Tracking without bnw (ICCV 2019)

Tracking without bells and whistles

philipp et al.

① Problem Description:

Problem: complex tracking scenarios

namely, small and occluded objects or missing detections.

Solution: tracking-by-detection

produced increasingly complex models

② Problem Solution:

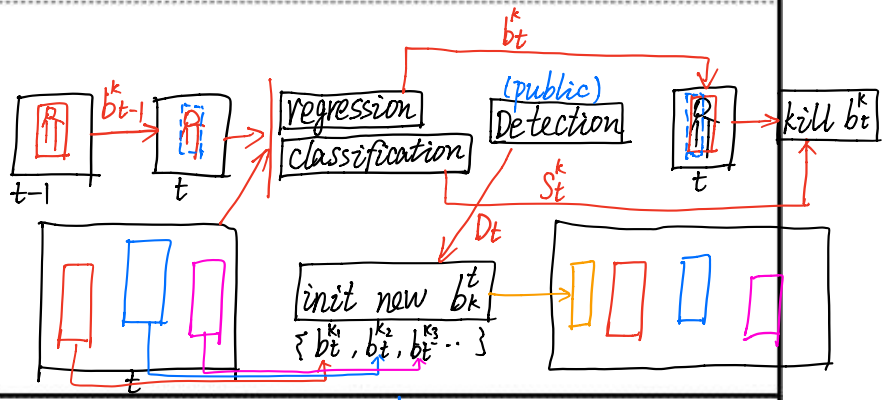
① introduce Tracktor

tackles multi-object tracking by exploiting the regression

② present two simple extensions

Siamese network / motion model

③ Conceptual Understanding



① trajectories: $T_k = \{b_{t_1}^k, b_{t_2}^k, \cdots\}$ $\begin{cases} b_t^k = (x, y, w, h) \\ t: \text{frame}, k: \text{object} \end{cases}$

② frame: $B_t = \{b_t^{k_1}, b_t^{k_2}, \cdots\}$

③ init: $D_0 = \{d_0^1, d_0^2, \cdots\} = B_0$

④ new: $\begin{cases} b_t^k \cap (T_0 \sim T_k) = 0 \\ \Rightarrow b_t^{k+1} \rightarrow T_{k+1} \end{cases}$; kill: $\begin{cases} ① S_t^k < \sigma_{active}, B_t \rightarrow b_t^k \\ ② (NMS)\ IoU < \lambda_{active}. \end{cases}$

Details of implementation

tracking muti-object:

① Tracking-by-detection step:

detecting object locations independently in each frame.

form tracks by linking corresponding detections.

② reID:

Siamese → appearance vectors.

FreID: store killed b_{t-1}^{K}.

compare embedding space and newly detected tracks → re-identify via threshold.

③ motion model:

{ large camera motion.
  low video frame rate.

solution:

{ CMC
  camera motion compensation
  ECC
  Enhanced Correlation Coefficient
  CVA
  constant velocity assumption

Experiments:

Detector: Faster R-CNN, FPN (ResNet-101)

dataset: MOT17Det

crop and resize pooling. (instead of RoI pooling)

TriNet (ResNet-50).

triplet loss.

batch hard strategy.

Improvement:

① model → { accuracy: Mask R-CNN
  speed: one-stage (YOLO)

② extension → specific scenarios (additional feature)