



Continual Test-Time Domain Adaptation

Qin Wang



Continual Learning vs Test Time Adaptation

2026.03.12

Overview



01 Introduction

02 Preliminary

03 CoTTA

04 Current & Future

05 Conclusion



Qin Wang

 팔로우

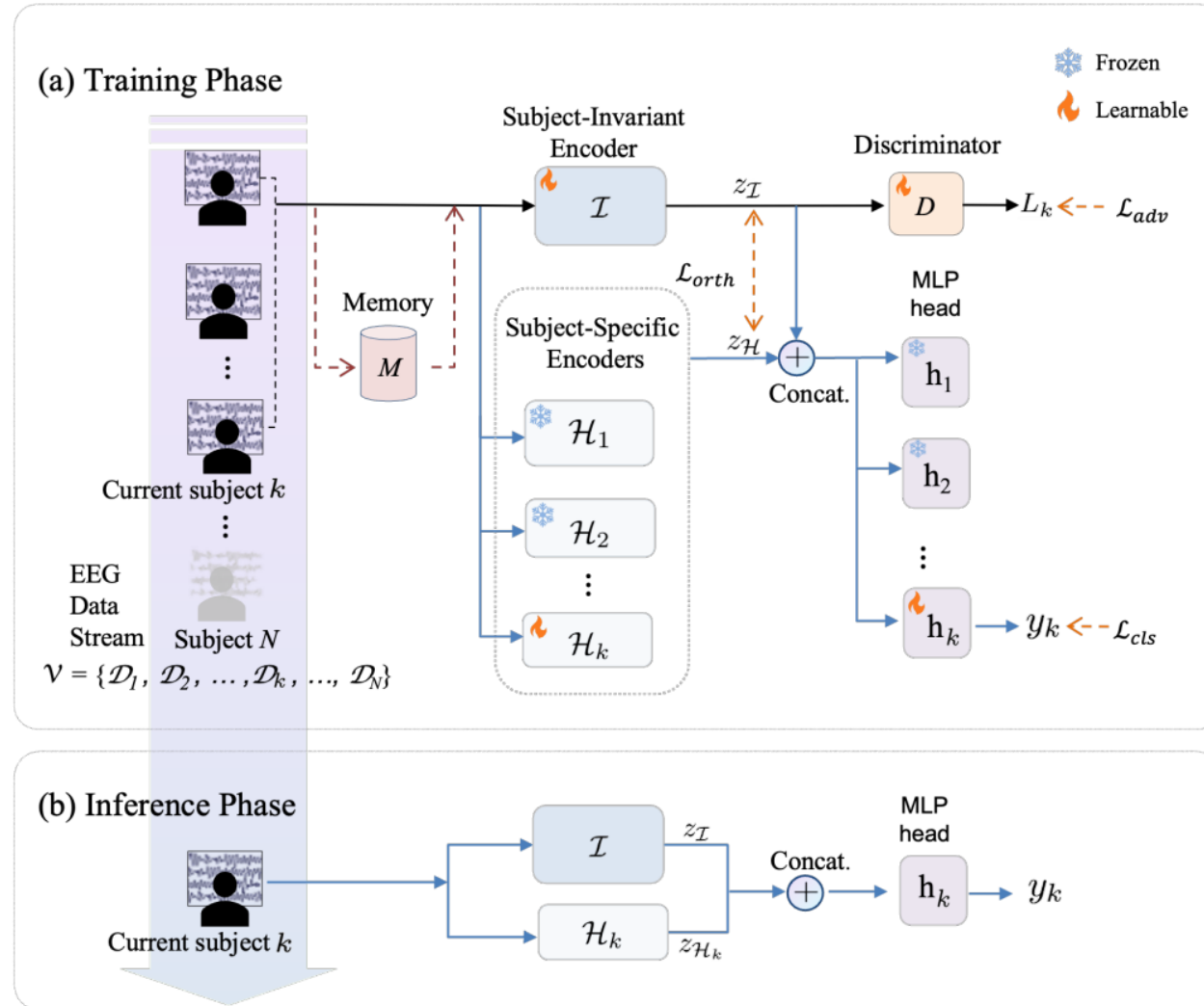
[ETH Zurich](#)

[ethz.ch의 이메일 확인됨 - 홈페이지](#)

