

Author's Accepted Manuscript

Learning to Segment with Image-level Annotations

Yunchao Wei, Xiaodan Liang, Yunpeng Chen,
Zequn Jie, Yanhui Xiao, Yao Zhao, Shuicheng Yan



PII: S0031-3203(16)00036-4
DOI: <http://dx.doi.org/10.1016/j.patcog.2016.01.015>
Reference: PR5617

To appear in: *Pattern Recognition*

Received date: 19 July 2015
Revised date: 17 January 2016
Accepted date: 18 January 2016

Cite this article as: Yunchao Wei, Xiaodan Liang, Yunpeng Chen, Zequn Jie, Yanhui Xiao, Yao Zhao and Shuicheng Yan, Learning to Segment with Image-level Annotations, *Pattern Recognition*, <http://dx.doi.org/10.1016/j.patcog.2016.01.015>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and a review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain

Learning to Segment with Image-level Annotations

Yunchao Wei^{a,c}, Xiaodan Liang^{b,c}, Yunpeng Chen^c, Zequn Jie^c, Yanhui Xiao^d,
Yao Zhao^{a,*}, Shuicheng Yan^c

^a*Institute of information Science, Beijing Jiaotong University, Beijing, 100044, China*

^b*School of Advanced Computing, Sun Yat-Sen University, Guangzhou, 510006, China*

^c*Vision and Machine Learning Lab, National University of Singapore, Singapore, 117583, Singapore*

^d*Peoples Public Security University of China, Beijing, 100038, China*

Abstract

Recently, deep convolutional neural networks (DCNNs) have significantly promoted the development of semantic image segmentation. However, previous works on learning the segmentation network often rely on a large number of ground-truths with pixel-level annotations, which usually require considerable human effort. In this paper, we explore a more challenging problem by learning to segment under image-level annotations. Specifically, our framework consists of two components. First, reliable hypotheses based localization maps are generated by incorporating the hypotheses-aware classification and cross-image contextual refinement. Second, the segmentation network can be trained in a supervised manner by these generated localization maps. We explore two network training strategies for achieving good segmentation performance. For the first strategy, a novel multi-label cross-entropy loss is proposed to train the network by directly using multiple localization maps for all classes, where each pixel contributes to each class with different weights. For the second strategy, the rough segmentation mask can be inferred from the localization maps, and then the network is optimized based on the single-label cross-entropy loss with the produced masks. We evaluate our methods on the PASCAL VOC 2012 segmentation benchmark. Extensive experimental results demonstrate the effectiveness

*Corresponding author

Email address: yzhao@bjtu.edu.cn (Yao Zhao)

Download English Version:

<https://daneshyari.com/en/article/4969951>

Download Persian Version:

<https://daneshyari.com/article/4969951>

[Daneshyari.com](https://daneshyari.com)