VL 2017

The 6th Workshop on Vision and Language

Proceedings of the Workshop

April 4, 2017
Valencia, Spain

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.

©2017 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 978-1-945626-51-7

Introduction

The 6th Workshop on Vision and Language 2017 (VL'17) took place in Valencia as part of EACL'17. 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: to provide a forum for reporting and discussing planned, ongoing and completed research that involves both language and vision; and to enable NLP and computer vision researchers to meet, exchange ideas, expertise and technology, and form new research partnerships.

Research involving both language and vision computing spans a variety of disciplines and applications, and goes back a number of decades. In a recent scene shift, the big data era has thrown up a multitude of tasks in which vision and language are inherently linked. The explosive growth of visual and textual data, both online and in private repositories by diverse institutions and companies, has led to urgent requirements in terms of search, processing and management of digital content. Solutions for providing access to or mining such data effectively depend on the connection between visual and textual content being made interpretable, hence on the 'semantic gap' between vision and language being bridged.

One perspective has been integrated modelling of language and vision, with approaches located at different points between the structured, cognitive modelling end of the spectrum, and the unsupervised machine learning end, with state-of-the-art results in many areas currently being produced at the latter end, in particular by deep learning approaches.

Another perspective is exploring how knowledge about language can help with predominantly visual tasks, and vice versa. Visual interpretation can be aided by text associated with images/videos and knowledge about the world learned from language. On the NLP side, images can help ground language in the physical world, allowing us to develop models for semantics. Words and pictures are often naturally linked online and in the real world, and each modality can provide reinforcing information to aid the other.

We would like to thank all the people who have contributed to the organisation and delivery of this workshop: the authors who submitted high quality papers; the programme committee for their prompt and effective reviewing; our keynote speakers; the ACL 2017 organising committee, and future readers of these proceedings for your shared interest in this exciting area of research.

April 2017,
Anja Belz, Erkut Erdem, Katerina Pastra and Krystian Mikolajczyk

Organizers:

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

Program Committee:

Raffaella Bernardi, University of Trento, Italy
Darren Cosker, University of Bath, UK
Aykut Erdem, Hacettepe University, Turkey
Jacob Goldberger, Bar Ilan University, Israel
Jordi Gonzalez, CVC UAB Barcelona, Spain
Douwe Kiela, University of Cambridge, UK
Arnau Ramisa, IRI UPC Barcelona, Spain
Josiah Wang, University of Sheffield, UK

Table of Contents

The BURCHAK corpus: a Challenge Data Set for Interactive Learning of Visually Grounded Word Meanings
Yanchao Yu, Arash Eshghi, Gregory Mills and Oliver Lemon 1

The Use of Object Labels and Spatial Prepositions as Keywords in a Web-Retrieval-Based Image Caption Generation System
Brandon Birmingham and Adrian Muscat 11

Learning to Recognize Animals by Watching Documentaries: Using Subtitles as Weak Supervision
Aparna Nurani Venkitasubramanian, Tinne Tuytelaars and Marie-Francine Moens 21

Human Evaluation of Multi-modal Neural Machine Translation: A Case-Study on E-Commerce Listing Titles
Iacer Calixto, Daniel Stein, Evgeny Matusov, Sheila Castilho and Andy Way 31

The BreakingNews Dataset
Arnau Ramisa, Fei Yan, Francesc Moreno-Noguer and Krystian Mikolajczyk 38

Automatic identification of head movements in video-recorded conversations: can words help?
Patrizia Paggio, Costanza Navarretta and Bart Jongejan 40

Multi-Modal Fashion Product Retrieval
Antonio Rubio Romano, LongLong Yu, Edgar Simo-Serra and Francesc Moreno-Noguer 43

Workshop Program

Tuesday, 4 April, 2017

9:15–9:30 ***Welcome and Opening Remarks***

9:30–10:30 *Invited Talk*
Prof. David Hogg

10:30–11:00 *The BURCHAK corpus: a Challenge Data Set for Interactive Learning of Visually Grounded Word Meanings*
Yanchao Yu, Arash Eshghi, Gregory Mills and Oliver Lemon

11:00–11:30 ***Coffee Break***

11:30–12:00 *The Use of Object Labels and Spatial Prepositions as Keywords in a Web-Retrieval-Based Image Caption Generation System*
Brandon Birmingham and Adrian Muscat

12:00–12:30 *Learning to Recognize Animals by Watching Documentaries: Using Subtitles as Weak Supervision*
Aparna Nurani Venkitasubramanian, Tinne Tuytelaars and Marie-Francine Moens

12:30–13:00 *Human Evaluation of Multi-modal Neural Machine Translation: A Case-Study on E-Commerce Listing Titles*
Iacer Calixto, Daniel Stein, Evgeny Matusov, Sheila Castilho and Andy Way

13:00–15:00 ***Lunch Break***

15:00–16:00 ***Poster Session including Quick Poster Presentations***

15:30–15:40 *The BreakingNews Dataset*
Arnau Ramisa, Fei Yan, Francesc Moreno-Noguer and Krystian Mikolajczyk

15:40–15:50 *Automatic identification of head movements in video-recorded conversations: can words help?*
Patrizia Paggio, Costanza Navarretta and Bart Jongejan

15:50–16:00 *Multi-Modal Fashion Product Retrieval*
Antonio Rubio Romano, LongLong Yu, Edgar Simo-Serra and Francesc Moreno-Noguer

16:00–16:30 ***Coffee Break***

Tuesday, 4 April, 2017 (continued)

16:30–17:30 *Invited Talk*
Prof. Mirella Lapata