מכון ויצמן למדע

THE WEIZMANN INSTITUTE OF SCIENCE

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Vision and Robotics Seminar

on Thursday, Jun 17, 2021
at 12:15

https://weizmann.zoom.us/j/97396405783?pwd=MUVWaElqWjVVSmtsQmU5MG9RMkZaQT09

Hila Levi
WIS

Combining bottom-up and top-down computations for image interpretation

Abstract:

Visual scene understanding has traditionally focused on identifying objects in images â learning to predict their presence and spatial extent. However, understanding a visual scene often goes beyond recognizing individual objects. In my thesis (guided by Prof. Ullman), I mainly focused on developing `task-dependent networkâ, that uses processing instructions to guide the functionality of a shared network via an additional task input. In addition, I also studied strategies for incorporating relational information into recognition pipelines to efficiently extract structures of interest from the scene. In the scope of high level scene understanding, which might be dominated by recognizing a rather small number of objects and relations, the above task-dependent scheme naturally allows goal-directed scene interpretation by either a single step or by a sequential execution with a series of different TD instructions. It simplifies the use of referring relations, grounds requested visual concepts back into the image-plane and improves combinatorial generalization, essential for AI systems, by using structured representations and computations. In the scope of âmulti-task learningâ the above scheme offers an alternative to the popular âmulti-branched architectureâ, which simultaneously execute all tasks using task-specific branches on top of a shared backbone, challenges capacity limitations, increases task selectivity, allows scalability and further tasks extensions. Results will be shown in various applications: object detection, visual grounding, properties classification, human-object interactions and general scene interpretation. Works included: 1. H. Levi and S. Ullman. Efficient coarse-to-fine non-local module for the detection of small objects. BMVC, 2019. https://arxiv.org/abs/1811.12152 2. H. Levi and S. Ullman. Multi-task learning by a top-down control network. ICIP 2021. https://arxiv.org/abs/2002.03335 3. S. Ullman et al. Image interpretation by iterative bottom-up top-down processing. http://arxiv.org/abs/2105.05592 4. A. Arbelle et al. Detector-Free Weakly Supervised Grounding by Separation. https://arxiv.org/abs/2104.09829. Submitted to ICCV. 5. Ongoing work â Human Object Interactions