SPFTN: A Self-Paced Fine-Tuning Network for Segmenting Objects in Weakly Labelled Videos

Dingwen Zhang¹, Le Yang¹, Deyu Meng², Dong Xu³ and Junwei Han¹
¹Northwestern Polytechnical University, ²Xian Jiaotong University, ³University of Sydney

**Problem**

*Goal:* learning to perform category-specific video object segmentation by only using video-level tags.

**Challenges:**
- Detecting!
- Associating!
- Recognizing!
- Segmenting!

**Conventional approaches:**
- Decompose positive and negative videos into spatial-temporal segments.
- Train segmentation-level classifiers or inference models under weak supervision.
- Identify the segments related to the given object categories in each video.

**Under studied problems:**
- Unclear how to address this problem via leveraging powerful DNNs.
- Explore scene context in each video frame rather than consider each spatial-temporal segment individually to provide helpful contextual priors.
- Alleviate the learning uncertainty brought by the negative videos due to the lack of principle ways to acquire them.

**Solution:** SPFTN!

**Our Approach**

**Main Idea:**
- Integrate SPL into the DNN learning objective to improve the learning capability of SPL and simultaneously perform weakly supervised training of DNN.
- Use a novel group curriculum self-paced term to encode helpful prior-knowledge.
- Capture object semantics only from positive videos to increase learning stability.
- Encode rich context information to help improve the segmentation accuracy.

**Self-paced Regularizer:**

\[
    f(V, p, k, y, t) = \sum_{t \in T} \sum_{k \in K} \sum_{w \in W} \frac{1}{v_i} \left[ \sum_{r \in R} r(W) + \sum_{s \in S} L(y, v_y, \theta(W)) \right] + \alpha \sum_{t \in T} \sum_{k \in K} \sum_{w \in W} \frac{1}{v_i} \left[ \delta \cdot (s + t) \right] \]

**Experiments**

**Dataset:** YouTube-Object dataset & DAVIS

**Algorithm 2:** The overall approach to apply our SPFTN for object segmentation in weakly labelled videos.

**Learning Objective:**

\[
    
    \min_{W, Y} \sum_{t \in T} \sum_{k \in K} \sum_{w \in W} \frac{1}{v_i} \left[ \sum_{r \in R} r(W) + \sum_{s \in S} L(y, v_y, \theta(W)) \right] + \alpha \sum_{t \in T} \sum_{k \in K} \sum_{w \in W} \frac{1}{v_i} \left[ \delta \cdot (s + t) \right] \]

**State-of-the-arts:**

**Ablation Study:**