供的开源代码和参数设置。如表 2 中所示,与使用 CGIntrinsics 训练相比,使用本文数据 集训练的 3 种方法的平均测试结果取得了更高的性 能。具体表现为:对于 3 种方法上的平均结果,使用 本文数据集,WHDR 降低了 8.87%,在 SAW 测试集 上的准确率 AP 提升了 2.74%。这表明在不同的网 络框架和损失函数设置的情况下,使用本文数据集进 行训练,均可以获得更高的分解质量。同时,IIW/ SAW 数据集中的输入图片是真实拍摄的图像,表明 在本文数据集上训练的网络具有更好的泛化能力。

在使用 Shi 等人(2017)的方法和 Li 和 Snavely (2018a)的方法提出的网络结构进行训练时,本文 数据集使得网络在 IIW 测试集、SAW 测试集上产生 的结果,相较于使用 CGIntrinsics 数据集训练的结 果,均有显著提升。在使用 Fan 等人(2018)的方法 进行训练时,本文数据集使得网络在IIW测试指标

表 2	7	下同本征分解	数据集	的有效性对比
Table	2	Comparison	of the	effectiveness of
		different	datase	te

训练数据	方法	在 IIW 数据 集上的 WHDR/%	在 SAW 数 据集上的 AP/%
	Shi 等人(2017)	34. 55	92.33
CCL	Fan 等人(2018)	27.13	93.81
CGIntrinsics	Li 和 Snavely(2018a)	19. 94	92.09
	平均	27.21	92.74
	Shi 等人(2017)	20. 53	95.84
本文数据集	Fan 等人(2018)	20. 89	92.96
	Li 和 Snavely(2018a)	18.34	97.66
	平均	19. 92	95.48

注:加粗字体表示平均值最优结果。

WHDR 有显著提升,但是在 SAW 测试集上与使用 CGIntrinsics 数据集训练的版本有较小的差距 (0.9%)。这是由于本文数据集提供的亮度图标签 图像为三通道的彩色图像,与使用 CGIntrinsics 的单 通道亮度标签图像相比,神经网络预测得到的亮度 图在局部区域内相邻像素的数值之间存在一些差异 的可能性增大,该现象可以通过简单引入现有的图 像滤波模块进行解决。

4.3 在 IIW/SAW 数据集上的测试结果

对本文本征分解网络框架与之前最优的本征分 解算法进行了对比。沿用Li和Snavely(2018a)中 的做法,对所提出的网络框架在IIW/SAW训练集 上进行了微调。在微调阶段中,使用了Fan等人 (2018)和Li和Snavely(2018a)方法中使用的训练/ 测试集划分,并使用了在Li和Snavely(2018a)方法 中提出的顺序反射损失和基于SAW数据集标注形 式设计的亮度损失。此外,为了充分利用IIW数据 集中的标注,使用了由Li和Snavely(2018a)方法提 供的数据增强之后的标签。最后,IIW/SAW测试集 上的测试结果汇总在表3中。

如表 3 所示, 在仅使用本文数据集训练的情况 下,本文方法在 IIW 测试集上的 WHDR 为 17.92%, 在 SAW 测试集上的准确率为 96.17%, 即取得了优 于之前所有传统方法的结果, 并且超越了大部分的 基于深度学习的方法。特别地, 在没有微调的情况 下,本文方法在SAW测试集上的结果接近于Li和

	表3	在 IIW/SAW 测试集上的定量结果
Fable 3	Qua	antitative comparison on IIW/SAW test sets

方法	在 IIW 数据 集上的 WHDR/%	在 SAW 数 据集上的 AP/%
Retinex(color) (Grosse 等,2009)	26. 89	85.26
Garces 等人(2012)	25.46	92. 39
Zhao 等人(2012)	23.20	89.72
Bell 等人(2014)	20. 64	92.18
Shi 等人(2017)	59.40	81.30
Narihira 等人(2015b)	37.30	86.08
Zhou 等人(2015)	19. 95	86.34
Nestmeyer 和 Gehler(2017)	17.69	88.64
Fan 等人(2018)	14. 4 5	85.13
Li 和 Snavely(2018a)	14.80	96. 57
本文	17.92	96. 17
本文(微调后)	15.10	96.65

