ACL 2020

Advances in Language and Vision Research

Proceedings of the First Workshop

July 9, 2020

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-952148-14-9 (Volume 1)

Introduction

Language and vision research has attracted great attention from both natural language processing (NLP) and computer vision (CV) researchers. Gradually, this area is shifting from passive perception, templated language, and synthetic imagery or environments to active perception, natural language, and photo-realistic simulation or real world deployment. Thus far, few workshops on language and vision research have been organized by groups from the NLP community. We organize the first workshop on Advances in Language and Vision Research (ALVR) in order to promote the frontier of language and vision research and to bring interested researchers together to discuss how to best tackle and solve real-world problems in this area.

Organizers:

Xin Wang, UC Santa Barbara
Jesse Thomason, University of Washington
Ronghang Hu, UC Berkeley
Xinlei Chen, Facebook AI Research
Peter Anderson, Georgia Tech
Qi Wu, Adelaide University
Asli Celikyilmaz, Microsoft Research
Jason Baldridge, Google Research
William Yang Wang, UC Santa Barbara

Program Committee:

Jacob Andreas, MIT
Angel Chang, Simon Fraser Univeristy
Devendra Chaplot, CMU
Abhishek Das, Georgia Tech
Daniel Fried, UC Berkeley
Zhe Gan, Microsoft
Christopher Kanan, Rochester Institute of Technology
Jiasen Lu, Georgia Tech
Ray Mooney, University of Texas, Austin
Khanh Nguyen, University of Maryland
Aishwarya Padmakumar, University of Texas, Austin
Hamid Palangi, Microsoft Research
Alessandro Suglia, Heriot-Watt University
Vikas Raunak, CMU
Volkan Cirik, CMU
Parminder Bhatia, Amazon
Khyathi Raghavi Chandu, CMU
Asma Ben Abacha, NIH/NLM
Thoudam Doren Singh, National Institute of Technology, Silchar, India
Dhivya Chinnappa, Thomson Reuters
Shailza Jolly, TU Kaiserslautern
Alok Singh, National Institute of Technology, Silchar, India
Mohamed Elhoseiny, KAUST
Marimuthu Kalimuthu, Saarland University
Simon Dobnik, University of Gothenburg
Shruti Palaskar, CMU

Invited Speaker:

Yoav Artzi, Cornell
Joyce Chai, University of Michigan
JJ (Jingjing) Liu, Microsoft
Louis-Philippe Morency, CMU
Mark Riedl, Georgia Tech
Lucia Specia, Imperial College London
Zhou Yu, UC Davis

Table of Contents

Workshop Program

Workshop schedule details: https://alvr-workshop.github.io

The workshop also holds the first Video-guided Machine Translation (VMT) challenge and the REVERIE challenge. The VMT challenge aims to benchmark progress towards models that translate source language sentence into the target language with video information as the additional spatiotemporal context. The challenge is based on the recently released large-scale multilingual video description dataset, VATEX. The VATEX dataset contains over 41,250 videos and 825,000 high-quality captions in both English and Chinese, half of which are English-Chinese translation pairs. The REVERIE challenge requires an intelligent agent to correctly localize a remote target object (cannot be observed at the starting location) specified by a concise high-level natural language instruction, such as "bring me the blue cushion from the sofa in the living room". Since the target object is in a different room from the starting one, the agent needs first to navigate to the goal location. When the agent determines to stop, it should select one object from a list of candidates provided by the simulator. The agent can attempt to localize the target at any step, which is totally up to algorithm design. But the agent is only allowed to output once in each episode, which means the agent only can guess the answer once in a single run.

Archival track papers presented at the workshop:

Extending ImageNet to Arabic using Arabic WordNet
Abdulkareem Alsudais

Toward General Scene Graph: Integration of Visual Semantic Knowledge with Entity Synset Alignment
Woo Suk Choi, Kyoung-Woon On, Yu-Jung Heo and Byoung-Tak Zhang

Visual Question Generation from Radiology Images
Mourad Sarrouti, Asma Ben Abacha and Dina Demner-Fushman

On the role of effective and referring questions in GuessWhat?!
Mauricio Mazuecos, Alberto Testoni, Raffaella Bernardi and Luciana Benotti

Latent Alignment of Procedural Concepts in Multimodal Recipes
Hossein Rajaby Faghihi, Roshanak Mirzaee, Sudarshan Paliwal and Parisa Kordjamshidi