



UNIVERSITY OF CENTRAL FLORIDA
CENTER FOR RESEARCH IN COMPUTER VISION

Dr. Ajay Divakaran SRI International

“Using Hierarchies of Skills to Assess and Achieve Automatic Multimodal Comprehension”

Wednesday, November 30, 2022 · 4:00PM · ENG1 327



ABSTRACT

Unlike current visual question answering (VQA), elementary school (K-5) teaching of reading comprehension has a graded approach based on a hierarchy of skills ranging from memorization to content creation. We take inspiration from such hierarchies to investigate both dataset creation and question answering techniques. First, we are currently creating a new visual question answering dataset that tests comprehension of VQA systems in a graded manner using hierarchical question answering with picture stories. Second, we investigate large language models such as GPT-Neo, the open version of GPT-3. We use Bloom’s Taxonomy of comprehension skills to analyze and improve the comprehension skills of large pre-trained language models. Our experiments focus on zero-shot question answering, using the taxonomy to provide proximal context that helps the model answer questions by being relevant to those questions. We show that targeting context in this manner improves performance across 4 popular common sense question answer datasets. Third, we propose conceptual consistency to measure a LLM’s understanding of relevant concepts. To compute it we extract background knowledge by traversing paths between concepts in a knowledge base and then try to predict the model’s response to the anchor query from the background knowledge. We investigate the performance of current LLMs in a commonsense reasoning setting using the CSQA dataset and the ConceptNet knowledge base. While conceptual consistency, like other metrics, does increase with the scale of the LLM used, we find that popular models do not necessarily have high conceptual consistency. Finally, we present work on detection and removal of bias in common multimodal machine comprehension datasets. We hypothesize that this naturally occurring bias present in the dataset affects even the best performing model. We verify our proposed hypothesis and propose an algorithm capable of modifying the given dataset to remove the bias elements.

BIOGRAPHY

Ajay Divakaran, Ph.D., is the Technical Director of the Vision and Learning Lab at the Center for Vision Technologies, SRI International, Princeton. Divakaran has been a principal investigator for several SRI research projects for DARPA, IARPA, ONR etc. His work includes multimodal analytics for social media, real-time human behavior assessment, event detection, and multi-camera tracking. He has developed several innovative technologies for government and commercial multimodal systems. He worked at Mitsubishi Electric Research Labs during 1998-2008 where he was the lead inventor of the world’s first sports highlights playback-enabled DVR, and several machine learning applications. Divakaran was named a Fellow of the IEEE in 2011 for his contributions to multimedia content analysis. He has authored two books, 140+ publications and 60+ issued patents. He received his Ph.D. degree in electrical engineering from Rensselaer Polytechnic Institute .