注:加粗字体表示每列最优值。

Snavely(2018a)方法在微调后的结果,这表明本文 方法具有较好的泛化能力。通过在 IIW 和 SAW 数 据集中的训练数据上微调,本文方法最终取得了较 好的数值结果。不同于 Li 和 Snavely(2018a)方法 及 Fan 等人(2018a)方法,本文方法仅在 IIW 和 SAW 数据集上微调,没有在整个训练阶段中使用 IIW 和 SAW 数据。这可能使得本文方法的 WHDR 误差略高于以上两种方法。

在图 6 中,展示了本文方法和另外两种方法 (Fan 等,2018;Li 和 Snavely,2018a)在 IIW 和 SAW 测试集上的分解结果。如图 6 所示,本文方法的本 征分解性能优于其他方法。不同于 Li 和 Snavely (2018a)方法、Fan 等人(2018)方法与 Nestmeyer 和 Gehler(2017)方法,本文方法产生的亮度图的结果 是具有彩色光照的。对于预测的反照率图像,如 图 6 所示,本文方法的结果颜色更自然。相较于本 文方法产生的结果,其他方法的结果没有将光照的 颜色和物体表面的材质颜色分开。因此,本文方法 产生的分解结果质量更高。

在 Fan 等人(2018)方法及 Nestmeyer 和 Gehler (2017)方法中,深度学习网络只预测反照率图像, 亮度图像由输入图像除以预测的反照率图像 得到。

反射率 图像

亮度 图像

反射率 图像





图 6 不同方法在 IIW/SAW 测试集上的分解结果

Fig. 6 Intrinsic decomposition results on images from IIW/SAW test sets for different methods((a) input image; (b) Nestmeyer and Gehler(2017);(c) Fan et al. (2018);(d) Li and Snavely(2018a);(e) ours;(f) ours (finetuned))

因此,反照率图像结果中不准确的区域,会影响 以上两种方法产生的亮度图像结果。由于 Li 和 Snavely(2018a)方法在损失函数中施加了平滑约束 项,使得该方法产生的结果十分平滑,甚至无法体现 场景中的很多几何信息。但是,在本文方法产生的 结果中,以上问题都未出现。

Gehler(2017)

此外,Nestmeyer和 Gehler(2017)以及 Zhou等人(2015)指出,在 IIW 数据集中的相对反照率标注中,由于标注为具有相同反照率关系的点对占据了所有标注的 2/3,使得该数据集中的标注向具有相同反照率的点对倾斜。例如,在不对输入图像进行本征分解的情况下,简单地将输入图像的数值从[0,1]范围缩放到[0.55,1]之间,即简单地将输入图像的像素之间的对比度降低,在 IIW 测试集上得到的 WHDR 误差值为 25.7%,甚至优于很多本征分解方法产生的结果。这个观察进一步表明了 IIW 测试集的标注中相等反照率的点对所占比例更高,如 Nestmeyer和 Gehler(2017)所述。因此,直接在 IIW

和 SAW 数据集上训练可能会导致深度学习网络过 拟合到其数据集所体现的数据分布,即倾向于预测 更平滑的反照率图像。

由于上述原因,本文方法首先在本文提出的合成数据集上训练,然后在 IIW 和 SAW 数据集上进行 微调,从而避免在 IIW 和 SAW 数据上的过度拟合。 使用本文的训练策略,提升了本征图像分解的视觉 结果,但可能使得本文方法在数值结果上的提升不 如在视觉结果上的提升显著。

4.4 在其他数据上的测试结果

基于深度学习的算法,在训练数据与测试数据 的分布相差较大时,性能极可能显著下降。由于本 文方法使用合成数据集训练,验证它在一般自然图 像上的本征分解结果是十分必要的。因此,本文在 图7中提供了3组不同方法在一般自然图像上的分 解结果。图7中显示,本文方法产生的结果明显优 于其他3种方法产生的分解结果。

