NAACL-HLT 2021

Advances in Language and Vision Research

Proceedings of the Second Workshop

June 11, 2021

©2021 The Association for Computational Linguistics

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 978-1-954085-37-4 (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 real-world environments. Thus far, few workshops on language and vision research have been organized by groups from the NLP community. This year, we are organizing the second workshop on Advances in Language and Vision Research (ALVR) in order to promote the frontier of language and vision research and bring interested researchers together to discuss how to best tackle real-world problems in this area.

This workshop covers (but is not limited to) the following topics:

- New tasks and datasets that provide real-world solutions in the intersection of NLP and CV;
- Language-guided interaction with the real world, e.g. navigation via instruction following or dialogue;
- External knowledge integration in visual and language understanding;
- Visually grounded multilingual study, e.g. multimodal machine translation;
- Fairness in multimodal machine learning;
- Shortcoming of existing language and vision tasks and datasets;
- Benefits of using multimodal learning in downstream NLP tasks;
- Self-supervised representation learning in language and vision;
- Transfer learning (including few/zero-shot learning) and domain adaptation;
- Cross-modal learning beyond image understanding, such as videos and audios;
- Multidisciplinary study that may involve linguistics, cognitive science, robotics, etc.

The details of our workshop can be found at https://alvr-workshop.github.io/.

Proceedings of the ALVR workshop from previous years can be found on ACL Anthology: https://www.aclweb.org/anthology/venues/alvr/

Organizers:

Xin (Eric) Wang, UC Santa Cruz Ronghang Hu, Facebook AI Research Drew Hudson, Stanford Tsu-Jui Fu, UC Santa Barbara Marcus Rohrbach, Facebook AI Research Daniel Fried, UC Berkeley

Program Committee:

Shubham Agarwal, Heriot-Watt University Arjun Akula, University of California, Los Angeles Asma Ben Abacha, NIH/NLM Luciana Benotti, The National University of Cordoba Khyathi Raghavi Chandu, Carnegie Mellon University Angel Chang, Stanford University Dhivya Chinnappa, Thomson Reuters Abhishek Das, Facebook AI Simon Dobnik, University of Gothenburg Thoudam Doren Singh, National Institute of Technology, Silchar, India Hamed Firooz, Facebook AI Zhe Gan, Microsoft Cristina Garbacea, University of Michigan Jack Hessel, AI2 Gabriel Ilharco, University of Washington Shailza Jolly, TU Kaiserslautern Germany Marimuthu Kalimuthu, Saarland University, Saarland Informatics Campus Noriyuki Kojima, Cornell University Christopher Kummel, Beuth University of Applied Sciences Berlin Loitongbam Sanayai Meetei, National Institute of Technology Silchar, India Khanh Nguyen, University of Maryland Yulei Niu, Renmin University of China Aishwarya Padmakumar, University of Texas, Austin Hamid Palangi, Microsoft Research Shruti Palaskar, Carnegie Mellon University Vikas Raunak, Carnegie Mellon University Arka Sadhu, University of Southern California Alok Singh, National Institute of Technology, Silchar India Alane Suhr, Cornell University Hao Tan, University of North Carolina Xiangru Tang, University of the Chinese Academy of Sciences, China Ece Takmaz, University of Amsterdam

Invited Speaker:

Jacob Andreas, MIT Jason Baldridge, Google Mohit Bansal, UNC Chapel Hill Yonatan Bisk, Carnegie Mellon University Joyce Y. Chai, University of Michigan Yejin Choi, University of Washington Raymond J. Mooney, University of Texas at Austin Anna Rohrbach, UC Berkeley Kate Saenko, Boston University William Wang, UC Santa Barbara

Table of Contents

<i>Feature-level Incongruence Reduction for Multimodal Translation</i> Zhifeng Li, Yu Hong, Yuchen Pan, Jian Tang, Jianmin Yao and Guodong Zhou	1
<i>Error Causal inference for Multi-Fusion models</i> Chengxi Li and Brent Harrison1	.1
Leveraging Partial Dependency Trees to Control Image Captions Wenjie Zhong and Yusuke Miyao1	.6
Grounding Plural Phrases: Countering Evaluation Biases by Individuation Julia Suter, Letitia Parcalabescu and Anette Frank2	22
PanGEA: The Panoramic Graph Environment Annotation Toolkit Alexander Ku, Peter Anderson, Jordi Pont Tuset and Jason Baldridge	29
Learning to Learn Semantic Factors in Heterogeneous Image Classification Boyue Fan and Zhenting Liu	54
Reference and coreference in situated dialogue Sharid Loáiciga, Simon Dobnik and David Schlangen	;9