



$$\sigma = \text{match}(\hat{y}^{main}, y)$$

$$\mathcal{L}^{main} = \mathcal{L}(\hat{y}^{main} |_{\sigma}, y) \quad \mathcal{L}^{aux} = \mathcal{L}(\hat{y}^{aux} |_{\sigma}, y)$$

$$C_1: \{q \in D_q | s(q) > \tau_q\} \quad C_2: \text{TopK}(\{q \in D_q | s(q) < \tau_q\}, N_{tr})$$

D T Detection / Track query (q : content, p : position)
 blue circle 1 Prediction of detection / track query (#1)
 flame icon Trainable
 N_q H_q Noisy query / History track query
 gray box \mathcal{F} : Global image feature
 checkered box \mathcal{P} : Positional encoding
 snowflake icon Frozen