



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

中国图象学报 图形学报

2020
05
VOL.25

ISSN1006-8961
CN11-3758/TB



智能抗疫
“小珈”助力



第25卷第5期（总第289期）
2020年5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



多相图像分割Vese-Chan模型连续最大流方法(第0926页)



保持区域平滑的3维动画形状高效编辑(第1019页)



树形结构卷积神经网络优化的城区遥感图像语义分割(第1043页)

学者观点

多源空一谱遥感图像融合方法进展与挑战

肖亮, 刘鹏飞, 李恒 0851

综述

中国图像工程: 2019

章毓晋 0864

图像处理和编码

相同编码参数HEVC视频重压缩检测

潘鹏飞, 姚晔, 王慧 0879

密集连接卷积网络图像去模糊

吴迪, 赵洪田, 郑世宝 0890

密集网络图像哈希检索

王亚鸽, 康晓东, 郭军, 李博, 张华丽, 刘汉卿 0900

图像分析和识别

域自适应城市场景语义分割

张桂梅, 潘国峰, 刘建新 0913

多相图像分割Vese-Chan模型连续最大流方法

王洁, 潘振宽, 魏伟波, 徐子森 0926

结合注意力机制和多属性分类的行人再识别

郑鑫, 林兰, 叶茂, 王丽, 贺春林 0936

多分辨率特征注意力融合行人再识别

沈庆, 田畅, 王家宝, 焦珊珊, 杜麟 0946

全卷积网络电线识别方法

刘嘉玮, 李元祥, 龚政, 刘心刚, 周拥军 0956

双核压缩激活神经网络艺术图像分类

杨秀芹, 张华熊 0967

图像理解和计算机视觉

高效检测复杂场景的快速金字塔网络SPNet

李鑫泽, 张轩雄, 陈胜 0977

结合自底向上注意力机制和记忆网络的视觉问答模型

闫茹玉, 刘学亮 0993

统筹图像变换与缝合线生成的无参数影像拼接

高炯笠, 吴军, 刘祺昌, 徐刚 1007

计算机图形学

保持区域平滑的3维动画形状高效编辑

冼楚华, 杨煜, 黄俊贤, 李桂清 1019

医学图像处理

肌骨超声图像特征检测及拼接

颜焕欢, 张培镇, 王伊依, 李璇, 阳维, 王青 1032

遥感图像处理

树形结构卷积神经网络优化的城区遥感图像语义分割

胡伟, 高博川, 黄振航, 李瑞瑞 1043

融合累积变异比和集成超限学习机的高光谱图像分类

尹玉萍, 魏林, 刘万军 1053

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Continuous max-flow method for multiphase segmentation Vese-Chan model(P0926)



Efficient 3D animation shape manipulation with region smoothness preservation(P1019)



Semantic segmentation of urban remote sensing image based on optimized tree structure convolutional neural network (P1043)

Scholar View

Progress and challenges in the fusion of multisource spatial-spectral remote sensing images

Xiao Liang, Liu Pengfei, Li Heng 0851

Review

Image engineering in China: 2019

Zhang Yujin 0864

Image Processing and Coding

Detection of double compression for HEVC videos with the same coding parameters

Pan Pengfei, Yao Ye, Wang Hui 0879

Motion deblurring method based on DenseNets

Wu Di, Zhao Hongtian, Zheng Shibao 0890

Image Hash retrieval with DenseNet

Wang Yage, Kang Xiaodong, Guo Jun, Li bo, Zhang Huali, Liu Hanqing 0900

Image Analysis and Recognition

Domain adaptation for semantic segmentation based on adaption learning rate

Zhang Guimei, Pan Guofeng, Liu Jianxin 0913

Continuous max-flow method for multiphase segmentation Vese-Chan model

Wang Jie, Pan Zhenkuan, Wei Weibo, Xu Zisen 0926

Improving person re-identification by attention and multi-attributes

Zheng Xin, Lin Lan, Ye Mao, Wang Li, He Chunlin 0936

Multi-resolution feature attention fusion method for person re-identification

Shen Qing, Tian Chang, Wang Jiabao, Jiao Shanshan, Du Lin 0946

Power line recognition method via fully convolutional network

Liu Jiawei, Li Yuanxiang, Gong Zheng, Liu Xingang, Zhou Yongjun 0956

Art image classification with double kernel squeeze-and-excitation neural network

Yang Xiuqin, Zhang Huaxiong 0967

Image Understanding and Computer Vision

Snap Pyramid Network: Real-time complex scene detecting system

Li Xinze, Zhang Xuanxiong, Chen Sheng 0977

Visual question answering model based on bottom-up attention and memory network

Yan Ruyu, Liu Xueliang 0993

Stitching of parametric-free images based on coordinated image transformation and seam-line generation

Gao Jiongli, Wu Jun, Liu Qichang, Xu Gang 1007

Computer Graphics

Efficient 3D animation shape manipulation with region smoothness preservation

Xian Chuhua, Yang Yu, Huang Junxian, Li Guiqing 1019

Medical Image Processing

Feature detection algorithm of musculoskeletal ultrasound image and its application of image stitching

Yan Huanhuan, Zhang Peizhen, Wang Yinong, Li Xuan, Yang Wei, Wang Qing 1032

Remote Sensing Image Processing

Semantic segmentation of urban remote sensing image based on optimized tree structure convolutional neural network

Hu Wei, Gao Bochuan, Huang Zhenhang, Li Ruirui 1043

Ensemble extreme learning machine with cumulative variation quotient for hyperspectral image classification

Yin Yuping, Wei Lin, Liu Wanjuan 1053

中图分类号: TN911.73 文献标识码: A 文章编号: 1006-8961(2020)05-0956-11

论文引用格式: Liu J W, Li Y X, Gong Z, Liu X G and Zhou Y J. 2020. Power line recognition method via fully convolutional network. Journal of Image and Graphics, 25(05):0956-0966 (刘嘉玮, 李元祥, 龚政, 刘心刚, 周拥军. 2020. 全卷积网络电线识别方法. 中国图象图形学报, 25(05):0956-0966) [DOI:10.11834/jig.190316]

全卷积网络电线识别方法

刘嘉玮¹, 李元祥¹, 龚政¹, 刘心刚², 周拥军¹

1. 上海交通大学航空航天学院, 上海 200240; 2. 中国航空工业集团公司雷华电子技术研究所, 无锡 214063

摘要: **目的** 电线预警对于直升机和无人飞行器的低空飞行安全至关重要, 利用可见光和红外图像识别电线是一个有效途径。传统识别方法需要人工设计的滤波器提取电线的局部特征, 再使用 Hough 变换等方法找出直线, 支持向量机和随机森林等机器学习方法仅给出图像中有无电线的识别结果。本文提出一种基于全卷积网络的电线识别方法, 能在自动学习特征提取器的同时得到电线的具体位置等信息。**方法** 首先利用复杂背景生成大量包含电线图像和像素标签的成对仿真数据; 然后改进 U-Net 网络结构以适应电线识别任务, 使用仿真数据进行网络训练。由于图像中电线所占的像素很少, 因此采用聚焦损失函数以平衡大量负样本的影响。**结果** 在一个同时包含红外图像和可见光图像各 4 000 幅的电力巡线数据集上, 与 VGG (visual geometry group) 16 等多种特征的随机森林方法相比, 本文方法的电线识别率达到了 99% 以上, 而虚警率不到 2%; 同时, 本文方法输出的像素分割结果中, 电线基本都能被识别出来。**结论** 本文提出的全卷积网络电线识别方法能够提取电线的光学图像特征, 而且与传统机器学习方法相比能将电线从场景中精确提取出来, 使得识别结果更加有判断的依据。

