ACL 2016

# 5th Workshop on Vision and Language
# (VL'16)

## Proceedings of the Workshop

August 12, 2016

This workshop is supported by ICT COST Action IC1307, the European Network on Integrating Vision and Language (iV&L Net): Combining Computer Vision and Language Processing For Advanced Search, Retrieval, Annotation and Description of Visual Data.

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 978-1-945626-11-1

# Introduction

The Fifth Workshop on Vision and Language 2016 (VL'16) took place in Berlin on the 12th August 2016, as part of ACL'16. The workshop is organised by the European Network on Integrating Vision and Language which is funded as a European COST Action. The VL workshops have the general aims: 1. to provide a forum for reporting and discussing planned, ongoing and completed research that involves both language and vision; and 2. to enable NLP and computer vision researchers to meet, exchange ideas, expertise and technology, and form new research partnerships.

The call for papers for VL'16 elicited a good number of submissions, each of which was peer-reviewed by three members of the programme committee. The interest in the workshop from leading NLP and computer vision researchers and the quality of submissions was high, so we aimed to be as inclusive as possible within the practical constraints of the workshop. In the end, we accepted five submissions as long papers, and eight as short papers. The resulting workshop programme packed a lot of exciting content into one day. We were delighted to be able to include in the programme a keynote presentation by Yejin Choi, University of Washington.

We would like to thank all the people who have contributed to the organisation and delivery of this workshop: the authors who submitted such high quality papers; the programme committee for their prompt and effective reviewing; our keynote speaker; the ACL 2016 organising committee, especially the workshops chairs; the participants in the workshop; and future readers of these proceedings for your shared interest in this exciting new area of research.

*August 2016,*  
*Anja Belz, Erkut Erdem, Krystian Mikolajczyk and Katerina Pastra*

**Organizers:**

Anya Belz, University of Brighton, UK
Erkut Erdem, Hacettepe University, Turkey
Krystian Mikolajczyk, Imperial College London, UK
Katerina Pastra, Cognitive Systems Research Institute, Greece

**Program Committee:**

Yannis Aloimonos, University of Maryland, US
Marco Baroni, University of Trento, Italy
Raffaella Bernardi, University of Trento, Italy
Ruken Cakici, Middle East Technical University, Turkey
Luisa Coheur, University of Lisbon, Portugal
Pinar Duygulu Sah'n, Hacettepe University, Turkey
Desmond Elliott, University of Amsterdam, Netherlands
Aykut Erdem, Hacettepe University, Turkey
Jordi Gonzalez, Autonomous University of Barcelona, Spain
Lewis Griffin, UCL, UK
David Hogg, University of Leeds, UK
Nazli Ikizler-Cinbis, Hacettepe University, Turkey
John Kelleher, UCD, Ireland
Frank Keller, University of Edinburgh, UK
Mirella Lapata, University of Edinburgh, UK
Fei Fei Li, Stanford University, US
Margaret Mitchell, Microsoft Research, US
Sien Moens, University of Leuven, Belgium
Francesc Moreno-Noguer, CSIC-UPC, Spain
Adrian Muscat, University of Malta, Malta
Ram Nevatia, University of Southern California, US
Barbara Plank, CST, University of Copenhagen, Denmark
Arnau Ramisa, INRIA Rhone-Alpes, France
Richard Socher, MetaMind Inc, US
Tinne Tuytelaars, University of Leuven, Belgium
Josiah Wang, University of Sheffield, UK
Fei Yan, University of Surrey, UK

**Invited Speaker:**

Yejin Choi, University of Washington, US

# Table of Contents

# Workshop Program

**Friday, August 12, 2016**

9:00–9:30 *Opening Remarks*

9:30–10:00 *Automatic Annotation of Structured Facts in Images*
Mohamed Elhoseiny, Scott Cohen, Walter Chang, Brian Price and Ahmed Elgammal

10:00–10:30 *Combining Lexical and Spatial Knowledge to Predict Spatial Relations between Objects in Images*
Manuela Hürlimann and Johan Bos

10:30–11:00 *Coffee Break*

**11:00–12:00** ***Invited talk: Yejin Choi***

12:00–12:30 *Focused Evaluation for Image Description with Binary Forced-Choice Tasks*
Micah Hodosh and Julia Hockenmaier

12:30–14:00 *Lunch Break*

14:00–14:30 *Leveraging Captions in the Wild to Improve Object Detection*
Mert Kilickaya, Nazli Ikizler-Cinbis, Erkut Erdem and Aykut Erdem

**14:30–15:30** ***Quick-fire presentations for posters (5mins each)***

15:30–16:00 *Coffee Break*

16:00–17:30 *Poster Session*

*Natural Language Descriptions of Human Activities Scenes: Corpus Generation and Analysis*
Nouf Alharbi and Yoshihiko Gotoh

*Interactively Learning Visually Grounded Word Meanings from a Human Tutor*
Yanchao Yu, Arash Eshghi and Oliver Lemon

**Friday, August 12, 2016 (continued)**

*Pragmatic Factors in Image Description: The Case of Negations*
Emiel van Miltenburg, Roser Morante and Desmond Elliott

*Building a Bagpipe with a Bag and a Pipe: Exploring Conceptual Combination in Vision*
Sandro Pezzelle, Ravi Shekhar and Raffaella Bernardi

*Exploring Different Preposition Sets, Models and Feature Sets in Automatic Generation of Spatial Image Descriptions*
Anja Belz, Adrian Muscat and Brandon Birmingham

*Multi30K: Multilingual English-German Image Descriptions*
Desmond Elliott, Stella Frank, Khalil Sima'an and Lucia Specia

*"Look, some Green Circles!": Learning to Quantify from Images*
Ionut Sorodoc, Angeliki Lazaridou, Gemma Boleda, Aurélie Herbelot, Sandro Pezzelle and Raffaella Bernardi

*Text2voronoi: An Image-driven Approach to Differential Diagnosis*
Alexander Mehler, Tolga Uslu and Wahed Hemati

*Detecting Visually Relevant Sentences for Fine-Grained Classification*
Olivia Winn, Madhavan Kavanur Kidambi and Smaranda Muresan