Motivation

- Exponential and constant growth of videos online.
- Unfeasible to manually label these data.
- Automatic methods for large-scale annotation are needed.
- New concepts are constantly appearing.
- Relations between concepts are change over time.
- VITS learns and updates concept relations on real time.

VITS Pipeline

- Automatic large scale video tagging system - video summarization.
- Vocabulary of over 2.5M Knowledge Graph concepts.
- Very specific multilingual tags.
- Indexing more than 150k videos per month.
- Updates concept relations based on social media events.

Knowledge Graph

- Concepts are universal semantic representation of worlds.
- 2M concepts vocabulary.
- Multilingual aspect.
- Freebase / Google Knowledge Graph + Wikidata [3]
- R = Concept relations matrix.
- The Knowledge Graph is useful for...
  - Synchronous keywords association.
  - Multilingual tagging.
  - Concept disambiguation.

Contextual Tagging

1. Keyword Extraction from Internet Crawling:
   - Download all related text sources.
   - Extract keywords with NLP techniques.
2. Keyword to Entity Mapping in the Knowledge Graph:
   - Get keyword concept candidates.
   - For each concept compute a Concept Score.
   - Concept Score.
     - Intra-score: usability, word similarity, year filtering, type filtering.
     - Inter-score: context applied with R matrix.

Experiments

- Quality of VITS tags?
- Subset from the YouTube-8M Dataset [1]
- Multilingual videos.
- Wide number of topics.
- Knowledge Graph labels.
- Contextual information.

Results

1. Tagging Statistics
   - Average number of tags in YouTube-8M subset is 3.64 tags/video VS 10.04 tags/video extracted by VITS.

2. Human Rating of Generated Tags
   - Top 40 most repeated tags in YouTube-8M Subset.
   - Top 40 most repeated tags extracted by VITS.
   - Top-40 tags evaluated as incorrect.
   - Top-40 tags discarded.

References