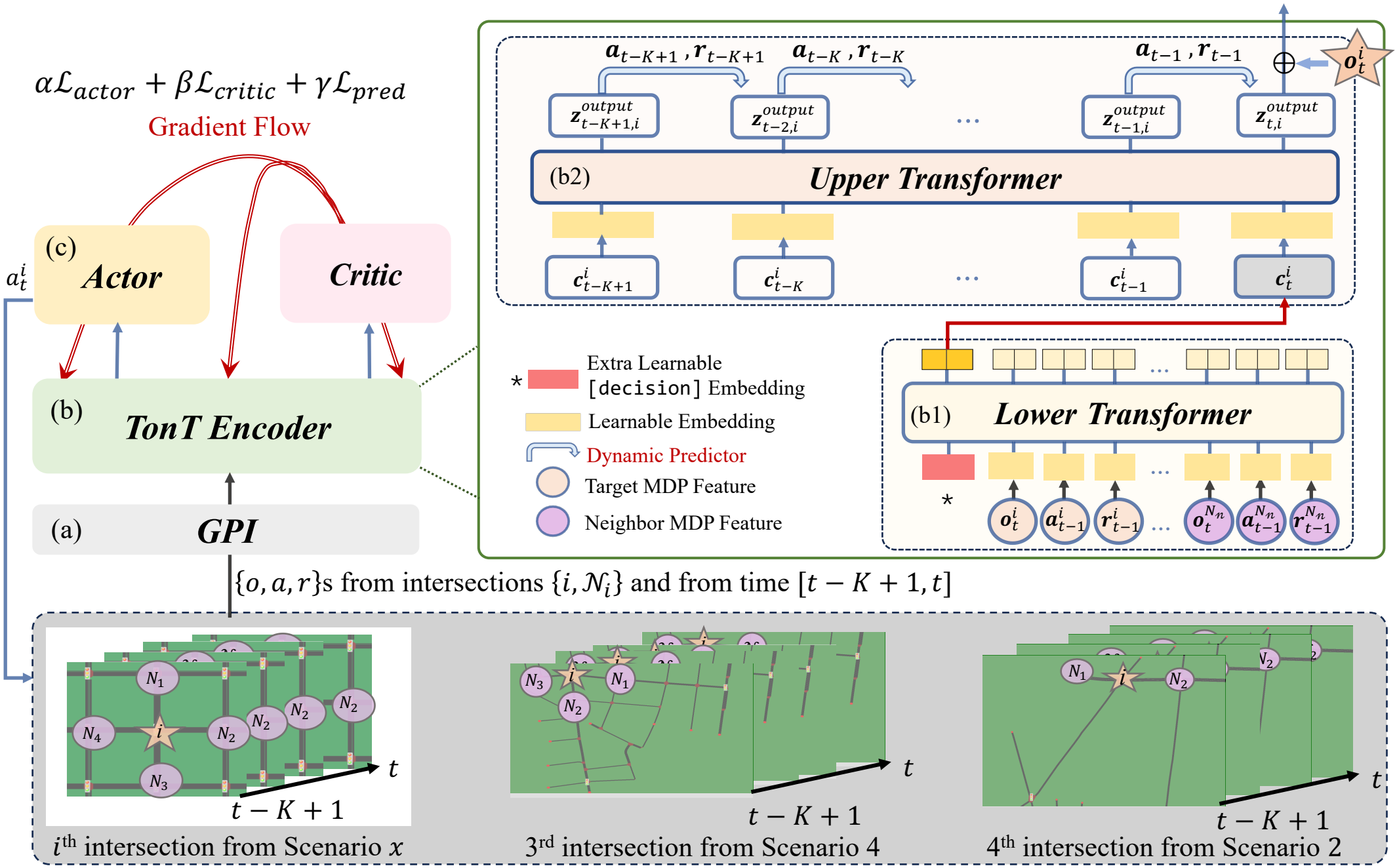
$\alpha\mathcal{L}_{actor} + \beta\mathcal{L}_{critic} + \gamma\mathcal{L}_{pred}$
Gradient Flow
(c) Actor
Critic
a_t^i
(b) TonT Encoder
(a) GPI
$\{o, a, r\}$s from intersections $\{i, \mathcal{N}_i\}$ and from time $[t-K+1, t]$
(b2) Upper Transformer
a_{t-K+1}, r_{t-K+1}
a_{t-K}, r_{t-K}
a_{t-1}, r_{t-1}
$z_{t-K+1,i}^{output}$
$z_{t-2,i}^{output}$
$z_{t-1,i}^{output}$
$z_{t,i}^{output}$
o_t^i
c_{t-K+1}^i
c_{t-K}^i
c_{t-1}^i
c_t^i
(b1) Lower Transformer
o_t^i a_{t-1}^i r_{t-1}^i $o_t^{N_n}$ $a_{t-1}^{N_n}$ $r_{t-1}^{N_n}$
* Extra Learnable [decision] Embedding
Learnable Embedding
Dynamic Predictor
Target MDP Feature
Neighbor MDP Feature
N_1 N_2 N_3 N_4 i t $t-K+1$
i^{th} intersection from Scenario x
3rd intersection from Scenario 4
4th intersection from Scenario 2
Each batch is a mix of intersections from various scenarios