Trajectory subnetwork

Keyframe$_t$ [3, H, W] → Feature extractor → Detector →

Box classes $\{c_t^i\}_{i=1}^N$ [N, C]

$\{c_t^i\}_{i=1}^N$

(a) trajectory indices [1, .., T] — FC$_{1a}$

[N, 4]
Box coordinates $\{B_t^i\}_{i=1}^N$

(b) [$\{B_t^i\}_{i=1}^N$] × T — FC$_{1b}$

Box features $\{f_t^i = f(B_t^i)\}_{i=1}^N$ [N, K]

(c) [$\{f_t^i\}_{i=1}^N$] × T — FC$_{1c}$

FC$_1$ concat → FC$_2$ → FC$_3$ → $\{(\mathbb{T}_t^i, c_t^i)\}_{i=1}^N$

([N, T, 4], [N, C])