JOURNAL OF IMAGE AND GRAPHICS

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

2022
04
VOL.27

ISSN1006-8961
CN11-3758/TB

口罩人脸姿态分类 P1125

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第27卷第4期（总第312期）
2022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

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

2022年4月 第27卷 第4期（总第312期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次

残差密集结构的东巴画渐进式重建(第1084页)

结合双模板融合与孪生网络的鲁棒视觉目标跟踪(第1191页)

注意力机制的曲面沉浸式投影系统补偿(第1238页)

Volume 27, Number 4
Published April 16, 2022

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS

Gradual model reconstruction of Dongba painting based on residual dense structure(P1084)

Double template fusion based siamese network for robust visual object tracking(P1191)

An attention mechanism based inter-reflection compensation network for immersive projection system(P1238)

E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2022)04-1204-12

论文引用格式:Xiao Z J, Zong J X, Lan H, Wei X and Tang X L. 2022. Image semantic segmentation based on manifold regularization constraint. Journal of Image and Graphics,27(04):1204-1215(肖振久,宗佳旭,兰海,魏宪,唐晓亮. 2022. 流形正则化约束的图像语义分割. 中国图象图形学报,27(04):1204-1215)[DOI:10.11834/jig.200527]

流形正则化约束的图像语义分割

肖振久[1],宗佳旭[1*],兰海[2],魏宪[2],唐晓亮[2]

1. 辽宁工程技术大学软件学院,葫芦岛　125105;2. 泉州装备制造研究所,泉州　362000

摘　要:**目的**　在基于深度学习的图像语义分割方法中,损失函数通常只考虑单个像素点的预测值与真实值之间的交叉熵并对其进行简单求和,而引入图像像素间的上下文信息能够有效提高图像的语义分割的精度,但目前引入上下文信息的方法如注意力机制、条件随机场等算法需要高昂的计算成本和空间成本,不能广泛使用。针对这一问题,提出一种流形正则化约束的图像语义分割算法。**方法**　以经过数据集 ImageNet 预训练的残差网络(residual network, ResNet)为基础,采用 DeepLabV3 作为骨架网络,通过骨架网络获得预测分割图像。进行子图像块的划分,将原始图像和分割图像分为若干大小相同的图像块。通过原始图像和分割图像的子图像块,计算输入数据与预测结果所处流形曲面上的潜在几何约束关系。利用流形约束的结果优化分割网络中的参数。**结果**　通过加入流形正则化约束,捕获图像中上下文信息,降低了网络前向计算过程中造成的本征结构的损失,提高了算法精度。为验证所提方法的有效性,实验在 Cityscapes 和 PASCAL VOC 2012(pattern analysis, statistical modeling and computational learning visual object classes)两个数据集上进行。在 Cityscapes 数据集中,精度值为 78.0%,相比原始网络提高了 0.5%;在 PASCAL VOC 2012 数据集中,精度值为 69.5%,相比原始网络提高了 2.1%。同时,在 Cityscapes 数据集中进行对比实验,验证了算法的有效性,对比实验结果证明提出的算法改善了语义分割的效果。**结论**　本文提出的语义分割算法在不提高推理网络计算复杂度的前提下,取得了较好的分割精度,具有极大的实用价值。

关键词:深度学习;语义分割;残差网络(ResNet);上下文信息捕捉;流形正则化

Image semantic segmentation based on manifold regularization constraint

Xiao Zhenjiu[1], Zong Jiaxu[1*], Lan Hai[2], Wei Xian[2], Tang Xiaoliang[2]

1. College of Software, Liaoning Technology University, Huludao 125105, China;

2. Quanzhou Institute of Equipment Manufacturing, Quanzhou 362000, China

Abstract: **Objective** Image semantic segmentation is one of the essential issues in computer vision and image processing. It aims to divide pixels in the image into different categories semantically, and to foresee pixel-level predictions. It has been widely used in various fields, such as scene information understanding, automatic driving and medical assisting diagnosis.

收稿日期:2020-09-08;**修回日期**:2020-12-08;**预印本日期**:2020-12-15

*通信作者:宗佳旭　jiaxu0017@163.com

基金项目:国家自然科学基金项目(61806186);福建省智能物流产业技术研究院建设项目(2018H2001);机器人与系统国家重点实验室项目(SKLRS-2019-KF-15);泉州市科技计划项目(2019C112)

Supported by: National Natural Science Foundation of China (61806186); the Program of "Fujian Intelligent Logistics Industry Technology Research Institute" (2018H2001); State Key Laboratory of Robotics and System (HIT) (SKLRS-2019-KF-15); the Program of Quanzhou Science and Technology Plan (2019C112)