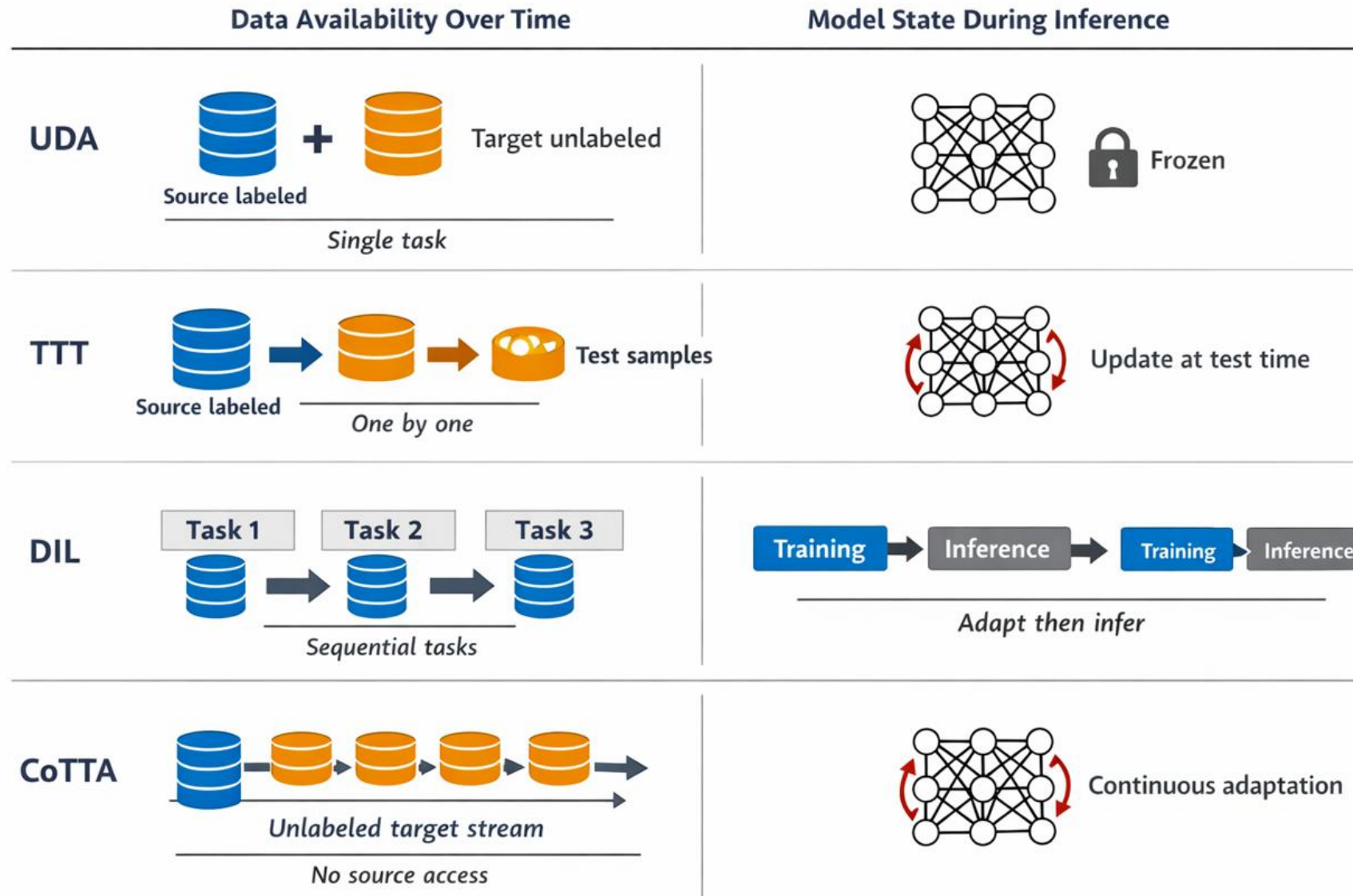
[Domain Adaptation](#) [Computer Vision](#)

제목	인용	연도
Continual Test-Time Domain Adaptation Q Wang, O Fink, L Van Gool, D Dai Conference on Computer Vision and Pattern Recognition (CVPR)	870	2022
Short-term load forecasting with deep residual networks K Chen, K Chen, Q Wang, Z He, J Hu, J He IEEE Transactions on Smart Grid 10 (4), 3943-3952	774	2018
Potential, challenges and future directions for deep learning in prognostics and health management applications O Fink, Q Wang, M Svensen, P Dersin, WJ Lee, M Ducoffe Engineering Applications of Artificial Intelligence 92, 103678	672	2020
Contextvp: Fully context-aware video prediction W Byeon, Q Wang, R Kumar Srivastava, P Koumoutsakos European Conference on Computer Vision (ECCV) Oral	208	2018

