



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

中国图象图形学报

2021
11
VOL.26

ISSN1006-8961
CN11-3758/TB



电力视觉前沿技术



第26卷第11期（总第307期）
2021年11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

主 编 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

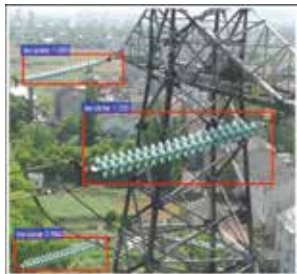
Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

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

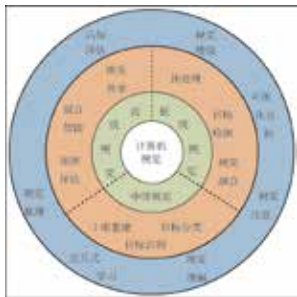
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

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



面向改进尺度缩放网络的绝缘子识别(第2561页)



混合增强视觉认知架构及其关键技术进展(第2619页)



激光点云的稀疏体素金字塔邻域构建与分类(第2703页)

编者按 I

电力视觉前沿技术

输电线路部件视觉缺陷检测综述

赵振兵, 蒋志钢, 李延旭, 威银城, 翟永杰, 赵文清, 张珂 2545

面向改进尺度缩放网络的绝缘子识别

赵文清, 张海明, 徐敏夫 2561

最优知识传递宽残差网络输电线路螺栓缺陷图像分类

威银城, 金超熊, 赵振兵, 丁洁涛, 吕斌 2571

可变形NTS-Net的螺栓属性多标签分类

张珂, 何颖宣, 赵凯, 冯晓晗, 赵振兵, 马占宇 2582

嵌入双注意力机制的Faster R-CNN航拍输电线路螺栓缺陷检测

威银城, 武学良, 赵振兵, 史博强, 聂礼强 2594

轻量化航拍图像电力线语义分割

许刚, 李果 2605

综述

混合增强视觉认知架构及其关键技术进展

王培元, 关欣 2619

深度学习人脸特征点自动定位综述

徐亚丽, 赵俊莉, 吕智涵, 张志梅, 李劲华, 潘振宽 2630

图像处理和编码

多尺度特征复用混合注意力网络的图像重建

卢正浩, 刘丛 2645

图像分析和识别

自监督深度离散哈希图像检索

万方, 强浩鹏, 雷光波 2659

L-UNet: 轻量化云遮挡道路提取网络

许苗, 李元祥, 钟娟娟, 左宗成, 熊伟 2670

基于多尺度特征多对抗网络的雾天图像识别

陈硕, 钟汇才, 李勇周, 王师峰, 杨建刚 2680

图像理解和计算机视觉

结合动态图卷积和空间注意力的点云分类与分割

宋巍, 蔡万源, 何盛琪, 李文俊 2691

激光点云的稀疏体素金字塔邻域构建与分类

陶帅兵, 梁冲, 蒋腾平, 杨玉娇, 王永君 2703

计算机图形学

单幅植物叶片图像的3维重建

任非儿, 刘通, 杨龙 2713

医学图像处理

深度学习多部位病灶检测与分割

黎生丹, 柏正尧 2723

遥感图像处理

对抗学习遥感图像场景识别

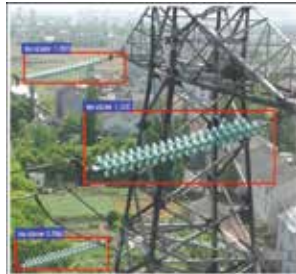
李彤, 张钧萍 2732

融合特征的卫星视频车辆单目标跟踪

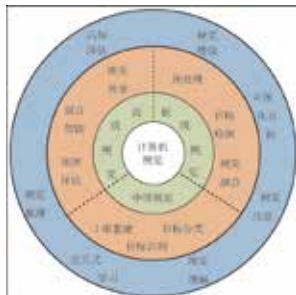
韩鸣飞, 李盛阳, 万雪, 轩诗宇, 赵子飞, 谭洪, 张万峰 2741

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Insulator recognition based on an improved scale-transferrable network(P2561)



Hybrid enhanced visual cognition framework and its key technologies(P2619)



Sparse voxel pyramid neighborhood construction and classification of LiDAR point cloud(P2703)

Frontier Technology of Power Computer Vision

Overview of visual defect detection of transmission line components

Zhao Zhenbing, Jiang Zhigang, Li Yanxu, Qi Yincheng, Zhai Yongjie, Zhao Wenqing, Zhang Ke 2545

Insulator recognition based on an improved scale-transferrable network

Zhao Wenqing, Zhang Haiming, Xu Minfu 2561

Image classification method of transmission line bolt defects using the optimal knowledge transfer wide residual network

Qi Yincheng, Jin Chaoxiong, Zhao Zhenbing, Ding Jietao, Lyu Bin 2571

Multi-label classification method of bolt attributes based on deformable NTS-Net

Zhang Ke, He Yingxuan, Zhao Kai, Feng Xiaohan, Zhao Zhenbing, Ma Zhanyu 2582

Bolt defect detection for aerial transmission lines using Faster R-CNN with an embedded dual attention mechanism

Qi Yincheng, Wu Xueliang, Zhao Zhenbing, Shi Boqiang, Nie Liqiang 2594

Research on lightweight neural network of aerial powerline image segmentation

Xu Gang, Li Guo 2605

Review

Hybrid enhanced visual cognition framework and its key technologies

Wang Peiyuan, Guan Xin 2619

Automatic facial feature points location based on deep learning: a review

Xu Yali, Zhao Junli, Lyu Zhihan, Zhang Zhimei, Li Jinhua, Pan Zhenkuan 2630

Image Processing and Coding

Multiscale feature reuse mixed attention network for image reconstruction

Lu Zhenghao, Liu Cong 2645

Image Analysis and Recognition

Self-supervised deep discrete hashing for image retrieval

Wan Fang, Qiang Haopeng, Lei Guangbo 2659

L-UNet: lightweight network for road extraction in cloud occlusion scene

Xu Miao, Li Yuanxiang, Zhong Juanjuan, Zuo Zongcheng, Xiong Wei 2670

Haze image recognition based on multi-scale feature and multi-adversarial networks

Chen Shuo, Zhong Huicai, Li Yongzhou, Wang Shizheng, Yang Jiangang 2680

Image Understanding and Computer Vision

Dynamic graph convolution with spatial attention for point cloud classification and segmentation

Song Wei, Cai Wanyuan, He Shengqi, Li Wenjun 2691

Sparse voxel pyramid neighborhood construction and classification of LiDAR point cloud

Tao Shuaibing, Liang Chong, Jiang Tengping, Yang Yujiao, Wang Yongjun 2703

Computer Graphics

3D reconstruction of a single plant leaf image

Ren Feier, Liu Tong, Yang Long 2713

Medical Image Processing

Multiorgan lesion detection and segmentation based on deep learning

Li Shengdan, Bai Zhengyao 2723

Remote Sensing Image Processing

Remote sensing image scene recognition based on adversarial learning

Li Tong, Zhang Junping 2732

Integrating multiple features for tracking vehicles in satellite videos

Han Mingfei, Li Shengyang, Wan Xue, Xuan Shiyu, Zhao Zifei, Tan Hong, Zhang Wanfeng 2741

中图法分类号: TN911.73 文献标识码: A 文章编号: 1006-8961(2021)11-2605-14

论文引用格式: Xu G and Li G. 2021. Research on lightweight neural network of aerial powerline image segmentation. Journal of Image and Graphics, 26(11):2605-2618(许刚,李果. 2021. 轻量化航拍图像电力线语义分割. 中国图象图形学报, 26(11):2605-2618)[DOI:10.11834/jig.200690]

轻量化航拍图像电力线语义分割

许刚, 李果*

华北电力大学电气与电子工程学院, 北京 102206

摘要: **目的** 电力线在航拍图像中的提取是智能巡检的重要研究内容, 基于深度学习的图像语义分割模型在此领域的应用已有较好的效果。然而, 图像训练集容量较小和预训练模型计算量过大是两个待解决的问题。**方法** 首先使用生成对抗网络模型结合圆锥曲线和色相扰动进行数据集增强, 对3种不同的损失函数以及两个色彩空间所训练的U-Net网络模型进行对比, 给出最优组合。然后提出了一种联合一阶泰勒展开和输出通道2范数的显著度指标, 对上述完整模型使用改进的通道级参数正则化方法来稀疏化模型权重, 并对稀疏模型进行网络剪枝和重训练以降低模型的计算量。最后, 在判决阈值的选择上, 使用自适应阈值替代固定值法以增强对亮度变化的鲁棒性。**结果** 实验结果表明, 提出的灰度输入轻量化模型IoU(intersection-over-union)指标为0.459, 但其参数量和计算量相当于IoU指标为0.573的可见光完整模型的0.03%和3.05%, 且自适应阈值法在合适的光照变化范围内能达到该条件下最优阈值的相似结果。**结论** 验证了不同数据集增强方法、损失函数、输入色彩空间组合对模型收敛性能、训练速度和过拟合程度的影响, 给出了各色彩空间内的最佳组合。同时, 采用网络剪枝的方式极大降低了电力线语义分割网络的参数量和运算量, 对网络模型的落地部署有积极的作用。