关键词: 全卷积网络 (FCN); 电线识别; 低空飞行安全; 仿真数据; 多源图像

Power line recognition method via fully convolutional network

Liu Jiawei¹, Li Yuanxiang¹, Gong Zheng¹, Liu Xingang², Zhou Yongjun¹

1. School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China;

2. Leihua Electronic Technology Research Institute, Aviation Industry Corporation of China, Ltd. (AVIC), Wuxi 214063, China

Abstract: **Objective** Tens of accidents involving helicopters occur every year owing to collisions with trees, wires, poles, and man-made buildings at low altitude. Just in 2014—2016, there were 96 crashes caused by hitting power lines around the world. Thus, warnings and avoiding wires are important for the low-altitude flight safety of helicopters and unmanned aerial vehicles. According to relevant studies, utilization of optical images is an effective way to identify wires. Traditional methods use manual filters to extract features of power lines and then use Hough transform to detect the lines. Machine learning methods, such as VGG (visual geometry group) 16 and random forest (RF), can only obtain a classification result for a picture, which makes confirming accuracy difficult. The full connection layer of the traditional convolutional neural network (CNN) is effective at classification tasks. However, it cannot carry out pixel segmentation tasks because of the loss of location information. By contrast, the fully convolutional network has no full connection layer, which misses location information. One kind of fully convolutional network, U-Net, is proposed to solve problems such as cell segmentation and retina segmentation. U-Net works well under the conditions of a small amount of samples and a small slice. A three-channel image is input into the network. Through the encoder and decoder, it finally becomes a one-channel feature

收稿日期: 2019-07-03; 修回日期: 2019-10-24; 预印本日期: 2019-10-31

基金项目: 工业和信息化部民用飞机专项科研项目 (MJZ-2016-S-44)

Supported by: Civil Aircraft Special Project of Ministry of Industry and Information Technology (MJZ-2016-S-44)

map via 1×1 kernel size convolution. To obtain the final value between 0 and 1, Sigmoid activation function is used before every convolution layer. In this study, a CNN recognition method based on U-net is proposed to detect power lines.

Method First, we obtain a power line data set containing 8 000 images with 4 000 pairs of visible and infrared images. The image size is 128×128 pixels, with each image having a pixel ground truth label. The network receptive field calculation formula is used to determine the depth of our network. Next, adjustments are made on this basis network to choose the best model. The basis network is named the U-Net-0 model. The U-Net-1 model removes the lower pooling layer in the U-Net-0 model and changes the step size of the convolution layer before the lower pooling layer to 2. It also removes the upper pooling layer and changes the convolution layer after the upper pooling layer to the inverse convolution layer with a step size of 2. Compared with U-Net-0, the U-Net-2 model eliminates the upper and lower pooling layers and the convolution layer in the middle, thereby reducing the network depth. In the U-Net-3 model, decoding is expected to be a dimensionality reduction process. Therefore, the number of convolution kernels of the decoding part is limited, and the number of parameters of feature graph output of each layer is not larger than that of the previous layer. Pictures with complex backgrounds are likewise used to generate a large number of paired synthetic data, including power line images with pixel labels. The generated synthetic data are then used for network training. For each image, the power line contains a small number of pixels. Thus, focal loss is used to balance the impact of a large number of negative samples. The four models use the same optimizer named “Adam”, which can automatically adjust the learning rate on the basis of SGD (stochastic gradient descent). The training procedure of each model is accelerated using an NVIDIA GTX 1080 TI device, which takes approximately 18 hours in 6 000 iterations with a batch size of 64. Loss, F1 score, and intersection-over-union (IoU) are the three evaluative criteria for trained models. The best model usually has low loss and high F1 score and IoU. Each model is used on visible and infrared images. The two results are combined to make a judgment. The power line, regardless of which of the same pair includes it, is finally considered detected in the mixed result. **Result** After these four models are tested on the data set, the number of correctly identified pixels and IoU on each image is counted. According to the statistical results, the IoU of most image recognition results exceeds 0.2, and the threshold of 30 pixels as the result classification is relatively good. If more than 30 pixels are identified on an image, this image might include a power line. By this standard, the proposed method achieves a recognition rate over 99%, while the false alarms are less than 2%. Moreover, VGG16, which is trained on 3 800 pairs of images and tested on 200 pairs of images, only obtains a recognition rate of 95% and a false alarm rate of 37%. RF is affected by feature extraction methods. Thus, the recognition rate and false alarm rate fluctuate greatly. For example, RF with local binary patterns has a recognition rate of 63.5% and a false alarm rate of 36.3% on infrared images. In addition, RF with discrete cosine transform obtains a recognition rate of 92.95% and a false alarm rate of 13.95% on infrared images. Although U-Net-3 has more learnable parameters than U-Net-2, its performance is substantially worse.

Conclusion Our models have higher recognition rates and lower false alarm rates than do other traditional methods on the same dataset. Results show that our models are more effective than other methods and can even clearly extract power lines from background. Our models are trained on synthetic data and tested on real data, which means better generalization performance. The comparison of the four models also shows that the number of parameters cannot completely determine the performance of the network and that the reasonable structure is important. However, our current models have a small receptive field and cannot be used for power line recognition in high-resolution images. In the future, the models will be further studied to increase their receptive field for adapting to larger images without greatly increasing the number of parameters.

Key words: fully convolutional network (FCN); power line recognition; low-altitude flight safety; synthetic data; multi-source images

0 引言

高压线是低空飞行中最容易引起事故的障碍之一,因其体积小,飞行员受注意力影响往往难以及时

发现,近年多有直升机撞线事故发生。2018年6月,安徽一架直升机喷洒农药时超低空飞行,触碰高压线坠毁,飞行员当场身亡。根据张娟等人(2017)的报告,仅2014—2016年期间全球撞击电线导致的坠机事故就有96起。同时高压线对无人飞行器的低

空飞行安全也有很大影响。为保证飞行安全,低空高压电线的自动识别和预警有很大的应用需求。

基于传统方法在电线识别方面的研究已初见成效,包括设计特殊滤波器、使用边缘检测算法后采取 Hough 变换、直线分割器(line segment detector, LSD)等方法。时圣革等人(2015)提出基于方向可调滤波的航拍图像电线检测算法,用多种滤波器提取类似电线的纹理;Song 和 Li(2014)提出先使用匹配滤波器和 1 阶高斯导数区分对称边缘与非对称边缘,再使用图论方法将电线分割区域连接成整条电线。操昊鹏等人(2018)使用 Hough 变换提取电线的大致位置,在小范围内使用总体最小二乘法直线拟合;Baker 等人(2016)在人为设定初值的情况下,使用抛物线的 Hough 变换跟踪图像中的电线。这些传统方法能够从图中找到电线的位置以及走势,但需要人工调整诸多参数,特别是滤波器的设计涉及很多参数,难以保证在不同场景都有满意的效果。

