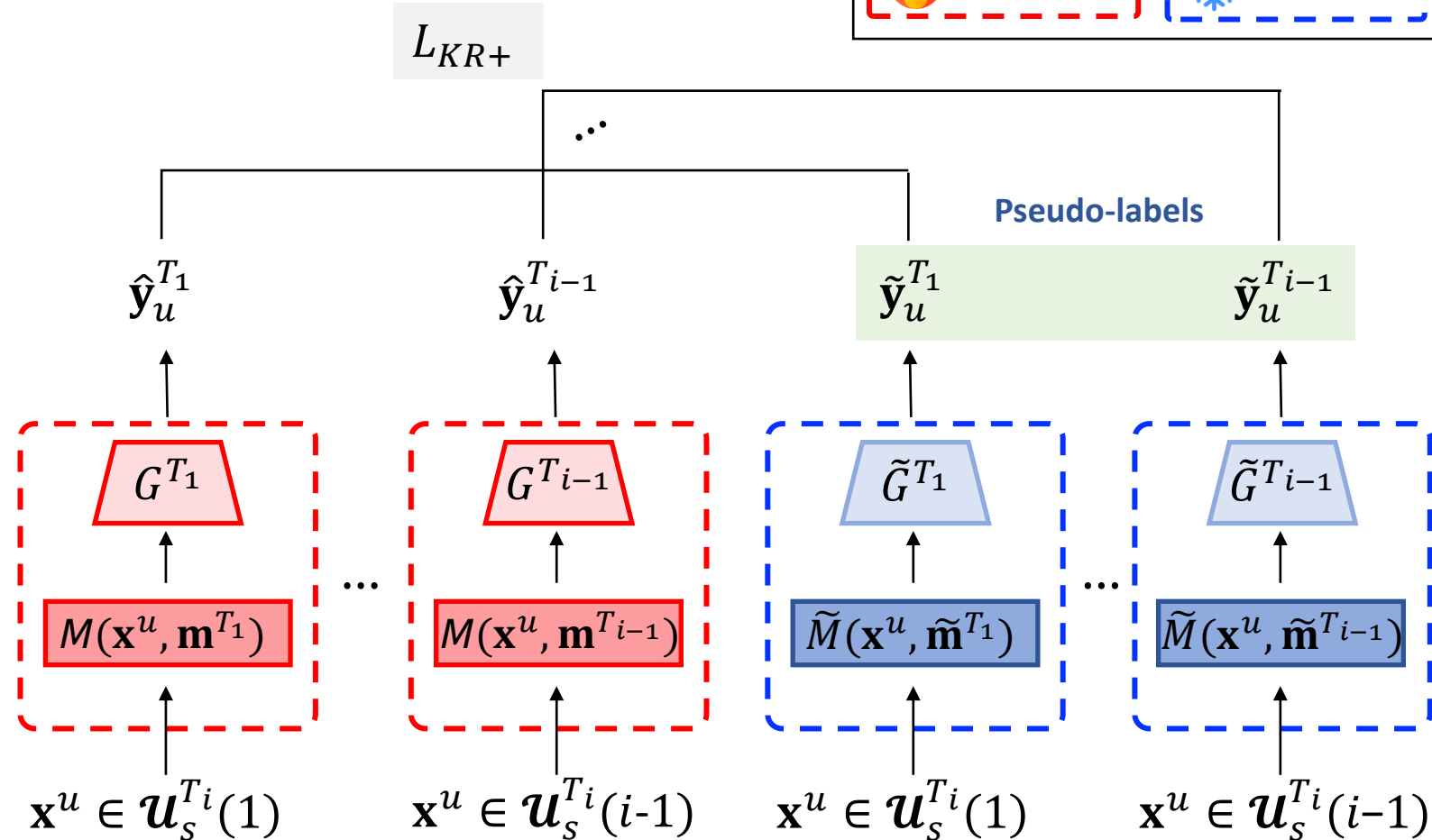
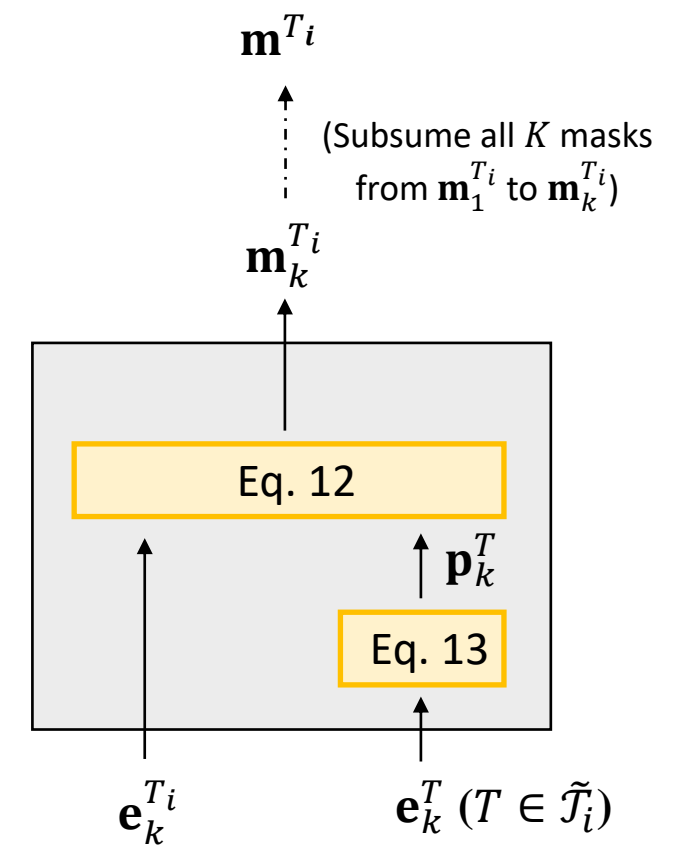


(a) Training on current task  $T_i$



(b) Knowledge retention



(c) Relation-aware task-specific mask

### Shared across tasks

$M(\mathbf{x}^u, \cdot)$  : Trained backbone  
 $\tilde{M}(\mathbf{x}^u, \cdot)$  : Frozen backbone

### Not shared across tasks

$G^{T_1} \dots G^{T_i}$  : Trained classifiers     $\mathbf{m}^{T_1} \dots \mathbf{m}^{T_i}$  : Trained masks  
 $\tilde{G}^{T_1} \dots \tilde{G}^{T_{i-1}}$  : Frozen classifiers     $\tilde{\mathbf{m}}^{T_1} \dots \tilde{\mathbf{m}}^{T_{i-1}}$  : Frozen masks