中国图象图形学报

王玉洁,樊庆楠,李坤,陈冬冬,杨敬钰,卢健智,Dani Lischinski,陈宝权 面向本征图像分解的高质量渲染数据集与非局部卷积网络



图 7 不同方法在其他图片数据上的分解结果

Fig. 7 Intrinsic decomposition results on unseen images for different methods

((a) input; (b) Nestmeyer and Gehter(2017); (c) Fan et al. (2018); (d) Li and Snavely(2018a); (e) ours))

如图 7 第 1 行所示,由对比方法(Fan 等,2018; Li 和 Snavely,2018a;Nestmeyer 和 Gehter,2017)所产 生的反照率图像,被场景中光照颜色严重影响,并且 丢失了瓷器表面的花纹细节。相反,本文方法产生 的结果将光照颜色从反照率图像中分离出去,并且 保留了瓷器表面的纹理细节。在图 7 中第 2 个例子 中,场景中有较强的相互反射,其他 3 种方法都没有 将这些彩色的相互反射分解到亮度图中,但本文方 法成功地将这些成分分解到亮度图中,但本文方 法成功地将这些成分分解到亮度图结果中。图 7 中 最后一个示例展示了室外场景的情况,这种类型的 数据是没有出现在训练数据中的。本文方法产生的 结果可以正确地将船体上的字母分解到反照率图像 中,避免它出现在亮度图中,而其他方法的结果均将 这些纹理错误地引入到对应的亮度图结果中。综 上,本文方法产生的本征图像分解结果体现出较高 的质量。

5 应 用

作为一个中级视觉任务,本征图像分解的结果 只有在用于下游应用时才更有价值。将在几个基于 本征图像分解结果的应用场景上,进一步验证本文 方法产生的本征分解结果的质量。

5.1 纹理编辑

在进行纹理编辑任务时,首先利用本征分解算 法将输入图像分解为反照率图像和亮度图像,然后 在反照率图像中将需要修改的纹理进行编辑,之后 用编辑后的反照率图像和原始的亮度图像一起合成

在图 8 的第 1 个示例中,在 3 种方法分解得到 的反照率图像中,3 只小猪的眉毛分别被修改,然后 使用修改过的反照率的图像乘以对应的亮度图像, 得到 对应 的编辑 结果。如图 8 所示, Fan 等人 (2018)方法由于错误地将眉毛的纹理引入到亮度 图像中,使得对应的编辑结果中存在原始眉毛图案 的伪影。Li 和 Snavely(2018a)方法所产生的编辑结 果中在原来图案的附近也存在伪影,是由该方法产 生的亮度图中存在的纹理复制现象引起的。对于 图 8 中第 2 个示例,墙上两幅画的位置在对应的反 照率图像中分别被交换,然后与对应的亮度图结果 相乘得到输出图像。图 8 中显示,第 2 个示例产生 了与第 1 个示例类似的现象:利用 Fan 等人(2018) 方法和 Li 和 Snavely(2018a)方法产生的分解结果 得到的编辑结果中,两幅画中原始内容依然存在于 原来的位置,这是由残留在亮度图像中的纹理造成 的。此外,Li 和 Snavely(2018a)方法产生的反照率 结果过度平滑,因此第 2 个示例中,基于该方法的分 解结果得到的编辑结果中缺少画中的细节。





Fig. 8 Comparison between results for texture editing using intrinsic components from different methods ((a) input; (b) Fan et al. (2018); (c) Li and Snavely(2018a); (d) ours)

5.2 光照编辑

光照编辑是通过修改本征分解结果中的亮度图像,然后与未经修改的反照率图像相乘得到输出结果来进行,它可以用来体现反照率图像的质量。

图9中展示了光照编辑任务的两个示例。在第1 个示例中,不同方法预测得到的亮度图中,由台灯照 亮的区域被平滑。由于 Fan 等人(2018)方法以及 Li和 Snavely(2018a)方法均未能去除其预测的反照 率图像中的光照,因此这两种方法产生的编辑结果 中仍然存在台灯照亮的区域。相反,本文方法产生 的编辑结果中,该区域的光照已被去除。在第 2个示例中,通过将产生的亮度图变为灰度图,再 与原始的反照率图像相乘得到输出结果,达到去 除场景中光照颜色的目的。显然,基于其他方法 产生的结果中仍然保留着原始输入图像中的整体 光照颜色。

