W21 – Shortcomings in Vision and Language

W21 – Shortcomings in Vision and Language

Shortcomings in Vision and Language (SiVL) was held on the 8th of September at ECCV in Munich. The workshop brought together experts at the intersection of vision and language to discuss modern approaches, tasks, datasets, and evaluation metrics for significant problems on the integration of these two modalities. The aim of the workshop was to facilitate discussion of novel research directions and to steer the community towards high-level challenges affecting the vision and language community broadly.

Inspiring talks were given by the invited speakers. Not surprisingly, Neural Networks dominated the scene, but interestingly the popular end-to-end one-big-black-box architecture has, in many cases, been replaced by a more modular one. On the one hand, traditional Computational Linguistics components – such as Part-of-Speech tags, question type detection etc. – have found their role in the NN architecture, and the importance to carry out qualitative analysis instead of only quantitative comparison of models' performance has been highlighted (Aishwarya Agrawal). On the other hand, traditional Computer Vision components – such as object localization – have been put back at work into end-to-end structures, showing the need for vision and language systems to focus on entities (Lucia Specia). Other interesting issues that emerged are the need to develop models that are not task specific and furthermore are able to deal with dataset bias (Vicente Ordóñez Román.) An example of the social impact of this research line has been shown by Danna Gurari, who presented her project to develop models to assist blind people. The workshop received 21 valid full-paper submissions. They were subjected to double-blind peer-review by at least two experts in vision and language. We also received 22 abstract submissions, which were selected by the workshop organizers based on topical relevance. In total, 10 full-papers and 19 abstracts were accepted to appear at the workshop. The workshop featured three poster sessions and one spotlight session where the authors of the accepted full-papers were given a chance to showcase their work in 4-minute spotlight talks.

The workshop also hosted the organizers of the 1st Visual Dialog Challenge (visualdialog.org/challenge/2018), who introduced a new robust evaluation metric for Visual Dialog and gave a detailed quantitative and qualitative overview of the systems that participated in the competition. Complete slides about the challenge are available at: https://goo.gl/SRkBEk. The workshop was sponsored by SAP (Moin Nabi) and was organized in collaboration with researchers from University of Amsterdam (Raquel Fernández), University of Edinburgh (Spandana Gella), University of Trento (Raffaella Bernardi), Georgia Institute of Technology (Dhruv Batra and Stefan Lee) and Rochester Institute of Technology (Kushal Kafle). Organizers in alphabetical order.

Dhruv Batra
Raffaella Bernardi
Raquel Fernández
Spandana Gella
Kushal Kafle
Stefan Lee
Moin Nabi