Noise adaptive weighting
training videos may be unrelated → we compute similarity between video frames and example images and use normalized similarity score to weight contribution of the video in the loss.

**METHOD**

**CHANGEIT DATASET**

- **Labelling step**
  - compute labels for a video using the state and action classifiers:
    - Look for the Change: Learning Object States and State-Modifying Actions from Untrimmed Web Videos
    - Tomáš Souček¹, Jean-Baptiste Alayrac², Antoine Miech², Ivan Laptev³, Josef Sivic¹
    - ¹Czech Technical University ²DeepMind ³ENS/INRIA

- **Contributions**
  - Self-supervised model for learning object states and state-modifying actions from long uncurated web videos.
  - Causal ordering signal (initial state → action → end state) is used as the supervision.
  - New uncurated dataset with 2600+ hours of video and 34 thousand changes of object states.

- **Challenges**
  - Visual variability of the objects and its states.
  - Thousands of objects with many more states, annotating is both difficult and expensive.
  - In-the-wild, noisy, uncurated, and long videos.

- **Goal**
  - Temporally localize object states together with the corresponding state-modifying actions in videos.
  - An example: the initial state plain corpus, the end state frosted cake, and the action cake frosting.
  - Use only uncurated videos from the internet with minimal supervision.

**OVERVIEW**

**RESULTS**

- Training on 44 categories of the ChangeIt dataset
- A separate model trained for each category
- A single frame predicted as the initial state, the action and the end state per video

**TEMPORAL LOCALIZATION**

- Gathered by searching Youtube for terms such as "How to cut an apple?"
- 44 state-changing action categories such as apple cutting, ball inflating, etc.
- 34,428 in-the-wild videos, in total 2,642 hours, average video length 4.6 minutes
- 667 videos per-frame annotated with labels: background, initial state, action, end state

**ABLATIONS**

- clear improvement with increasing dataset size
- model improves even when trained on more noisy low-ranking videos

**PROJECT PAGE**

- data.ciirc.cvut.cz/public/projects/2022LookForTheChange

**ACKNOWLEDGEMENTS**