关键词: 智能巡检; 图像语义分割; 稀疏正则化; 网络剪枝; 生成对抗网络(GAN)

Research on lightweight neural network of aerial powerline image segmentation

Xu Gang, Li Guo*

College of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China

Abstract: **Objective** Powerline semantic segmentation of aerial images, as an important content of powerline intelligent inspection research, has received widespread attention. Recently, several deep learning-based methods have been proposed in this field and achieved high accuracy. However, two major problems still need to be solved before deep learning models can be applied in practice. First, the sample size of publicly available datasets is small. Unlike target objects in other semantic segmentation tasks (e.g., cars and buildings), powerlines have few textures and structural features, which make powerlines easy to be misidentified, especially in scenes that are not covered by the training set. Therefore, constructing a training set that contains many different background samples is crucial to improve the generalization capability of the model. The second problem is the conflict between the amount of model computation and the limited terminal computing resources. Previous work has demonstrated that an improved U-Net model can segment powerlines from aerial images with satisfactory accuracy. However, the model is computationally expensive for many resource-constrained inference terminals (e.g.,

收稿日期: 2020-11-12; 修回日期: 2021-04-13; 预印本日期: 2021-04-20

* 通信作者: 李果 liguo@ncepu.edu.cn

unmanned aerial vehicles (UAVs)). **Method** In this study, the background images in the training set were learned using a generative adversarial network (GAN) to generate a series of pseudo-backgrounds, and curved powerlines were drawn on the generated images by utilizing conic curves. In detail, a multi-scale-based automatic growth model called progressive growing of GANs (PGGAN) was adopted to learn the mapping of a random noise vector to the background images in the training set. Then, its generator was used to generate serials of the background images. These background images and the curved powerlines generated by the conic curves were fused in the alpha channel. We created three training sets. The first one consisted of only 2 000 real background pictures, and the second was a mixture of 10 000 real and generated background images. The third training dataset was composed of 200 generated backgrounds and used to evaluate the similarity between the generated and original images. At the input of the segmentation network, random hue perturbation was applied to the images to enhance the generalization of the model across seasons. Then, the convergence accuracy of U-Net networks with three different loss functions was compared in RGB and grayscale color spaces to determine the best combination. Specifically, we trained U-Net with focal, soft-IoU, and Dice loss functions in RGB and gray spaces and compared the convergence accuracy, convergence speed, and overfitting of the six obtained models. Afterward, sparse regularization was applied to the pre-trained full model, and structured network pruning was performed to reduce the computation load in network inference. A saliency metric that combines first-order Taylor expansion and 2-norm metric was proposed to guide the regularization and pruning process. It provided a higher compression rate compared with the 2-norm that was used in the previous pruning algorithm. Conventional saliency metrics based on first-order expansion can change by orders of magnitude during the regularization process, thus making threshold selection during the iterative process difficult. Compared with these conventional metrics, the proposed metric has a more stable range of values, which enables the use of iteration-based regularization methods. We adopted a 0-norm-based regularization method to widen the saliency gap between important and unimportant neurons. To select the decision threshold, we used an adaptive approach, which was more robust to changes in luminance compared with the fixed-threshold method used in previous work. **Result** Experimental results showed that the convergence accuracy of the curved powerline dataset was higher than that of the straight powerline dataset. In RGB space, the hybrid dataset using GAN had higher convergence accuracy than the dataset using only real images, but no significant improvement in gray space was observed due to the possibility of model collapse. We confirmed that hue disturbance can effectively improve the performance of the model across seasons. The experimental results of the different loss functions revealed that the convergence intersection-over-union (IoU) of RGB and gray spaces under their respective optimal loss functions was 0.578 and 0.586, respectively. Dice and soft-IoU had a negligible difference in convergence speed and achieved the best accuracy in gray and RGB spaces, respectively. The convergence of focal loss was the slowest in both spaces, and neither achieved the optimal accuracy. At the pruning stage, by using the conventional 2-norm saliency metric, the proposed gray space lightweight model (IoU of 0.459) reduced the number of floating-point operations per second (FLOPs) and parameters to 3.05% and 0.03% of the full model in RGB space, respectively (IoU of 0.573). When the proposed joint saliency metric was used, the numbers of FLOPs and parameters further decreased to 0.947% and 0.015% of the complete model, respectively, while maintaining an IoU of 0.42. The experiment also showed that the Otsu threshold method worked stably within the appropriate range of illumination changes, and a negligible difference from the optimal threshold was observed. **Conclusion** Improvements in the dataset and loss function independently enhanced the performance of the baseline model. Sparse regularization and network pruning reduced the network parameters and calculation load, which facilitated the deployment of the model on resource-constrained inferring terminals, such as UAVs. The proposed saliency measure exhibited better compression capabilities than the conventional 2-norm metric, and the adaptive threshold method helped improve the robustness of the model when the luminance changed.

Key words: smart inspection; image semantic segmentation; sparse regularization; network pruning; generated adversarial network (GAN)

0 引言

电力线巡线是电网日常维护的重要组成部分,对保障电网健康运行起着至关重要的作用。然而,大量输电线路架设在交通不便的山区,人工巡线的效率较低,而使用无人机巡线可以克服以上缺点,但实时回传的画面清晰度不足,容易造成漏检,因此将故障识别部署到无人机上辅助人工判断是一个较好的方案。从图像中将电力线分割出来是进行电力线缺陷检测的基础,因此开发一种轻量级的电力线提取算法十分必要。

电力线提取可以分为传统方法和基于数据的深度学习方法。在传统方法方面,Zhang等人(2012)在Hough域中提取直线段,并同时进行了线段聚类 and 线追踪;Song和Li(2014)通过FDoG(flow-based difference-of-Gaussians)和匹配滤波器提取图像边缘信息,接着进行形态滤波提取线段区域,并用图理论进行线段聚类,该方法可以实现对曲线电力线的识别;Baker等人(2016)则先对地面拍摄的电力线图像进行预变换以凸显其抛物线特征,再使用Hough变换或者随机Hough变换进行抛物线识别,同时提出了一个基于临近像素的线追踪算法。传统方法在其特定的应用场景内有着较好的效果,但若涉及到该算法参数难以适应的应用场景,则需要专家知识对滤波器、变换域和特征提取进行调整,这使得算法迭代更新较为困难。赵乐等人(2021)对当前的可见光电力线提取算法研究现状进行了综述。基于数据的深度学习方法发展迅速,在多个领域的表现超越了传统算法。深度模型的网络架构一旦确定,就可以通过向数据集中添加数据进行增量训练的方式来提高在特定场景中的模型表现,具有相对较低的算法更新成本。

数据和模型是深度学习方法中至关重要的两个部分。电力线语义分割的具体任务中可供使用的数据集较小,因此选择一种小样本下也能取得较好性能的模型是一个较好的选择,同时,也有必要对数据集进行增强,以提高模型的泛化能力。在语义分割模型方面,将大型图像分类数据集的预训练模型作为主干网络进行特征提取是一种常见的网络设计思路,但此类预训练模型的容量和计算量本身已经较大,对于相对简单的电力线分割任务而言,其特征存

在大量冗余。另一方面,也有一些轻量化的语义分割模型相继提出,如Paszke等人(2017)提出的ENet(efficient neural network),Zhao等人(2018)提出的ICNet(image cascade network)和Yu等人(2018)提出的BiSeNet(bilateral segmentation network)等。电力线的分割任务粒度较粗,且属于较为简单的二分类任务,当前主要方法均采用自行设计或改进基础网络的方式进行电力线提取。Madaan等人(2017)以VGG(Visual Geometry Group)16网络为原型定义了一系列编解码模块,并搜索了最优的编解码模块组合。Zhang等人(2019)则以VGG16作为主干网络并自行设计了解码模块。王栩文(2019)对DeepLab v3+的解码器部分进行了改进,设计了更复杂的解码结构,同时选用ResNet-101或Xception作为主干网络提取特征。刘嘉伟等人(2020)则对U-Net网络进行改进,但该改进网络计算量依然较大,难以在无人机等资源受限的终端进行实际部署。对上述网络性能进行评估后,选择刘嘉伟等人(2020)的改进U-Net网络为基础模型,通过网络剪枝来进一步轻量化网络结构,以降低部署时的资源需求。

除进行模型轻量化之外,该工作在以下几个方面仍有可改进的空间:1)该工作训练集为由2 000幅生成的直线电力线图像,相对于其使用的4 000幅曲线电力线测试集而言容量较小,有必要研究数据集增强方法;2)语义分割研究中有许多其他的损失函数被提出,并且获得了比focal损失函数更好的结果,因此在损失函数选择上仍有改进的空间;3)实际情况中,模型的输出概率分布可能因光照条件和季节变化发生偏移,故有必要研究自适应的阈值选择方法。

