



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
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

2020
04
VOL.25

ISSN1006-8961
CN11-3758/TB



海岛岸线自动分割 P768



第25卷第4期 (总第288期)

2020年4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 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 顾行发
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@radi.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 | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa
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@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

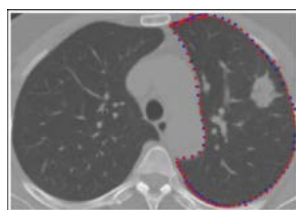
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

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



自适应在线判别外观学习的
分层关联多目标跟踪
(第0708页)



低秩和主动形状模型的肺区
分割(第0759页)



结合曲率滤波的HTM算法去
除遥感影像云雾(第0791页)

综述

深度学习目标检测方法综述

赵永强, 饶元, 董世鹏, 张君毅 0629

图像处理和编码

结合矩阵分解与差分隐私的人脸图像发布

张啸剑, 付聪聪, 孟小峰 0655

图像分析和识别

多尺度深度特征融合的变化检测

樊玮, 周末, 黄睿 0669

面向类内差距表情的深度学习识别

陈亮, 吴攀, 刘韵婷 0679

带后缀印刷体维吾尔文映射关系检索

伊克萨尼·普尔凯提, 阿布都萨拉木·达吾提, 艾斯卡尔·艾木都拉 0688

结合混沌鸟群算法的阴极铜板表面缺陷检测

王卓, 张长胜, 李伟, 钱俊兵, 唐都作, 蔡兵, 常以涛 0697

自适应在线判别外观学习的分层关联多目标跟踪

方岚, 于凤芹 0708

图像理解和计算机视觉

多方向Laplacian能量和与tetrolet变换的图像融合

沈瑜, 陈小鹏, 杨倩 0721

双重对偶生成对抗网络的跨年龄素描—照片转换

吴柳玮, 孙锐, 阚俊松, 高隽 0732

医学图像处理

残差密集相对平均CGAN的脑部图像配准

王丽芳, 张程程, 秦品乐, 蔺素珍, 高媛, 窦杰亮 0745

低秩和主动形状模型的肺区分割

孙申申, 田丹, 吴微, 康雁, 赵宏 0759

遥感图像处理

改进Deeplab网络的遥感影像海岸线分割

王振华, 钟元芾, 何婉雯, 曲念毅, 徐利智, 张文苹, 刘智翔 0768

卷积神经网络提取风云影像土壤湿度

韩颖娟, 周顺武, 王文文, 张学艺, 张承明, 栗佩琪 0779

结合曲率滤波的HTM算法去除遥感影像云雾

杨珂珂, 贾渊, 沈川 0791

ChinaMM 2019会议专栏

利用稳健非负矩阵分解实现无监督高光谱解混

宋晓瑞, 吴玲达, 孟祥利 0801

关键点深度特征驱动人脸表情识别

王善敏, 帅惠, 刘青山 0813

融合区域和边缘特征的水平集水下图像分割

孙杨, 陈哲, 王慧斌, 张振, 沈洁 0824

深度时空能量特征表示下的人体行为识别

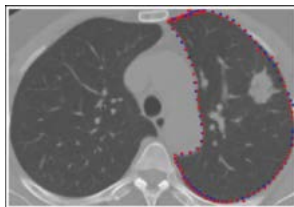
巢新, 侯振杰, 李兴, 梁久祯, 宦娟, 刘浩昱 0836

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Multi-object tracking based on adaptive online discriminative appearance learning and hierarchical association(P0708)



Lung segmentation by active shape model approach based on low rank theory(P0759)



Haze and cloud removal from remote sensing image using HTM algorithm based on curvature filtering(P0791)

Review

Survey on deep learning object detection

Zhao Yongqiang, Rao Yuan, Dong Shipeng, Zhang Junyi 0629

Image Processing and Coding

Private facial image publication through matrix decomposition

Zhang Xiaojian, Fu Congcong, Meng Xiaofeng 0655

Image Analysis and Recognition

Multiscale deep features fusion for change detection

Fan Wei, Zhou Mo, Huang Rui 0669

Depth learning recognition method for intra-class gap expression

Chen Liang, Wu Pan, Liu Yunting 0679

Mapping relationship retrieval of Uyghur-printed suffix word

Eksan Firkat, Abdusalam Dawut, Askar Hamdulla 0688

Cathodic copper plate surface defect detection based on bird swarm algorithm with chaotic theory

Wang Zhuo, Zhang Changsheng, Li Wei, Qian Junbing, Tang Duzuo, Cai Bing, Chang Yitao 0697

Multi-object tracking based on adaptive online discriminative appearance learning and hierarchical association

Fang Lan, Yu Fengqin 0708

Image Understanding and Computer Vision

Image fusion of multidirectional sum modified Laplacian and tetrolet transform

Shen Yu, Chen Xiaopeng, Yang Qian 0721

Double dual generative adversarial networks for cross-age sketch-to-photo translation

Wu Liuwei, Sun Rui, Kan Junsong, Gao Jun 0732

Medical Image Processing

Image registration method with residual dense relativistic average CGAN

Wang Lifang, Zhang Chengcheng, Qin Pinle, Lin Suzhen, Gao Yuan, Dou Jieliang 0745

Lung segmentation by active shape model approach based on low rank theory

Sun Shenshen, Tian Dan, Wu Wei, Kang Yan, Zhao Hong 0759

Remote Sensing Image Processing

Island shoreline segmentation in remote sensing image based on improved Deeplab network

Wang Zhenhua, Zhong Yuanfu, He Wanwen, Qu Nianyi, Xu Lizhi, Zhang Wenping, Liu Zhixiang 0768

Extracting soil moisture from Fengyun satellite images using a convolutional neural network

Han Yingjuan, Zhou Shunwu, Wang Wenwen, Zhang Xueyi, Zhang Chengming, Li Peiqi 0779

Haze and cloud removal from remote sensing image using HTM algorithm based on curvature filtering

Yang Keke, Jia Yuan, Shen Chuan 0791

Column of ChinaMM 2019

Unsupervised hyperspectral unmixing based on robust non-negative matrix factorization

Song Xiaorui, Wu Lingda, Meng Xiangli 0801

Facial expression recognition based on deep facial landmark features

Wang Shanmin, Shuai Hui, Liu Qingshan 0813

Level set method combining region and edge features for segmenting underwater images