Competitive performance has still suffered from challenges such as low contrast, uneven luminance and complicated scenarios currently. The performance of semantic segmentation algorithms have mainly constrained by the spatial context information. Current methods based on deep learning algorithms for image semantic segmentation has focused on harnessing the context information between pixels. For instance, the attention mechanism builds an element-wise weight matrix to capture the similarity between pixels which can be used as coefficient to summate the input. Meanwhile, probabilistic graphical models have been utilized in the spatial context as prior to enhance the classification confidence. However, these methodologies require massive computational resource (e. g. GPU memory). A contextual information capturing method is demonstrated based on manifold regularization. By assuming the data in the input image and the segmentation prediction share the same locally geometric structure in the low-dimensional manifold, this research illustrated possibility to harness the relevancy among pixels in more efficient way. As a result, the novel algorithm based on manifold regularization is issued to exploit the spatial context relation from a geometric perspective, which can be embedded into the deep learning framework to improve the performance with no increasing on both parameter amount and reasoning time. **Method** The contextual information analysis in the image can be effectively captured by manifold regularization. The DeepLab-v3 architecture is extracted the image features, which uses the residual network (ResNet) as the backbone network. The last two down-sampling layers of the model are pruned, and dilated convolution is employed in the subsequent convolutional layer to control the resolution of the features. For the methodology of regular segmentation, the cross-entropy of single pixel between prediction and ground truth is only involved in the cost function and sum up in total loss without any context information simply. A detailed manifold regularization penalty designation is integrated to single pixel information and the neighborhood context information. This geometric intuition for the initial image data has the same locally geometric shape with those in the segmented result. It indicates that the correspondences between clusters of data points in the input image and output result data points. For instance, when the distance of two input data points in the manifold sub-space is close, the corresponding segmentation result data points are close, and vice versa. Furthermore, the image into sub-image patches to capture the relationship between to customize the constraints between pixels. The hierarchical manifold regularization constraints are achieved via sub-image patch divides into different sizes. When the patch size is minimized, the constraint is between pixels substantially and the approach acts like other pixel-wise context aware algorithms such as fully connected conditional random field (CRF) model. On the contrary, the maximum patch size which equals to the input image size makes the approach become semi-supervised learning algorithm based on interconnected samples. The analyzed model gets improved on segmentation accuracy and achieves state-of-the-art performance. This model is based on two public datasets, Cityscapes and PASCAL VOC 2012 (pattern analysis, statistical modeling and computational learning visual object classes 2012). The performance is measured via mean intersection-over-union (mIoU) averaged across all the classes. The open source toolbox Pytorch is used to build the model. The stochastic gradient descent (SGD) method is adopted as the optimization. In addition, data augmentation is conducted by means of random cropping and inversion in accordance with probability levels. The operating system of the experimental platform is Centos7, with a GPU of model NVIDIA RTX 2080Ti and a CPU of Intel (R) Core (TM) i7-6850. **Result** The tests are conducted with the effect of manifold regularization. The algorithm achieves a good accuracy of the segmentation model without increasing computational complexity in the process of model implementation. On the benchmark, the ResNet50 backbone model improves the performance by 0. 8% with manifold regularization adopted on the PASCAL VOC 2012 dataset, while the ResNet101 backbone models bring 2. 1% mIoU gain. These results demonstrated that the manifold regularization get qualified performance with larger network model, and the analyszed results on the Cityscapes dataset also prove this inference, the ResNet50 model increases by 0. 3% while the ResNet101 model increases by 0. 5%. With the comparison of other context aggregation methods, we achieve mIoU of 78. 0% on the Cityscapes dataset and 69. 5% on the PASCAL VOC 2012 dataset. Furthermore, visualization of the segmentation results is implemented. The generated segmentation results are more accurate at the edges and have less error rate based on the algorithm with manifold regularization constraints. **Conclusion** This demonstration illustrates a novel algorithm for the context information image semantic segmentation via the manifold regularization constraints, which can be melted into the deep learning network model to improve the segmentation performance without changing the network structure. The results verify that the illustrated algorithm has good generalization capability in semantic segmentation.

Key words: deep learning; semantic segmentation; residual network (ResNet); capture of contextual information; manifold regularization

0 引 言

图像的语义分割是机器视觉领域一项必不可少的核心任务。语义分割可以广泛用于场景信息理解、自动驾驶和医疗辅助诊断等领域,且具有重要作用。目前,尽管对图像的语义分割展开了积极研究,但仍需重点解决精准区分不同尺寸的相同物体和克服遮挡、光照等问题(青晨 等,2020)。为了解决上述问题,有效完成不同场景的分割任务,有必要增强像素级别的识别能力,特别是全卷积神经网络(fully convolutional networks,FCN)(Long 等,2015)。但是受限于卷积操作的局域性,分割结果仅受限于局部的感受野中,无法融合长距离的上下文信息,从而限制了 FCN 类方法的发展。另外,在图像数据下采样过程中,会造成细节本征信息损失,尤其在 U 型网络结构中(韩慧慧 等,2020)。

为利用像素间的上下文信息,利用条件随机场(Zheng 等,2015)或注意力机制(Zhao 等,2018)使任意位置的单个特征可以感知其他位置的所有特征,从而使输出结果能够融合各像素点之间的关系。但是在条件随机场以及注意力机制中,仅利用单个像素点间的相似度建立势函数(Krähenbühl 和 Koltun,2012)或权重矩阵,难以在低对比度的图像中取得较好效果,另外需要生成巨大的注意力图来计算每个像素之间的关系,计算复杂度高且占用显存资源,从而限制了其在图像语义分割中的表现。基于以上问题,本文提出一种流形正则化约束的图像语义分割算法,通过将图像分割中的输入数据和输出结果视为两个不同流形并维持这两个流形之间对应关系来获取像素间的本征结构,有效利用了上下文的信息并提升了算法的分割精度。本文的主要贡献有:1)提出一种基于几何优化的流形正则化方法,通过将高维空间中相邻的输入数据点与输出结果维持在同样的流形结构上,引入了像素点间的上下文关系,提高了分割精度。2)与现有主流的图像分割算法相结合,本文提出的流形正则化算法能够很好地嵌入各类分割算法,并且提升分割精度,在多个数据集上的性能处于领先位置。本文算法代码已上传至 Github,共享网址为 https://github.com/jiaxu0017/Manifold_Segmentation。

1 相关工作

1.1 语义分割

近几年,针对语义分割的研究取得了新的进展,吸引了大量研究人员的关注。FCN 对语义分割进行了开拓性尝试,基于 FCN 思想的语义分割技术得到长足进展。然而开始的 FCN 由于多层的池化操作导致提取的特征信息在图像细节处丢失严重,从而影响了算法在细节处的表现。为解决这一问题,Chen 等人(2016)、Yu 和 Koltun(2016)通过较少的降采样操作以获得更精细的特征并利用膨胀卷积来增强感受野,DeepLabV2(Chen 等,2018)提出 ASPP(atrous spatial pyramid pooling)模块,利用不同的膨胀卷积捕获图像上下文中的信息,DeepLabV3(Chen 等,2017)设计则采用级联或并行的空洞卷积,通过不同速率的空洞卷积获取上下文信息,PSPnet(pyramid scene parsing network)(Zhao 等,2017)利用金字塔模型聚合上下文信息。此外,通过优化编码器—解码器结构,如 U-Net(Ronneberger 等,2015)、RefineNet(Lin 等,2017)和 SegNet(Badrinarayanan 等,2017)等,将信息融合在低层和高层结构中,获得不同尺寸的特征信息,预测分割后的图像,亦能够提升分割精度。另外,在分割结果上进行后处理,利用预测结果自身的上下文信息来优化分类结果也是一种提升精度的有效方法。常见的后处理方法包括利用条件随机场(conditional random field,CRF)(Chandra 等,2017)和马尔可夫随机场(Markov random field,MRF)(Liu 等,2015)等建立各像素间的图模型,进一步精确定位像素边界和分割结果。

1.2 上下文信息捕捉

利用上下文信息来增强图像语义分割在细节上的表现是目前的热门研究方向。自 FCN 类方法提出之后,研究人员也聚焦于上下文信息的捕捉上,利用概率图模型中的条件随机场和马尔可夫随机场捕获预测结果中的上下文依赖关系。和超等人

