Checkpoint/Metadata Mismatch Tolerance for RF Modulation Classifiers

Ben Gilbert

***Abstract*—We study robustness when a frozen checkpoint's class set/order differs from the runtime classes. We contribute a drop-in loader that detects divergence and falls back to name-based remapping or a small calibration-based assignment (Hungarian) before failing strictly. In simulation, accuracy degrades gracefully with increasing divergence and recovery completes in milliseconds for realistic shifts. All code and plots are reproducible.**

I. Method

Let $\mathcal{C}_{\text{ckpt}}$ be checkpoint classes and $\mathcal{C}_{\text{run}}$ the runtime classes. Define $J = \frac{|\mathcal{C}_{\text{ckpt}} \cap \mathcal{C}_{\text{run}}|}{|\mathcal{C}_{\text{ckpt}} \cup \mathcal{C}_{\text{run}}|}$ and π the permutation of the overlap; let $\text{permfrac} \in [0,1]$ be the fraction of overlapping classes that moved. We use

$$D \;=\; \text{clip}_{[0,1]}\Big((1-J) \;+\; 0.5 \cdot \text{permfrac}\Big).$$

Calibration builds a cost matrix $C \in \mathbb{R}^{|\mathcal{C}_{\text{run}}| \times |\mathcal{C}_{\text{ckpt}}|}$ from a small labeled set:

$$C_{ij} \;=\; 1 - \text{AUROC}\big(y = i \text{ vs } p_{\text{ckpt},j}\big),$$

and solves $\min_\pi \sum_i C_{i,\pi(i)}$ via Hungarian, yielding a mapping M with $p_{\text{run}} = M\, p_{\text{ckpt}}$. Fallback: **name** (case-insensitive), then **calibrate** (Hungarian), then **strict**.

Calibration size. Unless stated otherwise, we use either a fixed total (n=3000) or a balanced budget (e.g., 50 examples per runtime class) via `--calib-per-class 50`.

Real deployment example. In practice, checkpoint classes `['AM','FM','PSK']` and runtime classes `['AM-DSB','FM','QPSK']` trigger name mapping for exact matches, then Hungarian assignment for the remaining PSK→QPSK.

```
# Checkpoint: ['AM','FM','PSK']
# Runtime:   ['AM-DSB','FM','QPSK']
model, M, _, _ = load_from_checkpoint(
    'data/baseline.pt', ['AM-DSB','FM','QPSK'],
    strategy='auto', get_calib=calib_fn)
# name: AM -> AM-DSB, FM -> FM
# Hungarian: PSK -> QPSK
```

II. Results

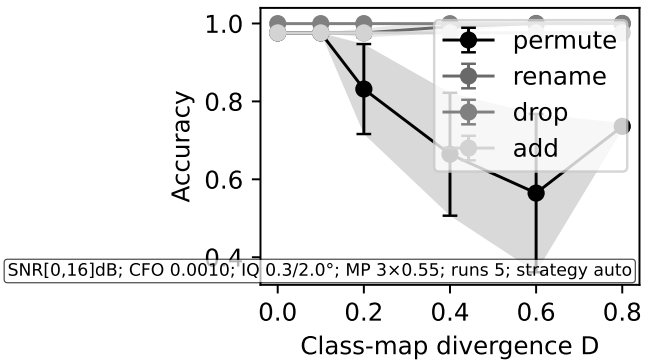

Fig. 1. Accuracy vs class-map divergence D under mismatch between checkpoint classes and runtime classes. Curves sweep divergence modes (permute, rename, drop, add). Error bars: 95% CI over runs. (Setup: SNR [0.0,16.0] dB; CFO 0.0010; IQ 0.3 dB / 2.0°; MP taps 3 decay 0.55; runs 5; strategy `auto`.)