总体流程图如图1所示。围绕提升基础模型性能 and 对其进行轻量化两方面,进行了以下改进和研究:

1)提出采用生成对抗网络结合圆锥曲线生成曲线电力线数据集的方法进行数据增强,同时进行色相扰动训练,提升了对季节变化的鲁棒性。

2)在灰度空间和RGB空间上分别对几种损失函数进行测试,证实了在RGB空间采用soft-IoU损失函数或在灰度空间中采用Dice损失函数是较好的组合。

3)提出一种联合通道范数与一阶梯度的显著度指标,并用其指导模型稀疏化;同时对稀疏模型进

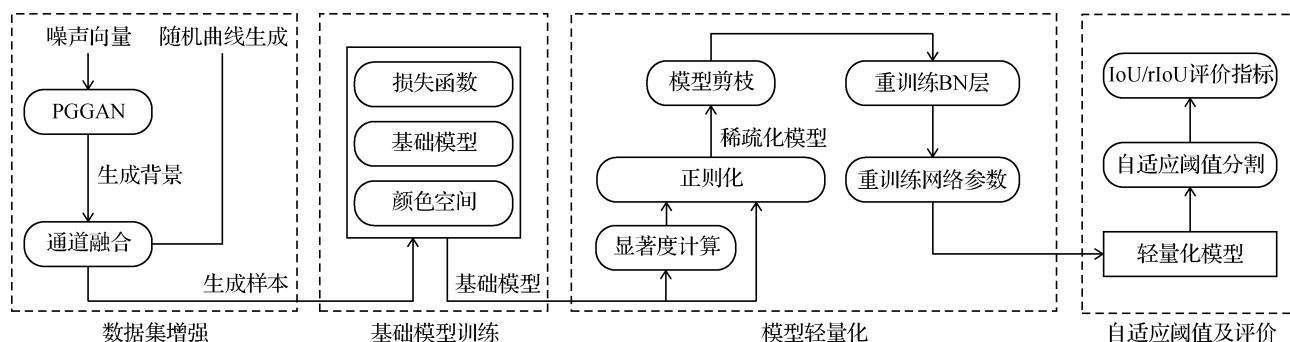


图1 算法整体流程图

Fig. 1 Overall flowchart of the algorithm

行网络剪枝,并采用改进的重训练流程以降低模型部署时的资源需求并尽量保护模型性能。实验中,轻量化灰度模型计算量为完整 RGB 模型的 3.05%,二者的 IoU (intersection-over-union) 指标分别为 0.459 和 0.573。

4)在轻量化网络上采用 Otsu 法进行判决阈值选择,此方法能有效提升模型对亮度变化的鲁棒性。

1 数据集

1.1 数据集

采用监督或半监督的方式训练语义分割网络需要像素级标注精度的数据集,而电力线所具有的纹理特征和像素数量远远少于常规语义分割任务中的目标物体,因此对标注精度的要求也就更高。实验采用 Yetgin 和 Gerek (2019a, b) 的两个数据集,总共包含 4 400 组图像,其中一半为不包含电力线的背景图像,而另一半中只有 200 组图像进行了语义分割任务的标注。若直接将 200 组图像进一步划分为训练、验证和测试集,训练集样本数量难以满足训练需求。故采用生成对抗网络 (generative adversarial networks, GAN) 生成 8 000 幅可见光图像对数据集进行增强。

1.2 数据增强

Goodfellow 等人 (2014) 提出生成对抗网络,其核心是设计一个可将随机生成的噪声信号映射到样本空间的生成网络,同时设计一个判别网络对生成网络的输出和真实数据进行区分,并在训练中对生成网络和判别网络进行交替训练。Arjovsky 和 Bottou (2017) 以及 Arjovsky 等人 (2017) 从理论上指出

先前生成对抗网络损失函数的缺陷,同时提出使用 Wasserstein 距离来衡量生成分布与样本分布的差异性。Karras 等人 (2018) 提出 PGGAN (progressive growing of GANs),该网络使用改进的 Wasserstein 损失函数,并采取了从小尺寸的图像开始逐渐提高到全尺寸的训练方式,同时在每次尺寸提升时对原网络进行自动生长,实现了高达 $1\,024 \times 1\,024$ 像素的高质量人像生成。

首先采用 PGGAN 网络生成部分背景数据进行数据集增强。生成网络输入为从正态分布中随机采样的 100 维噪声向量,输出为 128×128 像素的可见光图像。然后,对生成图像使用双线性插值进行 10 倍上采样,采用圆锥曲线生成电力线样本并进行高斯模糊。具体而言,使用随机数对圆锥曲线的焦点、焦半径、曲线的粗细、透明度以及线段条数进行扰动。最后在透明度通道上对二者进行图像融合,融合公式为

$$p_{\text{out}} = (1 - \alpha)p_{\text{bg}} + \alpha p_{\text{fg}}$$

$$\alpha = \frac{p_{\text{fg}}}{255}(0.8 - 0.5n) \quad (1)$$

式中, p_{out} 为融合像素, α 为混合系数, p_{bg} 为背景像素, p_{fg} 为前景像素 (即生成的电力线像素), n 为从 $0 \sim 1$ 均匀分布中采样的随机数 ($n \sim U(0, 1)$)。生成的图像及其标签如图 2 所示。

2 电力线语义分割网络的轻量化

2.1 基础模型

基础模型采用刘嘉玮等人 (2020) 提出的基于 U-Net 改进网络。U-Net 由 Ronneberger 等人 (2015) 提出,当前已成为语义分割网络的基本模型之一,其

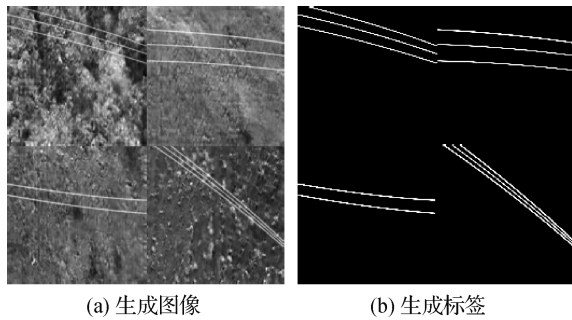


图2 生成图像及生成标签

Fig. 2 Generated images and their labels

((a) generated images; (b) generated labels)

各种变体在各类分割任务中均有广泛应用。实验部分将与 Madaan 等人(2017)以及 Zhang 等人(2019)提出的模型进行对比。

同时,采用 FocalLoss、soft-IoU 和 Dice 这3种损失函数分别训练 RGB 和灰度输入的基础模型,并对收敛性能进行研究,找出最优组合以提升基础模型的性能。3种损失函数的表达式分别为

$$L_{\text{focal}} = -y' \log y - (1 - y') \log(1 - y) \quad (2)$$

$$L_{\text{soft-iou}} = 1 - \frac{\sum_{i=1}^m y_i y'_i}{\sum_{i=1}^m (y_i + y'_i - y_i y'_i)} \quad (3)$$

$$L_{\text{dice}} = 1 - \frac{2 \sum_{i=1}^m y_i y'_i + 1}{\sum_{i=1}^m (y_i + y'_i) + 1} \quad (4)$$

式中, y' 为样本标签, y 为预测值, m 为标签图像的像素总数。

2.2 模型轻量化

在选取基础模型之后,采取模型压缩的方式对模型进行轻量化。基础模型参数数量约为 7.78 M, 处理 60 幅 128×128 像素大小的可见光图像(处理 1280×720 像素大小的视频,每帧需分为 60 个图像块)所需浮点运算数(floating-point operations per second, FLOPs)为 128 G。然而,航拍图像画面变化速度较快,且航拍设备对功率、重量体积等有较高的要求,若不对模型进行压缩直接使用将难以满足应用场景的硬件限制和实时性需求。

网络压缩主要有量化、低秩分解、剪枝和模型蒸馏等几种方法。这些方法通常可互相结合使用,如 Han 等人(2016)采用“迭代剪枝—量化—编码”的

方式。Liu 等人(2019)认为网络剪枝可能是一个网络搜索过程,网络结构搜索需要进行大量的架构尝试,计算代价大,是一个很有前景的研究方向。而选取的基础模型当前已经能够在此任务上取得较好的效果,因此对其进行压缩是一个更直接的选择。为避免模型性能在压缩过程中发生突变,采用保守的先进行稀疏正则化再进行模型剪枝的方法。

为最大限度地进行模型压缩并降低剪枝阶段阈值选取的难度,采用 0 范数正则化的方式进行模型稀疏,并对稀疏模型进行通道级结构剪枝。正则化方式可以最大限度进行模型压缩,同时 0 范数正则化有利于提升通道间神经元的显著度差异,利于修剪阈值的选择。相比于非结构性剪枝对硬件优化或编译器优化的依赖,结构剪枝对于各类未经优化的推理终端均有明显的加速作用。同时,为进一步进行模型压缩,提出了一种联合的显著度指标,可以实现更佳的压缩效果。