Sun Yang, Chen Zhe, Wang Huibin, Zhang Zhen, Shen Jie 0824

Action recognition under depth spatial-temporal energy feature representation

Chao Xin, Hou Zhenjie, Li Xing, Liang Jiuzhen, Huan Juan, Liu Haoyu 0836

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2020)04-0669-10

论文引用格式: Fan W, Zhou M and Huang R. 2020. Multiscale deep features fusion for change detection. Journal of Image and Graphics, 25(04): 0669-0678 (樊玮, 周末, 黄睿. 2020. 多尺度深度特征融合的变化检测. 中国图象图形学报, 25(04): 0669-0678) [DOI:10.11834/jig.190312]

多尺度深度特征融合的变化检测

樊玮, 周末, 黄睿

中国民航大学计算机科学与技术学院, 天津 300300

摘要: **目的** 图像的变化检测是视觉领域的一个重要问题, 传统的变化检测对光照变化、相机位姿差异过于敏感, 使得在真实场景中检测结果较差。鉴于卷积神经网络(convolutional neural networks, CNN)可以提取图像中的深度语义特征, 提出一种基于多尺度深度特征融合的变化检测模型, 通过提取并融合图像的高级语义特征来克服检测噪音。**方法** 使用VGG(visual geometry group)16作为网络的基本模型, 采用孪生网络结构, 分别从参考图像和查询图像中提取不同网络层的深度特征。将两幅图像对应网络层的深度特征拼接后送入一个编码层, 通过编码层逐步将高层与低层网络特征进行多尺度融合, 充分结合高层的语义和低层的纹理特征, 检测出准确的变化区域。使用卷积层对每一个编码层的特征进行运算产生对应尺度的预测结果。将不同尺度的预测结果融合得到进一步细化的检测结果。**结果** 与SC_SOBS(SC-self-organizing background subtraction)、SuBSENSE(self-balanced sensitivity segmenter)、FGCD(fine-grained change detection)和全卷积网络(fully convolutional network, FCN)4种检测方法进行对比。与性能第2的模型FCN相比, 本文方法在VL_CMU_CD(visual localization of Carnegie Mellon University for change detection)数据集中, 综合评价指标F1值和精度值分别提高了12.2%和24.4%; 在PCD(panoramic change detection)数据集中, F1值和精度值分别提高了2.1%和17.7%; 在CDnet(change detection net)数据集中, F1值和精度值分别提高了8.5%和5.8%。**结论** 本文提出的基于多尺度深度特征融合的变化检测方法, 利用卷积神经网络的不同网络层特征, 有效克服了光照和相机位姿差异, 在不同数据集上均能得到较为鲁棒的变化检测结果。

关键词: 变化检测; 特征融合; 多尺度; 孪生网络; 深度学习

Multiscale deep features fusion for change detection

Fan Wei, Zhou Mo, Huang Rui

Civil Aviation University of China, Computer Science and Technology, Tianjin 300300, China

Abstract: **Objective** Change detection aims at detecting the difference of the images captured from same scene in different time observations. This condition is an important research problem in computer vision. However, the traditional change detection methods, which use the handcrafted features and heuristic models, suffer from lighting variations and camera pose differences, resulting in bad change detection results. Recent deep learning-based convolutional neural networks (CNN) achieve huge success on several computer vision problems, such as image classification, semantic segmentation, and saliency detection. The main reason of the success of the deep learning-based methods is the abstract ability of CNN. To conquer the bad effects of lighting variations and camera pose differences, we can employ deep learning-based CNN in change detection problems. Unlike semantic segmentation, change detection inputs image pairs of two time observations. Thus, a key research problem is how to design an effective architecture of CNN, which can fully explore the intrinsic changes of the image pairs. To generate robust change detection results, we propose in this study a multiscale deep feature fusion-based

收稿日期: 2019-06-27; 修回日期: 2019-10-13; 预印本日期: 2019-10-20

基金项目: 国家自然科学基金项目(U1333109); 中国民航大学中央高校基金项目(3122018C021)

Supported by: National Natural Science Foundation of China(U1333109)

change detection (MDFCD). **Method** The proposed MDFCD network has two streams of feature extracting sub-networks, which share weight parameters. Each sub-network is responsible for learning to extract semantic features from the corresponding RGB image. We use VGG (visual geometry group) 16 as the basic backbone of the proposed MDFCD. The last fully connected layers of VGG16 are removed to save the spatial resolution of the features of the last convolutional layer. We adopt the features of convolutional blocks Conv3, Conv4, and Conv5 of VGG16 as our multiscale deep features because they can capture high-level, middle-level, and low-level features. Then, the Enc (encoding) module is proposed to fuse the deep features from the two time observations of the same convolutional block. We use "concat" operation to concatenate the features. The resulted features are input into Enc to generate change detection adaptive features at the corresponding feature level. The encoded features from the lower layer are upsampled twice in height and width. Then, we concatenate the deep features of the previous layers' convolutional blocks. Subsequently, Enc is used again to learn adaptive features. By progressively incorporating the features from Conv5 to Conv3, we obtain deep fusion of CNN features at multiple scales. To generate robust change detection, we add a convolutional layer with $2 \times 3 \times 3$ convolutional filters to generate change prediction at each scale encoding module. Then, the change predictions of all scales are concatenated together to produce final change detection results. Note that we use bicubic upsampling operation to upsample the change detection map at each scale to the size of the input image. **Result** We compared three benchmark datasets, namely, VL_CMU_CD (visual localization of Carnegie Mellon University for change detection), PCD (panoramic change detection), CDnet (change detection net), by the state-of-the-art change detection methods, that is, FGCD (fine-grained change detection), SC_SOBS (SC-self-organizing background subtraction), SuBSENSE (self-balanced sensitivity segmenter), and FCN (fully convolutional network). We employed F1-measure, recall, precision, specific, FPR (false positive rate), FNR (false negative rate), PWC (percentage of wrong classification) to evaluate the difference of the compared change detection methods. The experiments show that MDFCD are better than the other compared methods. Among the compared methods, deep learning-based change detection method FCD performs the best. On VL_CMU_CD, the F1-measure and Precision of MDFCD achieve 12.2% and 24.4% relative improvements over the second-placed change detection method FCN, respectively. On PCD, the F1-measure and precision of MDFCD obtain 2.1% and 17.7% relative improvements over FCN, respectively. On CDnet, compared with FCN, our F1-measure and precision achieve 8.5% and 5.8% relative improvements, respectively. From the experiments, we can find that MDFCD can detect the fine grained changes, such as telegraph poles. The proposed MDFCD are better in distinguishing the real changes with false changes caused by lighting variations and camera pose difference compared with FCN. **Conclusion** We studied how to effectively explore the deep convolutional neural networks for change detection problem. MDFCD network is proposed to alleviate the bad effects introduced by lighting variations and camera pose differences. The proposed method adopts a siamese network with VGG16 as the backbone. Each path is responsible for extracting deep features from reference and query images. We also proposed encoding module that fuses multiscale deep convolutional features and learn change detection adaptive features. The deep features are integrated together from high layers' semantic features with low layers' texture features. With this fusion strategy, the proposed method can generate more robust change detection results than other compared methods. The high layers' semantic features can effectively avoid the negative changes caused by lighting and season change. Meanwhile, the low layers' texture features help the proposed method obtain accurate changes at the object boundaries. Compared with deep learning method, FCN, where in the input is concatenate reference and query images, our method of extracting features with respect to each image can extract representatively features for change detection. However, as a general problem of deep learning-based methods, one should use large volume of training images to train CNNs. Another problem is that the present change detection methods pay considerable attention on region changes but not on object-level changes. In our future work, we plan to use weak supervised and unsupervised method to study the change detection to avoid using pixel-level labeled training images. We also plan to study incorporating object detection in change detection to generate object-level changes.