(2020)通过不同的膨胀卷积和池化等操作生成的特征图来聚合多尺度上下文信息,实现多尺度的上下文信息融合。Byeon 等人(2015)采用递归神经网络,通过捕获标签上的空间信息或局部特征丰富上下文的依赖关系。尽管采用上下文融合的方式有助于获取不同比例的对象,但是无法利用全局视图中对象或事物之间的关系。而基于图模型的方法又难以与现有图像处理的卷积神经网络模型完美融合,实现端到端学习。另外,通过递归神经网络隐式捕获全局关系,其有效性很大程度取决于长期记忆的结果。上述两种方法并不能满足不同像素对不同上下文信息的要求。

获取上下文之间的关系,对远程依赖进行建模仅靠上述技术还远远不够,注意力模块的引入为建立信息间的远程依赖提供了新的思路,并且在许多应用中取得了成功。Vaswani 等人(2017)利用注意力机制建立 Transformer 模型并用其替代循环神经网络,通过绘制全局依存关系用于机器翻译,效果得到大幅提升。目前,注意力机制越来越多地应用于计算机视觉领域。Wang 等人(2018b)利用一种非局部模块,通过计算特征图中每个空间点之间的相关矩阵生成巨大的注意力图,然后引导密集的上下文信息进行聚合。DANet(dual attention network)(Fu 等,2019)通过结合像素间的上下文信息和通道空间内的上下文信息来提高分割精度。CCNet(criss-cross attention network)(Huang 等,2019)通过新颖的交叉注意力模块在交叉路径上收集周围像素的上下文信息,并通过循环操作,使每一个像素最终可以捕获所有像素的远程依赖关系。

1.3　流形正则化

正则化约束思想有着丰富的数学历史,可回溯至 Tikhonov 求解不适定逆问题,其作为样条理论的核心思想已广泛用于机器学习领域(Evgeniou 等,2000),许多机器学习算法,如支持向量机均可视为正则化的特例。而流形正则化则利用数据分布的几何结构对具体学习任务的损失函数进行约束。假定数据的相关子集来自某种拓扑流形:带有不同标签的数据在流形曲面上距离较远。基于该假设,流形正则化广泛应用于半监督学习(Belkin 等,2006)。

在神经网络中,由于连续的合并操作或卷积步幅导致特征的分辨率下降,使得神经网络学习到越来越抽象的特征表示,即本征结构。但是,这种局部图像转化为固定的特征表示,往往会造成关键信息的损失。为解决本征结构的损失,控制复杂的几何分布,研究人员针对流形约束做了大量工作,提出多种应用流形正则化的半监督学习框架。Belkin 等人(2016)利用边际分布的几何形状,提出一种基于流形的正则化学习框架,主要针对半监督学习能够有效地使用未标记数据。另外,针对半监督学习,Geng 等人(2009)提出内在流形自动近似的算法,并开发了一个集合流形正则化的框架,结合一些初始的猜测来近似本征流形。为了更深一步地解释流形在半监督学习中的作用,Niyogi(2013)通过建立 minmax 框架,调查了多种学习的方式,通过对流形在半监督学习中的潜在用途来解释流形正则化及相关的几何算法。除了建立流形正则化的半监督学习中的框架,研究人员也将流形学习应用于图像分割领域中。Quispe 和 Petitjean(2015)利用先验对象的几何形状指导分割,该算法依赖扩散图来编码训练集的形状变化,并用对象分割提供相应的帮助。Luo 和 Huang(2014)提出一种针对刚性物体和非刚性物体运动分割的自适应流形降噪技术,通过流形约束的方式降低分割中的偏差。基于这一思路,本文利用流形正则化约束对图像语义分割模型参数进行优化,通过维持图像分割中的输入数据和输出结果两个流形之间对应关系,获取数据间的本征结构,从而提升算法模型的性能。

2　算法原理及实现

2.1　损失函数

深度学习算法通常使用随机梯度下降作为任务求解工具,为能保证求解结果快速而准确地收敛,需要保证类别信息的数学表达(损失函数)能够涵括各类情况,通常的多分类问题使用交叉熵作为类别间的损失函数,其损失函数 L_{cla} 定义为

$$L_{\mathrm{cla}}(\boldsymbol{y},\hat{\boldsymbol{y}}) = -\sum_{k=1}^{C} y_k \cdot \log(\hat{y}_k) \tag{1}$$

式中,C 为总类别数,k 为当前分类,y 为分类的概率真实值,y_k 为第 k 类结果的概率真实值,$\hat{y}$ 为该分类概率的模型预测结果,$\hat{y}_k$ 为第 k 类概率的模型预测结果。

由于图像分割可以定义为像素级别的分类任务,因此,分割任务中损失函数即对所有像素点的交叉熵损失函数 L_{seg} 进行求和,即

$$L_{\text{seg}}(\boldsymbol{y},\hat{\boldsymbol{y}}) = -\sum_{i=1}^{N}\sum_{k=1}^{C} y_k^i \log(\hat{y}_k^i) \tag{2}$$

式中，N 为图像的像素点总数，i 为当前像素点，y_k^i 表示第 i 个像素点为第 k 类结果的概率真实值，$\hat{y}_k^i$ 表示第 i 个像素点为第 k 类结果的模型预测结果。

通过式(2)可以看出，这类损失函数的缺点在于仅计算单个像素点的预测结果与真实值之间的惩罚值，没有考虑邻近像素点分类结果的影响。从直观上看，在邻近像素点预测为某个分类的情况下，该像素点预测为其他分类的损失惩罚应该增大。这类思想是多数上下文信息捕捉方法的前提假设，为解决这一问题，本文通过在损失函数中引入流形正则约束项来实现相邻像素间上下文信息的捕捉。

2.2 流形正则化

在本文中，假定输入数据与其对应的预测结果在高维原始数据空间内的低维流形曲面上有着相同的几何结构。基于这一假设，利用数据的几何结构构建正则化约束项是本文的主要创新点。在式(2)中加入流形正则约束项，总体损失 L_{tot} 为

$$L_{\text{tot}}(\boldsymbol{y},\hat{\boldsymbol{y}}) = L_{\text{seg}} + \lambda_a \|f\|_K^2 \tag{3}$$

式中，λ_a 是流形正则项的权重参数，用来控制流形约束对整体损失函数的贡献。$\|f\|_K^2$ 则对应流形正则项，用于控制输入数据与输出结果在流形曲面局部空间上的几何结构一致性，从而引入了不同位置像素间分类结果的上下文依赖。