2.2.1 显著度计算

显著度计算是剪枝的核心问题之一。因网络过于稀疏,采用输出通道卷积核的 2 范数作为显著度评价指标就可达到较好的压缩效果。需要注意的是,该指标在显著度差异较大的时候能较好发挥作用,但在差异较小时容易造成误剪。以某一层的卷积核为例,第 i 个输出通道的 2 范数为

$$\overline{\phi}_i = \|\mathbf{w}_i\|_2 = \sqrt{\sum_{h,w,c} \mathbf{W}(h,w,c,i)^2} \quad (5)$$

式中, $\mathbf{w}_i$ 为第 i 个输出通道的参数集合, $\mathbf{W}$ 为该层卷积的卷积核, h, w, c, i 分别代表卷积核的高、宽、输入通道和输出通道。

除范数之外,基于泰勒展开的指标因其具有明确的数学基础受到了较多的关注。该方法始于 Le 等人(1989)提出的 OBD(optimal brain damage)算法, Molchanov 等人(2017, 2019)将其应用到深度网络上并对其进行改进,仅进行一阶泰勒展开以规避高阶导数的求解,其输出通道 i 的显著度为

$$\overline{\phi}_i = \sum_{h,w,c} \left(\frac{\partial L(\mathbf{D}, \mathbf{W}) \mathbf{W}(h,w,c,i)}{\partial \mathbf{W}(h,w,c,i)} \right)^2 \quad (6)$$

式中, $L, \mathbf{D}$ 分别为模型损失函数和输入数据集。

然而,模型训练过程中梯度数值变化较大,数量级相当不稳定。因此,通过上述展开法获得的显著度数值难以通过预先定义缩放系数的方式来保持其在迭代过程中的数值范围稳定。这对于迭代过程中

的阈值选择相当不利。相反,2 范数在层内显著度计算的精确度上不及一阶展开法,但从实验中看,2 范数能提供各层显著度的相对参考,具有更稳定的数值范围。因此,提出一种联合指标,先采用一阶展开法计算显著度,并将各层显著度分别放缩到 0~1 区间,再用各层输出通道 2 范数的最大值分别进行加权,以稳定数值范围,便于正则化算法的迭代更新。结合式(5)(6),其表达式如下

$$\Phi = \left\{ \phi_i \mid \phi_i = \frac{\max_{j \in [1, I]} (\bar{\phi}_j) \cdot \bar{\phi}_i}{\max_{j \in [1, I]} (\bar{\phi}_j)}, i \in [1, I] \right\} \quad (7)$$

式中, Φ 为由同一层各输出通道显著度 ϕ_i 组成的显著度向量, I 为该层的输出通道总数。

2.2.2 稀疏正则化

正则化算法的主干采用基于半二次方分裂法的 0 范数正则化算法。其正则化流程如算法 1 所示。

算法 1 0 范数稀疏正则化

输入: 数据集 D , 模型参数 $W^{(0)}$ 。

参数: 初始正则半径 $r^{(0)}$, 终止正则半径 $r^{(N)}$, 迭代步长 σ , 损失倍数 k 。

输出: 模型参数 $W^{(N)}$ 。

1) 令 $r^{(n)} = r^{(0)}$

2) While $r^{(n)} < r^{(N)}$ do

3) 计算各通道显著度 $\phi_i^{(n)}$

4) 根据 $\phi_i^{(n)}$ 计算辅助变量 $T^{(n)} = \{t_1^{(n)}, t_2^{(n)}, \dots\}$

$$t_i^{(n)} = \begin{cases} w_i^{(n)} & \phi_i^{(n)} \geq r^{(n)} \\ 0 & \text{其他} \end{cases}$$

5) 最小化下式计算 $W^{(n)}$ ($\alpha^{(n)}$ 可由 $r^{(n)}$ 算得):

$$L = \min_{W^{(n)}} \{ L(D, W^{(n)}) + \alpha^{(n)} \|W^{(n)} - T^{(n)}\|_2^2 \}$$

6) 更新半径 $r^{(n+1)} = r^{(n)} / \sqrt{\sigma}$

7) End while

返回: 模型参数 $W^{(N)}$ 。

算法 1 中, n 为当前迭代次数, $r^{(0)}, r^{(N)}$ 分别代表起始与结束时的正则半径, 起始半径越大搜索范围越广。迭代终止半径越小则低显著度区域的显著度越低, 剪枝的准确度越高。为稳定二者的取值范围, 在进行正则化之前先将显著度进行归一化, 使得各层的显著度主要分布在 0~1 区间内。在此前提下, $r^{(0)}, r^{(N)}$ 典型值分别为 1 和 0.1。 σ 与半径收敛速度有关, 通常设置为 1.2。 k 用于控制正则项与训

练函数的数值比例, 该比例过大可能造成网络性能严重下降, 一般设置为 8。 $T^{(n)}$ 与 $W^{(n)}$ 形状相同, t_i, w_i 均为第 i 个输出通道的通道参数。

2.2.3 剪枝和重训练

稀疏化后, 通常会将卷积层或批正则化层对应的通道滤波器置 0 以模拟通道被减去的情况, 此时的批正则化层所携带的均值与方差信息统计了大量零输入通道。而实际简化模型中, 零输入通道将被直接移除, 因此批正则化层的统计均值方差与实际值相差较远, 故参数需重新调整。Li 等人(2020)观察到该现象, 同时提出在剪枝后进行前向传播来重新初始化批归一化层的方差和均值, 进而提升子网性能的评价准度。为减少重训练过程中对原网络参数的扰动, 防止破坏已学到的特征信息, 采用以下重训练步骤:

1) 使用大学习率或 Adam, 单独对批归一化层进行重训练, 直到损失函数稳定。

2) 使用小学习率的动量梯度下降法对整个模型参数进行微调。

在重初始化是否会破坏收敛精度的问题上, Liu 等人(2019)以较大学习率进行重训练, 结果显示重训练时重初始化与否不会影响收敛精度; 但 Frankle 和 Carbin(2018)较小学习率的结果显示初值对收敛精度有较大影响。本文实验部分所达到的模型压缩率较高且重训练时学习率较小, 收敛精度体现出初值敏感性, 但若在重训练阶段采用较大学习率, 则会出现相当大的性能振荡。值得一提的是, 在振荡过程中模型仍有可能取得较好的性能。

2.3 自适应阈值及评价指标

语义分割网络的输出部分是一个 Sigmoid 层, 用于将输出特征图映射到 0~1 概率区间。判决阈值采用 Otsu 法进行自适应选择。Otsu 法可找到使得类内方差最小的分割阈值, 二分类即最小化下式

$$\min_t (\sigma_0^2(t) \sum_{i=1}^t p(i) + \sigma_1^2(t) \sum_{i=t+1}^{+\infty} p(i)) \quad (8)$$

式中, t 为分割阈值, $\sigma_0^2(t), \sigma_1^2(t)$ 为二分类情况下两类各自的方差, $p(i)$ 为概率分布直方图。若阈值精度为 0.01, 则 $p(1)$ 为数值在 $[0, 0.01]$ 范围的像素占总数的比例。

然而, 测试集中的部分数据标注误差较大。如图 3 所示, 样本标签的覆盖范围远小于实际电力线。在此情形下, 当且仅当预测结果同样较细且与标签

高度重合时才可能发生测得的 IoU 大于真实 IoU 的情况,而在其他大部分情况下测得的 IoU 小于真实 IoU,故认为数据集所测得的平均 IoU 指标将低于真实的 IoU 值。因此,将 IoU 视为性能的近似下界。

为解决此问题,设计了一个允许一定误差的 rIoU 指标进行辅助评价,其定义和 IoU 指标的定义分别为

$$\begin{cases} f_{rIoU} = \frac{(Y \otimes K) \cap Y'}{Y \cup ((Y \otimes K) \cap Y')} \\ f_{IoU} = \frac{Y \cap Y'}{Y \cup Y'} \end{cases} \quad (9)$$

式中, Y 为样本标签, Y' 为二值化的网络输出, K 为一个膨胀核, $\bar{Y}$ 表示对 Y 进行逻辑逆运算, $\otimes$ 为形态学膨胀操作。

由式(9)可知, rIoU 在分子中膨胀核 K 的作用下,将电力线标签周边被误分类为电力线的像素当成正确分类,这使得该分子一般大于实际 IoU 的分子。而在标签较细或合适的情况下, rIoU 分母小于实际 IoU 的分母,因此数据集的平均 rIoU 大于实际 IoU,可用于表示模型性能的近似上界。

从另一方面,较细和较粗的预测结果能取得相同的 IoU 值,而在电力线检测任务中,后者在一定范围内是一种可以容忍的错误。图4中预测标签随着膨胀次数的增加逐渐变粗,同时 rIoU 指标也逐渐上升。在 IoU 指标相同的情况下, rIoU 指标鼓励选取超出标签的范围,以此避免较细的预测结果。