Key words: change detection; feature fusion; multiscale; siamese network; deep learning

0 引言

变化检测针对同一场景不同时间获取的图像中的变化区域, 广泛用于土地覆盖 (Bruzzone 等, 2004)、植被覆盖 (Khan 等, 2017)、城市监测 (Taneja 等, 2011)、道路交通安全、异常检测 (刘华平等, 2013)、视频监控 (Liang 等, 2017) 和自动驾驶 (吴沫等, 2007) 等。

传统的变化检测方法包括代数法 (Malila, 1980)、聚类法 (Ghosh 等, 2009) 和空间转换法 (Li 和 Yeh, 1998; Deng 等, 2008), 这些方法大都采用阈值选取、区域划分、空间转化等技术直接对图像特征进行对比, 检测结果受光照和相机位姿等因素的影响较大 (Liang 等, 2017)。

深度卷积神经网络 (deep convolutional neural network, DCNN) 在计算机视觉领域取得了巨大成功, 代表性成就有 AlexNet (Krizhevsky 等, 2012), 全卷积网络 (fully convolutional network, FCN) (Shelhamer 等, 2014), VGGNet (visual geometry group network) (Simonyan 和 Zisserman, 2014), GoogLeNet (Szegedy 等, 2017)。不同于传统方法直接将检测图像的特征进行比较, 深度学习方法从大量样本数据中学习图像的抽象特征, 直接建立原始图像与变化结果之间的映射, 使得基于深度学习的变化检测方法对光照和相机位姿差异具有较好的鲁棒性。本文基于深度卷积神经网络, 提出一种基于多尺度深度特征融合的变化检测方法 (multiscale deep features fusion-based change detection network, MDFCD), 采用孪生网络 (siamese network) 结构, 分别从参考图像和查询图像中提取不同网络层的深度特征。提取出的高层语义特征可以有效解决光照和相机位姿差异引起的非真实变化, 低层纹理特征可以较好刻画变化区域的边缘及细节。MDFCD 通过将两幅图像对应网络层的深度特征拼接后送入一个编码层, 在编码层中逐步将高层与低层网络特征进行多尺度融合, 充分结合高层的语义和低层的纹理特征, 生成准确的变化检测结果。为了有效训练网络, MDFCD 采用多个损失函数分别对每个编码层的预测结果进行评价, 将多个编码层的预测结果进行融合得到更为鲁棒的变化检测。本文使用 VL_CMU_CD (visual localization of Carnegie Mellon University

for change detection) (Alcantarilla 等, 2018)、PCD (panoramic change detection) (Sakurada 和 Okatani, 2015) 和 CDnet (change detection net) (Wang 等, 2014) 3 个变化检测数据集, 分别训练和测试 MDFCD, 并与 4 种变化检测方法进行对比。结果表明, MDFCD 在存在光照变化、相机位姿差异的图像上取得了优于其他对比方法的变化检测结果。

1 相关工作

深度学习在图像变化检测中逐步得到应用, 主要采用卷积神经网络 (convolutional neural network, CNN)、生成对抗网络 (generative adversarial networks, GAN) 和循环神经网络 (recurrent neural network, RNN) 等方法。

Sakurada 和 Okatani (2015) 提出将 CNN 与超像素结合的变化检测方法, 使用 CNN 提取变化前后的图像特征进行比较, 得到粗略的变化图, 再用超像素分割估计出一个更加准确的变化边界。Fujita 等人 (2017) 使用 CNN 对海啸前后的建筑图像分别提取特征, 在全连接层做相似度计算来检测房屋受损等级。Alcantarilla 等人 (2018) 将变化前后的图像拼接后输入面向图像分割的 FCN 网络应用于城市街道变化检测, 取得了优异的检测结果。Huang 等人 (2017) 提出基于 CNN 的相机位姿校正网络 and 变化检测网络相结合的变化检测模型, 使用位姿校正网络解决由于图像拍摄角度带来的差异。

GAN 将变化检测视为给定参考图像和查询图像后生成变化图的过程。GAN 由生成器和判别器组成。生成器建立原始图像与变化图之间的联系, 判别器判断生成结果图的正确性。经过数轮对抗训练, 得到生成器与判别器的最优模型。Gong 等人 (2017) 提出多光谱图像无监督变化检测的 GAN 模型, 用传统变化检测方法 (如主成分分析法 (principal component analysis, PCA)) 得到图像差异图作为先验, 训练一个生成器。该模型能够有效利用差异图中的丰富信息, 在对抗训练中消除输入图像的噪音干扰。

如果将变化检测中的参考图像和查询图像看成时间序列, 考虑图像间的时间依赖, 可以将变化检测的过程建模成深度时序模型。Lyu 和 Lu (2016) 使用一个端到端的 RNN 来解决高光谱图像的变化检测任务, 这个基于长短期记忆网络 (long short term memo-

ry, LSTM)模型的 RNN 网络表现出了良好的检测效果和迁移能力。Mou 等人(2019)将 CNN 与 LSTM 模型相结合,提出一个统一的框架来学习图像的光谱—空间—时间的特征表示,在学习图像光谱—空间特征的同时,能够自适应地学习多时相图像的依赖性。

