Call for Submissions

The How2 Challenge - New Tasks for Vision and Language

Research at the intersection of vision and language has attracted an increasing amount of attention over the last ten years. Current topics include the study of multi-modal representations, translation between modalities, bootstrapping of labels from one modality into another, visually-grounded question answering, embodied question-answering, segmentation and storytelling, and grounding the meaning of language in visual data. Still, these tasks may not be sufficient to fully exploit the potential of vision and language data.

To support research in this area, we recently released the How2 data-set, containing 2000 hours of how-to instructional videos, with audio, subtitles, Brazilian Portuguese translations, and textual summaries, making it an ideal resource to bring together researchers working on different aspects of multimodal learning. We hope that a common dataset will facilitate comparisons of tools and algorithms, and foster collaboration.

We are organizing a workshop, “The How2 Challenge - New Tasks for Vision and Language” at ICML 2019, to bring together researchers and foster the exchange of ideas in this area. We seek submissions in the following two categories:

- Papers that describe work on the How2 data, either on the “shared challenge tasks”, e.g. multi-modal speech recognition, machine translation, or video summarization challenge, described on the How2 web site (a leader-board will be provided), or creating “un-shared”, novel tasks.
- Papers that describe other related and relevant work to further vision and language ideas by proposing new tasks, or analyzing the utility of existing tasks and data sets in interesting ways

The organizers encourage both the publication of novel work that is relevant to the topics of discussion, and late-breaking results on the How2 tasks in a single format. The workshop will also feature a number of invited talks, and a moderated discussion around the challenges and opportunities that current tasks in vision and language present. We aim to stimulate discussion around new tasks that go beyond image captioning and visual question answering, and which could form the basis for future research in this area. We seek to create a venue to encourage collaboration between different sub-fields and help establish new research directions that we believe will sustain multimodal machine learning research for years to come.

Invited speakers:

- Katerina Fragkiadaki (Carnegie Mellon University)

1 https://github.com/srvk/how2-dataset
2 https://srvk.github.io/how2-challenge/
- Lisa Anne Hendricks (UC Berkeley)
- Qin Jin (Renmin University)
- Angeliki Lazaridou (DeepMind)
- Devi Parikh (Georgia Tech)
- Kate Saenko (Boston University)
- Bernt Schiele (Max Plank Institute for Informatics)

**Important dates:**

- Challenge starts: March 15, 2019
- Paper submission: May 15, 2019
- Notification: May 22, 2019
- Workshop date: June 14 or 15, 2019

For more information, visit [https://srvk.github.io/how2-challenge/](https://srvk.github.io/how2-challenge/)

Contact us at [how2challenge@gmail.com](mailto:how2challenge@gmail.com).
Research at the intersection of vision and language has been attracting a lot of attention in recent years. Topics include the study of multi-modal representations, translation between modalities, bootstrapping of labels from one modality into another, visually-grounded question answering, segmentation and storytelling, and grounding the meaning of language in visual data. An ever-increasing number of tasks and datasets are appearing around this recently-established field.

At NeurIPS 2018, we released the How2 data-set, containing more than 60,000 (2000h) videos, with audio, transcriptions, translations, and textual summaries. We believe it presents an ideal resource to bring together researchers working on the previously mentioned separate tasks around a single, large dataset. This rich dataset will facilitate the comparison of tools and algorithms, and hopefully foster the creation of additional annotations and tasks. We want to foster discussion about useful tasks, metrics, and labeling techniques, in order to develop a better understanding of the role and value of multi-modality in vision and language. We seek to create a venue to encourage collaboration between different sub-fields, and help establish new research directions and collaborations that we believe will sustain machine learning research for years to come.

- The usual suspects might work (DE: can we do better than the usual suspects? Are there any speakers that you don’t usually see but we might want to see.)
  - Bernt Schiele
  - Trevor Darrel, Kate Saenko, Markus Rohrbach, Jeff Donahue,
  - Devi Parikh, Dhruv Patra.
  - SP: Richard Socher, Caiming Xiong (Salesforce)
  - SP: Yejin Choi, Zach Lipton, Jim Glass, Herman Kamper
  - RS: Antonio Torralba, Tali Dekel, David Harwath.
  - DE: Jamie Ryan Kiros

Yes, we should do better than the usual suspects, I agree. We asked:
- Lisa Anne Hendricks (ok)
- Qin Jin (ok)
- Devi Parikh (ok)
- Katerina Fragkiadaki (ok)
- Bernt Schiele (ok)
- Angeliki Lazaridou (ok)
- Kate Saenko (ok)
- Jamie Kiros (no response)
- Herman Kamper (no response)
- Yejin Choi (declined)
- Erkut Erdem (have not asked yet)

This is from https://icml.cc/Conferences/2019/CallForWorkshops

Below, we include the criteria by which workshop submissions will be evaluated:
1. Fit for ICML (connections to rest of program and past workshops)
2. Potential impact (promising topic)
3. Novelty and originality (emerging topic)
4. Quality of the abstract & clarity of purpose
5. Confirmed invited speakers with sufficient coverage of the topic
6. Organizers’ relevant expertise (please avoid excessive self-promotion)
7. Organizers & speakers diversity
8. Room for contributed work (posters and contributed talks)
9. Open discussion/panel

- our requirements are:
  ● Workshop title
  ● A brief description of the topics to be covered, and an explanation as to why the workshop will appeal to ICML audiences
  ● A short description and rough timetable of the planned activities (talks, posters, panels)
  ● List of invited speakers, specifying who is confirmed and who is unconfirmed.
  ● A description of the history of the workshop (if it previously took place then when/where)
  ● Similar past and current events at ICML and NeurIPS in the last 1-2 years, even if not organised by the present workshop organisers. New workshops are welcome to build on prior workshops, if a good case is made; completely original workshops are also welcome.
  ● A list of organizers with email addresses, web page URLs, pointers to Google Scholar or other similar citation service pages, and a one-paragraph bio for each organizer, describing research expertise, and previous experience organizing scientific meetings.

How2 Dataset

Email: how2challenge@gmail.com [password: groundeds2s]

TODO:

Mar 5

1. Elaborate on distribution list
   a. Corpora-List - Desmond
   b. Speech-List - Florian
   c. Computer Vision-List - Lucia
   d. Dagstuhl-List - Florian / Lucia  
      seminar-19021@dagstuhl.de or seminar-19021@portal.dagstuhl.de
   e. Twitter!
2. Update call and website by Thursday / Friday

Feb 26
- create website -> under https://github.com/srvk/how2-dataset
  - Host the proceedings on the website
  - Create CodaLab interface for competition? Kaggle?
    - Add baseline results for the initial tasks in the leaderboard
- use easychair for reviewing? (soft reviewing, mainly about the form)
- Ask ICML if they will keep track of who registers for our conference
- Promote the workshop!
  - Decide date: June 14th and/or 15th
  - Final decision : DATE - 3/4 weeks => ~May 22nd
  - DL for paper submission: DATE - 4/5 weeks => ~ May 15th
  - Publish call: March 15th BY THE LATEST? Ok tomorrow then ;) OK RAMON, LET’S GO
    - Corpora-List, Speech-centric lists, Dagstuhlers, Everyone who downloaded the How2 corpus
    - Twitter, Facebook
- Aims of the workshop:
  - Add to the current tasks leaderboard (+ mark constrained / unconstrained)
  - Suggest new tasks + baseline results


Commented [1]: announce that DL are 1 week earlier so that we can extend them for 1 more week afterwards. This has the advantage to have information on the number of submissions 1 week earlier