# LANTERN 2021

Proceedings of the Third Workshop on Beyond Vision and LANguage: inTEgrating Real-world kNowledge (LANTERN)

> April 20, 2021 Kyiv, Ukraine (Online)



©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 978-1-954085-15-2

## Introduction

Welcome to LANTERN 2021, the Third Workshop Beyond Vision and Language: Integrating Real-World Knowledge, co-located with EACL 2021. Building on the successes of the first edition co-located with EMNLP-IJCNLP 2019 and the second edition co-located with COLING 2020, the third edition of LANTERN aims to again bring together and interconnect researchers focusing on natural language research from a multimodal perspective. In particular, the main goal of the workshop is to promote and foster research which uses machine- and deep-learning techniques to interconnect language with vision and other modalities by leveraging external knowledge. We encourage contributions which exploit very diverse sources of external knowledge like knowledge graphs, fixed and dynamic environments, cognitive and neuroscience data, etc. This results in the workshop being open to all research directions which acknowledge the importance of knowledge in acquiring, using, and evaluating language in real-world settings.

In this third edition, we called for both long and short papers. All the accepted contributions are published in these Proceedings. LANTERN 2021 received 11 submissions, all of which were double-blindly reviewed by three highly-qualified reviewers. In total, 5 papers (4 long, 1 short) were accepted to appear in the Proceedings of the workshop, with an acceptance rate of around 55% (comparable to that of the first and second edition, which was around 53% and 57%, respectively).

Contributions are representative of a broad range of current problems and approaches and include among others a novel approach to image captioning, an approach for zero-shot image classification leveraging external knowledge from Wikipedia as well as a new approach for injecting structured knowledge into vision-and-language transformer models. Such richness of approaches and perspectives is in line with the purpose of the workshop, and confirms the growing interest for the integration of external knowledge for language and vision problems.

The program of the workshop, besides 5 oral paper presentations, includes invited talks by Aida Nematzadeh, Brenden Lake, Danna Gurari, and Sebastian Riedel. The workshop received sponsorship by iDeaL SFB 1102 and HuggingFace, which sponsor a best paper award.

The LANTERN Workshop Organizers

#### **Organizers:**

Marius Mosbach, Saarland University (Germany) Michael A. Hedderich, Saarland University (Germany) Sandro Pezzelle, University of Amsterdam (The Netherlands) Dietrich Klakow, Saarland University (Germany) Marie-Francine Moens, Katholieke Universiteit Leuven (Belgium) Zeynep Akata, University of Tübingen (Germany)

#### **Program Committee:**

Afra Alishahi, Tilburg University Alane Suhr, Cornell University Alessandro Suglia, Heriot-Watt University Ashutosh Modi, IIT Kanpur Carina Silberer, University of Stuttgart Claudio Greco, University of Stuttgart Claudio Greco, University of Trento David Schlangen, University of Potsdam Hao Tan, University of North Carolina Iacer Calixto, UvA/NYU Jacob Goldberger, Bar Ilan Miaoran Zhang, Saarland University Ravi Shekhar, Queen Mary University London Sina Zarriess, University of Jena Somak Aditya, Microsoft Research India Stanislav Frolov, DFKI

#### **Invited Speaker:**

Aida Nematzadeh, DeepMind Brenden Lake, NYU & Facebook AI Research Danna Gurari, University of Texas at Austin Sebastian Riedel, UCL & Facebook AI Research

## **Table of Contents**

Reasoning over Vision and Language: Exploring the Benefits of Supplemental Knowledge Violetta Shevchenko, Damien Teney, Anthony Dick and Anton van den Hengel
Visual Grounding Strategies for Text-Only Natural Language Processing Damien Sileo
<i>Exploiting Image–Text Synergy for Contextual Image Captioning</i> Sreyasi Nag Chowdhury, Rajarshi Bhowmik, Hareesh Ravi, Gerard de Melo, Simon Razniewski and Gerhard Weikum
Large-Scale Zero-Shot Image Classification from Rich and Diverse Textual Descriptions Sebastian Bujwid and Josephine Sullivan
What Did This Castle Look like before? Exploring Referential Relations in Naturally Occurring Multi- modal Texts Ronja Utescher and Sina Zarrieß53

### **Conference Program**

The program of LANTERN 2021 consists of 4 keynotes by invited speakers and 5 contributed talks. The contributed talks will be followed by a virtual break-out session in Gather town. The full program of LANTERN 2021 can be found at https://www.lantern.uni-saarland.de/2021/.

The five accepted papers are:

Reasoning over Vision and Language: Exploring the Benefits of Supplemental Knowledge

Violetta Shevchenko, Damien Teney, Anthony Dick and Anton van den Hengel

Visual Grounding Strategies for Text-Only Natural Language Processing Damien Sileo

Exploiting Image–Text Synergy for Contextual Image Captioning

Sreyasi Nag Chowdhury, Rajarshi Bhowmik, Hareesh Ravi, Gerard de Melo, Simon Razniewski and Gerhard Weikum

Large-Scale Zero-Shot Image Classification from Rich and Diverse Textual Descriptions Sebastian Bujwid and Josephine Sullivan

What Did This Castle Look like before? Exploring Referential Relations in Naturally Occurring Multimodal Texts Ronja Utescher and Sina Zarrieß