上述检测模型可以较好地提取图像特征,在一定程度上避免光照差异和相机位姿差异的干扰,但是在计算差异图时只使用了提取到的深层图像特征,较浅层的特征并没有得到充分利用。

本文提出一种基于多尺度深度特征融合的变化检测方法 MDFCD,不同于 Huang 等人(2017)直接将参考图像和查询图像拼接后输入到网络产生变化检测结果,MDFCD 使用孪生网络结构(Zagoruyko 和 Komodakis,2015)。孪生网络最初用来比较图像间

相似度,如笔迹(Bromley 等,1993)和人脸识别(Nair 和 Hinton,2010)。本文沿用孪生网络的思想,分别从参考图像和查询图像中提取深度网络特征,并将高层网络特征逐步融合到低层网络特征中。实验证明以孪生网络结构为基础的 MDFCD 模型能够产生具有较好细节的变化检测结果。

2 本文方法

MDFCD 采用孪生网络方式分别提取参考图像 X 和查询图像 Y 的不同尺度的深度特征,通过融合高层的语义特征和低层的纹理特征得到鲁棒的变化检测结果,结构如图 1 所示。本节从网络结构和网络训练两方面介绍 MDFCD。

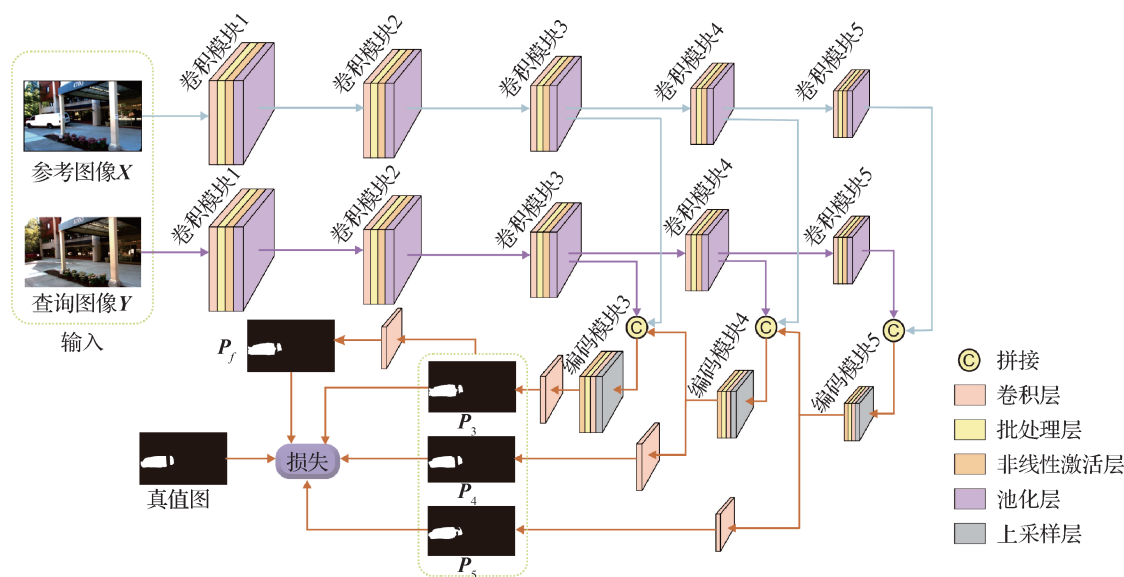


图 1 基于多尺度深度特征融合的变化检测网络

Fig. 1 Framework of multiscale deep feature fusion based change detection network

2.1 MDFCD 的网络结构

MDFCD 使用 VGG (visual geometry group) 16 (Simonyan 和 Zisserman,2015)作为孪生网络的基本网络,通过移除 VGG16 的全连层,保留卷积模块 Conv1—Conv5 用于提取参考图像和查询图像的不同尺度的深度特征。在同一尺度,将参考图像和查询图像的特征拼接后使用编码(encoding, Enc)模块进行降维,然后将降维后的特征与前一层特征进行拼接。

2.1.1 Conv 模块

卷积(convolution, Conv)模块是进行特征提取的基本模块,由卷积层(convolution)、批处理层

(batch normalization, BN)、非线性激活层(rectified linear unit, ReLU)和池化层(pool)组成,如图 1 所示。卷积层通过卷积运算提取图像特征,批处理层对神经网络的输入进行归一化,使其保持均值为 0、方差为 1 的标准正态分布,加快网络训练速度。非线性激活层作为激活函数,能够增加网络的非线性因素,增强模型的表达能力。最后由池化层对图像特征进行 2 倍的下采样,这样既能保留特征图的主要特征,又能减少下一层的计算量。

2.1.2 Enc 模块

如图 1 所示,Enc 模块由上采样(upsample)层、卷积层、批处理层和非线性激活层组成。为了保证

拼接时图像特征具有相同尺寸,使用上采样层进行2倍的上采样,对图像特征进行缩放。此外,由于在特征提取阶段提取的特征维数较高,将两幅图像的对应特征直接拼接会导致计算量过大,进而影响训练速度,所以使用卷积层减少图像的输出通道数对图像特征进行降维,在得到丰富图像特征的同时提高计算效率,缩短训练时间。卷积层的输出通道数统一设置为64。Enc模块的具体计算过程为

$$F_i = \phi(\text{cat}(F_{i+1}, \text{cat}(F_i^x, F_i^y))), i = 3, 4 \quad (1)$$

式中, $\text{cat}()$ 表示拼接操作,将图像特征在通道方向上进行拼接。 F_i 是参考图像和查询图像的第 i 层特征图与编码后的第 $i+1$ 层特征图拼接后的编码特征图。 F_i^x 和 F_i^y 分别为参考图像和查询图像在网络第 i 层的深度特征。注意到 Conv5 层是特征提取的最后一层,需要单独计算,其编码后的特征为