给定一组数据 $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^N$，$\boldsymbol{x}_i \in \mathbf{R}^m$，$\boldsymbol{Y} = \{\boldsymbol{y}_i\}_{i=1}^N$，$\boldsymbol{y}_i \in \mathbf{R}^m$，$\boldsymbol{Y}$ 为 $\boldsymbol{X}$ 的预期结果。用图 $G(\boldsymbol{X},\boldsymbol{\Omega})$ 表征数据点之间的几何结构关系。在此，$\boldsymbol{\Omega} = [\omega_{ij}]$ 是一个矩阵，ω_{ij} 表征了图上任意两个节点 x_i 和 x_j 之间的联系紧密程度，使用高斯热核的欧氏距离对矩阵中的元素 ω_{ij} 进行定义，具体为

$$\omega_{ij} = \begin{cases} \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2) & j \in \boldsymbol{N}_i \\ 0 & j \notin \boldsymbol{N}_i \end{cases} \tag{4}$$

式中，$\boldsymbol{N}_i$ 表示 $\boldsymbol{x}_i$ 的近邻数据点集合，j 为范围 $\boldsymbol{N}_i$ 内的任意一点。当 $\boldsymbol{x}_j$ 不处于 $\boldsymbol{x}_i$ 的邻域内时，ω_{ij} 为 0，即不考虑非近邻点之间的相互影响，仅考虑邻域内的流形结构。

通过 ω_{ij} 建立输入数据 $\boldsymbol{x}_i$ 之间的相似度，为保证相邻的 $\boldsymbol{x}$ 输入数据对应的预测结果 $\boldsymbol{y}$ 也具有高相关性，给出流形正则项的定义，具体为

$$\|f\|_K^2 = \sum_{i=1}^{N}\sum_{j\in N_i}\sum_{k=1}^{C} \|\hat{\boldsymbol{y}}_k^i - \hat{\boldsymbol{y}}_k^j\|^2 \omega_{ij} \tag{5}$$

从式(5)可以看出，当数据 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 具有高相似度时，其对应的预测结果 $\hat{\boldsymbol{y}}_k^i$ 与 $\hat{\boldsymbol{y}}_k^j$ 之间的差值也应该尽可能小，否则该惩罚项就会增大损失函数。从而维持了输入数据和预测结果直接的几何结构，如图 1 所示。

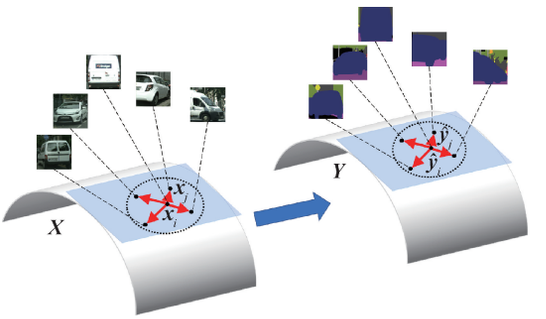


图 1 几何结构示意图

Fig. 1 Geometric structure diagram

因此，针对图像分割问题的求解，即对引入流形正则约束项后的总体损失函数 $L_{\text{tot}}(\boldsymbol{y},\hat{\boldsymbol{y}})$ 进行优化求解，即

$$\arg\min_{\theta}\left[\sum_{i=1}^{N}\sum_{k=1}^{C} -\boldsymbol{y}_k^i \log(\hat{\boldsymbol{y}}_k^i) + \lambda_a \sum_{i=1}^{N_p}\sum_{j\in N_i}\sum_{k=1}^{C} \|\hat{\boldsymbol{y}}_k^i - \hat{\boldsymbol{y}}_k^j\|^2 \omega_{ij}\right] \tag{6}$$

式中，N 表示所有图像点，N_p 表示图像中所有不相交的子集。θ 为模型所有参数，代表了损失函数 L_{tot} 的解空间。

2.3 子图像块划分

在实际图像分割问题中，给定训练数据和标签真实值 $\{(\boldsymbol{P}_m,\boldsymbol{T}_m)\}_{m=1}^M$，$\boldsymbol{P}_m \in \mathbf{R}^{w\times h\times 3}$ 和 $\boldsymbol{T}_m \in \mathbf{R}^{w\times h\times c}$，$M$ 为训练样本总数，$\boldsymbol{P}_m$ 表示第 m 个训练图像，$\boldsymbol{T}_m$ 表示第 m 个图像标签，w 和 h 代表图像的宽和高，c 代表对应的分类数。本文引入分割系数 s，即将原始图像分割为 $s\times s$ 个子图像块，并将整个子图像块的原始数据和预测结果作为流形正则项中数据点 $\boldsymbol{x}_i$ 的取值，即

$$\begin{cases} \{\boldsymbol{x}_i\}_{i=1}^{s\times s} & \boldsymbol{x}_i \in \mathbf{R}^{(w/s)\cdot(h/s)\cdot 3} \\ \{\hat{\boldsymbol{y}}_i\}_{i=1}^{s\times s} & \hat{\boldsymbol{y}}_i \in \mathbf{R}^{(w/s)\cdot(h/s)\cdot c} \end{cases} \tag{7}$$

从式(7)可以看出，当分割的子图像块取最小值 1×1 时，其形式与 CRFasRNN(conditional random fields as recurrent neural networks)(Zheng 等，2015)中的势函数十分接近，核心思想均为给具有相同 RGB 取值的像素点分配相同的预测结果，在概率图模型中，即增大 $P(\boldsymbol{y}\mid\boldsymbol{x})$，从几何结构的角度看，即

减小高相似度的输入数据对应的输出结果在高维空间中的欧氏距离。当所选子图像块逐渐增大，可视为在高维数据空间中相邻近的子图像块对应的分割结果在其高维空间中亦十分接近。当子图像块的大小等于输入图像尺寸 $h \times w$ 时，此时可在训练批数据之间建立关联，从而将本文算法用于流行正则约束常见的半监督学习（Belkin 等，2006）。

2.4　算法实现

将流行正则约束项加入到现有的深度学习网络模型，可实现端对端的图像语义分割模型的训练。

在深度学习分割模型上，给定输入图像以及标签真实值 $\{(\boldsymbol{P}_m, \boldsymbol{T}_m)\}_{m=1}^{M}$，用卷积操作实现权重矩阵 $\boldsymbol{\Omega}$ 的快速计算，如图 2 所示。具体步骤如下：

1）将原始图像分割为 $s \times s$ 个子图像块作为 $\boldsymbol{x}_i$；

2）利用式（4）计算各子图像块之间的权重矩阵 $\boldsymbol{\Omega}$；

3）将原始图像送入深度学习图像分割网络，计算预测结果 $\hat{\boldsymbol{y}}$；

4）根据式（6）计算包含流形正则项约束的总体损失函数 L_{tot}；

5）利用随机梯度下降法对模型参数进行更新；

6）重复步骤 3）—5），直至结果收敛。

上述步骤中，确定子图像块的数目 N 后，步骤 2）与训练过程无关，可在训练前完成所有子图像块间权重矩阵 $\boldsymbol{\Omega}$ 的计算，从而大幅减少运算时间。