近年来,机器学习理论在目标检测和识别应用中取得了很大成功。Yetgin 和 Gerek(2018)提出在使用不同尺度的图像切片进行离散余弦变换提取特征后,采用支持向量机(support vector machine, SVM)或随机森林(random forest, RF)进行分类。Madaan 等人(2017)使用图形学技术制造仿真图像提升数据量,先使用 POV-Ray(persistence of vision raytracer)技术渲染大量仿真图像,再采用神经网络架构搜索寻找比较好的检测框架,但由于目前架构搜索技术不成熟,实际识别效果没有达到较高水平,同时由于这类算法只输出一个分类结果,不能用容易理解的方式展现得出结果的过程,难以令人信服。传统的卷积神经网络的全连接层擅长分类任务,但是由于会丢失位置信息,对像素分割任务无能为力。

全卷积网络由于没有全连接层,可保证分割结果都是基于没有缺失的上下文特征得到,能够使图中的每一个像素都得到一个对应的分类结果。U-Net(Ronneberger 等,2015)作为一种全卷积网络,在少量样本、小切片的情况下,在眼底视网膜分割和细胞分割等应用场景得到了非常优秀的结果。

鉴于 U-Net 在生物医学图像分割上的成功,本文提出一种基于 U-Net 卷积神经网络的电线识别方法。为应对电线数据难以标注及获取成本高等问题,本文方法先在仿真数据上进行训练,输入电线图像切片,得到像素级的分割结果,最后在真实数据集

上进行实验,验证结果。

本文创新点如下:1)改进 U-Net 网络结构以适应电线识别任务。与传统方法和其他分类方法不同,本文方法能够自行学习特征,减少人工经验对算法的影响,最终给出电线的具体位置信息,使结果更有依据。2)使用大量仿真数据进行训练,增强网络的泛化能力。3)同时使用可见光和红外图像进行判断,提高识别能力。

1 数据准备

1.1 真实数据集

本文使用的原始视频数据由土耳其研究团队与土耳其电力传输公司合作在不同季节的土耳其 21 个地区在拍摄得到,红外和可见光图像的原始分辨率分别为 576×325 像素和 $1\,920 \times 1\,080$ 像素(Yetgin 和 Gerek,2018),Yetgin 和 Gerek(2018)从中整理出 8 000 幅有价值图像,包括红外和可见光图像各 4 000 幅,其中“有电线”和“没有电线”的图像各 2 000 幅,统一缩放为 128×128 像素,且红外与可见光图像一一对应。图 1 展示了可见光图像和红外图像各两幅与其对应的真值。

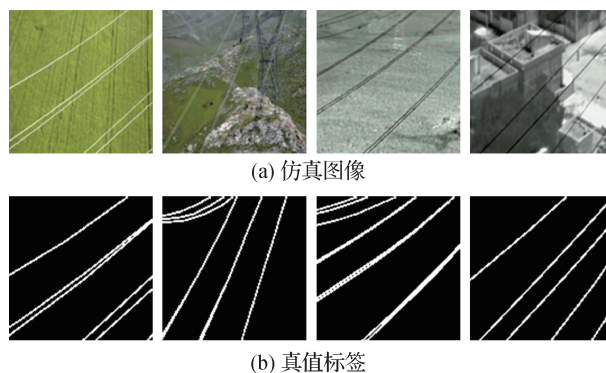


图 1 数据集样例

Fig. 1 Samples of the dataset

((a) synthetic images; (b) ground truth)

由于背景、温度、天气及照明条件不同,最终的数据集包含部分难以识别电线的图像,这些图像因与背景的对比度低导致电线几乎不可见。该数据集没有像素级的标注,基于实验需要,本文对 4 000 幅包含电线的图像进行了像素级的标注。

1.2 仿真数据

真实数据获取成本高,但神经网络需要更加多

样的数据保持泛化能力,于是本文采用仿真数据进行训练,仿真数据可随时生成,多样性大大多于真实数据。

为了得到更多的仿真图像,使用真实数据集中2 000幅没有电线的可见光图像作为背景,在此基础上随机添加类似电线的线段,并且产生对应的像素分类标签。虽然平时很难观察到竖直方向的电线,但是在统计真实数据集中的电线走向后,确定生成的电线倾角在 $\pm 80^\circ$ 之间,一幅图像中的多条电线倾角相差在 $\pm 5^\circ$ 以内,并且不能只在图像边界10%的

区域中生成。仿真过程中,使用16倍超采样抗锯齿保证仿真电线与真实电线的相似性,特别是降低仿真电线的锯齿感。具体来说,创建一个16倍边长的超大型缓冲区,在此缓冲区中随机添加0~5根16倍粗细的电线,对缓冲区高斯模糊后,选取每个 16×16 区域中的256个像素混合成一个像素,使得整个仿真图像中的电线色彩过渡趋于平滑。调整亮度并且获得对应的标签后,叠加到选取的背景图像上,就生成一幅仿真图像。生成过程及更多结果示例如图2和图3所示。

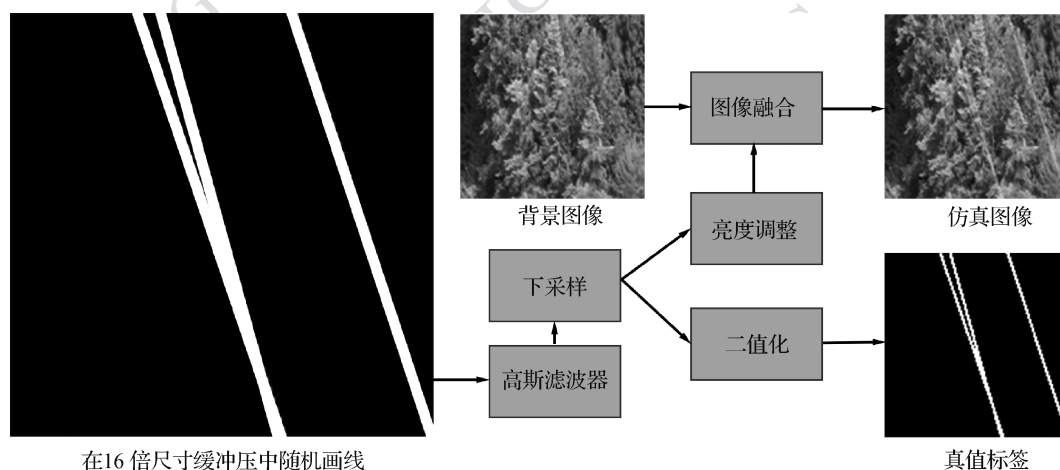


图2 仿真图像生成流程

Fig. 2 Flow chart of the synthetic image generator

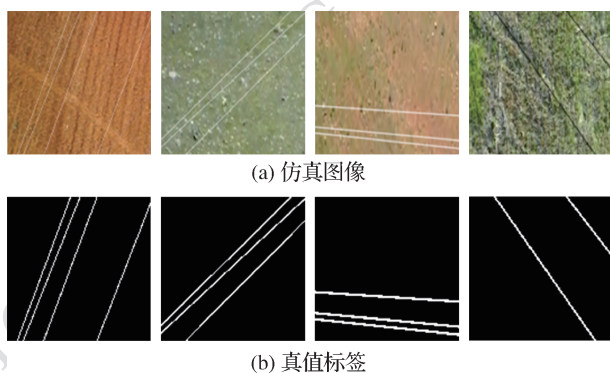


图3 仿真图像与真值标签

Fig. 3 Synthetic images and ground truth labels

((a) Synthetic images; (b) ground truth)