$$F_5 = \phi(\text{cat}(F_5^x, F_5^y)) \quad (2)$$

2.1.3 多尺度变化估计

基于编码模块 Enc 产生的编码特征 F_i , MDFCD 使用一个卷积层产生相应尺度的变化检测结果 P_i , 可表示为

$$P_i = \text{conv}(F_i), i = 3, 4, 5 \quad (3)$$

不同尺度的变化检测结果拼接后,通过卷积层进行融合得到最终的变化检测结果 P_f , 可形式化为

$$P_f = \text{conv}(\text{cat}(P_3, P_4, P_5)) \quad (4)$$

式中, P_3 、 P_4 、 P_5 分别为第 3、4、5 层对应的预测结果, P_f 为 3、4、5 层经过拼接并进入卷积层融合后得到的预测结果。

2.1.4 损失函数

MDFCD 采用交叉熵损失函数, 计算式为

$$\text{Loss} = -[y \ln \hat{y} + (1 - y) \ln(1 - \hat{y})] \quad (5)$$

式中, Loss 表示损失值, y 表示图像的标签值, $\hat{y}$ 表示预测结果值。

MDFCD 采用多重损失策略, 根据式(5), 分别对第 3、4、5 层的预测图 P_3 、 P_4 、 P_5 逐层求得损失 Loss_3 、 Loss_4 、 Loss_5 , 再对融合后的预测图 P_f 求出融合损失 Loss_f 。MDFCD 整体的损失由不同尺度对应的损失相加得到, 整体损失为

$$\text{Loss} = \text{Loss}_3 + \text{Loss}_4 + \text{Loss}_5 + \text{Loss}_f \quad (6)$$

2.2 网络训练

使用 VGG16 的网络参数对图 1 中 MDFCD 孪生网络的 Conv1—Conv5 前 5 个卷积模块进行初始化。对新添加的网络层, 如 Enc 模块中的卷积层参数采

用均值为 0、标准差为 0.06 的正态分布进行初始化, 对于产生预测结果的卷积层参数按照均值为 0、标准差为 0.3 的正态分布进行初始化。网络的初始学习率为 0.001, 随着迭代次数的增加, 学习率每一轮降低为前一轮的 1/10。批处理数量为 16, 冲量为 0.9。使用 Adam 优化器进行网络参数更新。为了避免过拟合以及与对比方法进行公平对比, 所有网络训练 30 次迭代后终止。

3 实验

3.1 数据集

采用 VL_CMU_CD (Alcantarilla 等, 2018)、PCD (Sakurada 和 Okatani, 2015) 和 CDnet (Wang 等, 2014) 3 个变化检测数据集进行实验。PCD 是一个全景变化检测数据集, 分为 TSUNAMI 和 GSV (Google street view) 两个子集。TSUNAMI 子集包含 100 对海啸受灾区的变化全景图, GSV 子集包含 100 对谷歌街景全景图。VL_CMU_CD 是从用于拓扑定位的 VL_CMU (Badino 等, 2011) 数据集中选取的 152 对 RGB 图像序列组合而成, 用来研究城市宏观变化。CDnet 是用于变化检测任务的大型真实场景视频数据集, 包括 11 个具有挑战性的场景类别和 53 个子类别。各数据集详细信息见表 1。

表 1 各数据集详细信息
Table 1 Details of the datasets

数据集	图像数	分辨率/像素	年份	训练集	测试集
PCD	400	1 024 × 224	2015	16 000	80
VL_CMU_CD	2 724	1 024 × 768	2016	16 005	332
CDnet	159 278	720 × 480	2014	40 148	10 061

由于深度网络需要大量的训练数据, 为达到理想的训练效果, 对部分数据做了相应扩展。在 VL_CMU_CD 数据集中, 采用裁剪和翻转的方式扩展数据, 在变化区域的最小包围矩形之外, 随机选取 5 对点进行剪裁, 生成包含变化区域的图像对。在 PCD 数据集中, 图像的长宽比过大, 直接进行训练容易导致图像结果失真, 所以按一定的长宽比对图像进行随机裁剪, 从每组图像对中裁剪出 50 组子图像对。在 CDnet 数据集中, 从每个子数据集中选取 1 幅图像作为参考图像, 其他图像作为查询图像, 形成图像

对,由于产生的图像对数量较多,因此没有进行数据扩展,所有图像均被缩放到 448×320 像素进行训练和测试。在各数据集,随机选取 80% 的图像对用于训练 MDFCD,剩余的 20% 用于测试。各数据集中的训练集和测试集的图像对数量见表 1。

3.2 对比方法

采用 SC_SOBS (SC-self-organizing background subtraction) (Maddalena 和 Petrosino, 2012)、SuB-SENSE (self-balanced sensitivity segmenter) (St-Charles 等, 2015)、FGCD (fine-grained change detection) (Feng 等, 2015) 和 FCN (fully convolutional network) (Alcantarilla 等, 2018) 4 种变化检测模型作为对比方法,在 VL_CMU_CD、PCD 和 CDnet 数据集上进行实验。SC_SOBS 和 SuBSENSE 在变化检测基准数据集 CDnet 上有优异表现,FGCD 通过相机位姿校正和光照校正来对齐图片,在微变检测中取得了良好效果,FCN 采用深度网络对变化区域进行端到端的学习,是目前变化检测领域较为先进的检测方法。

3.3 评价标准

采用综合评价指标 $F1$ ($F1$ -measure)、召回率 R (recall)、精确率 P (precision)、特异性 Sp (specific)、假正例率 FPR (false positive rate)、假负例率 FNR (false negative rate) 和误分类百分比 PWC (percentage of wrong classification) 7 个评价指标进行定量分析。

$F1$ 、 Sp 和 PWC 具体计算为

$$F1 = \frac{2P \times R}{P + R} \quad (7)$$

$$Sp = \frac{TN}{TN + FP} \quad (8)$$

$$PWC = \frac{FN + FP}{TP + FN + FP + TN} \times 100\% \quad (9)$$

式中, P 为精确率, R 为召回率, TP 为真正例个数, FN 为假负例个数, FP 为假正例个数, TN 为真负例个数。

3.4 实验结果

各检测方法在 VL_CMU_CD 数据集上的量化和检测结果如表 2 和图 2 所示。可知 MDFCD 模型的 $F1 = 0.791$, $P = 0.771$, 比位于第 2 的 FCN ($F1 = 0.705$, $P = 0.620$) 分别高 12.2% 和 24.4%, 其他 3 种方法的 $F1$ 和 P 值都没有超过 0.2。MDFCD 模型的特异性、假正例率和误分类百分比也均优于其余 4 种方法。从图 2 可以看出,FCN 对变化区域的检测效果明显优于其他 3 种传统变化检测方法。但当场景中的光照变化较大、变化物体细小时,FCN 会出现错误的检测结果。如图 2 第 1 行,FCN 多检测出 1 片未变化区域。在第 3 行中,较小的变化区域未被 FCN 检测到,而 MDFCD 能够准确检测到相应变化。在变化细节方面,MDFCD 同样更胜一筹,如第 2、4 行变化物体的轮廓细节,FGCD 的检测结果较为模糊且不够准确,SuBSENSE 和 SC_SOBS 基本失效。因此在准确度和精确度上,MDFCD 均高于对比方法。

