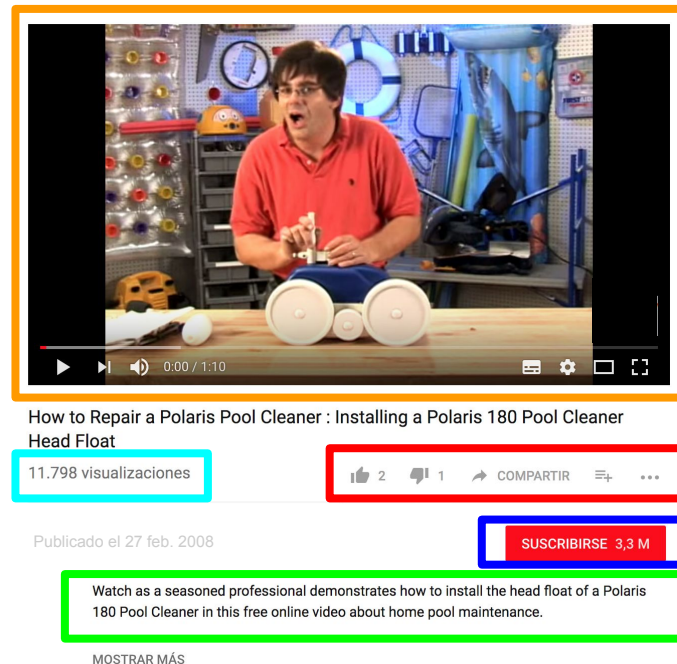


Dataset

- 2000h of **how-to** videos (Yu et al., 2014)
 - 300h for MT, 480h for ASR (as of today)
 - Shared splits, held-out data
- Ground truth captions
- Metadata
 - Number of likes / dislikes
 - Visualizations
 - Uploader, Date
 - Tags
- Video descriptions (“summaries”)
 - 80K descriptions for 2000h
- Very different topics
 - Cooking, fixing things, playing instruments, etc.
- 300,000 segments translated into Portuguese



GROUND



