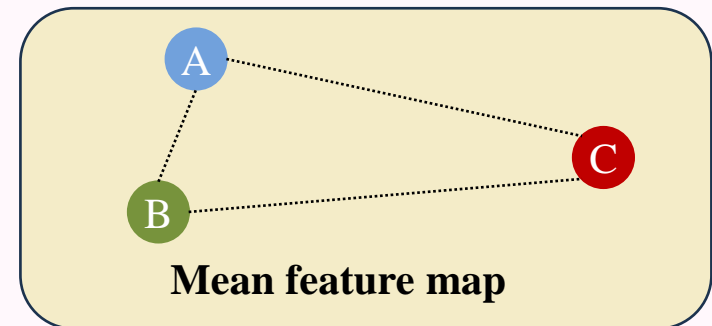
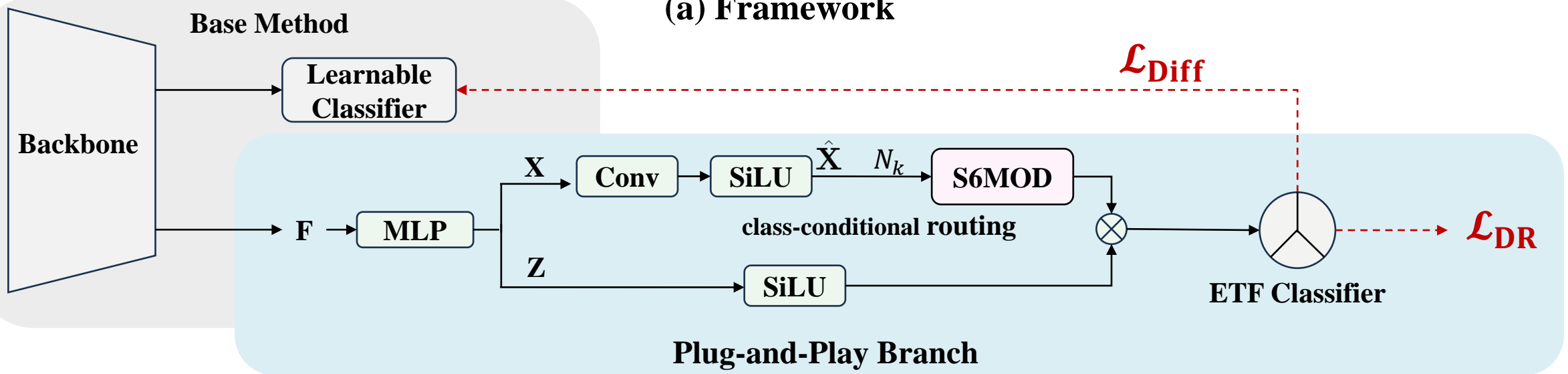
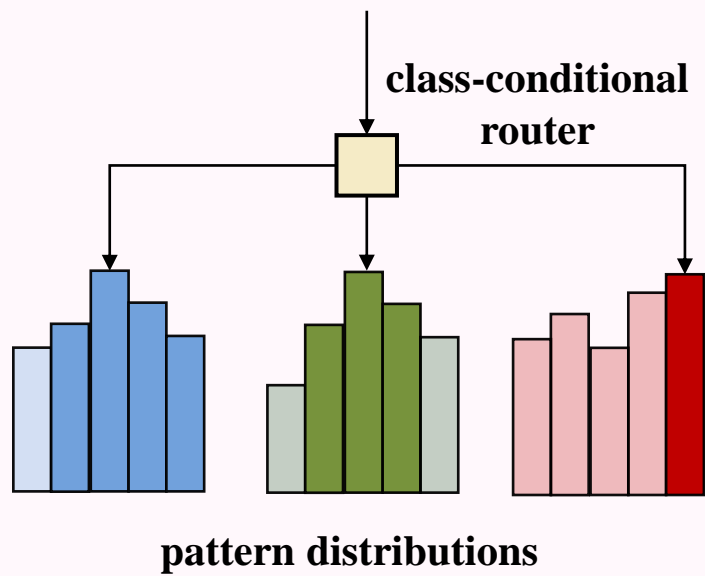