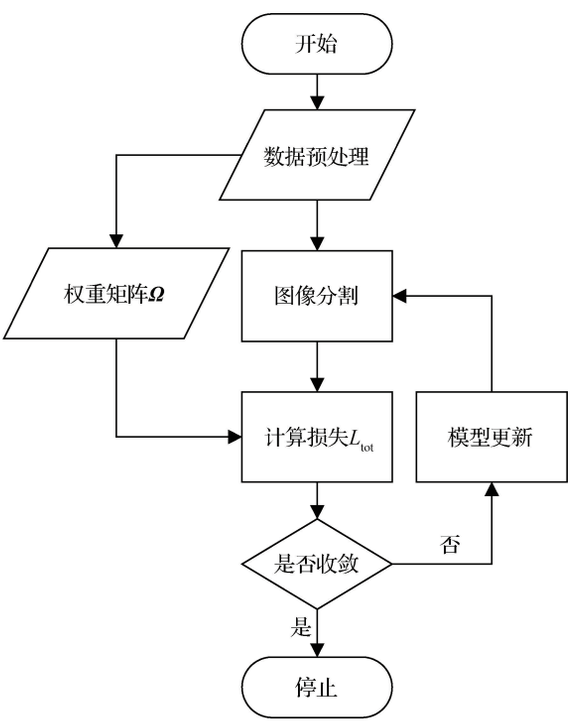


图 2　算法流程图

Fig. 2　Algorithm flow chart

3　实验分析

为了验证本文算法的有效性，通过实验比较了算法在 Cityscapes（Cordts 等，2016）和 PASCAL VOC 2012（pattern analysis, statistical modeling and computational learning visual object classes）（Everingham 等，2012）两个数据集上的准确性，并与当前先进算法进行比较。实验结果表明，本文算法在 Cityscapes 和 PASCAL VOC 2012 两个数据集上均达到了最佳效果。

实验平台的操作系统为 Centos7，硬件配置为 NVIDIA RTX 2080Ti GPU 和 Intel（R）Core（TM）i7-6850 CPU。本文算法实现采用开发语言为 Python3. 7. 6，基于开源的工具箱 Pytorch1. 5. 0。采用优化函数 SGD（stochastic gradient descent）进行小批量训练，采取一种多元的学习率 $\left(1 - \frac{E}{E_{\text{max-epoch}}}\right)^{0.9}$，式中，$E_{\text{max-epoch}}$ 为最大迭代次数，E 为当前迭代次数。采用随机裁剪和依概率水平反转的方式进行数据增强以强化模型的性能。

本文算法采用基于 ImageNet 上的 ResNet（residual network）预训练网络模型（He 等，2016），并使用 DeepLab v3 作为图像分割骨架网络，并在骨架网络基础上加入子图像块的提取以及流形正则约束项作为本文算法的最终实现。

3.1　评价标准和数据集

采用平均交并比（mean intersection over union，mIoU）作为语义分割的评价标准，其计算为

$$mIoU = \frac{1}{C}\sum_{i=1}^{C}\frac{tp}{tp + fp + fn} \tag{8}$$

式中，tp 表示某一类别中正确的正样本分类结果，fp 表示错误的正样本分类结果，fn 表示错误的负样本分类结果，对所有类别的交并比求平均，即为平均交并比 $mIoU$。

实验比较了不同分割算法在 Cityscapes 和 PASCAL VOC 2012 两个数据集上的平均交并比。

Cityscapes 数据集用于城市语义分割任务，该数据集中包含来自 50 个城市的 5 000 幅高质量像素级精细标注的图像和 20 000 幅粗略标注的图像，并

且在分割图像中存在大量的遮挡、目标尺寸大小不一及光照不均的情况，如图 3 所示。每一幅图像的分辨率都为 1 024 × 2 048 像素，共分为 35 个类，其中 19 类用于语义分割的评估，实验仅采用 5 000 幅精细标注的图像。PASCAL VOC 2012 数据集主要针对目标对象进行分割，非目标对象视为背景。共包含 10 582 幅图像，涉及 20 个前景对象和 1 个背景类的分割。

图 3　Cityscapes 分割示意图

Fig. 3　Cityscapes segmentation example diagram

3.2　与骨架网络的对比实验

本文通过引入流形正则化约束项对图像中的上下文信息进行捕获，同时维持语义分割过程中数据的几何结构一致性。本文算法中分割尺度 s 为实验中一项关键的超参数，将直接影响实验结果。为确定分割尺度 s 的大小，基于 ResNet101 网络对分割尺度进行快速实验，对比了平均交并比（mIoU）和运行时间（time），超参数 λ_a 设为 0.01，实验结果如表 1 所示。可以看出，随着分割尺度 s 的逐渐增大，模型效果也得到提升，当 s 增大到 10 之后，继续增大 s 对模型的提升效果已趋于饱和。同时，随着 s 的增大，训练时长成倍增加。为平衡实验精度和训练时长，后续实验中的 s 均设为经验值 10，即将原始图像平均分割为 100 个子图像块。

表 1　本文算法在不同分割尺度的平均交并比和运行时间

Table 1　The mIoU and time by proposed algorithm with different segmentation scales

分割尺度	mIoU/%	运行时间/h
6	77.3	**20.0**
8	77.6	25.9
10	**77.7**	36.7
12	**77.7**	65.4

注：加粗字体为各列最优结果。

为进一步验证本文算法的有效性，将本文算法与基础骨架网络进行实验对比。设置不同的约束权重 λ_a 作为超参数，对流形正则项的影响进行经验性分析，实验结果如表 2 所示。可以看出，流形正则化约束的图像语义分割算法显著提升了模型性能。λ_a 取值在[0, 0.001]时，模型精度逐步提高，λ_a = 0.001 时模型精度最高；λ_a 取值在[0.001, 0.01]时，模型精度逐步下降，但仍优于骨架网络；λ_a > 0.01 时，模型精度持续下降并且性能弱于骨架网络。故参数 λ_a 应控制在[0, 0.01]，选取 0.001 时，模型取得最优值。针对这一实验结果，本文认为这一数值与模型训练数据的内在维度（intrinsic dimension）有关，当网络模型将数据从输入数据的流形形状转换至输出结果的流形形状时，两者的流形维度即各自内在维度的差异会对超参数的选取造成一定影响。当两者流形维度相差较大时，过强的形状约束会对模型的网络性能造成影响，此时应该减小流形约束项的影响，即减小 λ_a，反之亦然。与骨架网络相比，采用流形正则化约束算法的 mIoU 最高为 78.0%，在对网络模型推理过程不引入额外计算量的前提下，最终结果提高了 0.5%。

