

# **Second Workshop on Advances in Language and Vision Research 2021**

Online  
11 June 2021

ISBN: 978-1-7138-3014-6

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2021) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## 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 Harrison . . . . .	11
<i>Leveraging Partial Dependency Trees to Control Image Captions</i>	
Wenjie Zhong and Yusuke Miyao . . . . .	16
<i>Grounding Plural Phrases: Countering Evaluation Biases by Individuation</i>	
Julia Suter, Letitia Parcalabescu and Anette Frank . . . . .	22
<i>PanGEA: The Panoramic Graph Environment Annotation Toolkit</i>	
Alexander Ku, Peter Anderson, Jordi Pont Tuset and Jason Baldridge . . . . .	29
<i>Learning to Learn Semantic Factors in Heterogeneous Image Classification</i>	
Boyue Fan and Zhenting Liu . . . . .	34
<i>Reference and coreference in situated dialogue</i>	
Sharid Loáiciga, Simon Dobnik and David Schlangen . . . . .	39
<i>Language-based Video Editing via Multi-Modal Multi-Level Transformer</i>	
Tsu-Jui Fu, Xin Wang, Scott Grafton, Miguel Eckstein and William Yang Wang . . . . .	45
<i>Pathdreamer: A World Model for Indoor Navigation</i>	
Jing Yu Koh, Honglak Lee, Yinfei Yang, Jason Baldridge and Peter Anderson . . . . .	57