2 基于U-Net卷积神经网络的电线识别

2.1 U-Net卷积神经网络简介

全卷积网络与卷积神经网络类似,对卷积层和

池化层进行级联,降低输入图像的空间维度,生成更多抽象特征。不同点在于卷积神经网络后面有全连接层,可以全局预测感兴趣的对象是否存在。全卷积网络后面没有全连接层,而是使用多重转置卷积或者进行上采样,直到特征图的高宽与输入图像一致,得出每一个像素位置的特征。因此,全卷积网络是端到端的网络。

U-Net是全卷积网络中较为著名的一种,首次出现在生物细胞分割中,对图像进行先解码再编码操作,因结构形似字母U而得名,如图4所示。

卷积神经网络主要由收缩路径(编码器)和扩展路径(解码器)组成。U-Net在编码时,每做一次下采样,便在解码结构的相应位置做一次上采样,同时在两个结构中间使用跳跃连接以保证特征定位准确。U-Net结构对称、整洁明了,便于修改成适合某个任务的形式,有很多改进的种类,网络深度是其中一个主要关注点且已有相关研究。Yu等人(2018)提出的深层聚合网络可以看成多个不同深度的U-Net叠加,

使得网络具有提取不同尺寸特征的能力,因此深的网络对小尺寸目标也有一定作用;Zhou 等人(2018)提出 U-Net ++,在深层聚合网络基础上加入同深度模块间

的跳跃连接,让网络的收敛速度更快,并且可以剪枝优化,删去较深的节点,最终确认网络适宜的深度。图 5 是 3 种 U-Net 网络结构的大致示意。

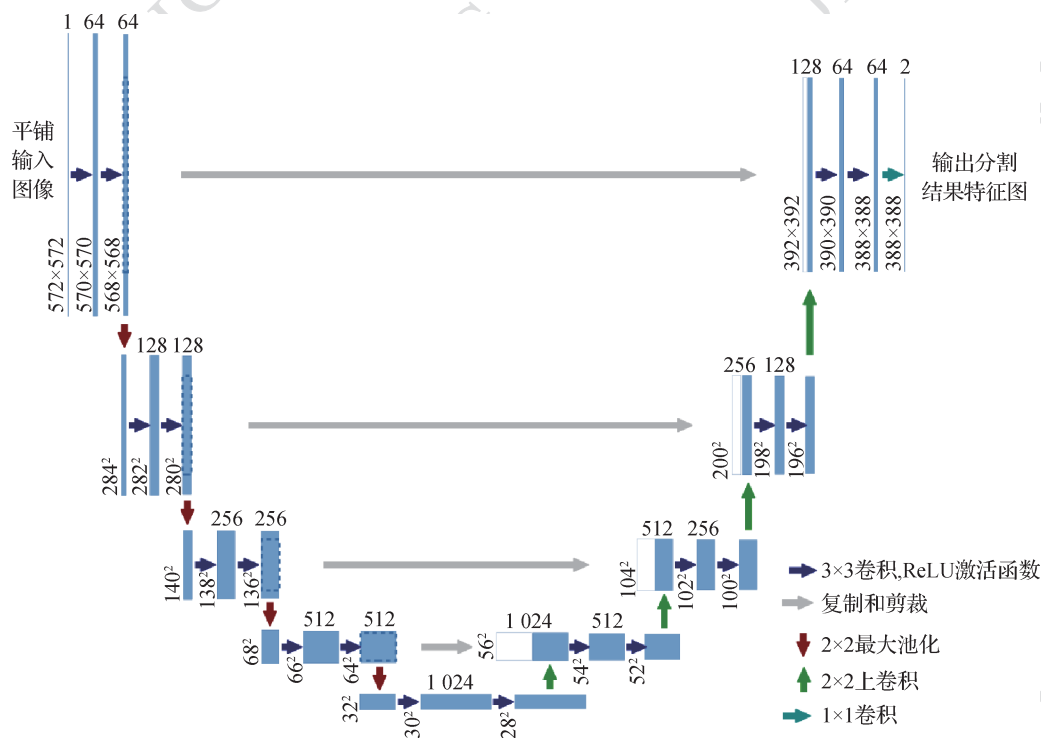


图 4 U-net 分割算法详细结构

Fig. 4 Detailed structure of U-net segmentation algorithm

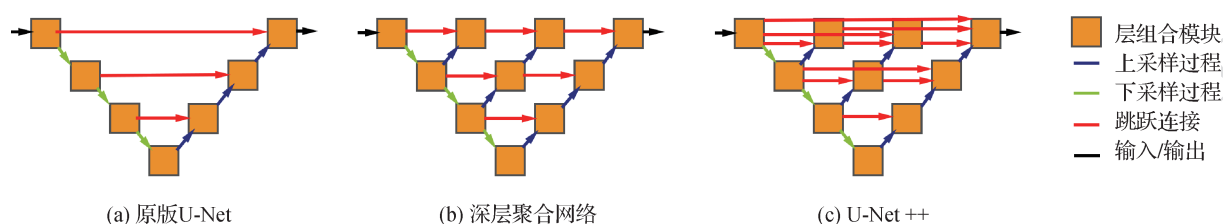


图 5 几种 U-Net 模型结构

Fig. 5 Several U-Net model structures ((a) original U-Net; (b) deep layer aggregation; (c) U-Net ++)

2.2 电线识别网络模型

2.2.1 网络模型结构

针对红外和可见光图像的电线识别问题,对 U-Net 进行修改,以适应电线识别任务。

首先用感受野(Hiên,2017)确定神经网络的深度。在卷积神经网络中,感受野指特征图上某点受输入空间影响的区域,具体为

$$l_k = l_{k-1} + ((f_k - 1) \times \prod_{i=1}^{k-1} s_i) \quad (1)$$

式中, l_k 是第 k 层的感受野, f_k 是第 k 层的卷积核大小, s_i 是第 i 层的步长。如果电线识别网络模型全

部卷积核大小均设定为 3×3 ,每两个卷积层后就有有一个池化层,可以计算层数与感受野尺寸的关系,如表 1 所示。

表 1 网络层数与感受野尺寸

Table 1 Sizes for network layers and their receptive fields

变量	取值							
k	11	12	13	14	15	16	17	18
s_k	1	2	1	1	2	1	1	2
f_k	3	2	3	3	2	3	3	2
l_k	54	62	78	110	126	158	222	254