Introduction



01 Problem Setting: Overview of Adaptation Settings

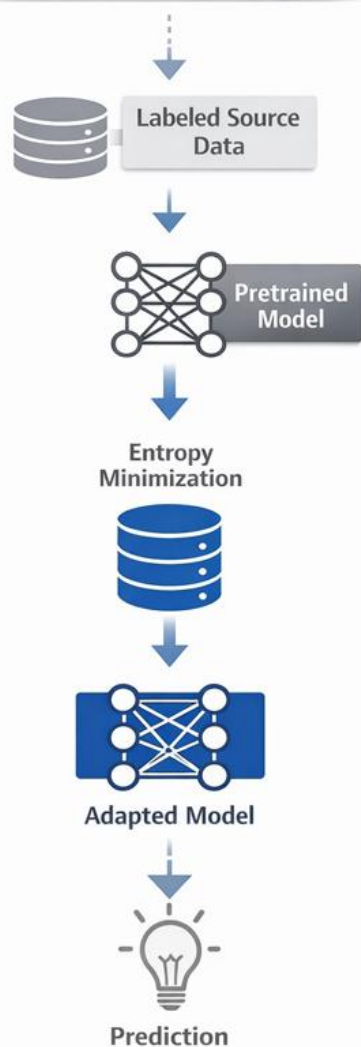


01 Problem Setting: Overview of Adaptation Settings

Criteria	Unsupervised Domain Adaptation	Domain Incremental Learning	Continual TTA
Domain Shift	Single Target	Sequential Targets	Continuous Streaming
Training Data	Pre-collected Target	Provided per New Domain	None (Test Data Only)
Adaptation Phase	Offline (Pre-deployment)	Offline (Incremental)	Online (On-the-fly)
Main Challenge	Bridging Domain Gap	Preventing Forgetting	Error Accumulation & Forgetting

Preliminary

Source Training



No Source Data Access

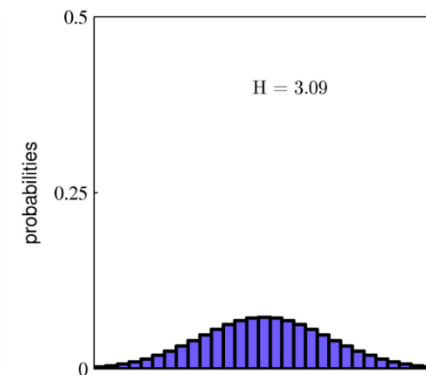
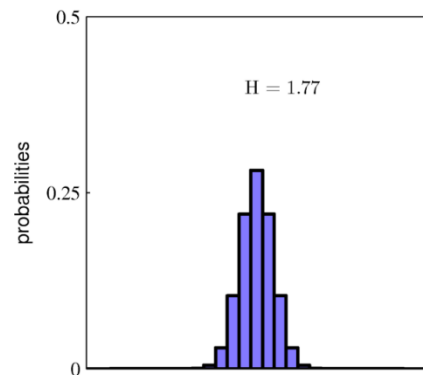
✓ Core objective

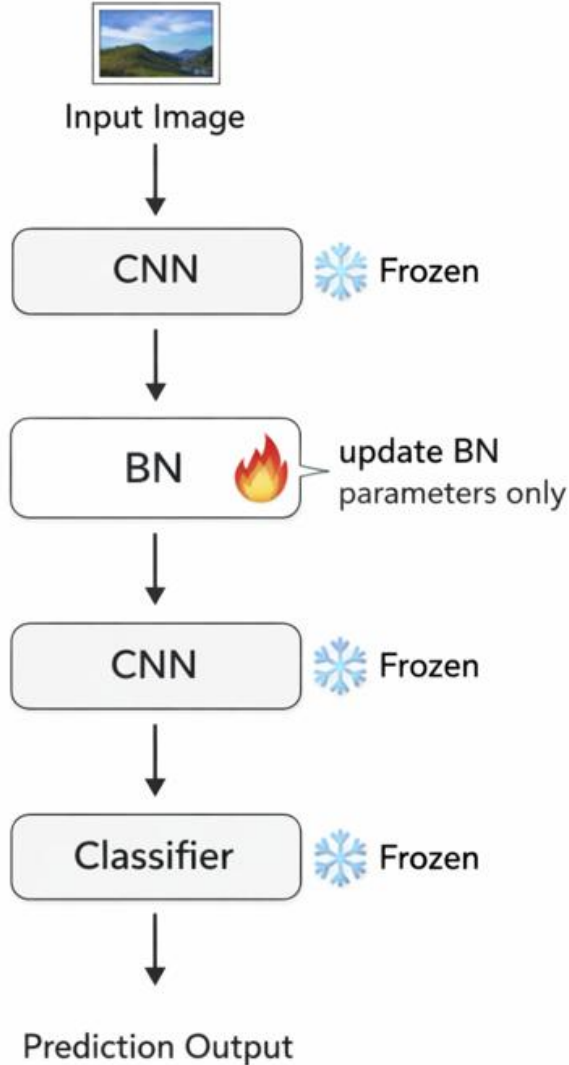
$$L_{\text{ent}}(x) = - \sum p(y|x) \log p(y|x)$$

- High entropy: uncertain
- Low entropy: confident
- No target labels required

✓ Why entropy?

- Label-free objective
- Computed from predictions only
- Usable during inference





✓ Why BN-only adaptation?

