

Continual Learning Solution Blueprint

Setting. Known task IDs and discrete concept drifts; efficiency required.

Method. Hybrid: tiny replay buffer ($\approx 1\text{--}2\%$ /task, balanced) + LwF-style logit distillation; optional lightweight EWC on backbone.

Architecture. Shared backbone + multi-head classifier (per-task softmax) or single head with task embedding.

Loss. $L = L_{\text{CE}}(\text{current}) + \lambda_{\text{KD}} L_{\text{KD}}(\text{prev logits}, T > 1) + \lambda_{\text{ER}} L_{\text{ER}}(\text{replay}) + \lambda_{\text{EWC}} \Omega(\theta)$.
Interleave batches current:replay $\approx 2:1$.

Forgetting mitigation. Replay gives direct rehearsal; distillation preserves boundaries; EWC anchors high-Fisher weights.

Ops. Reservoir sampling, per-class caps, early stop on current validation while tracking old-task buffer accuracy.