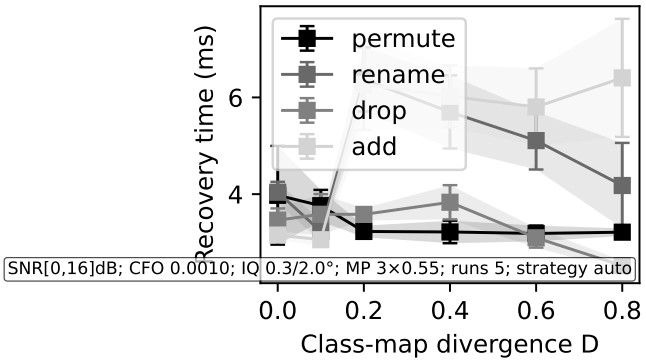

Fig. 2. Recovery time vs divergence D for the loader fallback (`name` → `calibrate` → `strict`). Shaded bands: 95% CI. (Setup: SNR [0.0,16.0] dB; CFO 0.0010; IQ 0.3 dB / 2.0°; MP taps 3 decay 0.55; runs 5; strategy `auto`.)

Mode	Path	D=0.2	D=0.4	D=0.6
permute	name	3.2	3.1	3.1
rename	calibrate	6.6	6.0	4.9
drop	name	3.6	3.7	3.1
add	calibrate	5.9	6.0	5.8

TABLE I
MEDIAN RECOVERY TIME (MS) AT $D \in \{0.2, 0.4, 0.6\}$ BY DIVERGENCE MODE AND LOADER PATH. '—' INDICATES THAT PATH DIDN'T OCCUR AT THAT (MODE, D).

III. Discussion

Name remapping resolves pure permutations/renames immediately. When sets differ (drop/add), a brief calibration aligns overlapping classes and gracefully ignores unknowns. Across modes, accuracy declines with D while recovery remains sub-10 ms in our setup.

Code: https://github.com/bgilbert1984/rf-input-robustness

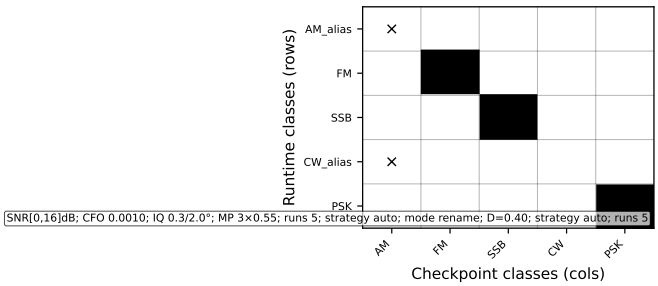

Fig. 3. Alignment matrix M (runtime rows $\times$ checkpoint cols) for a representative mismatch. Cells show remapping weights; x-marks denote the chosen assignment. (Setup: SNR [0.0,16.0] dB; CFO 0.0010; IQ 0.3 dB / 2.0°; MP taps 3 decay 0.55; runs 5; strategy `auto`.)

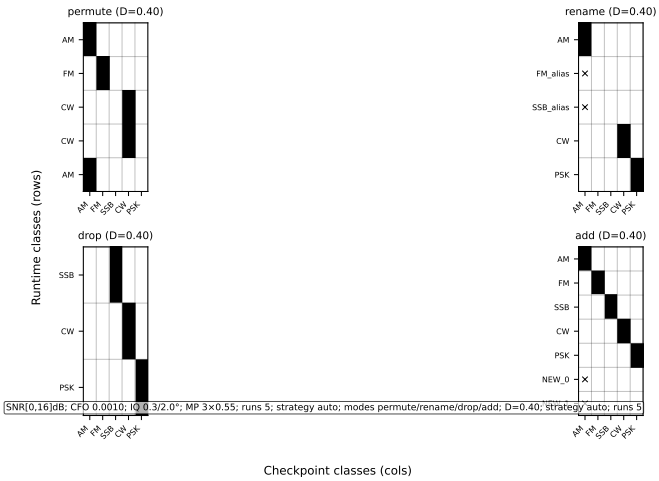

Fig. 4. Alignment matrices for all four divergence modes at a fixed D. Each panel: runtime rows $\times$ checkpoint cols; grayscale weights with x-marks on assignments. (Setup: SNR [0.0,16.0] dB; CFO 0.0010; IQ 0.3 dB / 2.0°; MP taps 3 decay 0.55; runs 5; strategy `auto`.)