- Addressing Domain Shift
- Direct Normalization
- Increased Stability
- Preventing Overfitting

✓ Limitation

- **continual domain drift**
- **Error accumulate**

02 Why TENT Fails in Continual Streams

Continual Target Stream



- no labels
- no source data
- no task boundary

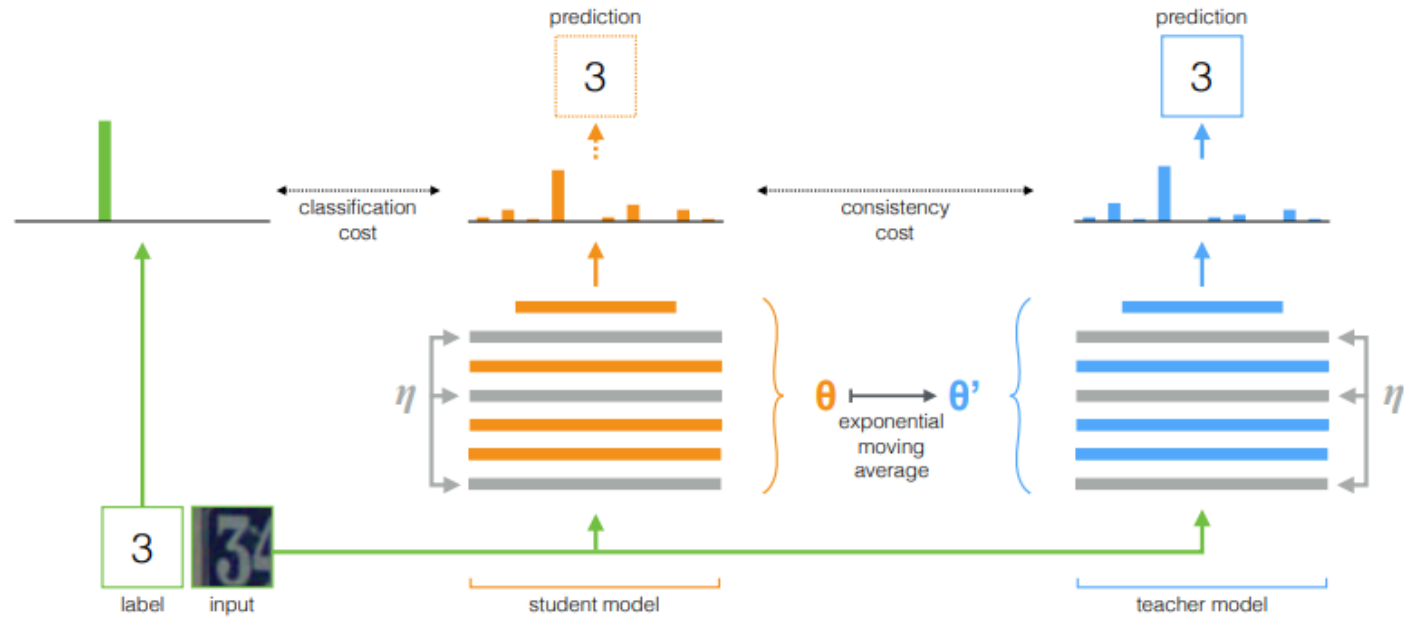
Error accumulation



- Errors compound over time
- No corrective supervision

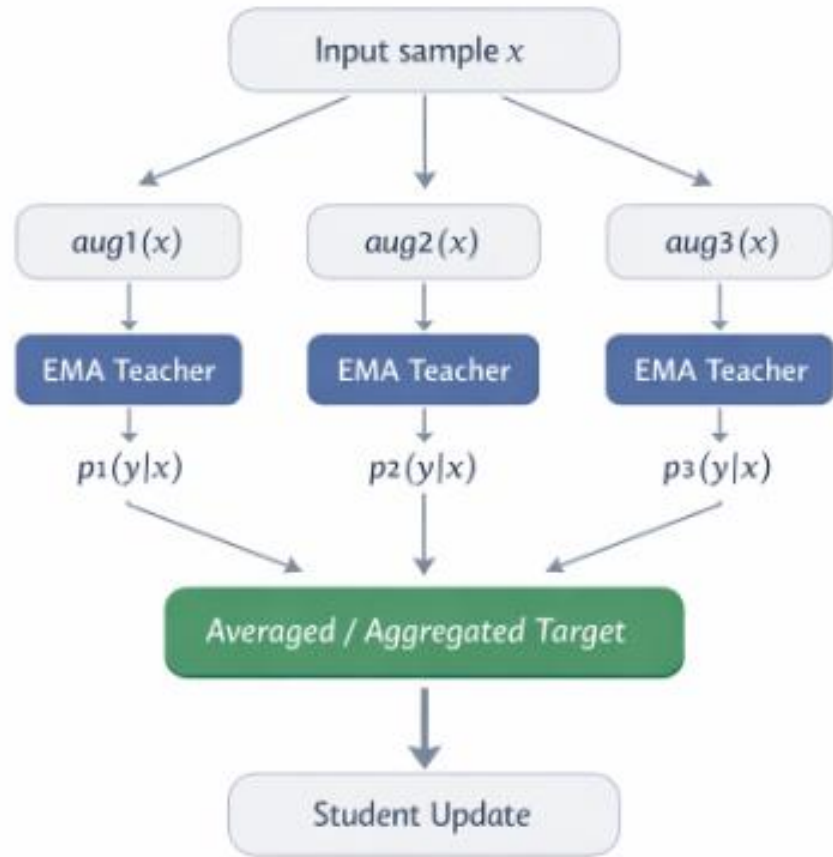
In continual test-time adaptation, the model must remain **plastic** enough to follow evolving domains, while staying **stable** enough to avoid drifting toward **erroneous self-confirming** updates.

