UNIVERSITY OF SOUTH FLORIDA

Defense of a Doctoral Dissertation

Beyond Labels and Captions: Contextualizing Grounded Semantics for Explainable Visual Interpretation

by

Sathyanarayan Aakur Narasimhan

For the Ph.D. degree in Computer Science and Engineering

While significant milestones have been achieved in the field of computer vision, majority of the work has been concentrated on *supervised* visual recognition. This implies a closed world where the underlying assumption is that all environments contain the same objects and events, which are in one-to-one correspondence with the ground evidence in the image. Hence, the learned knowledge is limited to the annotated training set. Increasingly complex models require massive amounts of training data and offer little to no explainability due to the lack of transparency in the decision-making process. In this dissertation, we develop an inherently explainable approach for generating rich interpretations of visual scenes. We move towards an *open world, open-domain* visual understanding by decoupling the ideas of recognition and reasoning. We show that the proposed approach is able to advance the state-of-the-art results in complex benchmarks to handle data imbalance, complex semantics and complex visual scenes without the need for vast amounts of domain-specific training data. Extensive experiments on several publicly available datasets show the efficacy of the proposed approaches.

Examining Committee