然而,基于本文方法产生得到的编辑结果中,原 始场景中的整体光照颜色则被全部去除。这是由于 其他两种方法估计的亮度图本身是灰度图,默认亮 度图中的光照为白光,导致输入图中的彩色光照信 息都被错误地分解到了反照率图像中。

此外,按照 Zhou 等人(2015)方法中的实验设 置,从 Boyadzhiev 等人(2013)以及 Li 和 Snavely (2018a)方法提供的数据集中选取了特殊的图像序 列数据,序列中的每幅图像中的纹理是严格一致 的,只有光照不同。对于提取到的图像序列,基于 不同的本征分解方法,进行了重光照任务,将序列 中的两幅图像首先进行本征分解,然后对两幅图 的亮度图像进行交换,再与两幅图像对应的原始 反照率图像进行有插乘,得到最终的光照编辑结果。 由图 10 中的两个示例显示,在视觉上,基于本文 方法的分解结果所得到的重光照结果,最接近与 交换后的亮度图所对应的原始输入图像。而对于 其他两种方法,重光照的结果总是存在颜色偏差, 这是由于以上两种方法没有将光照颜色从反照率 图像去除。 王玉洁,樊庆楠,李坤,陈冬冬,杨敬钰,卢健智,Dani Lischinski,陈宝权 面向本征图像分解的高质量渲染数据集与非局部卷积网络



图 10 基于不同方法的本征分解结果的光照交换结果对比

Fig. 10 Comparison between results for light exchange using intrinsic components from different methods ((a) input;(b) Fan et al. (2018);(c) Li and Snavely (2018a);(d) ours

6 消融实验

通过消融实验对本文方法的各个模块的作用进行分析。在消融实验中,所有的变体都在不经过微调的情况下,在IIW/SAW测试集上进行测试,数值结果汇总在表4中。

6.1 非局部图卷积模块的有效性

为了验证本文方法中图卷积模块 h_{nlm} 的有效性,将 h_{nlm} 从本文设计的图卷积神经网络中去掉后

进行训练,该变体表示为表 4 中的"本文(去 掉 h_{nlm})"。

如表4所示,与本文方法的完整模型相比,去掉 *h_{nlm}*之后,在反照率图像和亮度图像上的定量结果 均显著下降,验证了图卷积模块*h_{nlm}*的有效性。

6.2 反照率图像优化模块的有效性

表4中第2行展示了变体"本文(去掉 h_r)"的 测试结果,该变体是通过将反照率图像优化模块 h_r 从本文方法的框架中去掉得到的。如表4所示,去 掉 h_r之后,WHDR误差增加至20.89%,即反照率图

中国	图象	图形	学报
		GE AND G	RAPHICS

表4 洋	锏实验结果	
Table 4 Ablation study is	esults of the pro	posed method
方法	在 IIW 数据 集上的 WHDR/%	在 SAW 数据 集上的 AP/%
本文(去掉 h _{nlm})	19.08	92. 19
本文(去掉 h _r)	20. 89	96.17
本文(无其他光照成分)	17.65	94.94
本文(CGIntrinsics)	18.13	93.10
本文	17.92	96. 17

注:加粗字体表示最优值。

像的预测结果显著下降,这体现了本文网络框架中 反照率优化模块 h. 对提升反照率图像结果的贡献。

6.3 在输入图像中渲染其他光照成分的作用

如第3节中所述,本文数据集在输入图像中渲 染了多种光照成分,以增加输入图像的逼真度。为 了验证渲染这些成分对提升本征分解结果的作用, 此处设计了一个变体——"本文(无其他光照成 分)":即使用标签图像的反照率图像和亮度图像进 行相乘合成出只有漫反射成分的输入图像,然后对 本文方法在这些数据上进行训练。该变体的测试结 果汇总于表 4, 其中 SAW 测试集上的结果明显下 降。这表明,在输入图像中渲染非漫反射光照成分 对亮度图的预测的有效性。