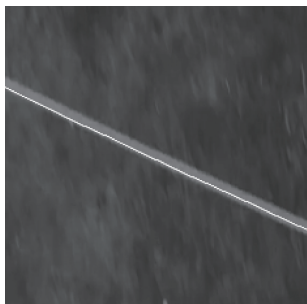


图3 不准确的标签和图像

Fig. 3 Inaccurate label and image pair

3 实验

实验主要由以下4部分组成:1)对数据集、实验设置进行说明,并将本文基础模型与两个相似工

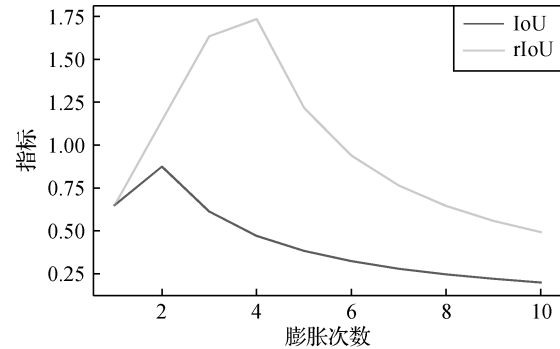


图4 指标随膨胀次数的变化情况

Fig. 4 Variation of indicators with the number of expansions

作进行对比,同时验证数据集增强方法的有效性;2)进行基础模型的训练,并寻找色彩空间和损失函数的最优组合;3)对 RGB 输入和灰度输入的完整网络进行压缩,实现模型轻量化;4)对自适应阈值法在不同亮度下的效果进行验证。实验部分模型均在 Colab 平台上完成,使用 Tensorflow 1.15.2 静态图进行网络搭建及测试工作。

3.1 数据集及实验设置

实验部分有两个背景数据集,一个由原始的 2 000 幅不包含电线的真实可见光图像组成,标记为 2 k,另一个在此基础上额外增加了由 PPGAN 生成的 8 000 幅伪背景图像,标记为 10 k。采用圆锥曲线在两个背景集上随机生成曲线电力线构成两个曲线训练集。另外,在 2 k 背景集上额外生成了一组直线电力线训练集用于与曲线集进行对比。验证集为从 200 组有标签的样本中随机选择的 50 组,余下的 150 组则作为测试集。训练集图像大小为 128×128 像素,而验证与测试集的图像大小均为 512×512 像素,因此测试时先分割为 4 块 128×128 像素大小进行输入,得到输出后将其拼接为 512×512 像素大小。同时在训练时均使用了水平、垂直翻转以及色相扰动进一步进行数据增强。

实验部分 rIoU 指标中的膨胀核长宽为 5,如无特别说明,进行语义分割时的判决门限均为 0.5,训练批大小均为 64,数据集迭代次数为 600 次。

首先在 RGB 空间的 2 k 曲线数据集上训练 Madaan 等人(2017)、Zhang 等人(2019)和本文基础模型,其结果如表1所示。在此处,训练时均采用和原文一致的网络架构、损失函数和优化器设置。为方便对比,以本文的基础模型作为计算量和参数量的对比标准。判决阈值的选择上,本文基础模型的

表 1 不同模型的收敛精度

Table 1 Convergence accuracy of different models			
模型	计算量/%	参数量/%	IoU
Madaan 等人(2017)	85.7	0.72	0.497
Zhang 等人(2019)	478	190	0.555
本文基础模型	100	100	0.556

阈值选为 0.5,而其他模型则为最优阈值。

表 1 中,Madaan 等人(2017)定义了一系列编解码器,在特定数据集下采用穷举的方式寻找最优的编解码器组合。此方法虽然在计算量与参数量上均有优势,但因不同数据集的数据分布可能存在较大差异,故在本数据集上性能指标不及本文基础模型。Zhang 等人(2019)方法采用 VGG16 作为主干网络,此类网络在各类任务中使用较为广泛,因此在不同数据集上具有较好的泛化性能,但此类主干网络的计算量较大是其主要缺陷,这使得对其进行压缩也相对耗时,因此认为本文基础模型是一个较好的轻量化起点。

为验证不同数据增强方法的有效性,使用 focal 损失函数和 Adam 优化器在不同数据集下进行测试。同时,为了评估生成样本与真实样本的差距,从 8 000 组生成的样本中随机抽取 200 组构成生成集,并用该集对 2 k 数据集训练的模型进行测试。

表 2 为不同数据集的训练结果。刘嘉玮等人(2020)自行对数据集中所有电力线图像均进行了标注并作为测试集,表 2 中最后一行为其论文结果。

根据综合指标,无论是在 RGB 还是灰度空间,曲线数据集的 IoU 与直线数据集相似,但 rIoU 明显高于后者,即前者在选取电力线周围像素的同时,也增加了标签范围内电力线像素的选取,具有更好的分割效果。RGB 色彩空间的 rIoU 值均高于灰度空间,而 IoU 值则正好与此相反,这说明 RGB 色彩空间中语义分割网络的结果倾向于选择更多的电力线周围像素,而在灰度空间中的语义分割精度则更高。对于电力线分割任务而言,在 IoU 值相似的情况下,选取较多的电线周围像素有助于增加动态场景分割的稳定性。使用生成对抗网络进行数据集增强后,RGB 空间的 rIoU 值降低,IoU 值提升,但对灰度空间影响较小。

表 2 不同数据集的收敛精度

Table 2 Convergence accuracy on different datasets						
训练集 色彩 训练集 空间 线形			rIoU (IoU)			
			验证集	测试集	综合	生成集
2 k	RGB	曲线	1.508 (0.543)	1.503 (0.561)	1.504 (0.556)	(0.639)
2 k	RGB	直线	1.566 [*] (0.565 [*])	1.410 [*] (0.546 [*])	1.449 [*] (0.551 [*])	(0.640 [*])
2 k	灰度	曲线	1.440 (0.584)	1.416 (0.580)	1.422 (0.581)	(0.915)
2 k	灰度	直线	1.356 (0.566)	1.323 (0.573)	1.331 (0.571)	(0.915)
10 k	RGB	曲线	1.498 (0.596)	1.302 (0.560)	1.351 (0.569)	—
10 k	灰度	曲线	1.449 (0.598)	1.326 (0.572)	1.357 (0.578)	—
刘嘉玮等人(2020)			(0.6)	(0.6)	(0.6)	—

注:“综合”指验证集与测试集的综合指标,加粗字体为每列指标的最优值,“-”代表未进行实验,“*”标注数据可认为是该工作在本数据集上的复现结果。

生成集测试在不同色彩空间的识别精度具有明显差异的原因是:1)生成对抗网络训练过程中出现了局部的模式坍塌造成生成样本的纹理多样性降低,这使得灰度空间的生成集精度远超测试集,故该空间数据集增强后的收敛精度没有明显提升。2)生成模型仍保留有较好的颜色多样性,因此 RGB 空间的生成集测试结果与测试集更接近,故数据集增强后 IoU 指标有所提升。

图 5 为相同参数下不同数据集的训练曲线。“2 k”与“10 k”为样本容量;“Gray”和“RGB”分别代表使用灰度或可见光图像作为输入;标注“Str”的项表明对应数据集中仅生成了直线电力线。可见,曲线数据集收敛速度明显快于直线数据集。

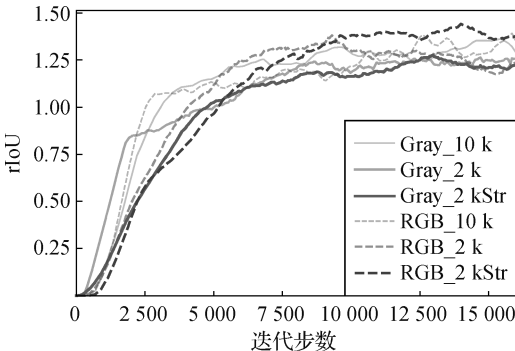


图 5 不同数据集的训练曲线

Fig. 5 Training loss in different datasets

图6(a)为表2中10 k模型在采用不同的输入色相位偏移和判决阈值时的指标值变化。可见,即使采用了色相扰动进行数据增强,对于不同的色相,其最佳判决阈值不断变化。而图6(b)则是一个未添加色相扰动的训练结果。可见,为提高模型对颜色(季节变化)的鲁棒性,在训练阶段进行色相扰动非常必要。考虑到在前文中灰度空间的收敛精度与RGB空间相似,采用灰度输入来丢弃颜色特征也是一个选择,同时也可获得更好的压缩比。

综上所述,曲线数据集的训练结果指标和收敛速度均优于直线数据集。生成对抗网络的增强结果有效提升了RGB空间收敛精度,而在灰度空间则差异不大。对生成集的验证证实生成图像较少增加灰度样本的多样性。而色相扰动则有效提升了模型对