02 EMA Teacher for Stable Online Supervision

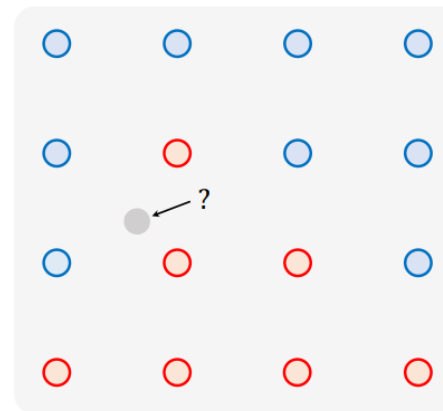


$$\theta_{teacher} \leftarrow \alpha \theta_{teacher} + (1 - \alpha) \theta_{student}$$

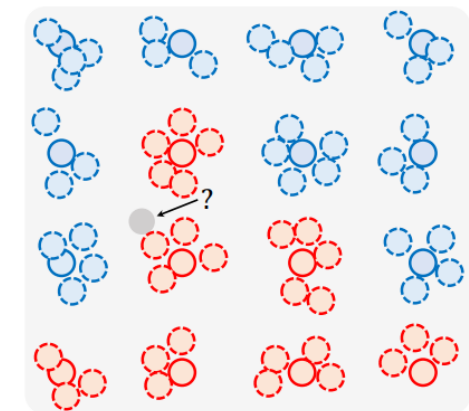
02 Augmentation Consistency for Robust Predictions



- **Mitigating Noise**
- **Variance Reduction**
- **Reliable Supervision**

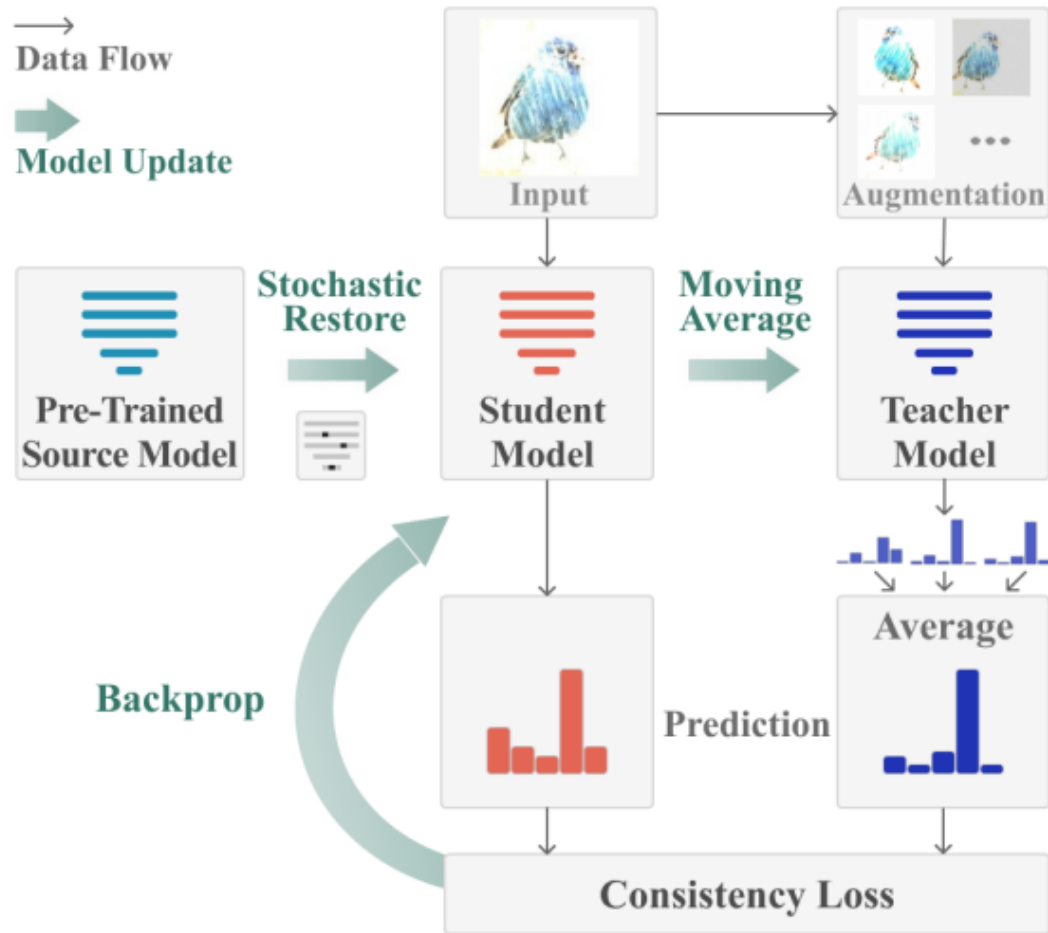


Empirical risk minimization
(Without data augmentation)



