One-Shot Segmentation in Clutter

Claudio Michaelis [1,2], Matthias Bethge [1-4] and Alexander Ecker [1,2,4]

1 Centre for Integrative Neuroscience, University of Tübingen, Germany 2 Bernstein Center for Computational Neuroscience, Tübingen, Germany 3 Max Planck Institute for Biological Cybernetics, Tübingen, Germany 4 Center for Neuroscience and Artificial Intelligence, Baylor College of Medicine, Houston, USA

contact: claudio.michaelis@uni-tuebingen.de

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Abstract

Task: Find and segment a previously unseen character in a cluttered scene based on a single instruction example.

Dataset: Cluttered Omniglot is composed of characters from the Omniglot [1] dataset which are augmented with affine transformations and random coloring.

Difficulty: Using oracle models with access to various amounts of ground-truth information, we can show that clutter poses a serious problem to current state-of-the-art computer vision models.

Proposed solution: To counter the problem of clutter we suggest segmenting objects before comparing them to the target. To this end, we introduce MaskNet, a model that attends to multiple candidate locations, generates segmentation proposals to mask out background clutter and then selects among the segmented objects .

Results:
1. Clutter is a serious problem for current state-of-the-art convnets.
2. Removing clutter makes the same task very easy.
3. Segmenting objects first to separate them from the cluttered background before deciding among objects is beneficial.

Task

Cluttered Omniglot: Segmentaion task.

Traditional Omniglot: 20-way choice task [2]

Cluttered Omniglot examples

Baseline

Siamese-U-Net Baseline: Scene and target are encoded using a Siamese network architecture [2]. A pixel wise comparison between both encodings is performed. The scene is then segmented from this comparison map using a U-Net architecture. [3]

Siamese-U-Net Baseline: One-Shot performance similar to validation performance. Clutter increases difficulty

Conclusion
- The model learns the task rather than the training classes.
- Clutter increases the task difficulty.

Oracles

Oracle: Characters are extracted by their color and cropped into patches, either with or without the background of distractors. A decision is then made beween the extracted patches.

Oracle with and without clutter: Without clutter the performance is almost perfect even for a large number of patches . With clutter the performance drops.

Conclusion
Clutter indeed causes a problem.

MaskNet

MaskNet: Following the principle "Segment first, Decide later" segmentation proposals are generated at multiple target locations. Afterwards a decision among these segmentations is made.

MaskNet: Growing performance gains over the baseline with increasing clutter leves.

Conclusion
Segmenting object candidates before deciding among them improves performance.

Additional Analysis

MaskNet Ablations:
1. Clutter impedes the ability to generate good segmentation proposals.
2. Initializing the proposal generation process with information about the target is essential.

Localization: MaskNets performance gain over the baseline is even bigger, when only localization accuracy is evaluated.

Conclusion
Localization and segmentation are intertwined problems which are both influenced by clutter.

Model Architecture

Conclusions

- ★ Clutter is a serious problem for current computer vision algorithms
- ★ One-Shot segmentation without clutter can be solved with current computer vision models
- ★ Segmenting objects first before comparing them helps to deal with clutter
- ★ Object localization and segmentation are connected tasks and can't be solved separately

Funding

This work was supported by the German Research Founda- tion (DFG) through Collaborative Research Center (CRC 1233) "Robust Vision" and DFG grant EC 479/1-1, and by the Advanced Research Projects Activity (IARPA) via Department of Interior/Interior Business Center (DoI/IBC) contract number D16PC00003. The U.S. Government is authorized to reproduce and distribute reprints for Govern- mental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorse- ments, either expressed or implied, of IARPA, DoI/IBC, or the U.S. Government.

References

[1] Lake, B. M., Salakhutdinov, R., and Tenenbaum, J. B. Human-level concept learning through probabilistic program induction. Science, 350(6266):1332–1338, 2015.

[2] Koch, G., Zemel, R., and Salakhutdinov, R. Siamese Neural Networks for One-shot Image Recognition ICML, 2015.

[3] Ronneberger, O., Fischer, P., and Brox, T. U- Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer, 2015.

github.com/michaelisc/cluttered-omniglot

CIN Werner Reichardt Centre for Integrative Neuroscience

bccn tübingen

imprs-is