(a) Framework

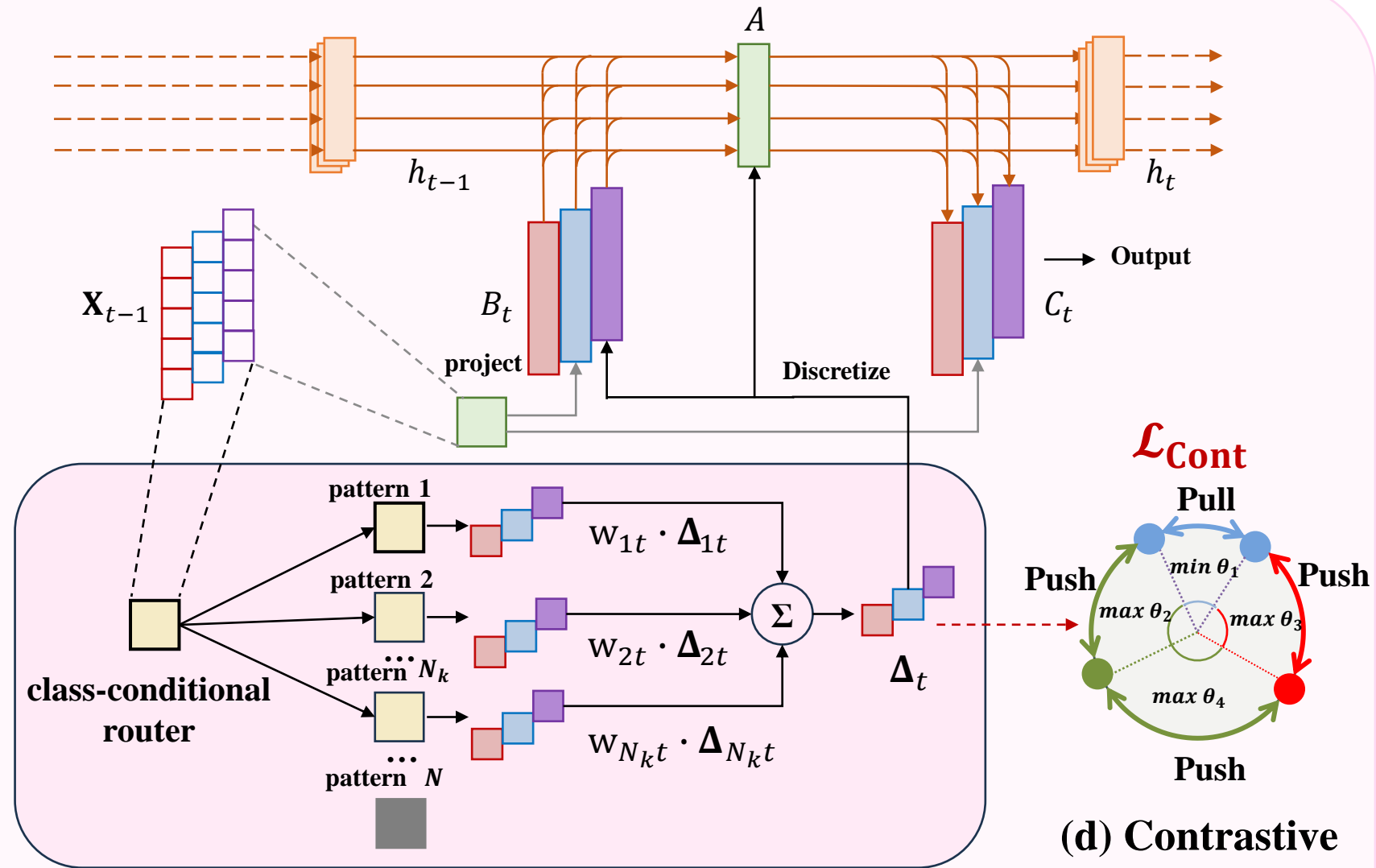


$$\sigma_k = \text{Avg}_c \{ \exp(-\lambda_0 \|\mathbf{M}_k - \mathbf{M}_c\|_2) \}$$

$$N_k = \text{ceil}(N \cdot \sigma_k)$$



(b) Class-Conditional Routing



(c) S6 with Mixture of Discretization (S6MOD)

(d) Contrastive Discretization Loss