Annotating Object Instances with a Polygon-RNN

Lluís Castrejón[1], Kaustav Kundu[1,2], Raquel Urtasun[1,2,3], Sanja Fidler[1,2]

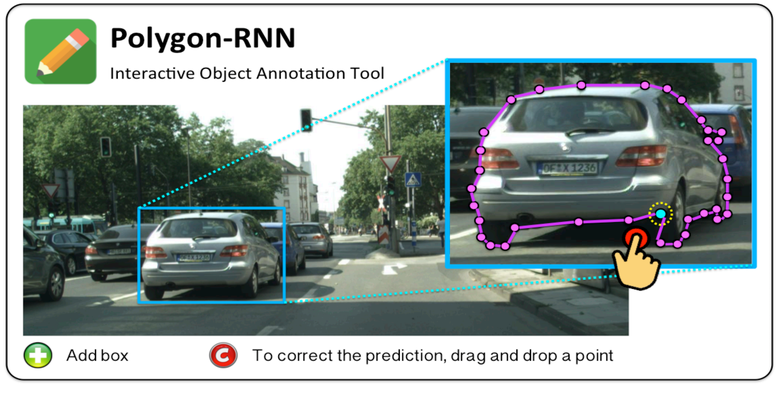
University of Toronto[1], Vector Institute[2], Uber ATG Toronto[3]

Visit our website:

IEEE 2017 Conference on Computer Vision and Pattern Recognition

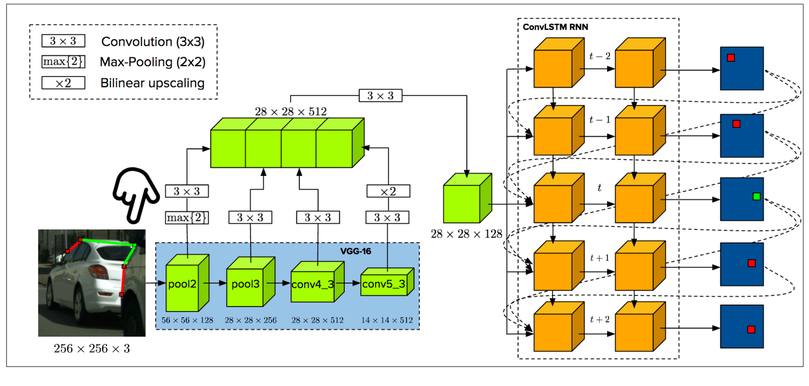
CVPR July 21-26 2017 HONOLULU

Object Instance Annotation



- Annotating object instances is expensive and time-consuming
- We propose a model that segments instances automatically using polygons
- Our model easily integrates user corrections and can be used as an annotation tool

Polygon-RNN



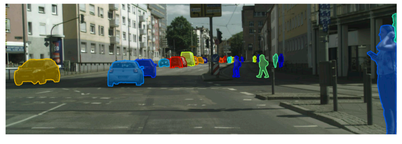
Prediction Mode

We investigate what is the agreement (in % IOU) with the GT annotations in Cityscapes

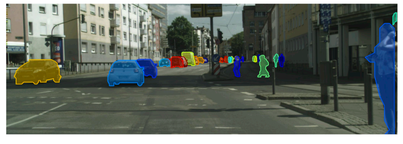
- Our **inputs** are the GT instance boxes
- **No user corrections** are needed

Model	Bicycle	Bus	Person	Train	Truck	Motorcycle	Car	Rider	Mean
Square Box	35.41	53.44	26.36	39.34	54.75	39.47	46.04	26.09	40.11
Dilation10	46.80	48.35	49.37	44.18	35.71	26.97	61.49	38.21	43.89
DeepMask	47.19	69.82	47.93	62.20	63.15	47.47	61.64	52.20	56.45
SharpMask	52.08	**73.02**	53.63	**64.06**	65.49	51.92	65.17	56.32	60.21
Ours	**52.13**	69.53	**63.94**	53.74	**68.03**	**52.07**	**71.17**	**60.58**	**61.40**

GT Cityscapes (813 clicks)



Ours (28 boxes, no corrections)

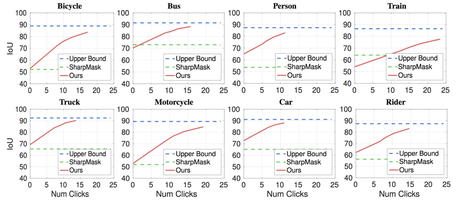


Annotation Mode

Determine what is the user interaction required for different annotation quality levels

- We **simulate a user** by correcting predicted vertices that deviate further than a **threshold T** from the corresponding GT vertex in the output space

Method	Num. Clicks	IoU	Annot. Speed-Up
Cityscapes GT	33.56	100	-
Ann. full image	79.94	69.5	-
Ann. crops	96.09	78.6	-
Ours (Automatic)	0	73.3	No ann.
Ours (T=1)	9.3	87.7	x3.61
Ours (T=2)	6.6	85.7	x5.11
Ours (T=3)	5.6	84.0	x6.01
Ours (T=4)	4.6	82.2	x7.31



Annotation mode: Comparison with GrabCut

GrabCut is a method to segment images commonly used as an annotation tool

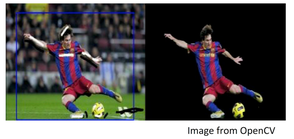
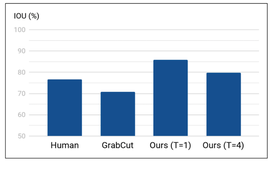


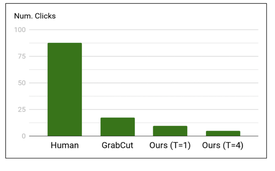
Image from OpenCV

Key differences from Polygon-RNN:

- Produces a dense pixel-wise annotation
- Users segment instances by painting strokes

Comparison on 54 random car instances



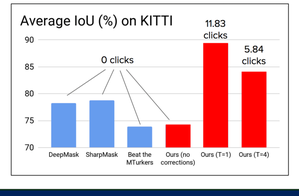


Higher IOU than GrabCut and comparable to Human Ann.

2x speed-up compared to GrabCut

Cross-Domain Results on KITTI

We test on KITTI models trained on Cityscapes without fine-tuning



- Without any user corrections, our model performs comparably to SOTA
- With <6 clicks on average our model is at the human agreement level

Conclusions

- Our method is competitive with SOTA instance segmentation models given GT boxes
- x4.7 speed-up when annotating Cityscapes with our model
- With Polygon-RNN an annotator can get more accurate annotations with less clicks compared to GrabCut
- Our model can be applied out-of-the-box in other domains