

Translation-aware Fully Convolutional Instance Segmentation

Jifeng Dai*, Haozhi Qi*, Yi Li*

Microsoft Research Asia

Visual Computing Group

(*Equal contribution. This work was done when Haozhi Qi and Yi Li were interns at MSRA)



Highlights

- The first pure FCN-based method for instance segmentation
- 1st place in COCO segmentation challenge 2016, 11% better than 2nd
 - 33% better than the 2015 championship entry (MNC)
 - We won the challenge back-to-back!
- Fastest CNN-based method for instance segmentation
 - 0.24 sec/img using ResNet-101 on K40 GPU
 - ~6x faster than MNC

Jifeng Dai, Kaiming He, Jian Sun. Instance-aware Semantic Segmentation via Multi-task Network Cascades. In CVPR 2016 Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. In CVPR 2016



Conventional FCNs

- Successful end-to-end solution for semantic segmentation
 - Each pixel on score maps is a per-pixel category classifier



semantic segmentation result

Jonathan Long*, Evan Shelhamer*, Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. In CVPR 2015.



Conventional FCNs

- Unclear application in instance segmentation
 - Instance segmentation: *translation variant* output
 - Conventional FCNs: by nature *translation invariant*
 - Some endeavors exist, but no pure FCN-based end-to-end solution



instance segmentation result

Jifeng Dai, Kaiming He, Yi Li, Shaoqing Ren, Jian Sun. Instance-sensitive Fully Convolutional Networks. In ECCV 2016. Xiaodan Liang, Yunchao Wei, Xiaohui Shen, Jianchao Yang, Liang Lin, Shuicheng Yan. Proposal-free Network for Instance-level Object Segmentation. Arxiv, 2015. Shu Liu, Xiaojuan Qi, Jianping Shi, Hong Zhang, Jiaya Jia. Multi-scale Patch Aggregation (MPA) for Simultaneous Detection and Segmentation. In CVPR 2016.

Prevalent Methods for Instance Segmentation

Microsoft°

- Composed of two sub-networks divided by Rol pooling & masking
 - Issues
 - Detailed spatial info is damaged and even lost in feature warping & fc sub-network
 - Region-specific fc sub-network computation



Xiaodan Liang, Yunchao Wei, Xiaohui Shen, Zequn Jie, Jiashi Feng, Liang Lin, Shuicheng Yan. Reversible recursive instance-level object segmentation. In CVPR 2016 Jifeng Dai, Kaiming He, Jian Sun. Instance-aware Semantic Segmentation via Multi-task Network Cascades. In CVPR 2016 Jifeng Dai, Kaiming He, Jian Sun. Convolutional Feature Masking for Joint Object and Stuff Segmentation. In CVPR 2015



Can we make FCNs aware of translation for end-to-end instance segmentation?

Translation-aware Fully Convolutional Networks (TA-FCNs)

- Translation-aware fg/bg score maps + position-sensitive RoI pooling
- Shared score maps between foreground mask estimation and categorization



Microsoft[®]

Research

translation-aware fg/bg score maps

Jifeng Dai, Yi Li, Kaiming He, Jian Sun. R-FCN: Object Detection via Region-based Fully Convolutional Networks. In NIPS 2016. Jifeng Dai, Kaiming He, Yi Li, Shaoqing Ren, Jian Sun. Instance-sensitive Fully Convolutional Networks. In ECCV 2016.



TA-FCNs



Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In NIPS 2015.



TA-FCNs

- Backbone architecture
 - Based on ResNet-101
 - "Algorithme atrous" on conv5 (output stride: 32 pixels->16 pixels)

Inference

- Negligible per-Rol computation
- Well preserved spatial info: no feature warping/fc layers
- Learning
 - Cost-free online hard example mining (OHEM)

Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. In ICLR 2015.

Abhinav Shrivastava, Abhinav Gupta, Ross Girshick. Training Region-based Object Detectors with Online Hard Example Mining. In CVPR 2016.



TA-FCNs

- Implementation details (very few hacks)
 - Finetuning on COCO 2014 trainval from ImageNet pretrained net
 - Image shorter side of 600 pixels in training
 - 12 anchors, 300 proposals per image in RPN
 - Mask voting on the initial/refined Rols





Experiments

• Comparison with MNC on COCO test-dev (using ResNet-101)

method	OHEM on initial/refined RoIs?	train time /img	test time/img	mAP	mAP@0.5	mAP (small)	mAP (mid)	mAP (large)
MNC	No	2.05s	1.37s	24.6	44.3	4.7	25.9	43.6
TA-FCN	No	0.53s	0.24s	<u>28.8</u>	<u>48.7</u>	<u>6.8</u>	<u>30.8</u>	<u>49.5</u>
MNC	Yes	3.22s	1.37s	N/A	N/A	N/A	N/A	N/A
TA-FCN	Yes	0.54s	0.24s	29.2	49.5	7.1	31.3	50.0



Experiments

• COCO Segmentation Challenge 2016 Entry

	mAP	mAP@0.5
FAIRCNN (2015 2nd)	25.0	45.6
MNC+++ (2015 1st)	27.9	51.2
G-RMI (2016 2nd)	33.8	56.9
TA-FCN	29.2	49.5
+ multiscale testing	32.0	51.9
+ horizontal flip	32.7	52.7
+ multiscale training	33.6	54.5
+ ensemble (2016 1st)	37.6	59.9

Research



Results on the first 5k images from the COCO test set is available at <u>https://github.com/daijifeng001/TA-FCN</u>

Microsoft[®]



Results on the first 5k images from the COCO test set is available at <u>https://github.com/daijifeng001/TA-FCN</u>



- TA-FCN
 - The first pure FCN-based end-to-end solution for instance segmentation
 - Fast and accurate

• Our team



Haozhi Qi*

Jifeng Dai*

Yi Li*

(*Equal contribution. This work was done when Haozhi Qi and Yi Li were interns at MSRA)

- Resources
 - More details would be available in our tech report
 - Code of TA-FCN would be released
 - Results on the first 5k images from the COCO test set is available now!

https://github.com/daijifeng001/TA-FCN