6.4 本文所提数据集对本文方法的作用

为了探究本文数据集对本文方法的贡献,变体 "本文(CGIntrinsics)"被训练和测试,即在 Li 和 Snavely(2018a)提出的数据集 CGIntrinsics 上训练本 文提出的图卷积神经网络。如表4所示,相较于使 用本文提出的数据集训练相比,该变体在反照率图 像和亮度图像的数值结果上均变差,这进一步表明 了本文所提数据集对提升本征图像分解质量的的有 效性。

结 论 7

提出了一个基于非局部图卷积神经网络的本征 分解算法,引入的图卷积模块以显式的方式利用特 征图中的非局部线索。为了克服现有数据集的 局 限性,提出了一个新的高质量的本征图像数据集,其 中提供了反照率图像和亮度图像的逐像素标签。

并且,本文数据集中在亮度图中考虑了非白光光 照,使得本文本征分解模型可以更好地将输入图像 中的材质颜色和光照成分分开,特别是由漫反射表 面之间相互反射所引入的光照效果。通过将不同方 法在本文数据集与之前的数据集上分别训练并进行 测试与对比,验证了本文数据集对提升本征图像分 解质量的有效性。此外,通过将本文本征图像分解 网络与之前的方法比较,在定量结果上实现了较好 的结果,在视觉效果上有较大的提升。此外,通过将 本文方法和其他两种前沿的本征分解方法的分解结 果应用于一系列图像编辑任务中,进一步直观地展 示了每种方法产生的分解结果的质量及其应用 价值。

在场景中有大面积的颜色变化剧烈的复杂纹理 时,本文方法所产生的亮度图结果中仍存在纹理信 息。这是由于在不利用几何信息的情况下,网络对 由纹理产生的边缘和由阴影产生的边缘的区分能力 有限。同时,由于将图像分解成为反照率图像和亮 度图像所能支持的编辑任务有限,未来的工作中考 虑将场景的几何信息、光照信息和材质信息进行联 合估计或优化,提升分解的质量并支持更丰富的 应用。

参考文献(References)

- Barron J T and Malik J. 2015. Shape, illumination, and reflectance from shading. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37 (8): 1670-1687 [DOI: 10. 1109/TPAMI. 2014. 2377712]
- Barrow H G and Tenenbaum J M. 1978. Recovering intrinsic scene characteristics from images//Hanson A and Riseman E, eds. Computer Vision Systems. New York: Academic Press
- Baslamisli A S, Groenestege T T, Das P, Le H A, Karaoglu S and Gevers T. 2018. Joint learning of intrinsic images and semantic segmentation//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 289-305 [DOI: 10.1007/978-3-030-01231-1_18]
- Bell S, Bala K and Snavely N. 2014. Intrinsic images in the wild. ACM Transactions on Graphics, 33(4): #159 [DOI: 10.1145/2601097. 2601206
- Bi S, Han X G and Yu Y Z. 2015. An L1 image transform for edge-preserving smoothing and scene-level intrinsic decomposition. ACM Transactions on Graphics, 34(4): #78 [DOI: 10.1145/2766946]
- Boyadzhiev I, Paris S and Bala. 2013. User-assisted image compositing for photographic lighting. 2013. ACM Transactions on Graphics,

32(4):1-12 [DOI: 10.1145/2461912.2461973]