表 2　本文算法与基础骨架网络在不同权重下的 mIoU 对比

Table 2　Comparison of mIoU by bone network and our algorithm among different weights

方法	基础网络	λ_a	mIoU/%
骨架网络	ResNet101	0	77.5
本文	ResNet101	0.000 1	77.6
本文	ResNet101	0.001	**78.0**
本文	ResNet101	0.01	77.7
本文	ResNet101	0.1	77.4

注：加粗字体为最优结果。

3.3　ResNet50 和 ResNet101 模型泛化实验

为验证本文算法的泛化性，在不同数据集上分别以 ResNet50 和 ResNet101 为基础网络，与骨架网络 BoneNet（bone network）进行对比实验，结果如表 3 所示。可以看出，与采用骨架网络的分割结果相比，本文算法在 Cityscapes 数据集上以 ResNet50 和 ResNet101 为基础网络的 mIoU 分别提升了 0.3% 和

0.5%，在 PASCAL VOC 2012 数据集上以 ResNet50 和 ResNet101 为基础网络的 mIoU 分别提升了 0.8% 和 2.1%。通过实验对比可以发现，与参数量较少的 ResNet50 网络相比，本文提出的流形正则约束项在 ResNet101 骨架网络上的精度提升幅度更多，原因在于流形正则化主要对模型参数的优化过程起约束作用，因此在模型参数量更大的网络模型中能取得更好的效果。

表 3　本文算法与骨架网络在不同数据集和不同基础网络上的 mIoU 对比

Table 3　Comparison of mIoU with different base networks between BoneNet and our algorithm on different datasets

/%

算法	Cityscapes		PASCAL VOC 2012	
	ResNet50	ResNet101	ResNet50	ResNet101
骨架网络	77.0	77.5	67.1	67.4
本文	**77.3**	**78.0**	**67.9**	**69.5**

注：加粗字体表示各列最优结果。

3.4　各项分类结果及可视化

为验证流形正则化约束的图像语义分割算法对网络的影响，对 Cityscapes 数据集中每个单独的语义类别，在以 ResNet101 为基础的骨架网络和本文算法上进行定量和定性实验对比，结果如表 4 所示。

从表 4 可以发现，本文算法对大多数语义类别的精度均有所提升，主要原因有：1）采用流形正则化约束的图像语义分割算法增加了图像分割过程中的上下文信息，使分割模型在学习过程中不再局限于局部信息；2）采用本文算法使得图像在源域和目标域之间保留了更多的本征结构，使得目标图像更加贴近原始图像中的几何形态；3）通过添加约束项，提高了网络的学习能力，优化了模型性能。

本文提出的流形正则化约束的图像语义分割算法和骨架网络在 Cityscapes 数据集上的分割效果如图 4 所示，不同颜色表示不同的分割目标，标准分割图像中的黑色区域为未标记区域，对比部分采用不同色彩的窗口标出，并对第 4 行和第 5 行的细节部分进行了放大展示。可以发现，本文算法的精度较骨架网络有了较大提升，可以纠正一些误分类别，如第 1 行的人行道和第 3 行的巴士。对未标记部分分类更加平滑，分割图像整洁连续，如第 2 行的道路，虽然在标准分割图像中未进行标记，但是通过本文

表 4　模型语义类别实验精度对比表

Table 4　Comparison of experimental accuracy of model semantic categories

/%

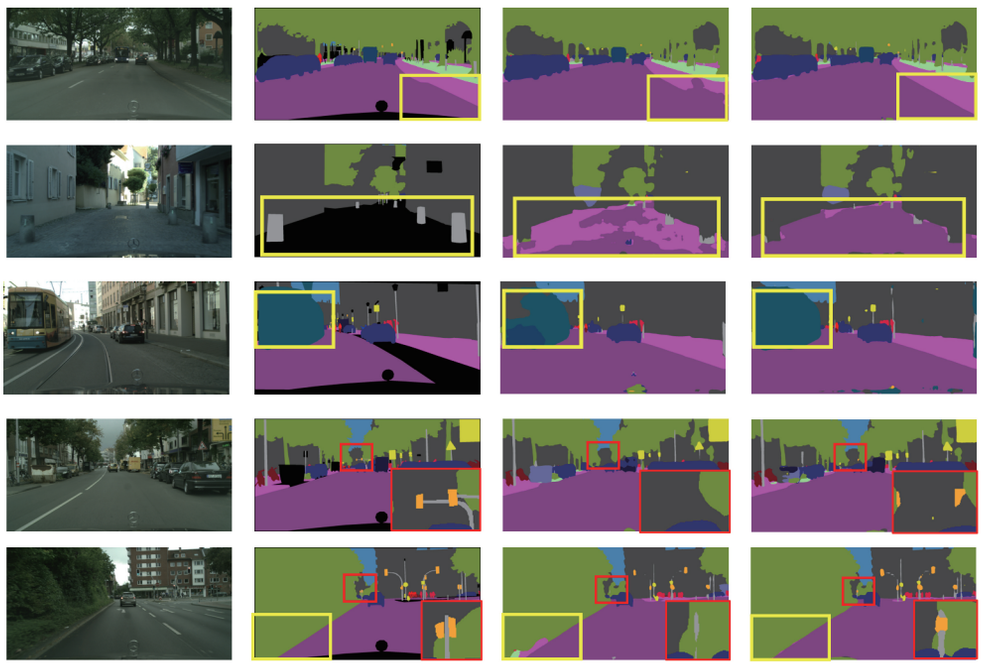
类别	骨架网络	本文算法
道路	**97.9**	**98.0**
人行道	83.6	84.2
建筑物	92.3	92.3
墙壁	53.5	54.9
围栏	59.6	61.1
杆	62.4	63.3
交通灯	67.8	69.4
交通标志	76.5	77.8
植物	92.4	92.5
地表	62.9	64.4
天空	95.0	94.8
人	81.3	81.7
骑士	63.3	63.6
小型车	94.8	94.9
卡车	81.7	81.0
巴士	87.5	87.2
动车	76.3	76.5
摩托车	67.5	67.7
自行车	76.1	76.3
平均交并比	77.5	78.0