从表1可以看出,在第15层位置感受野尺寸为 126×126 像素,基本上完全覆盖了输入的 128×128 像素的数据,因此确定最多使用5个池化层构建网络。考虑到尺寸为 3×3 的sobel算子

可以通过旋转和反向得到8个卷积核,而卷积核数量越多,越能提取更多的特征。因此决定第1层用两倍数量,即16个卷积核,之后每深一层卷积核数量翻1倍,最终得到网络结构如图6所示。

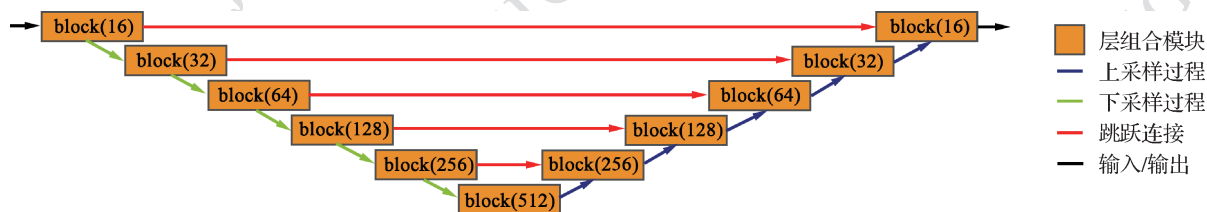


图6 电线识别网络模型结构示意图

Fig. 6 Structure sketch of the power line detection network model

为了保证网络在训练过程中的稳定,在处理之前,将输入数据归一化到 $0 \sim 1$ 之间,并且在每一个卷积层之后使用sigmoid激活函数将输出映射到 $0 \sim 1$ 之间。但是sigmoid函数的饱和区会使得反向传播时梯度消失,训练效果变差。而批规范化(batch normalization, BN)方法能够极大缓解该问题(Ioffe 和 Szegedy, 2015)。批规范化层将输入数据强

行映射到一个均值为0、方差为1的标准正态分布,避开了sigmoid函数的饱和区,使每一层的输入数据都在一个固定范围,不仅防止网络剧烈振荡,而且大幅提升训练速度,优化训练结果。此外,在两个卷积核中间使用参数为0.2的dropout层作为网络中的噪声干扰因素,增加网络的泛化能力。图7为一个块的内部结构示意图。

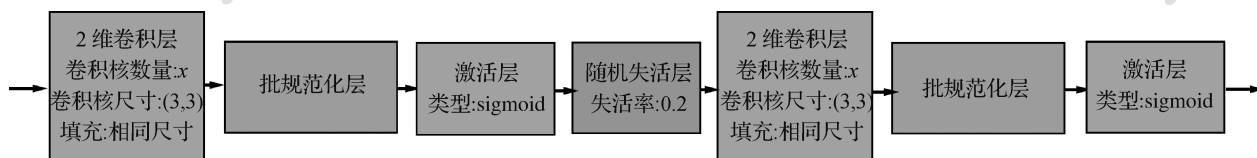


图7 包含两个2维卷积层参数为 x 的块

Fig. 7 A block including 2 conv 2D layers and x parameter

2.2.2 网络训练参数

每幅图上电线像素的数量远小于非电线像素,样本的极度不平衡使得采用传统的交叉熵等损失函数难以训练出正确的电线识别网络。因此网络训练采用焦点损失(focal loss)(Lin等, 2017)作为损失函数,根据正负样本的数量分别加权,抑制数量多的一方,增强数量少的一方,最终计算每幅图上每个像素损失函数的均值。

由于输出图像为 128×128 像素,像素值又归一化到 $0 \sim 1$ 之间,取均值计算出的损失会比较小,手工调节随机梯度下降法的参数较为繁琐,步长太小容易陷入局部最优,步长太大会振荡,最终发散。因此训练选择Adam(Kingma和Ba, 2017)优化算法进行优化,该算法是随机梯度下降法的扩展,结合了适应性梯度算法(AdaGrad)和均方根传播算法(RMSProp)的优点,能够自适应地调整学习速

率,并且使用默认参数能在绝大多数情况下取得较优效果。

为了让网络尽可能地训练充分,默认设置最大训练迭代次数为2000,每一代包含30步,每一个批次包含64个样本,但测试结果在200代时仍未变得更优则终止训练。训练中采用损失函数 L 、F1分数值 S_{F1} 、重叠度(intersection over union, IoU)值 S_{IoU} 作为评价标准,具体计算为

$$L = -y \log y' - (1 - y) \log (1 - y') \quad (2)$$

$$S_{F1} = 2 \cdot \frac{T_p}{2T_p + F_p + F_n} \quad (3)$$

$$S_{IoU} = \frac{T_p}{T_p + F_p + F_n} \quad (4)$$

式中, y 为标签值, y' 为网络输出值, T_p 为电线被识别为电线的像素数量, F_p 为背景被识别为电线的像素数量, F_n 为电线被识别为背景的像素数量。

3 实验结果与分析

3.1 网络结构实验

网络深度确定后仍有一些参数需要调节。例如,使用带步长的卷积层还是池化层进行降采样、是否在上采样时设置不同的卷积核数量等。为了验证这些变化的影响,选择细节有差异的网络模型进行对照实验。U-Net-0 即 2.2.1 节中的网络模型, U-Net-1 将 U-Net-0 的下池化层和上池化层去掉,将下池化层前的卷积层步长改为 2,将上池化层后的卷积层改为步长为 2 的反置卷积层。U-Net-2 在 U-Net-0 的基础上,删去一次上下池化和中间的卷积层,降低了网络的深度。U-Net-3 希望解码过程是一个降维过程,在 U-Net-0 的基础上限制了解码部分的卷积核数量,使得每一层输出的特征图参数数量都不大于上一层的数量。

将这 4 个网络模型用同样的参数进行训练,训练时实时产生仿真数据输入网络,训练过程中评价

标准 L 、 S_{F1} 和 S_{IoU} 随迭代次数的变化趋势如图 8 所示。可以看出, U-Net-1 效果最好, U-Net-0 次之。虽然使用带步长的卷积下采样和上采样会导致特征偏移(Dai 等, 2017), 每次向图像左上方偏移约 0.5 像素。但此处的偏移实际不到 3 像素, 通过学习应该能够消除一定影响。U-Net-2 由于参数量较少, 实际效果不如 U-Net-0, 但相差不大, 对嵌入式等场景可以使用这种删减层数的网络模型。U-Net-3 虽然比 U-Net-2 深, 但各项指标均不如 U-Net-2。U-Net-3 参数约为 23 M, 远多于 U-Net-2 的 1.9 M。因此模型深浅, 参数多少并不能单纯地决定模型的好坏, 需要实际对模型进行验证。