- Bruna J, Zaremba, Szlam A and Lecun Y. 2013. Spectral networks and locally connected networks on graphs. [EB/OL]. [2021-08-16]. https://arxiv.org/pdf/1312.6203.pdf
- Butler D J, J Wulff, Stanley G B and Black M J. 2012. A naturalistic open source movie for optical flow evaluation. //Proceedings of 2012 European Conference on Computer Vision (ECCV). Florence, Italy;Springer;611-625 [DOI:10.1007/978-3-642-33783-3_44]
- Chen Q F and Koltun V. 2013. A simple model for intrinsic image decomposition with depth cues// Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, Australia; IEEE; 241-248 [DOI: 10.1109/ICCV.2013.37]
- Fan Q N, Yang J L, Hua G, Chen B Q and Wipf D. 2018. Revisiting deep intrinsic image decompositions//Proceedings of 2018 IEEE/ CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8944-8952 [DOI: 10.1109/CVPR.2018. 00932]
- Garces E, Munoz A, Lopez-Moreno J and Gutierrez D. 2012. Intrinsic images by clustering. Computer Graphics Forum, 31 (4): 1415-1424 [DOI: 10.1111/j.1467-8659.2012.03137.x]
- Gehler P V, Rother C, Kiefel M, Zhang L M and Schölkopf B. 2011. Recovering intrinsic images with a global sparsity prior on reflectance//Proceedings of the 24th International Conference on Neural Information Processing Systems. Granada, Spain: Curran Associates Inc: 765-773
- Grosse R, Johnson M K, Adelson E H and Freeman W T. 2009. Ground truth dataset and baseline evaluations for intrinsic image algorithms//Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE: 2335-2342 [DOI: 10. 1109/ICCV. 2009. 5459428]
- Hamilton W L, Ying R and Leskovec J. 2017. Inductive representation learning on large graphs//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA; Curran Associates Inc; 1025-1035
- Henaff M, Bruna J and LeCun Y. 2015. Deep convolutional networks on graph-structured data [EB/OL]. [2021-08-16]. https://arxiv. org/pdf/1506.05163.pdf
- Kovacs B, Bell S, Snavely N and Bala K. 2017. Shading annotations in the wild//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 850-859 [DOI: 10.1109/CVPR.2017.97]
- Land E H and McCann J J. 1971. Lightness and retinex theory. Journal of the Optical Society of America, 61(1): 1-11 [DOI: 10.1364/JOSA.61.000001]
- Li R Y, Wang S, Zhu F Y and Huang J Z. 2018a. Adaptive graph convolutional neural networks//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI Press: 3546-3553
- Li W B, Saeedi S, McCormac J, Clark R, Tzoumanikas D, Ye Q,

Huang Y Z, Tang R and Leutenegger S. 2018b. InteriorNet: megascale multi-sensor photo-realistic indoor scenes dataset//Proceedings of 2018 British Machine Vision Conference. Newcastle, UK: BMVA Press

- Li Z Q and Snavely N. 2018a. CGIntrinsics: better intrinsic image decomposition through physically-based rendering//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 381-399 [DOI: 10.1007/978-3-030-01219-9_23]
- Li Z Q and Snavely N. 2018b. Learning intrinsic image decomposition from watching the world//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 9039-9048 [DOI: 10.1109/CVPR.2018.00942]
- Ma W C, Chu H, Zhou B L, Urtasun R and Torralba A. 2018. Single image intrinsic decomposition without a single intrinsic image//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 211-229 [DOI: 10.1007/978-3-030-01264-9_13]
- Meka A, Zollhöfer M, Richardt C and Theobalt C. 2016. Live intrinsic video. ACM Transactions on Graphics, 35(4): #109 [DOI: 10. 1145/2897824.2925907]
- Monti F, Boscaini D, Masci J, Rodolà E, Svoboda J and Bronstein M M. 2017. Geometric deep learning on graphs and manifolds using mixture model cnns//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5425-5434 [DOI: 10.1109/CVPR.2017.576]
- Narihira T, Maire M and Yu S X. 2015a. Learning lightness from human judgement on relative reflectance//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 2965-2973 [DOI: 10. 1109/CVPR. 2015. 7298915]
- Narihira T, Maire M and Yu S X. 2015b. Direct intrinsics: learning albedo-shading decomposition by convolutional regression//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 2992 [DOI: 10.1109/ICCV. 2015.342]
- Nestmeyer T and Gehler P V. 2017. Reflectance adaptive filtering improves intrinsic image estimation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 1771-1780 [DOI: 10.1109/CVPR.2017. 192]
- Sha H and Liu Y. 2021. Review on deep learning based prediction of image intrinsic properties. Journal of Graphics, 42(3): 385-397 (沙浩, 刘越. 2021. 基于深度学习的图像本征属性预测方法综 述. 图学学报, 42(3): 385-397)
- Shen L, Tan P and Lin S. 2008. Intrinsic image decomposition with nonlocal texture cues//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage: USA: IEEE: 1-7 [DOI: 10.1109/CVPR.2008.4587660]

