$$\mathcal{L}_{\text{NCE}}(\{x_i, y_i\}_{i=1}^K; f, h) = \sum_{i=1}^K \log\left(\frac{e^{f(y_i)^T h(x_i)}}{\sum_{j=1}^K e^{f(y_i)^T h(x_j)}}\right) + \sum_{j=1}^K \log\left(\frac{e^{f(y_i)^T h(x_i)}}{\sum_{i=1}^K e^{f(y_i)^T h(x_j)}}\right) \quad (8)$$

$$\mathcal{L}_{\text{BC}}\left(\{s_i, a_i, s_i^+, \ell_i\}_{i=1}^K; \pi\right) = \sum_{i=1}^K \log \pi\left(a_i \mid s_i, \xi(\ell_i)\right) + \log \pi\left(a_i \mid s_i, \psi(s_i^+)\right) \quad (9)$$

$$\mathcal{L}_{\text{TRA}}\left(\{s_i, a_i, s_i^+, g_i, \ell_i\}_{i=1}^K; \pi, \phi, \psi, \xi\right) \quad (10)$$
$$= \underbrace{\mathcal{L}_{\text{BC}}\left(\{s_i, a_i, s_i^+, \ell_i\}_{i=1}^K; \pi, \psi, \xi\right)}_{\text{behavioral cloning}}$$
$$+ \underbrace{\mathcal{L}_{\text{NCE}}\left(\{s_i, s_i^+\}_{i=1}^K; \phi, \psi\right)}_{\text{temporal alignment}} + \underbrace{\mathcal{L}_{\text{NCE}}\left(\{g_i, \ell_i\}_{i=1}^K; \psi, \xi\right)}_{\text{task alignment}}$$

Algorithm 1: Temporal Representation Alignment (TRA)

input: dataset $\mathcal{D} = \left(\{s_{t,i}, a_{t,i}\}_{t=1}^H, \ell_i\right)_{i=1}^N$
initialize networks $\Theta \triangleq (\pi, \phi, \psi, \xi)$
while training **do**
    sample a batch of transitions $\left\{(s_{t,i}, a_{t,i}, s_{t+k,i}, \ell_i)\right\}_{i=1}^K \sim \mathcal{D}$ for $k \sim \text{Geom}(1-\gamma)$
    $\Theta \leftarrow (\pi, \phi, \psi, \xi) - \alpha \nabla_\Theta \mathcal{L}_{\text{TRA}}\left(\{s_{t,i}, a_{t,i}, s_{t+k,i}, \ell_i\}_{i=1}^K; \Theta\right)$
output: language ℓ-conditioned policy $\pi(a_t|s_t, \xi(\ell))$
        goal g-conditioned policy $\pi(a_t|s_t, \psi(g))$