注：加粗字体表示各列最优结果。

算法依旧取得了良好效果，分类正确且边缘平滑。原因是采用流形正则化约束的图像分割算法摆脱了局部信息的限制，可以利用图像中的上下文信息进行学习，在边缘处理和像素点分类时可以获得更加优异的效果。另外，采用本文算法可以加强局部特征中细节方面的描述，如第 4 行和第 5 行橙色部分，图中的信号灯等细节信息成功标记出来。因为采用本文算法减少了图像分割过程中本征结构的损失，使得图像中的细节信息得到更好展示。综上所述，本文算法无论对图像中边界及区域信息的描述还是对局部信息细节的区分都是有帮助的，分割图像的语义一致性得到明显改善。

为了进一步评估本文算法的有效性，在 PASCAL VOC 2012 数据集上进行测试，结果如图 5 和表 3 所

示。可以看出,本文算法在 PASCAL VOC 2012 数据集上的精度明显提高,与在 Cityscapes 数据集上的实验结果相似,本文算法减少了误分和漏分现象,如图 5 第 1、2、4 行所示,并且分割图像的边缘更加平



(a) 原始图　(b) 标准分割图　(c) 骨架网络分割图　(d) 本文算法分割图

图 4　本文算法与骨架网络在 Cityscapes 数据集上的分割效果对比

Fig. 4　Comparison of segmentation results between backbone network and ours on Cityscapes dataset

((a) original images; (b) ground truth; (c) segmentation results by backbone network; (d) segmentation results by ours)

(a) 原始图　(b) 标准分割图　(c) 骨架网络分割图　(d) 本文算法分割图

图 5　本文算法与骨架网络在 PASCAL VOC 2012 数据集上的分割效果对比

Fig. 5　Comparison of segmentation results between backbone network and ours on PASCAL VOC 2012 dataset

((a) original images; (b) ground truth; (c) segmentation results by backbone network; (d) segmentation results by ours)

滑,如图 5 第 3 行和第 4 行所示。通过在不同数据集上的实验,不仅证明了算法可以提高模型分割的精度,也证明了算法具有良好的泛化性,表明本文算法可以广泛用于不同的应用场景,具有一定的实用价值。

3.5　与先进算法模型的对比实验

将本文算法在 Cityscapes 和 PASCAL VOC 2012 数据集上与目前先进算法进行横向比较。

在 Cityscapes 数据集上,对比方法有 FCN 8s(Long 等,2015)、DeepLab-V2(Chen 等,2018)、RefineNet(Lin 等,2017)、SSED(semantic segmentation of encoder-decoder structure)(韩慧慧 等,2020)、HarDNet(harmonic DenseNet)(Chao 等,2019)和 DUC(dense upsampling convolution)(Wang 等,2018a),结果如表 5 所示。可以看出,在数据集 Cityscapes 上,本文算法的精度较 FCN 8s 等基础模型有了显著提升,并且优于 DeepLab-V2、RefineNet、SSED、HarDNet 和 DUC 等先进模型。

表 5　不同模型在 Cityscapes 数据集上的横向实验对比

Table 5　Comparison of accuracy experiments of different models on Cityscapes dataset

/%

模型	mIoU
FCN 8s(Long 等,2015)	65.3
DeepLab-V2(Chen 等,2018)	70.4
RefineNet(Lin 等,2017)	73.6
SSED(韩慧慧 等,2020)	74.1
HarDNet(Chao 等,2019)	75.9
DUC(Wang 等,2018a)	77.6
本文	**78.0**

注:加粗字体表示最优结果。

在 PASCAL VOC 2012 数据集上,与 SegNet(Badrinarayanan 等,2017)、FCN 8s(Long 等,2015)、Hypercolumn(Hariharan 等,2015)和 ESPNetv2(efficient spatial pyramid network v2)(Mehta 等,2019)等方法进行对比,结果如表 6 所示。可以看出,在 PASCAL VOC 2012 数据集上,本文算法的精度优于 SegNet、FCN 8s、Hypercolumn 和 ESPNetv2 等模型。

通过以上两个对比实验可以看出,本文算法适用于多种不同场景,并取得了较好结果,在一定程度上解决了误分、漏分问题,并使得分割图像的边缘更加光滑。上述实验结果表明,本文算法在图像分割问题上优于对比方法。

表 6　不同模型在 PASCAL VOC 2012 数据集上的横向实验对比

Table 6　Comparison of accuracy experiments of different models on PASCAL VOC 2012 dataset

/%

模型	mIoU
SegNet(Badrinarayanan 等,2017)	59.9
FCN 8s(Long 等,2015)	62.2
Hypercolumn(Hariharan 等,2015)	62.6
ESPNetv2(Mehta 等,2019)	68.0
本文	**69.5**

注:加粗字体表示最优结果。

4　结　论

本文提出一种流形正则化约束的图像语义分割算法,假定输入图像与预测结果在低维流形空间上存在相同几何结构并以此作为约束,促使网络自适应地捕获数据间的上下文信息。在无需生成巨大特征矩阵并不引入任何推理过程中的额外计算量的前提下,建立了图像分割网络中像素点间的依赖关系。实验结果表明,本文算法十分有效并且适用于不同的骨架网络,尤其在参数量更大的网络模型上表现更为出色,在 Cityscapes 和 PASCAL VOC 2012 两个图像语义分割数据集上的分割性能均优于其他模型。但是,本文算法仅引入了权重系数 λ_a 和分割尺度 s 作为超参数,在下一步工作中,将探求验证各超参数和算法模型之间的关系,并引入模型训练中的其他变量,如像素点的位置信息等,进一步优化算法。

参考文献(References)

Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]

Belkin M, Niyogi P and Sindhwani V. 2006. Manifold regularization: a geometric framework for learning from labeled and unlabeled exam-

ples. The Journal of Machine Learning Research, 7: 2399-2434 [DOI: 10.1007/s10846-006-9077-x]

Byeon W, Breuel T M, Raue F and Liwicki M. 2015. Scene labeling with LSTM recurrent neural networks//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 3547-3555 [DOI: 10.1109/CVPR.2015.7298977]

Chandra S, Usunier N and Kokkinos I. 2017. Dense and low-rank Gaussian CRFs using deep embeddings//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 5113-5122 [DOI: 10.1109/ICCV.2017.546]

Chao P, Kao C Y, Ruan Y S, Huang C H and Lin Y L. 2019. HarDNet: a low memory traffic network//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 3551-3560 [DOI: 10.1109/ICCV.2019.00365]

Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2016. Semantic image segmentation with deep convolutional nets and fully connected CRFs. [EB/OL]. [2020-08-20]. http://arxiv.org/pdf/1412.7062.pdf [DOI: 10.1080/17476938708814211]

Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L. 2018. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4): 834-848 [DOI: 10.1109/TPAMI.2017.2699184]