表 2 不同方法在 VL_CMU_CD 数据集上的量化结果

Table 2 Quantitation results of different methods on VL_CMU_CD dataset

方法	$F1$	R	P	Sp	FPR	FNR	PWC
FGCD	0.120	0.135	0.19	0.965	0.035	0.865	9.742
SC_SOBS	0.171	0.939	0.105	0.377	0.623	0.061	58.462
SuBSENSE	0.144	0.999	0.085	0.144	0.856	0.001	79.704
FCN	0.705	0.883	0.620	0.963	0.037	0.117	3.867
MDFCD(本文)	0.791	0.860	0.771	0.987	0.013	0.140	2.140

注:加粗字体为每列最优值。

各检测方法在 PCD 数据集上的量化和检测结果如表 3 和图 3 所示。从表 3 可以看出, MDFCD 的 $F1 = 0.678$, FCN 的 $F1 = 0.664$, 其他 3 种方法的 $F1$ 值都没有超过 0.5, 相对于 FCN, MDFCD 的 $F1$ 提高了 2.1%。MDFCD 的精度 (0.764) 比 FCN 的精度 (0.649) 高 17.7%, 特异性 (0.941) 比 FCN

(0.857) 高 9.8%, 假正例率和误分类百分比值均低于其他方法, 验证了 MDFCD 优于其他对比方法。从图 3 也可以看出 MDFCD 优于其他对比方法, 如图 3 中第 3、4 行变化的电线杆, FCN 的检测结果较为粗糙, 忽略了很多轮廓细节, 其余方法的检测结果也均不够准确甚至失效, 而 MDFCD 可以很好地捕

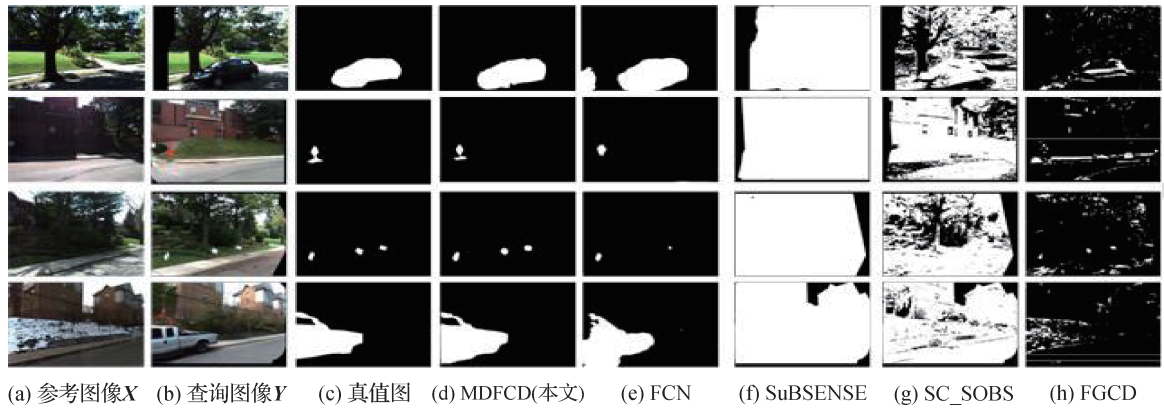


图 2 不同方法在 VL_CMU_CD 数据集上的检测结果

Fig. 2 Test results of different methods on VL_CMU_CD dataset ((a) reference image X ; (b) query image Y ; (c) ground truth; (d) MDFCD(ours); (e) FCN; (f) SuBSENSE; (g) SC_SOBS; (h) FGCD)

表 3 不同方法在 PCD 数据集上的量化结果

Table 3 Quantitation results of different methods on PCD dataset

方法	$F1$	R	P	Sp	FPR	FNR	PWC
FGCD	0.227	0.379	0.190	0.432	0.568	0.621	57.090
SC_SOBS	0.378	0.882	0.262	0.132	0.868	0.118	66.678
SuBSENSE	0.406	0.898	0.286	0.230	0.770	0.102	59.708
FCN	0.664	0.745	0.649	0.857	0.143	0.255	16.105
MDFCD(本文)	0.678	0.659	0.764	0.941	0.059	0.341	13.094

注:加粗字体为每列最优值。

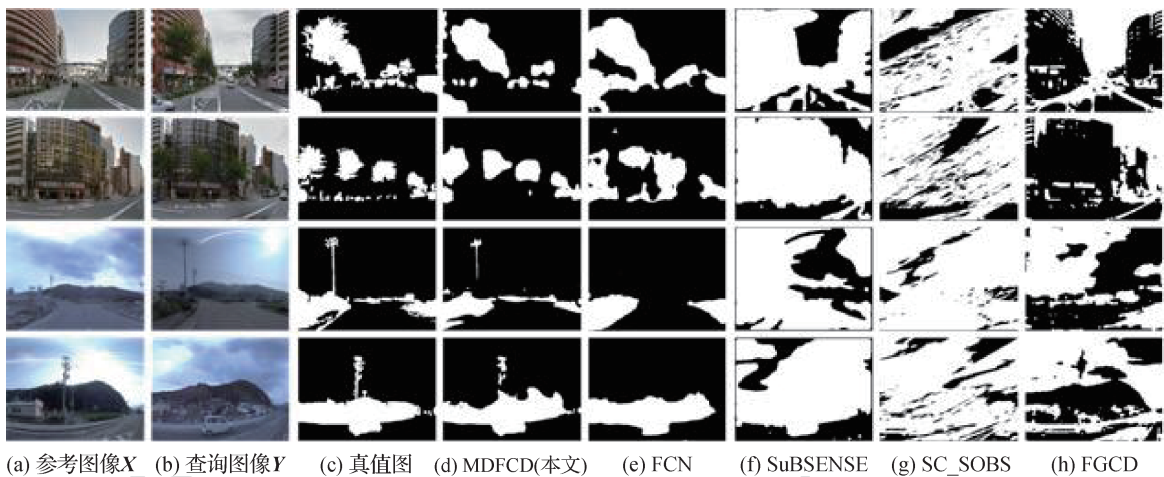


图 3 不同方法在 PCD 数据集上的检测结果

Fig. 3 Test results of different methods on PCD dataset ((a) reference image X ; (b) query image Y ; (c) ground truth; (d) MDFCD(ours); (e) FCN; (f) SuBSENSE; (g) SC_SOBS; (h) FGCD)