为了测试第 1 层 16 个卷积核是否冗余, 将 4 个网络的卷积核数量减半, 然后使用同样的参数进行训练, 结果如图 9 所示。与图 8 进行对比, 可以看出所有卷积层的卷积核数量减半后, 几个模型的性能有了不同程度的下降。其中 U-Net-0 表现最好, U-Net-1 的下降幅度最大, 而 U-Net-3 的下降幅度最小。因此, 第 1 层使用 16 个卷积核对于模型性能的提高是有意义的。

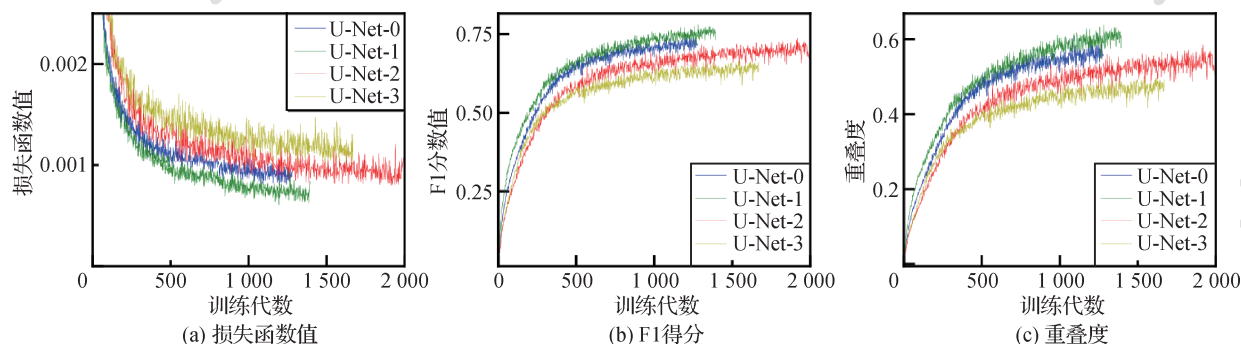


图 8 不同模型的实时评价指标

Fig. 8 Real-time evaluation index of different models ((a) loss function value; (b) F1 score; (c) IoU)

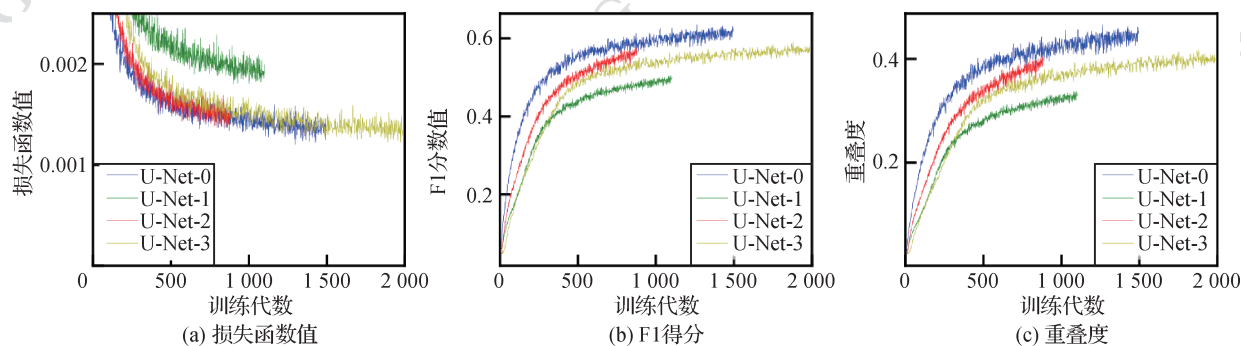


图 9 卷积核减半的不同模型的实时评价指标

Fig. 9 Real-time evaluation index of different models with half filters ((a) loss function value; (b) F1 score; (c) IoU)

3.2 真实数据集测试

本文认为如果一个像素有 50% 以上概率为电线, 则认为这个像素为电线, 反之则不是电线。选取上一节中的模型分别在真实数据集的 8 000 幅图像上进行测试, 可以得到网络在可见光图像和红外图像上的识别结果。本文认为只要在对的一组可见光图像或者红外图像任意一种上识别出电线, 则这一组都识别出了电线, 以此作为综合识别结果。

根据本文自行标注的标签, 在 4 000 幅有电线的图像上识别出电线像素再分别统计, 内容包括每幅图上识别正确的电线像素数量与 IoU 值, 虚警的电线像素数量与 IoU 值, 选择 IoU 值是因为方便统计。图 10—图 12 是 U-Net-0 模型在可见光图像 (visible light image, 图中标注为 VL)、红外图像

(infrared image, 图中标注为 IR) 和综合图像 (mix image, 图中标注为 MIX) 上的统计结果。

从图 10 和图 11 可以看出, 大部分图像的 IoU 值很高, 即便将正确像素比例的阈值设置为 0.2 也可以得到较好的分割效果。但是实际检测过程中没有标签作为参考, 只能使用像素绝对数量作为判断依据。需要说明, 在图 10 (a) (b) 中可以看到很多识别结果为 0 像素的图像, 这是因为可见光图像中有一部分本身没有拍摄清楚的电线, 而对应的红外图像有电线, 但在这一部分图像中没有识别出电线, 真值中的像素数量为 0, 于是认为 IoU 也为 0。此外, 选择不同阈值分别计算可见光图像、红外图像和综合图像上的识别率和误检率, 结果如图 13 和图 14 所示。

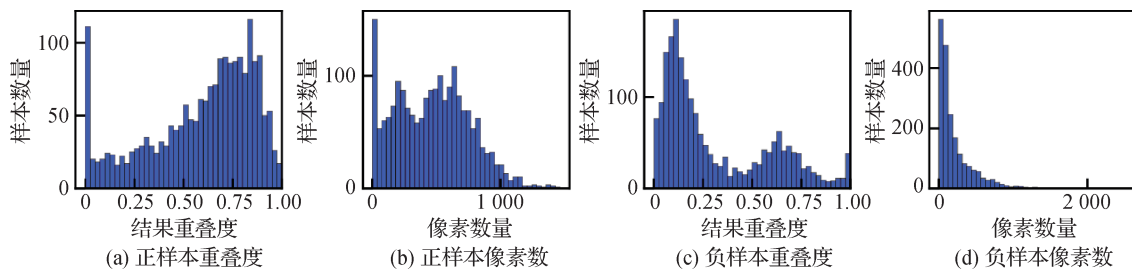


图 10 U-Net-0 模型在可见光图像上的结果统计

Fig. 10 Results of statistical analysis of U-Net-0 on visible images

((a) positive IoU count; (b) positive pixels count; (c) negative IoU count; (d) negative pixels count)