- Shen L and Yeo C. 2011. Intrinsic images decomposition using a local and global sparse representation of reflectance//Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE: 697-704 [DOI: 10. 1109/CVPR. 2011.5995738]
- Shi J, Dong Y, Su H and Yu S X. 2017. Learning non-lambertian object intrinsics across ShapeNet categories//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5844-5853 [DOI: 10.1109/CVPR.2017.619]
- Shi J, Dong Y, Tong X and Chen Y Y. 2015. Efficient intrinsic image decomposition for RGBD images//Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology. Beijing, China: ACM: 17-25 [DOI: 10.1145/2821592.2821601]
- Simonovsky M and Komodakis N. 2017. Dynamic edge-conditioned filters in convolutional neural networks on graphs//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA; IEEE; 29-38 [DOI: 10.1109/CVPR. 2017.11]
- Sinha P and Adelson E. 1993. Recovering reflectance and illumination in a world of painted polyhedra//Proceedings of the 4th International Conference on Computer Vision. Berlin, Germany: IEEE: 156-163 [DOI: 10.1109/ICCV.1993.378224]
- Song S R, Yu F, Zeng A, Chang A X, Savva M and Funkhouser T. 2017. Semantic scene completion from a single depth image//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 190-198 [DOI: 10.1109/CVPR.2017.28]
- Wang N Y, Zhang Y D, Li Z W, Fu Y W, Liu W and Jiang Y G. 2018. Pixel2Mesh: generating 3D mesh models from single RGB images// Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 55-71 [DOI: 10.1007/ 978-3-030-01252-6_4]
- Wang Y J, Li K, Yang J Y and Ye X C. 2017. Intrinsic decomposition from a single RGB-D image with sparse and non-local priors//Proceedings of 2017 IEEE International Conference on Multimedia and Expo (ICME). Hong Kong, China: 1201-1206 [DOI: 10.1109/ ICME. 2017. 8019390]
- Yi L, Su H, Guo X W and Guibas L. 2017. SyncSpecCNN: synchronized spectral CNN for 3D shape segmentation//Proceedings of 2017
 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hawaii, USA: IEEE: 6584-6592 [DOI: 10.1109/ CVPR.2017.697]

- Zhao Q, Tan P, Dai Q, Shen L, Wu E and Lin S. 2012. A closed-form solution to retinex with nonlocal texture constraints. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(7): 1437-1444 [DOI: 10.1109/TPAMI.2012.77]
- Zhou T H, Krahenbuhl P and Efros A A. 2015. Learning data-driven reflectance priors for intrinsic image decomposition//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 3469-3477 [DOI: 10.1109/ICCV.2015. 396]

作者简介



王玉洁,1994年生,女,博士研究生,主要研 究方向为本征图像分解、全息图像生成与压 缩及三维深度学习。 E-mail:yujiew. cn@gmail.com



陈宝权,通信作者,男,教授,主要研究方向 为计算机图形与可视化。 E-mail:baoquan@pku.edu.cn

樊庆楠,男,博士,主要研究方向为计算机视觉、计算机图形 学。E-mail: fqnchina@gmail.com 李坤,女,副教授,主要研究方向为计算机视觉、计算机图形 学、图像/视频处理。E-mail:lik@tju.edu.cn 陈冬冬,男,博士,主要研究方向为图像生成和表示学习。 E-mail:cddlyf@gmail.com 杨敬钰,男,教授,主要研究方向为计算机视觉、智能图像/ 视频处理。E-mail: yjy@tju.edu.cn 卢健智,男,学士,主要研究方向为分布式计算、渲染。 E-mail: ljz@3vjia.com Dani Lischinski,男,教授,主要研究方向为计算机图形学、图 像视频处理和计算机视觉。 E-mail: danix@mail.huji.ac.il