捉这些细节。

各检测方法在 CDnet 数据集上的量化和检测结果如表 4 和图 4 所示。从表 4 可以看出, MDFCD 的 $F1$ 值 (0.880)、召回率 (0.936) 和精度 (0.842) 比第 2 位的 FCN 分别高 8.5%、9.6% 和

5.8%, 其他指标也均超过对比方法。从图 4 可以看出, FGCD 有一定的检测效果, SC_SOBS 和 SuBSENSE 可基本检测出图像变化, 但变化区域较模糊, FCN 的检测结果好于以上 3 种方法, 但在检测细节上仍有欠缺, 而 MDFCD 能准确检测出细节清

表 4 不同方法在 CDnet 数据集上的量化结果
Table 4 Quantitation results of different methods on CDnet dataset

方法	$F1$	R	P	Sp	FPR	FNR	PWC
FGCD	0.195	0.491	0.162	0.737	0.263	0.509	27.215
SC_SOBS	0.497	0.739	0.499	0.910	0.090	0.261	9.632
SuBSENSE	0.571	0.751	0.608	0.931	0.069	0.249	7.599
FCN	0.811	0.854	0.796	0.992	0.008	0.146	1.085
MDFCD(本文)	0.880	0.936	0.842	0.994	0.006	0.064	0.647

注:加粗字体为每列最优值。

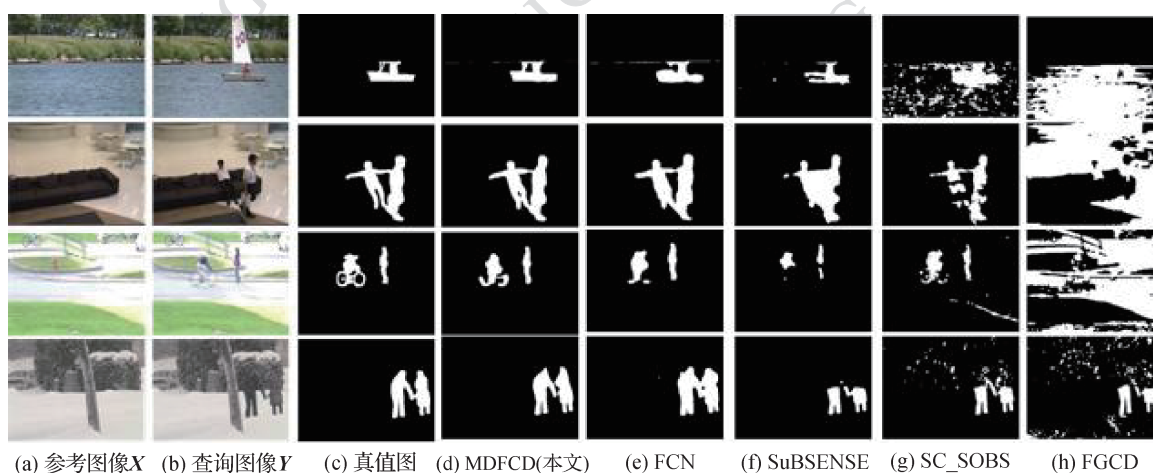


图 4 不同方法在 CDnet 数据集上的检测结果

Fig. 4 Test results of different methods on CDnet dataset ((a) reference image X ;

(b) query image Y ; (c) ground truth; (d) MDFCD(ours); (e) FCN; (f) SuBSENSE; (g) SC_SOBS; (h) FGCD)

晰的变化区域,如第2、4行的人物边缘和第3行的车轮。

由实验结果可知,在 VL_CMU_CD 和 PCD 数据集上,SC_SOBS、SuBSENSE 和 FGCD 等传统方法的效果并不理想,而在 CDnet 数据集上有一定的检测效果,原因是前两个数据集中场景的光照变化和相机位姿差异较大,传统方法不能很好地克服光照变化和相机位姿差异带来的噪音,而 CDnet 数据集中的光照变化和位姿差异较小,传统方法能表现出更好的检测能力,这也从侧面印证了 MDFCD 在克服场景噪音、提高检测准确度和对变化细节的描述上都优于对比方法。

4 结 论

针对传统的变化检测方法对光照变化、相机位姿差异过于敏感,在真实场景中检测结果较差这一

问题,提出了一种基于多尺度深度特征融合的变化检测网络 MDFCD。使用 VGG 作为基本模型,采用孪生网络结构,分别从参考图像和查询图像中提取不同网络层的深度特征,再用一个编码模块逐步将高层语义特征与低层纹理特征融合,产生较为鲁棒的变化检测结果。通过融合不同尺度的深度特征,在克服光照变化和相机位姿差异、检测细小变化区域方面取得了较好效果。在 VL_CMU_CD、PCD 和 CDnet 3 个变化检测数据集上, $F1$ 值分别为 79.1%、67.8% 和 88%,精度值分别为 77.1%、76.4%、84.2%,相较于 4 个对比方法均有不同程度的提高。

由于监督学习需要大量的标注数据进行网络训练,而实际应用中大量的标注数据较难获取。为了解决基于深度学习的变化检测对数据的依赖问题,在未来的研究工作中将着重研究如何使用弱监督或无监督方法产生鲁棒的变化检测结果。

参考文献 (References)