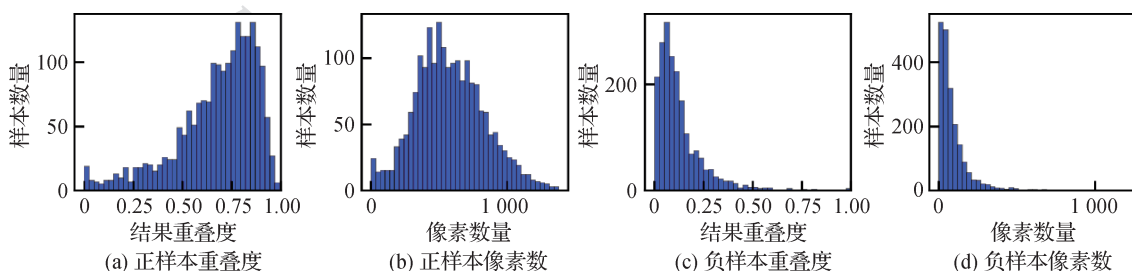


图 11 U-Net-0 模型在红外图像上的结果统计

Fig. 11 Results of statistical analysis of U-Net-0 on infrared images

((a) positive IoU count; (b) positive pixels count; (c) negative IoU count; (d) negative pixels count)

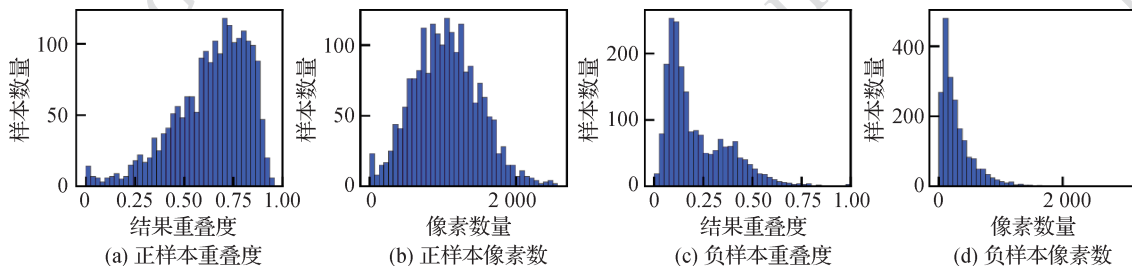


图 12 U-Net-0 模型综合可见光和红外图像上的结果统计

Fig. 12 Results of statistical analysis of U-Net-0 on both visible images and infrared images

((a) positive IoU count; (b) positive pixels count; (c) negative IoU count; (d) negative pixels count)

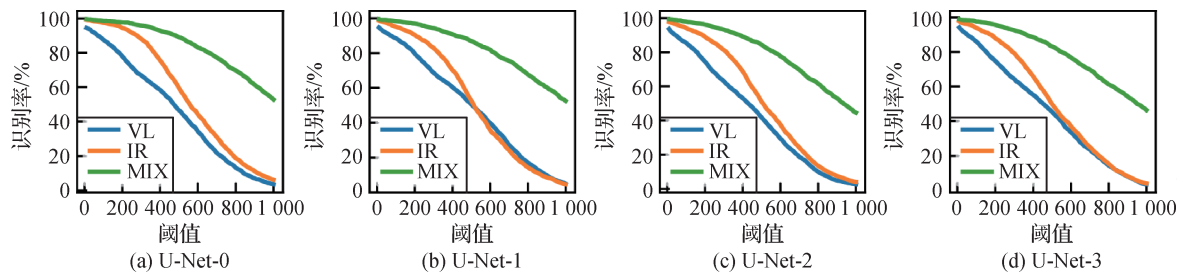


图 13 4 个模型的识别率与阈值的关系

Fig. 13 Relations between threshold and recognition rate of four models ((a) U-Net-0; (b) U-Net-1; (c) U-Net-2; (d) U-Net-3)

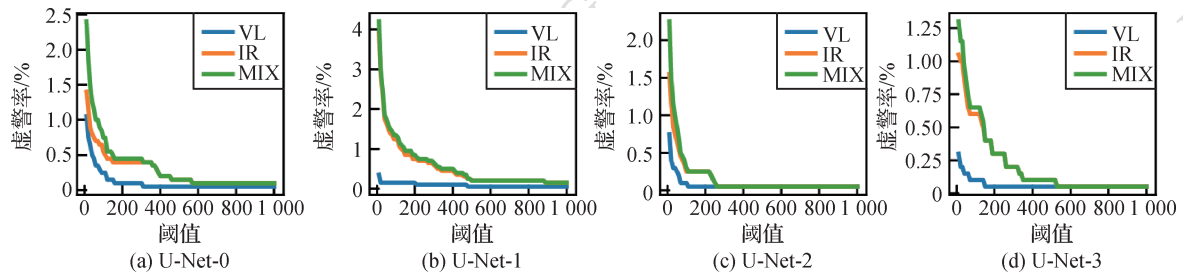


图 14 4 个模型的虚警率与阈值的关系

Fig. 14 Relations between threshold and false alarm rate of four models ((a) U-Net-0; (b) U-Net-1; (c) U-Net-2; (d) U-Net-3)

根据上述结果,选择阈值为 30 的 U-Net-0 模型与其他两类方法在本文使用数据集上的测试结果进行比较。第 1 类本文选择离散余弦变换(discrete cosine transform, DCT)、局部二值模式(local binary patterns, LBP)和方向梯度直方图(histogram of oriented gradient, HOG)等结果较好的方法,使用随机森林对提取的特征进行分类,这类方法采用多种特征提取方

法,使用多种分类传统机器学习方法比较效果(Yetgin 和 Gerek, 2018),但测试数据和训练数据有重合,无法保证泛化能力。第 2 类使用 VGG(visual geometry group) 16 对图像直接分类(Zhang 等, 2018),将 1 900 幅可见光图像用于训练,最后的 100 幅用于测试,因此测试样本较小。这两类方法都只是直接识别 1 幅图像切片中是否包含电线,结果如表 2 所示。

表 2 电线场景识别率和虚警率

Table 2 Scene recognition rate and false alarm rate of power wires

方法	图像类别	有电线样本总数	识别数	识别率/%	无电线样本总数	虚警数	虚警率/%
DCT + RF	可见光	2 000	1 859	92.95	2 000	279	13.95
	红外	2 000	1 982	99.1	2 000	87	4.35
LBP + RF	可见光	2 000	1 242	62.1	2 000	916	4.58
	红外	2 000	1 269	63.45	2 000	725	36.25
HOG + RF	可见光	2 000	1 550	77.5	2 000	519	25.95
	红外	2 000	1 629	81.45	2 000	326	16.30
VGG16	可见光	100	95	95	100	37	37.00
本文	可见光	2 000	1 870	93.5	2 000	12	0.60
	红外	2 000	1 978	98.9	2 000	17	0.85
	混合	2 000	1 986	99.3	2 000	29	1.45

注:加粗字体为各列最优值。

从表2可以看出,本文方法在识别率和误检率上均有非常不错的综合表现,结合可见光和红外两种数据源的综合结果更是比两种源单独识别结果都要好。与其他方法对比,本文方法在每幅图像上具有像素级的识别结果,能够供飞行员确认是否真地检测出了电线,而不是仅得出一个“有无电线”的结论,使得模型更加有说服力。4种模型在图1的4个样本上的识别结果如图15所示。