Vicinal risk minimization
(With data augmentation)

CoTTA



✓ Base adaptation

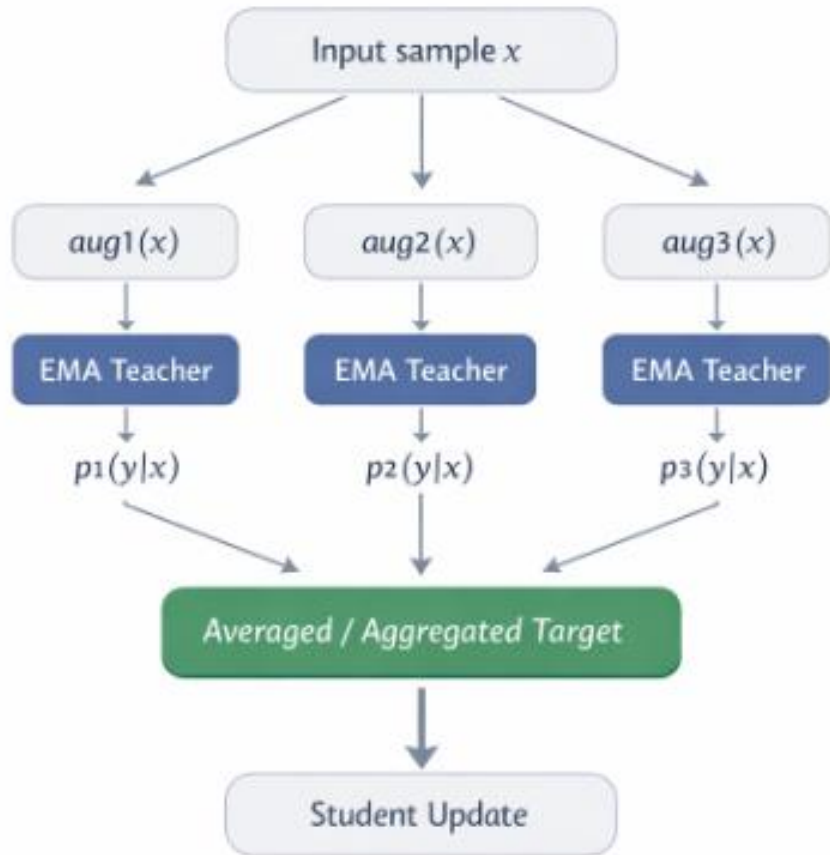
1. Target sample arrives
2. Student is updated online
3. Entropy-based adaptation

✓ Prediction-level stabilization

- **EMA teacher:** temporal smoothing
- **Augmentation:** variance reduction

✓ Parameter-level stabilization

- **Stochastic Restore:**
Prevents drifts & Preserves source knowledge



$$\tilde{p}(y|x) = \frac{1}{K} \sum_{k=1}^K p_{teacher}(y|a_k(x))$$

✓ The Problem

- Instantaneous student predictions can be noisy

✓ Stable Target Generation

- **EMA teacher:** smooths supervision over time
- **Augmentation Averaging:**
Reduces variance across multiple views

✓ Key Benefit

- Provides reliable supervision for **robust online adaptation**

$$M \sim \text{Bernoulli}(p) \quad x^{l+1}$$
$$= W_{t+1} * x^l$$

$$W_{t+1} = M \odot W_0 + (1 - M) \odot W_{t+1}$$

✓ **Interpretation**

- M : restore mask
- W_0 : source weights
- partial reset only

✓ **Motivation**

- continual updates cause drift
- source knowledge may fade

✓ **Effect**

- keeps a source anchor
- reduces long-term forgetting

Algorithm 1 The proposed continual test-time adaptation

Initialization: A source pre-trained model $f_{\theta_0}(x)$, teacher model $f_{\theta'_0}(x)$ initialized from $f_{\theta_0}(x)$.

Input: For each time step t , current stream of data x_t .

- 1: Augment x_t and get weight and augmentation-averaged pseudo-labels from the teacher $f_{\theta'_t}$ by Equation 4.
- 2: Update student f_{θ_t} by consistency loss in Equation 5.
- 3: Update teacher $f_{\theta'_t}$ by moving average in Equation 2.
- 4: Stochastically restore student f_{θ_t} by Equation 8.

Output: Prediction $f_{\theta'_t}(x_t)$; Updated student model $f_{\theta_{t+1}}(x)$; Updated teacher model $f_{\theta'_{t+1}}(x)$.