- Alcantarilla P F, Stent S, Ros G, Arroyo R and Gherardi R. 2018. Street-view change detection with deconvolutional networks. *Autonomous Robots*, 42 (7): 1301-1322 [DOI: 10.1007/s10514-018-9734-5]
- Badino H, Huber D and Kanade T. 2011. Visual topometric localization//2011 IEEE Intelligent Vehicles Symposium (IV). Baden-Baden, Germany; IEEE: 794-799 [DOI: 10.1109/IVS.2011.5940504]
- Bromley J, Guyon I, LeCun Y, Säckinger E and Shah R. 1993. Signature verification using a "Siamese" time delay neural network//Proceedings of the 6th International Conference on Neural Information Processing Systems. Denver, Colorado; Morgan Kaufmann Publishers Inc: 737-744
- Bruzzone L, Cossu R and Vernazza G. 2004. Detection of landcover transitions by combining multivariate classifiers. *Pattern Recognition Letters*, 25 (13): 1491-1500 [DOI: 10.1016/j.patrec.2004.06.002]
- Deng J S, Wang K, Deng Y H and Qi G J. 2008. PCA-based landuse change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing*, 29 (16): 4823-4838 [DOI: 10.1080/01431160801950162]
- Feng W, Tian F P, Zhang Q, Zhang N, Wan L and Sun J Z. 2015. Fine-grained change detection of misaligned scenes with varied illuminations//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile; IEEE: 1260-1268 [DOI: 10.1109/ICCV.2015.149]
- Fujita A, Sakurada K, Imaizumi T, Ito R, Hikosaka S and Nakamura R. 2017. Damage detection from aerial images via convolutional neural networks//Proceedings of the 15th IAPR International Conference on Machine Vision Applications (MVA). Nagoya, Japan; IEEE: 5-8 [DOI: 10.23919/MVA.2017.7986759]
- Ghosh S, Mishra N S and Ghosh A. 2009. Unsupervised change detection of remotely sensed images using fuzzy clustering//Proceedings of the 7th International Conference on Advances in Pattern Recognition. Kolkata, India; IEEE: 385-388 [DOI: 10.1109/ICAPR.2009.82]
- Gong M G, Niu X D, Zhang P Z and Li Z T. 2017. Generative adversarial networks for change detection in multispectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 14 (12): 2310-2314 [DOI: 10.1109/LGRS.2017.2762694]
- Huang R, Feng W, Wang Z Z, Fan M Y, Wan L and Sun J Z. 2017. Learning to detect fine-grained change under variant imaging conditions//Proceedings of 2017 IEEE International Conference on Computer Vision Workshops. Venice, Italy; IEEE: 2916-2924 [DOI: 10.1109/ICCVW.2017.344]
- Khan S H, He X M, Porikli F and Bennamoun M. 2017. Forest change detection in incomplete satellite images with deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55 (9): 5407-5423 [DOI: 10.1109/TGRS.2017.2707528]
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks//Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada; Curran Associates Inc: 1097-1105
- Li X and Yeh A G O. 1998. Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the Pearl River Delta. *International Journal of Remote Sensing*, 19(8): 1501-1518 [DOI: 10.1080/014311698215315]
- Liang D, Kaneko S I, Sun H and Kang B. 2017. Adaptive local spatial modeling for online change detection under abrupt dynamic background//Proceedings of 2017 IEEE International Conference on Image Processing (ICIP). Beijing, China; IEEE: 2020-2024 [DOI: 10.1109/ICIP.2017.8296636]
- Liu H P, Li J M, Hu X L and Sun F C. 2013. Recent progress in detection and recognition of the traffic signs in dynamic scenes. *Journal of Image and Graphics*, 18(5): 493-503 (刘华平, 李建民, 胡晓林, 孙富春. 2013. 动态场景下的交通标识检测与识别研究进展. *中国图象图形学报*, 18(5): 493-503) [DOI: 10.11834/jig.20130502]
- Lyu H and Lu H. 2016. Learning a transferable change detection method by recurrent neural network//Proceedings of 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Beijing, China; IEEE: 5157-5160 [DOI: 10.1109/IGARSS.2016.7730344]
- Maddalena L and Petrosino A. 2012. The SOBS algorithm: what are the limits? //Proceedings of 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Providence, RI, USA; IEEE: 21-26 [DOI: 10.1109/CVPRW.2012.6238922]
- Malila W A. 1980. Change vector analysis: an approach for detecting forest changes with Landsat//Remotely Sensed Data Symposium. West Lafayette; Purdue University: 326-335
- Mou L C, Bruzzone L and Zhu X X. 2019. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57 (2): 924-935 [DOI: 10.1109/TGRS.2018.2863224]
- Nair V and Hinton G E. 2010. Rectified linear units improve restricted Boltzmann machines//Proceedings of the 27th International Conference on International Conference on Machine Learning. Haifa, Israel; Omnipress: 807-814
- Sakurada K and Okatani T. 2015. Change detection from a street image pair using CNN features and superpixel segmentation//Proceedings of the British Machine Vision Conference (BMVC). Swansea, UK; BMVC: 61.1-61.12 [DOI: 10.5244/C.29.61]
- Shelhamer E, Long J and Darrell T. 2014. Fully convolutional networks

- for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4): 640-651 [DOI: 10.1109/TPAMI.2016.2572683]
- Simonyan K and Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2019-06-12]. <https://arxiv.org/pdf/1409.1556.pdf>
- St-Charles P L, Bilodeau G A and Bergevin R. 2015. SuBSENSE: a universal change detection method with local adaptive sensitivity. IEEE Transactions on Image Processing, 24(1): 359-373 [DOI: 10.1109/TIP.2014.2378053]
- Szegedy C, Ioffe S, Vanhoucke V and Alemi A A. 2017. Inception-v4, inception-ResNet and the impact of residual connections on learning//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, California, USA: AAAI: 4278-4284
- Taneja A, Ballan L and Pollefeys M. 2011. Image based detection of geometric changes in urban environments//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE: 2336-2343 [DOI: 10.1109/ICCV.2011.6126515]
- Wang Y, Jodoin P M, Porikli F, Konrad J, Benezeth Y and Ishwar P. 2014. CDnet 2014: an expanded change detection benchmark dataset//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus, OH, USA: IEEE: 393-400 [DOI: 10.1109/CVPRW.2014.126]
- Wu M, An X J and He H G. 2007. On vision-based lane departure detection approach. Journal of Image and Graphics, 12(1): 110-115 (吴沫, 安向京, 贺汉根. 2007. 基于视觉的车道跑偏检测方法研究及仿真. 中国图象图形学报, 12(1): 110-115) [DOI: 10.3969/j.issn.1006-8961.2007.01.019]

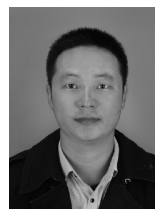
- Zagoruyko S and Komodakis N. 2015. Learning to compare image patches via convolutional neural networks.//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, Massachusetts. IEEE: 4353-4361 [DOI: 10.1109/CVPR.2015.7299064]

作者简介



樊玮, 1968年生, 男, 教授, 主要研究方向为智能信息处理、机器学习。

E-mail: wfancauc@163.com



黄睿, 通信作者, 男, 博士, 讲师, 主要研究方向为计算机视觉、机器学习。

E-mail: ruihuang@tju.edu.cn

周末, 女, 硕士研究生, 主要研究方向为计算机视觉。

E-mail: 1648420377@qq.com