场景色相变化的鲁棒性。

3.2 损失函数

在2 k曲线数据集上使用与上述相同的Adam优化器验证不同损失函数的收敛效果,结果如表3和图7所示。其中,图7中的sIoU为soft-IoU的缩写。

表3 不同损失函数的收敛精度

Table 3 Convergence accuracy of different loss functions

色彩空间	损失函数	rIoU(IoU)			
		验证集	测试集	综合	曲线集
RGB	focal	1.508 (0.543)	1.503 (0.561)	1.504 (0.556)	(0.639)
RGB	Dice	1.461 (0.568)	1.494 (0.576)	1.486 (0.574)	(0.567)
RGB	soft-IoU	1.570 (0.569)	1.527 (0.581)	1.538 (0.578)	(0.656)
灰度	focal	1.440 (0.584)	1.416 (0.580)	1.422 (0.581)	(0.915)
灰度	Dice	1.557 (0.601)	1.470 (0.581)	1.491 (0.586)	(0.912)
灰度	soft-IoU	1.414 (0.566)	1.411 (0.566)	1.412 (0.566)	(0.886)

注:“综合”指验证集与测试集的综合指标,加粗字体为每列指标的最优值。

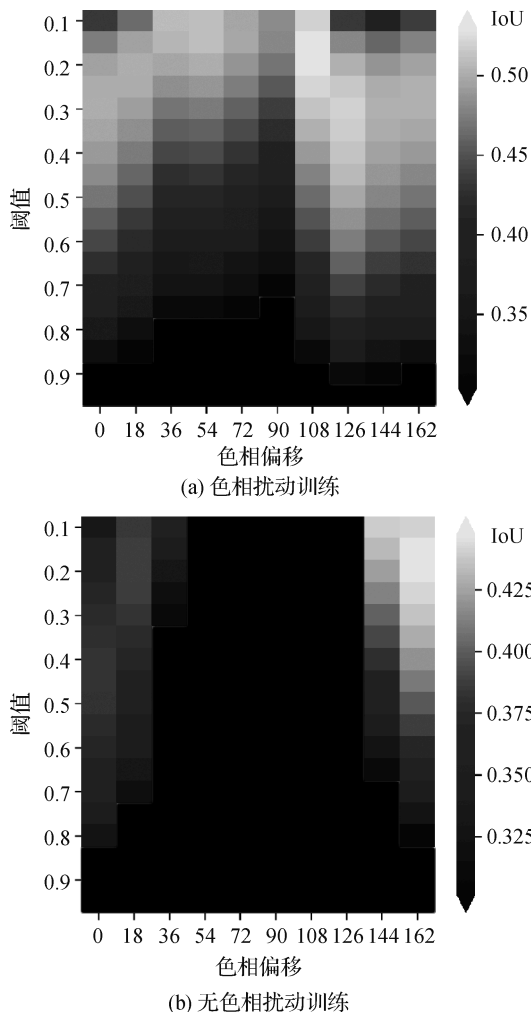


图6 RGB空间下色相扰动对结果的影响

Fig.6 The influence of hue shift on accuracy in RGB space

((a) training result with hue disturbance;
(b) training result without hue disturbance)

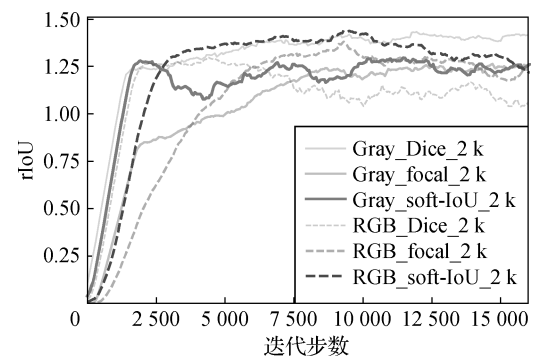


图7 不同损失函数的训练曲线

Fig.7 Training loss using different loss functions

表3结果显示:1)从收敛精度上看,soft-IoU和Dice分别在RGB和灰度空间具有最佳的收敛精度。同时由图7所示,二者在对应的色彩空间中均达到最高值,且曲线稳定高于其他方法;2)从收敛速度上看,在两个色彩空间中,Dice均最快收敛,soft-IoU次之但差异不大,focal收敛最慢且差异明显;3)从过拟合角度看,在RGB色彩空间,三者达到最优

值之后均出现了不同程度的过拟合,其中 Dice 最为严重,而在灰度空间则并不明显。

综上所述,RGB 空间采用 soft-IoU 损失函数和灰度空间中采用 Dice 损失函数是两个最佳组合。

3.3 模型轻量化

在此部分均以 RGB 空间的基础模型为 100% 基线计算参数量和计算量。其模型参数数量为 7.78 M,批大小为 60 时的计算量为 128 G。

3.3.1 RGB 空间

对 RGB 空间下 focal 损失函数的完整模型进行压缩。为了避免网络在正则化过程中出现较大的性能变化,在网络压缩部分均采用动量梯度下降法。正则化算法中的 $r^{(0)}$, $r^{(N)}$ 分别为 1 和 0.1, σ 和 k 分别为 1.4 和 10。在正则化过程中,正则半径随着正则次数增加以比例 $1/\sqrt{\sigma}$ 进行收缩。均采用 2 范数作为显著度指标。

图 8 描述了各阶段的通道显著度曲线。在 1~7 阶段时,网络的显著度指标均减小,但分布相对均匀且基本不具备区分度。从阶段 8 开始,低显著度的通道值在正则化效果下逐渐减小至 0,而高显著度的通道仍维持较高的值。因模型的显著度曲线基本为阶梯状,在下一步中简单使用 0.1 和 0.01 两个全局阈值对整个模型进行修剪,然后在当前参数值的基础上对轻量化模型进行重训练。

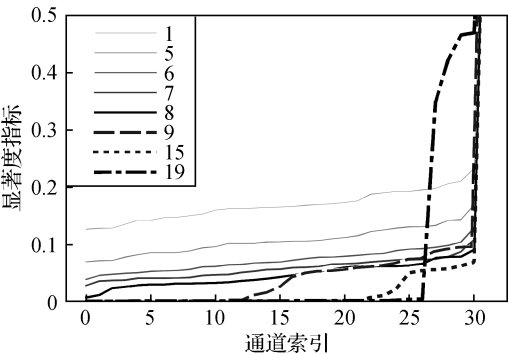


图 8 各正则化阶段的特征通道显著度曲线
Fig. 8 The saliency curves in each stage of sparse regularization

表 4 展示了上述稀疏模型在不同损失函数、修剪阈值下的重训练结果。以 soft-IoU 为损失函数的重训练结果仍然优于 focal 的结果,即使完整模型是由 focal 损失函数训练得出。然而,focal 损失函数的两个模型在训练集上的 IoU 值均能达到 0.92 以上,但 soft-IoU 损失函数的两个 2 k 训练集模型在训练

表 4 RGB 空间下的轻量化模型精度
Table 4 Accuracy of lightweight model in color space

数据集	损失函数	修剪阈值	计算量 /%	参数量 /%	rIoU (IoU)	
					固定阈值	Otsu 阈值
10 k	-	0	100	100	1.351 (0.569)	1.375 (0.569)
2 k	focal	0.1	3.6	0.17	0.917 (0.402)	0.912 (0.395)
2 k	focal	0.01	9.62	0.23	0.780 (0.297)	0.959 (0.400)
2 k	soft-IoU	0.1	3.6	0.17	1.162 (0.433)	1.306 (0.444)
2 k	soft-IoU	0.01	9.62	0.23	0.964 (0.354)	1.006 (0.379)
10 k	soft-IoU	0.01	9.62	0.23	0.983 (0.447)	1.041 (0.484)

注:损失函数指在重训练时的损失函数,加粗字体为轻量化模型中每列相同指标的最优值。

集上的 IoU 值均不足 0.8,若继续训练,则会出现严重的过拟合情况,此现象与图 7 中过拟合的现象一致。而 soft-IoU 的 10 k 训练集模型在训练集精度达到 0.88 时仍未出现严重过拟合,这使得模型能继续训练并最终使 0.01 模型的 IoU 指标达到了表内最高值。对 Otsu 法的测试显示该方法所获得的阈值较为可靠,能极大提升部分模型的性能。主要原因是某些情况下部分颜色和电力线相近的物体在概率图上也会产生很高的响应,而此现象通常发生在电力线本身具有较强响应的情况下。此类错误可通过采用自适应阈值来缓解。在随后的亮度敏感性分析中将对 Otsu 法的效果进行详细分析。

3.3.2 灰度空间

对表 3 中灰度空间内 Dice 损失函数的收敛模型进行压缩,其结果如表 5 所示。在此部分中为实现更加极限的压缩性能,除 2 范数外,还采用了本文提出的联合指标进行剪枝,其结果见最后一行。

对比表 4 和表 5 可知,0.05 模型在 Otsu 阈值下的性能指标超越了彩色空间内最好的结果(表 4 最后一行),同时计算量与参数量低于后者,而 2 范数的 0.1 模型在 Otsu 阈值下的指标超越了彩色空间中的所有 0.1 模型,联合指标的结果则在计算量上具有明显优势。