Chen L C, Papandreou G, Schroff F and Adam H. 2017. Rethinking atrous convolution for semantic image segmentation [EB/OL]. [2020-08-20]. https://arxiv.org/pdf/1706.05587.pdf

Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S and Schiele B. 2016. The cityscapes dataset for semantic urban scene understanding//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 3213-3223 [DOI: 10.1109/CVPR.2016.350]

Everingham M, Van G L, Williams C K I, Winn J and Zisserman A. 2012. PASCAL visual object classes challenge 2012 [DB/OL]. [2020-08-20]. http://host.robots.ox.ac.uk/pascal/VOC/voc2012/

Evgeniou T, Pontil M and Poggio T. 2000. Regularization networks and support vector machines. Advances in Computational Mathematics, 13(1): #1 [DOI: 10.1023/A:1018946025316]

Fu J, Liu J, Tian H J, Li Y, Bao Y J, Fang Z W and Liu H Q. 2019. Dual attention network for scene segmentation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 3141-3149 [DOI: 10.1109/CVPR.2019.00326]

Geng B, Xu C, Tao D C, Yang L J and Hua X S. 2009. Ensemble manifold regularization//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 2396-2402 [DOI: 10.1109/CVPR.2009.5206695]

Han H H, Li W T, Wang J P, Jiao D and Sun B S. 2020. Semantic segmentation of encoder-decoder structure. Journal of Image and Graphics, 25(2): 255-266 (韩慧慧, 李帷韬, 王建平, 焦点, 孙百顺. 2020. 编码—解码结构的语义分割. 中国图象图形学报, 25(2): 255-266) [DOI: 10.11834/jig.190212]

Hariharan B, Arbeláez P, Girshick R and Malik J. 2015. Hypercolumns for object segmentation and fine-grained localization//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 447-456 [DOI: 10.1109/CVPR.2015.7298642]

He C, Zhang Y H and He Z F. 2020. Semantic segmentation of workpiece target based on multiscale feature fusion. Journal of Image and Graphics, 25(3): 476-485 (和超, 张印辉, 何自芬. 2020. 多尺度特征融合工件目标语义分割. 中国图象图形学报, 25(3): 476-485) [DOI: 10.11834/jig.190218]

He K, Zhang X, Ren S and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Huang Z L, Wang X G, Huang L C, Huang C, Wei Y C and Liu W Y. 2019. CCNet: criss-cross attention for semantic segmentation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 603-612 [DOI: 10.1109/ICCV.2019.00069]

Krähenbühl P and Koltun V. 2012. Efficient inference in fully connected CRFs with Gaussian edge potentials [EB/OL]. [2020-08-20] https://arxiv.org/pdf/1210.5644.pdf

Lin G S, Milan A, Shen C H and Reid I. 2017. RefineNet: multi-path refinement networks for high-Resolution semantic segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 5168-5177 [DOI: 10.1109/CVPR.2017.549]

Liu Z W, Li X X, Luo P, Loy C C and Tang X O. 2015. Semantic image segmentation via deep parsing network//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 1377-1385 [DOI: 10.1109/ICCV.2015.162]

Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]

Luo D J and Huang H. 2014. Video motion segmentation using new adaptive manifold denoising model//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 65-72 [DOI: 10.1109/CVPR.2014.16]

Mehta S, Rastegari M, Shapiro L and Hajishirzi H. 2019. ESPNetv2: a light-weight, power efficient, and general purpose convolutional neural network//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 9182-9192 [DOI: 10.1109/CVPR.2019.00941]

Niyogi P. 2013 Manifold regularization and semi-supervised learning:

some theoretical analyses. The Journal of Machine Learning Research, 14(1): 1229-1250 [DOI: 10.5555/2567709.2502619]

Qing C, Yu J, Xiao C B and Duan J. 2020. Deep convolutional neural network for semantic image segmentation. Journal of Image and Graphics, 25(6): 1069-1090 (青晨, 禹晶, 肖创柏, 段娟. 2020. 深度卷积神经网络图像语义分割研究进展. 中国图象图形学报, 25(6): 1069-1090) [DOI: 10.11834/jig.190355]

Quispe A M and Petitjean C. 2015. Shape prior based image segmentation using manifold learning//Proceedings of 2015 International Conference on Image Processing Theory, Tools and Applications (IPTA). Orleans, France: IEEE: 137-142 [DOI: 10.1109/IPTA.2015.7367113]

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]

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need [EB/OL]. [2020-08-20]. https://arxiv.org/pdf/1706.03762.pdf

Wang P Q, Chen P F, Yuan Y, Liu D, Huang Z H, Hou X D and Cottrell W. 2018a. Understanding convolution for semantic segmentation//Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe, USA: IEEE: 1451-1460 [DOI: 10.1109/WACV.2018.00163]

Wang X L, Girshick R, Gupta A and He K M. 2018b. Non-local neural networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: 7794-7803 [DOI: 10.1109/CVPR.2018.00813]

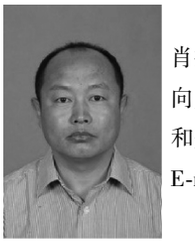
Yu F and Koltun V. 2016 Multi-scale context aggregation by dilated convolutions [EB/OL]. [2020-08-20]. https://arxiv.org/pdf/1511.07122.pdf

Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]

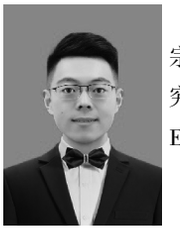
Zhao H S, Zhang Y, Liu S, Shi J P, Loy C C, Lin D H and Jia J Y. 2018. PSANet: point-wise spatial attention network for scene parsing//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 270-286 [DOI: 10.1007/978-3-030-01240-3_17]

Zheng S, Jayasumana S, Romera-Paredes B, Vineet V, Su Z Z, Du D L, Huang C and Torr P H S. 2015. Conditional random fields as recurrent neural networks//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 1529-2537 [DOI: 10.1109/ICCV.2015.179]

作者简介



肖振久,1968 年生,男,副教授,主要研究方向为图像与视觉信息计算、网络与信息安全和数字水印。
E-mail:xiaozhenjiu@lntu.edu.cn



宗佳旭,通信作者,男,硕士研究生,主要研究方向为机器学习和计算机视觉。
E-mail:jiaxu0017@163.com

兰海,男,助理研究员,主要研究方向为机器学习和计算机视觉。E-mail:lanhai09@fjirsm.ac.cn
魏宪,男,研究员,主要研究方向为机器学习、机器视觉和模式识别。E-mail:xian.wei@fjirsm.ac.cn
唐晓亮,男,副研究员,主要研究方向为机器学习、计算机视觉和数据挖掘。E-mail:xltang@fjirsm.ac.cn