Method	Weight-avg.	Aug-avg.	Stochastic Restore	$t \rightarrow$															Mean
				Gaussian	shot	impulse	defocus	glass	motion	zoom	snow	frost	fog	brightness	contrast	elastic_trans	pixelate	jpeg	
Source				72.3	65.7	72.9	46.9	54.3	34.8	42.0	25.1	41.3	26.0	9.3	46.7	26.6	58.5	30.3	43.5
BN Stats Adapt				28.1	26.1	36.3	12.8	35.3	14.2	12.1	17.3	17.4	15.3	8.4	12.6	23.8	19.7	27.3	20.4
Pseudo-label				26.7	22.1	32.0	13.8	32.2	15.3	12.7	17.3	17.3	16.5	10.1	13.4	22.4	18.9	25.9	19.8
TENT-online* [61]				24.8	23.5	33.0	12.0	31.8	13.7	10.8	15.9	16.2	13.7	7.9	12.1	22.0	17.3	24.2	18.6
TENT-continual [61]				24.8	20.6	28.6	14.4	31.1	16.5	14.1	19.1	18.6	18.6	12.2	20.3	25.7	20.8	24.9	20.7
CoTTA (Ours)	✓			27.2	22.8	30.8	12.1	30.1	13.9	11.9	17.2	16.0	14.3	9.4	13.1	19.9	15.4	19.9	18.3
CoTTA (Ours)	✓	✓		24.5	21.0	26.0	12.3	27.9	13.9	12.0	16.6	15.9	14.7	9.4	13.6	19.8	14.7	18.7	17.4
CoTTA (Ours)	✓	✓	✓	24.3	21.3	26.6	11.6	27.6	12.2	10.3	14.8	14.1	12.4	7.5	10.6	18.3	13.4	17.3	16.2 (0.1)

Conclusion

✓ Strengths

- formulates the CTTA problem clearly
- introduces a simple but effective combination of **EMA teacher, augmentation averaging, and stochastic restoration**
- shows strong robustness under long continual shift