表 5 灰度空间下的轻量化模型精度

Table 5 Accuracy of lightweight model in gray space						
数据集	损失函数	修剪 阈值	计算量 /%	参数量 /%	rIoU (IoU)	
					固定 阈值	Otsu 阈值
2 k	-	0	100	100	1.491 (0.586)	1.494 (0.587)
2 k	soft-IoU	0.05	3.97	0.04	1.137 (0.509)	1.174 (0.519)
2 k	soft-IoU	0.1	3.05	0.03	0.913 (0.409)	1.404 (0.459)
2 k	soft-IoU	0.1	0.947*	0.015*	1.027* (0.419*)	1.029* (0.420*)

注: * 标注数据为联合指标的剪枝结果,加粗字体为轻量化模型中每列相同指标的最优值。

3.4 亮度敏感性

数据集中图像的亮度分布较为相似,而实际情况中的可见光图像易出现过曝或欠曝的现象,这可能造成整体概率分布偏移,进而影响固定阈值的效果。选取 RGB 空间完整模型、RGB 空间和灰度空间的最佳轻量化模型进行对比,通过 γ 变换的方式改变图像亮度以模拟不同曝光条件对阈值选择的影响。实验中, γ 取值从 0.1 ~ 10(图像亮度从变亮过渡到变暗, $\gamma = 1$ 时不改变亮度),评价指标采用 IoU,据此绘出不同曝光条件下选取不同判决阈值的模型性能,如图 9(a)所示,图例为 IoU 值与颜色的对应关系。为进一步对比不同模型和阈值选择方式的性能差异,在部分子图中将两个 IoU 指标分布图做差,如图 9(b)所示,图例为指标差值与颜色的对应关系。但若该位置上两个 IoU 指标均小于 0.3,则无论指标差距多大均不纳入绘图范围,因为过低的 IoU 值模型本身已不可使用,如图 9(b)中 γ 小于 0.4 的部分。

图 9 是 RGB 色彩空间下的完整模型,由图 9(a)可知:1)完整模型具有较强的分割能力,输出概率分布具有较好的区分度;2)模型对低亮度具有更强鲁棒性,这是由于数据集中电力线亮度明显高于背景,降低亮度可增加电力线与背景的对比度。由图 9(b)可知,Otsu 阈值法较固定阈值法在绝大部分区域差异在 0.02 内。

图 10 是两个色彩空间下轻量化模型的测试结果,对比图 10(a)(b)可知,Otsu 法在低亮度情况下相比最佳阈值性能有略微下降,而在较为广泛的高

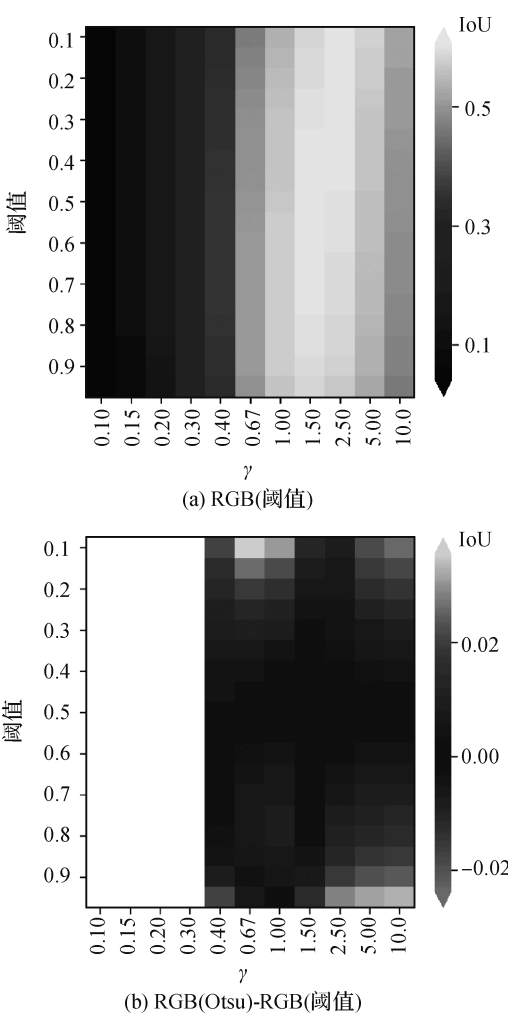


图 9 完整模型不同阈值方法下的亮度敏感度对比

Fig.9 Comparison of brightness sensitivity of the full model under different threshold methods
((a) RGB (threshold); (b) RGB(Otsu)-RGB(threshold))

亮度区间 Otsu 法性能好于阈值法。由图 10(c)(d)可知,在灰度模型中,Otsu 法在各种亮度条件下的性能与固定阈值法差异基本在 0.05 以内。由图 10(a)(c)可知,在过低亮度时($\gamma \geq 10$),RGB 模型的阈值选择相比灰度模型更加宽松,即前者的输出概率图在分布上具有更大的区分度。灰度模型仅依据灰度信息进行语义分割,相比 RGB 模型更易受亮度变化干扰。对比图 9(a)和图 10(a)可知,图像亮度越高,轻量化模型性能对阈值越敏感,而此现象在完整模型上则不明显。

在灰度空间中,Otsu 法在轻量化模型下的效果与固定阈值法相似,故将在灰度空间 Otsu 法的结果与 RGB 空间中两种方法的结果进行对比,对比结果如图 11 所示。

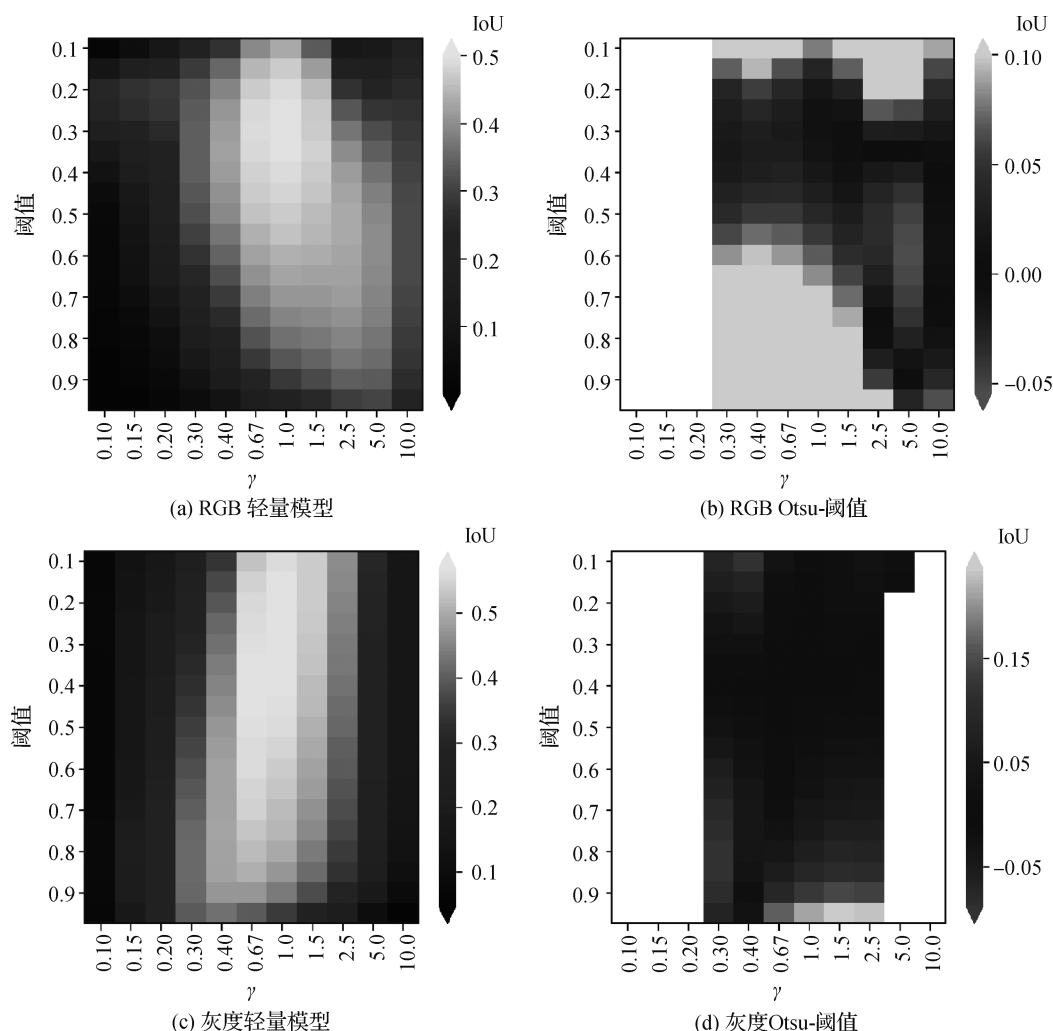


图 10 轻量模型的亮度敏感度分析

Fig. 10 Luminance sensitivity analysis of lightweight model

((a) RGB lightweight model; (b) RGB Otsu-threshold; (c) gray lightweight model; (d) gray Otsu-threshold)

