VITAL: Visual Tracking via Adversarial Learning
Yibing Song¹, Chao Ma², Xiaohui Wu³, Lijun Gong⁴, Linchao Bao¹, Wangmeng Zuo⁵, Chunhua Shen⁶, Rynson Lau⁵, and Ming-Hsuan Yang⁶
¹Tencent AI Lab, ²The University of Adelaide, ³Harbin Institute of Technology, ⁴Tencent, ⁵CityU HK, ⁶UC Merced

Source code: https://ybsong00.github.io/cvpr18_tracking/index

Introduction:

- Two-stage tracking-by-detection framework:
  1. Drawing samples around target object.
  2. Classifying each sample as either target object or background.

- Limitations:
  1. Limited positive samples fail to capture rich appearance variations.
  2. Class imbalance between positive and negative samples.

- Our motivations:
  1. We diversify positive samples through adaptively dropout redundant CNN features. Adversarial learning helps our tracker exploit the most robust features over a long temporal span in the classifier training, rather than overfitting to discriminative features in single frames.
  2. We propose cost sensitive loss to decrease the effect of easy negative samples.

Framework:

- Feature extractor
- Adversarial feature generator
- Binary discriminator

Tracking Pipeline:

- Model initialization:
  - Offline pretraining: we train D from scratch in a few iterations and incorporate G for adversarial learning. During each iteration, we first train D and then G.
  - Online finetuning: we online finetune the classifier using samples from the first frame.
- Online detection: we remove G and follow the two-stage tracking-by-detection framework for target localization.
- Model update: we online collect samples and update the model.

Visualization:

- (a) Input Frame #003
- (b) Entropy map without GAN
- (c) Entropy map with GAN
- (d) Input Frame #160
- (e) Entropy map without GAN
- (f) Entropy map with GAN

Experiments:

Evaluations on the OTB 2013 dataset. More evaluations are presented in the paper.