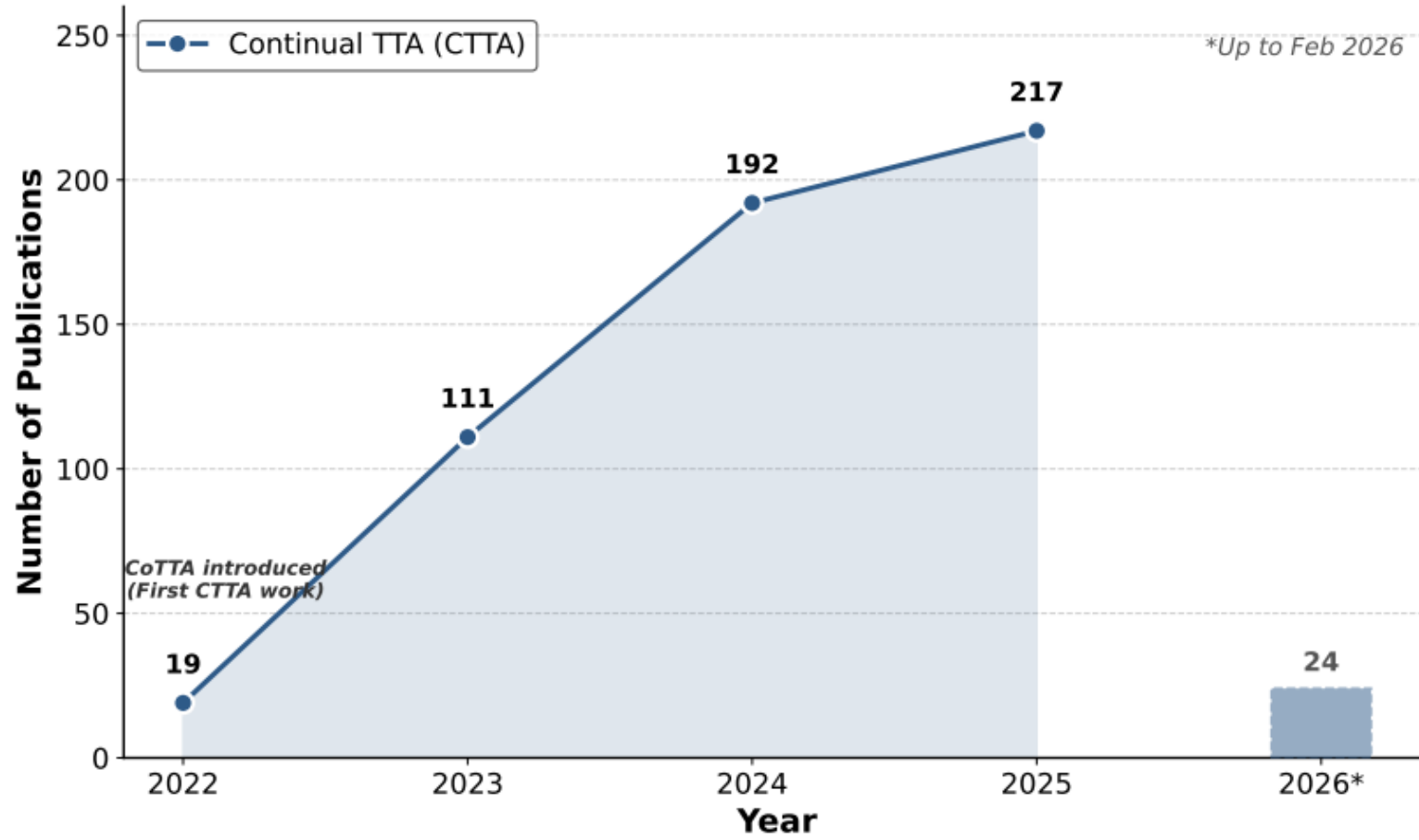
✓ Overall evaluation

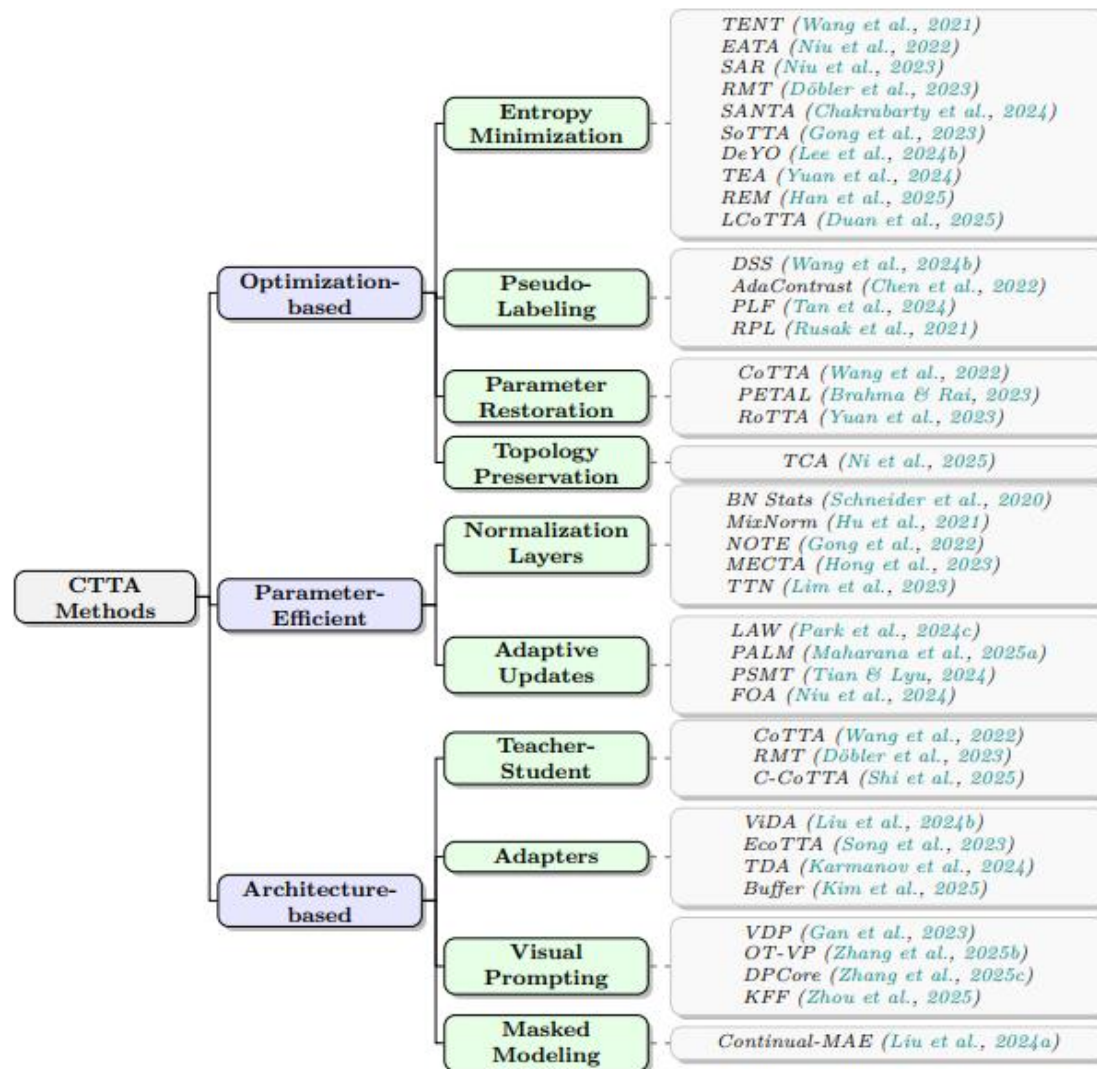
- a meaningful early baseline for **continual test-time adaptation**
- not just a performance improvement, but a clear step toward **stable online adaptation**

“CoTTA is a meaningful early CTTA method that clearly shows stable continual adaptation requires both robust supervision and drift control.”

Current & Future

05 Current





- ✓ **Benchmark Bias**
 - limited benchmark diversity
 - inconsistent protocols
 - backbone/tuning sensitivity

- ✓ **Foundation Models**
 - growing interest in adapters/prompting
 - still underexplored for large pretrained models
 - reliability remains a concern

- ✓ **Theoretical Foundation**
 - weak theory for drift and forgetting
 - limited understanding of stable adaptation
 - mostly empirical progress so far



Thank You



2026.01.05

BrainLAB Journal Club
인공지능응용학과 허세민