Revisiting Entropy Minimization for *Long-Sequence* Continual Test-Time Adaptation

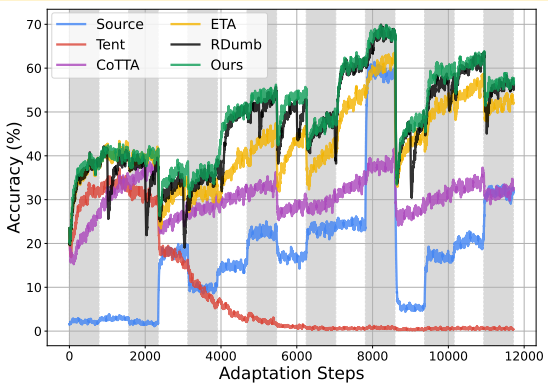
Wei Qin Chuah, Ruwan Tennakoon, Ali Bab-Hadiashar
RMIT University, Australia

BMVC 2025

RMIT UNIVERSITY

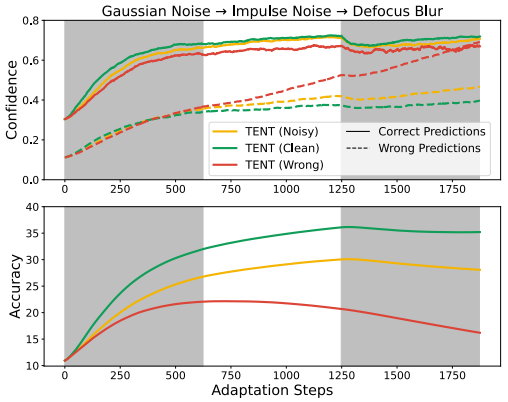
Introduction

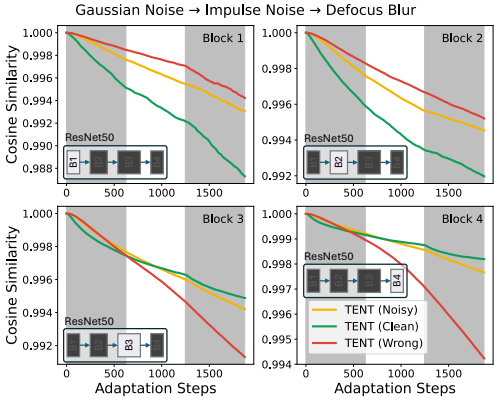
- Deep neural networks struggle with performance when the test data distribution is different from training, especially with continual and long-sequence shifts.
- Existing Test-Time Adaptation (TTA) methods suffer in "*long-sequence continual adaptation*" as models overfit to self-generated noisy pseudo-labels, degrading accuracy.



Label Noise Memorization

- Early Learning Phenomenon – model initially learns intrinsic patterns but eventually starts to memorize label noise.
- As adaptation steps increase, models start overfitting to incorrect predictions, especially in prolonged adaptation scenarios.





Proposed Methodology: Strategic Memorization Regularization (SMR)

- **Goal**: Prevent noisy label memorization during long-sequence continual test-time adaptation.
- **Core Idea**: Constrain layer-wise weight updates using cosine similarity to their initialization, limiting harmful drift due to adapting to noisy pseudo-labels.
- For each layer l, compare current weights W_l with initialized weights W_l^0:

$$\mathcal{L}_{SMR} = \sum_{l=1}^{L} \psi_l^2 \left\| W_l - W_l^0 \right\|^2 \qquad \psi_l = \cos\left(W_l, W_l^0\right) = \frac{W_l \cdot W_l^0}{\|W_l\| \|W_l^0\|}$$

- Total loss:

$$\mathcal{L} = \mathcal{L}_{entropy} + \lambda \mathcal{L}_{SMR}$$

- Adaptation Protocols:
  - Update only batch normalization parameters.
  - No source domain or calibration set required.
  - Works for multiple architectures (ResNet, ViT).

Experiments

Error Rates (%) Comparison:

Method	CCC	CIFAR100C	ImageNet-C (50k)	
			ResNet-50	VIT_B_16
TENT	98.9	62.2	94.5	94.2
ETA	98.9	32.2	57.1	48.4
RDumb	64.6	32.2	53.2	44.3
EATA	61.1	32.2	52.0	46.1
Ours (SMR)	**53.3**	**30.3**	**51.0**	**43.0**