Video Object Segmentation Using Adversarial Networks and Mathematical Morphology

Amin FEHRI
MINES ParisTech, PSL Research University, CMM - Center of Mathematical Morphology

Problem

Video object segmentation is a two classes labelling problem desiring to separate foreground objects from the background region of a video.

- The choice of the object to follow across the video is highly dependent of the application.
- Thus this object is often specified by the user on one or a few frames, making the video object segmentation a semi-supervised problem. It can also be seen as a mask propagation problem.

Representation Learning using Adversarial Networks

- Fully ConvNets approach as in SOTA [1, 3].
- Input: RGB + previous mask ; Output: probability map.
- Adversarial networks as a regularizer to label variables (foreground/background) are no more predicted independently from each other (as in [2]).
- Loss minimization w.r.t. \( \theta \), the segmentor model parameters.
- Loss maximization w.r.t. \( \theta_a \), the adversarial model parameters.

Following the object using MM

- Levelings: morphological filtering techniques allowing to filter an image without creating any new contours. We do a leveling of the filtered CNN output using the previous frame segmentation as a reference.
- Morphological reconstruction from one frame to the next one. Safeguard condition: the area of the object must not vary too much.

Ensuring temporal coherency using levelings and morphological reconstruction.

Preliminary results

Perspectives

- More thorough evaluation on reference databases (such as DAVIS [4]).
- Transfer learning to take advantage of powerful existing architectures.
- Learning the more suited morphological operators within the CNN (difficult because MM operators are non-differentiable for the most part).

References