The How2Sign Dataset*

Amanda Duarte, Shruti Palaskar, Lucas Ventura, Deepti Ghadiyaram, Kenneth DeHaan
Florian Metze, Jordi Torres and Xavier Giro-i-Nieto

CVPR VIRTUAL JUNE 14-19 · GALLAUDET UNIVERSITY · BSC · UPC · Carnegie Mellon University

How2Sign

Download

https://github.com/how2sign

Highlights

- The first large-scale continuous American Sign Language dataset.
- More than **80 hours** of multimodal and multiview videos of ASL with sentence-level alignment for more than **35k sentences**.
- A rich set of annotations including **gloss**, **category labels**, and automatically extracted **2D keypoints**.
- Contain a 3-hour subset recorded in the Panoptic studio[2] with 500+ cameras enabling high-quality 3D keypoints estimation.
- We conduct a study with ASL signers that gave insights on challenges that can be addressed in the field of SL.

Statistics

	Green Screen Studio	Panoptic Studio [2]
ASL videos	2,529	124
Duration (h)	79.12	2.96
English sentences	35,191	1,582
Vocabulary size	16,609	3,260
Body Pose	2D	3D
Camera views	3 HD + 1 RGB-D	520 views
# signers	9	6

Please refer to Table 2 of the paper for complete statistics information.

The Dataset

How2 Dataset [1]

Instructional videos

Speech signal

English Transcription

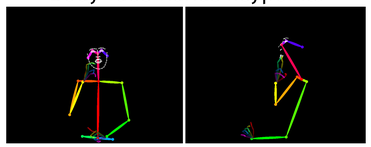
Hi, I'm Amelia and I'm going to talk to you about how to remove gum from hair.

How2Sign Dataset

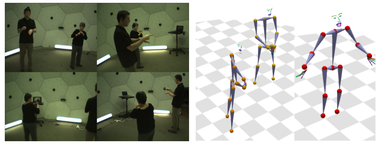
Green Screen Studio RGB videos

Green Screen studio RGB-D videos

Body-face-hands keypoints



Multi-view studio data[2] (for a subset)



Gloss Annotations

HI, ME FS-AMELIA WILL ME TALK GUM IX-LOC-HAIR STUCK

How to use?

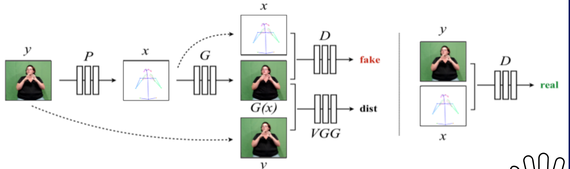
- ASL translation (from ASL to English);
- ASL production (from English to ASL videos);
- ASL recognition;
- Sign segmentation;
- Video topic classification;
- End-to-end English speech to ASL;
- 3D pose estimation/reconstruction;
- and more...

Application

Synthesizing sign language videos

Source video · 2D keypoints · Generated video

1. Video Generation



2. Training a style transfer GAN [3]

Acknowledgments

"la Caixa" Foundation · Supported by facebook research · Excelencia Severo Ochoa · Gobierno de España Ministerio de Economía y Competitividad · Generalitat de Catalunya

References

[1] Sanabria, Ramon, et al. "How2: a large-scale dataset for multimodal language understanding." arXiv preprint arXiv:1811.00347 (2018).
[2] Joo, Hanbyul, et al. "Panoptic studio: A massively multiview system for social motion capture." In ICCV 2015.
[3] Chan, Caroline, et al. "Everybody dance now." In ICCV 2019.

***How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language**