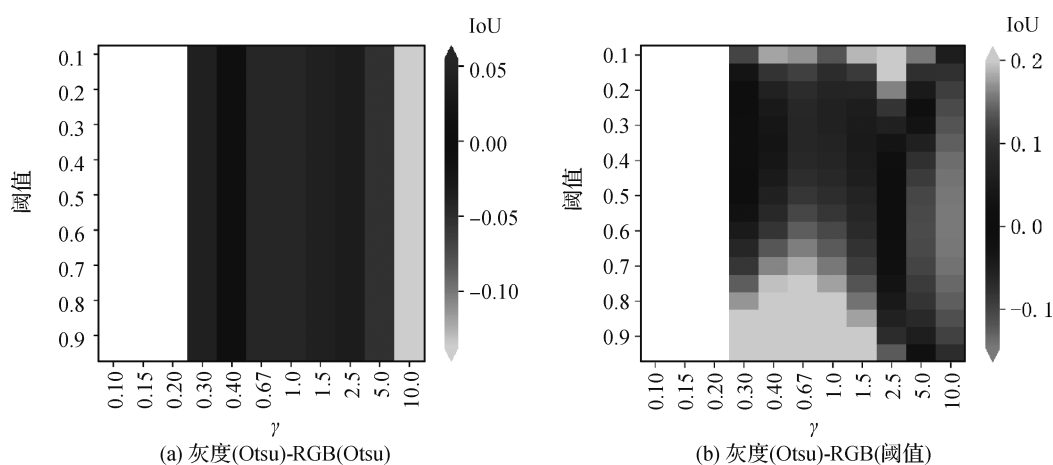


图 11 不同色彩空间的轻量化模型的亮度敏感度对比

Fig. 11 Comparison of brightness sensitivity of lightweight models in different color spaces

((a) gray(Otsu)-RGB(Otsu); (b) gray(Otsu)-RGB(threshold))

图11结果显示,在采用Otsu阈值法且亮度变化较小(γ 取0.4~2.5)时,灰度模型较RGB模型有略微优势,而其他区域则采用RGB模型较好。若RGB模型采用阈值法,RGB模型在右侧 $\gamma > 2.5$ 的区域有性能优势。考虑到实际情况,可以很容易地通过减小光圈或者降低感光度来降低图片亮度,同时电力线巡检时的光照条件较为充足,实际上难以达到 $\gamma = 2.5$ 的低光照情形。因此认为灰度空间的轻量化模型可以用于替代RGB空间的轻量化模型。

4 结 论

首先对现有基于U-Net的航空电力线图像语义分割网络进行了改进研究,包括数据集增强方法、色彩空间和损失函数的选择,并得出了两套最佳组合。然后采用传统通道范数和提出的联合显著度指标对不同色彩空间的网络模型进行了稀疏正则化,并采用网络剪枝大幅降低了网络参量和运算量。最后对固定阈值法和自适应阈值法在不同光照条件下的性能表现进行了研究,证实了在一定范围的光照变化下,自适应阈值法能取得在该条件下最优阈值相似的性能。

当前研究工作仍有部分不足有待改进。首先,采用生成对抗网络进行数据集增强在灰度空间没有出现性能提升,可能与生成对抗网络的模式坍塌以及测试样本过少有关,因此该数据集增强方法在完整模型训练时的作用还需进一步验证。其次,在此工作中使用的验证与测试集样本量不足,为使网络性能更好,需要进行更多的标注工作。同时,从数据集集中的红外图像来看,红外图像的电力线特征相比可见光更为明显,且几乎不存在颜色干扰,将可见光与红外图像共同输入模型可成为进一步提升模型性能的一个研究方向。最后,网络剪枝能有效提升模型在通用终端上的推理速度,但推理加速是一个从底层硬件设计到软件上模型结构优化的全栈任务,若能根据终端特性,如传输带宽、管脚数目和计算单元数量等进行更具针对性的剪枝操作,则可在终端执行速度相同的情况下执行更多的计算,有效提升轻量网络的效果。

参考文献 (References)

Arjovsky M and Bottou L. 2017. Towards principled methods for training

- generative adversarial networks//Proceedings of the 5th International Conference on Learning Representations. Toulon, France; [s. n.]
- Arjovsky M, Chintala S and Bottou L. 2017. Wasserstein GAN[EB/OL]. [2020-10-27]. <https://arxiv.org/pdf/1701.07875.pdf>
- 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]
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial nets//Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2. Montreal, Canada; MIT Press: 2672-2680 [DOI: 10.5555/2969033.2969125]
- Han S, Mao H and Dally W J. 2016. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding//Proceedings of the 4th International Conference on Learning Representations Conference Track Proceedings. San Juan, Puerto Rico; [s. n.]
- Karras T, Aila T, Laine S and Lehtinen J. 2018. Progressive growing of GANs for improved quality, stability, and variation//Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada; [s. n.]
- Le Cun Y, Denker J S and Solla S A. 1989. Optimal brain damage//Proceedings of the 2nd International Conference on Neural Information Processing Systems. Denver, USA; MIT Press: 598-605 [DOI: 10.5555/2969830.2969903]
- Li B L, Wu B W, Su J and Wang G R. 2020. Eagleeye: fast sub-net evaluation for efficient neural network pruning//Proceedings of 2020 European Conference on Computer Vision (ECCV). Glasgow, UK; Springer: 639-654 [DOI: 10.1007/978-3-030-58536-5_38]
- 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(5): 956-966 (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(5): 956-966) [DOI: 10.11834/jig.190316]
- Liu Z, Sun M J, Zhou T H, Huang G and Darrell T. 2019. Rethinking the value of network pruning//Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA; [s. n.]
- 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, Canada; IEEE: 3487-3494 [DOI: 10.1109/IROS.2017.8206190]
- Molchanov P, Tyree S, Karras T, Aila T and Kautz J. 2017. Pruning convolutional neural networks for resource efficient inference//Proceedings of the 5th International Conference on Learning Representations. Toulon, France; [s. n.]

- Molchanov P, Mallya A, Tyree S, Frosio I and Kautz J. 2019. Importance estimation for neural network pruning//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 11256-11264 [DOI: 10.1109/CVPR.2019.01152.]
- Paszke A, Chaurasia A, Kim S and Culurciello E. 2017. ENET: a deep neural network architecture for real-time semantic segmentation [EB/OL]. [2021-01-06]. <https://arxiv.org/pdf/1606.02147.pdf>
- 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]
- 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]
- Wang X W. 2019. Research on Semantic Segmentation of Power Line Based on Image. Hangzhou: Zhejiang University (王翔文. 2019. 基于图像的输电线路语义分割技术研究. 杭州: 浙江大学)
- Yetgin Ö E and Gerek Ö N. 2019a. Powerline Image Dataset (Infrared-IR and Visible Light-VL) [DB/OL]. [2020-10-18]. <https://data.mendeley.com/datasets/n6wrv4ry6v/8> [DOI: 10.17632/n6wrv4ry6v.8]
- Yetgin Ö E and Gerek Ö N. 2019b. Ground Truth of Powerline Dataset (Infrared-IR and Visible Light-VL) [DB/OL]. [2020-10-18]. <https://data.mendeley.com/datasets/twpx8xccsw/9> [DOI: 10.17632/twpx8xccsw.9]
- Yu C Q, Wang J B, Peng C, Gao C X, Yu G and Sang N. 2018. BiSeNet: bilateral segmentation network for real-time semantic segmentation//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 334-349 [DOI: 10.1007/978-3-030-01261-8_20]
- Zhang H, Yang W, Yu H, Zhang H J and Xia G S. 2019. Detecting power lines in UAV images with convolutional features and structured constraints. *Remote Sens*, 11(11): #1342 [DOI: 10.3390/RS11111342]
- Zhang J J, Liu L, Wang B H, Chen X G, Wang Q and Zheng T R. 2012. High speed automatic power line detection and tracking for a UAV-based inspection//Proceedings of 2012 International Conference on Industrial Control and Electronics Engineering. Xi'an, China: IEEE: 266-269 [DOI: 10.1109/ICICEE.2012.77]
- Zhao H S, Qi X J, Shen X Y, Shi J P and Jia J Y. 2018. ICNet for real-time semantic segmentation on high-resolution images//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 418-434 [DOI: 10.1007/978-3-030-01219-9_25]
- Zhao L, Wang X P, Yao H T and Tian M. 2021. Survey of power line extraction methods based on visible light aerial image. *Power System Technology*, 45(4): 1536-1546 (赵乐, 王先培, 姚鸿泰, 田猛. 2021. 基于可见光航拍图像的电力线提取算法综述. *电网技术*, 45(4): 1536-1546) [DOI: 10.13335/j.1000-3673.pst.2020.0300a]

作者简介



许刚, 1963年生, 男, 教授, 主要研究方向为智能信息处理。

E-mail: xugang@ncepu.edu.cn



李果, 通信作者, 男, 硕士研究生, 主要研究方向为模型压缩。

E-mail: liguo@ncepu.edu.cn