

Recurrent Semantic Instance Segmentation

Amaia Salvador*, Manel Baradad, Xavier Giró-i-Nieto and Ferran Marqués
Universitat Politècnica de Catalunya

We present a recurrent model for end-to-end instance-aware semantic segmentation that is able to sequentially generate pairs of masks and class predictions. Our proposed system is trainable end-to-end for instance segmentation, does not require further post-processing steps on its output and is conceptually simpler than current methods relying on object proposals [1]. While recent works [4, 3] have proposed recurrent architectures for instance segmentation, these are trained and evaluated for a single category.

Our model (depicted in Figure 1) is composed of a series of Convolutional LSTMs [5] that are applied in chain with upsampling layers in between to predict a sequence of binary masks and associated class probabilities. Skip connections are incorporated in our model by concatenating the output of the corresponding convolutional layer in the base model with the upsampled output of the ConvLSTM. Binary masks are finally obtained with a 1×1 convolution with sigmoid activation. We concatenate the side outputs of all ConvLSTM layers and apply a per-channel max-pooling operation followed by a single fully-connected layer with softmax activation to obtain the category for each predicted mask.

We train and evaluate our models with the Pascal VOC 2012 dataset [2]. Figure 2 shows some example predictions from our model, where object coloring indicates the order in which objects were found by the network (i.e. 0: dark blue, 1: green, 2: red, 3: light blue). Future work will aim at analyzing and understanding the behavior of the network on other datasets, comparing the system with state of the art solutions and study the relationship of the learned object discovery patterns of our model with those of humans.

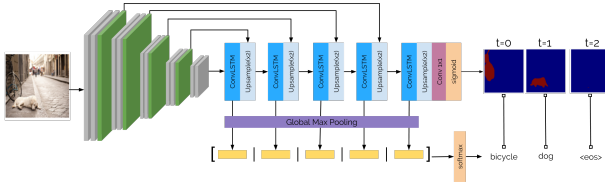


Figure 1: Proposed architecture

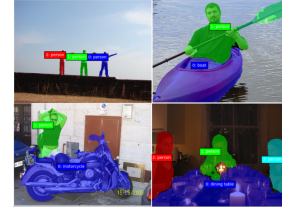


Figure 2: Sample predictions.

References

- [1] Jifeng Dai, Kaiming He, and Jian Sun. Instance-aware semantic segmentation via multi-task network cascades. In *CVPR*, 2016.
- [2] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge. *IJCV*, 2010.
- [3] Mengye Ren and Richard S. Zemel. End-to-end instance segmentation with recurrent attention. In *CVPR*, 2017.
- [4] Bernardino Romera-Paredes and Philip Hilaire Sean Torr. Recurrent instance segmentation. In *ECCV*, 2016.
- [5] SHI Xingjian, Zhouong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation. In *NIPS*, 2015.

*This work has been developed in the framework of projects TEC2013-43935-R and TEC2016-75976-R, financed by the Spanish Ministerio de Economía y Competitividad and the European Regional Development Fund (ERDF).

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.