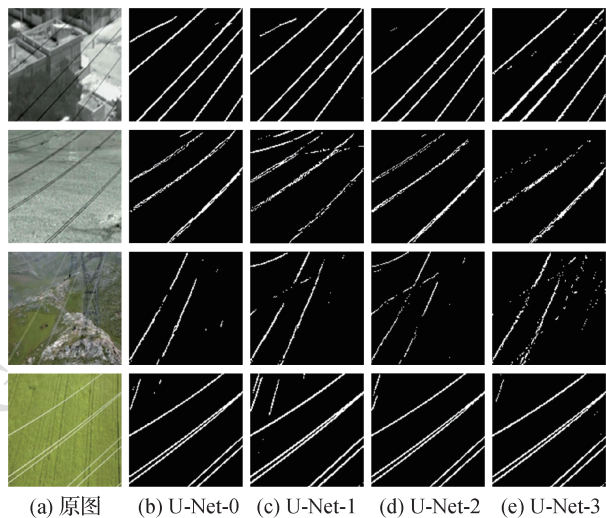


图15 结果样例

Fig. 15 Samples of results ((a) original images; (b) U-Net-0; (c) U-Net-1; (d) U-Net-2; (e) U-Net-3)

4 结论

针对传统方法和其他机器学习方法在电线识别中存在的问题,提出一种全卷积网络模型,能够从大量数据中自行提取电线特征,避免了通过人工经验提取特征带来的误差,同时与其他一些识别方法仅得出“有无电线”的结果不同,本文方法的结果是得到一幅包含电线位置信息的二值图。为使模型泛化能力更强,先使用大量仿真数据对有细微差异的几种模型进行训练,然后在真实拍摄的数据上进行测试,并使用可见光图像和红外图像综合检测,使结果更具鲁棒性。

在真实数据集上的测试及与其他方法的对比表明,本文方法在自动提取特征的情况下达到了更好的识别效果及更低的虚警率,并且像素级检测结果使得“有无电线”的判断有更强的依据。

但是本文方法由于模型结构问题,暂不能用于

高清图像,这是因为计算过程中的参数量与图像尺寸正相关,且感受野覆盖面积相对较小,随着图像尺寸增大,硬件性能成为瓶颈。同时本文的仿真数据是基于来自土耳其的真实数据集制作,生成的电线都是小切片的俯拍,不能随意生成大尺寸图像的电线。今后将在这些问题上继续开展研究。

参考文献 (References)

- Baker L, Mills S, Langlotz T and Rathbone C. 2016. Power line detection using Hough transform and line tracing techniques//Proceedings of 2016 International Conference on Image and Vision Computing New Zealand (IVCNZ). Palmerston North, New Zealand: IEEE: 1-6 [DOI: 10.1109/IVCNZ.2016.7804438]
- Cao H P, Zeng W M, Shi Y H and Xu P. 2018. Power line detection based on Hough transform and total least squares method. Computer Technology and Development, 28(10): 164-167 (操昊鹏, 曾卫明, 石玉虎, 徐鹏. 2018. 基于 Hough 变换和总体最小二乘法的电力线检测. 计算机技术与发展, 28(10): 164-167) [DOI: 10.3969/j.issn.1673-629X.2018.10.034]
- Dai J F, Qi H Z, Xiong Y W, Li Y, Zhang G D, Hu H and Wei Y C. 2017. Deformable convolutional networks//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 764-773 [DOI: 10.1109/ICCV.2017.89]
- Hiên Đ H T. 2017. A guide to receptive field arithmetic for Convolutional neural networks [EB/OL]. (2017-06-04) [2019-06-05]. <https://syncedreview.com/2017/05/11/a-guide-to-receptive-field-arithmetic-for-convolutional-neural-networks/>
- Ioffe S and Szegedy C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-03-02) [2019-06-05]. <https://arxiv.org/pdf/1502.03167v3.pdf>
- Kingma D P and Ba J L. 2017. Adam: a method for stochastic optimization [EB/OL]. (2017-01-30) [2019-06-05]. <https://arxiv.org/pdf/1412.6980v9.pdf>
- Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2017. Focal loss for dense object detection//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 2999-3007 [DOI: 10.1109/ICCV.2017.324]
- Madaan R, Maturana D and Scherer S. 2017. Wire detection using synthetic data and dilated convolutional networks for unmanned aerial vehicles//Proceedings of 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vancouver BC, Canada: IEEE: 3487-3494 [DOI: 10.1109/IROS.2017.8206190]
- Ronneberger O, Fischer P and Brox T. 2015. U-net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241

[DOI: 10.1007/978-3-319-24574-4_28]

Shi S G, Wang M and Dong L W. 2015. Power line detection algorithm of aerial images based on steerable filter. Optics and Optoelectronic Technology, 13(6): 45-48 (时圣革, 王淼, 董力文. 2015. 一种基于方向可调滤波的航拍图像电力线检测算法. 光学与光电技术, 13(6): 45-48)

Song B Q and Li X L. 2014. Power line detection from optical images. Neurocomputing, 129: 350-361 [DOI: 10.1016/j.neucom.2013.09.023]

Yetgin Ö E and Gerek Ö N. 2018. Automatic recognition of scenes with power line wires in real life aerial images using DCT-based features. Digital Signal Processing, 77: 102-119 [DOI: 10.1016/j.dsp.2017.10.012]

Yu F, Wang D Q, Shelhamer E and Darrell T. 2018. Deep layer aggregation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA; IEEE: 2403-2412 [DOI: 10.1109/CVPR.2018.00255]

Zhang J, Zhan Y M and Wang Y M. 2017. Helicopter accident statistics and analysis all over the world during 2014 and 2016. Helicopter Technique, 193(3): 68-72 (张娟, 詹月玫, 王咏梅. 2017. 2014-2016 年世界直升机事故统计及分析[J]. 直升机技术, 193(3): 68-72) [DOI: 10.3969/j.issn.1673-1220.2017.03.015]

Zhang X C, Xiao G, Gong K, Zhao J H and Bavirisetti D P. 2018. Automatic power line detection for low-altitude aircraft safety based on deep learning//Proceedings of 2018 International Conference on Aerospace System Science and Engineering. Singapore: Springer: 169-183 [DOI: 10.1007/978-981-13-6061-9_11]

Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2018. UNet ++: a nested U-net architecture for medical image segmentation//Proceedings of the 4th International Workshop Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada, Spain: Springer: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]

作者简介



刘嘉玮, 1995 年生, 男, 硕士研究生, 主要研究方向为图像处理与机器学习的理论和应用。E-mail: eiwqueqie@sjtu.edu.cn



李元祥, 通信作者, 男, 副教授, 博士生导师, 主要研究方向为图像识别、图像压缩、遥感图像定位。

E-mail: yuanxli@sjtu.edu.cn

龚政, 男, 本科生, 主要研究方向为图像处理。

E-mail: gz666666@sjtu.edu.cn

刘心刚, 男, 高级工程师, 主要研究方向为机载监视防撞系统。E-mail: liuxingang@chinaleihua.com

周拥军, 男, 讲师, 主要研究方向为自然地理学和测绘学。

E-mail: yjzhou@sjtu.edu.cn