

**W22 – 2nd YouTube-8M Large-Scale  
Video Understanding Workshop**

# W22 – 2nd YouTube-8M Large-Scale Video Understanding Workshop

The 2nd YouTube-8M Large-Scale Video Understanding Workshop was hosted at ECCV'18 by the Video Understanding Team at Google AI Perception. This workshop is based on the YouTube-8M video classification challenge hosted by Kaggle, composed of 4 invited talks, presentations by top performers at the competition, and a poster session for deeper discussion. We ran two tracks:

- **Classification Challenge Track** was organized as a Kaggle competition for large-scale video classification based on the YouTube-8M dataset. Researchers were invited to participate in the classification challenge by training a model on the public YouTube-8M training and validation sets and submitting video classification results on a blind test set. Top-ranking submissions in the challenge leaderboard were invited to the workshop to present their method.
- **General Paper Track** invited researchers to submit any papers involving research, experimentation, or applications on the YouTube-8M dataset. We encouraged participants to explore any relevant topics of interest using YouTube-8M dataset; for example, large-scale video recommendations, search, and discovery, joining the dataset with other publicly available resources (e.g, video metadata), or data visualization.

Paper submissions were reviewed by the workshop organizers and accepted papers were invited for oral or poster presentations at the workshop. Submissions were evaluated in terms of potential impact (e.g. performance on the classification challenge), technical depth, scalability, novelty, and presentation. At least two reviewers were assigned to review each paper, and two program chairs made the final decision. We did not require blind submissions. We did not restrict submissions of relevant work that is under review or will be published elsewhere. Previously published work was also acceptable as long as it was re-targeted towards YouTube-8M.

We hosted four invited talks:

- **Human action recognition and the Kinetics dataset** by Andrew Zisserman, *Oxford University & Google DeepMind*
- **Segmental Spatio-Temporal Networks for Discovering the Language of Surgery** by Rene Vidal, *Johns Hopkins University*
- **Learning video representations for physical interactions and language-based retrieval** by Josef Sivic, *Inria and Czech Technical University*
- **Towards Video Understanding at Scale** by Manohar Paluri, *Facebook*

We would like to express our gratitude to all our colleagues for submitting papers to the 2nd YouTube-8M Large-Scale Video Understanding Workshop, as well as to the members of the Program Committee for organizing this year's attractive program.

September 2018

Joonseok Lee  
George Toderici  
Paul Natsev  
Rahul Sukthankar