Private and Stable Test-Time Adaptation with Differential Privacy

Zefeng Li*[12], Qiaoyue Tang*[1], Mathias Lécuyer†[1], Evan Shelhamer†[12]

UBC[1] Vector Institute[2]

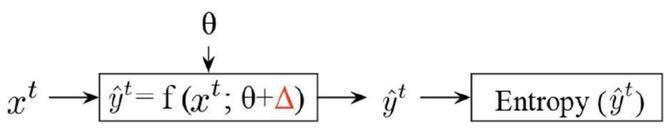
We make test-time adaptation private with differential privacy, and show that per-sample clipping improves both privacy and stability.

1 MOTIVATION

Test-time adaptation (TTA) updates a model on unlabeled test data to recover accuracy under distribution shift.

These updates can leak information about past test samples because model parameters depend on the test data.

Goal: enable private, stable TTA with minimal loss in adaptation accuracy.



Fully Test-time Adaptation

2 MAIN CONTRIBUTIONS

Develop DP versions of Tent, EATA, SAR, DeYO, and COME.

Show that per-sample gradient clipping is a general mechanism that improves TTA performance even without DP noise.

Measure the privacy/accuracy trade-off for test-time adaptation in episodic and continual settings.

3 DP-TTA FRAMEWORK

Test batch

↓

Per-sample TTA loss

↓

Per-sample gradients

↓

Clip each gradient to norm C

↓

Add Gaussian noise

↓

Update model

DP-TTA Update (per batch B_t)

$$g_i(x_i) = \nabla_\theta \ell_{\mathrm{TTA}}(x_i; \theta_t)$$

$$\tilde{g}_i(x_i) = \frac{g_i(x_i)}{\max(1, ||g_i(x_i)||_2 / C)}$$

$$\Delta_t^{PP} = \frac{1}{|B_t|}\left[\sum_i \tilde{g}_i(x_i) + \mathcal{N}(0, \sigma^2 C^2 I)\right]$$

$$\theta_{t+1} = \theta_t - \eta \Delta_t^{PP}$$

4 EXPERIMENTAL SETUP

Model: ViT-Base/16

Data: ImageNet-C (severity 5) + clean ImageNet-1k validation

Settings: episodic and continual adaptation

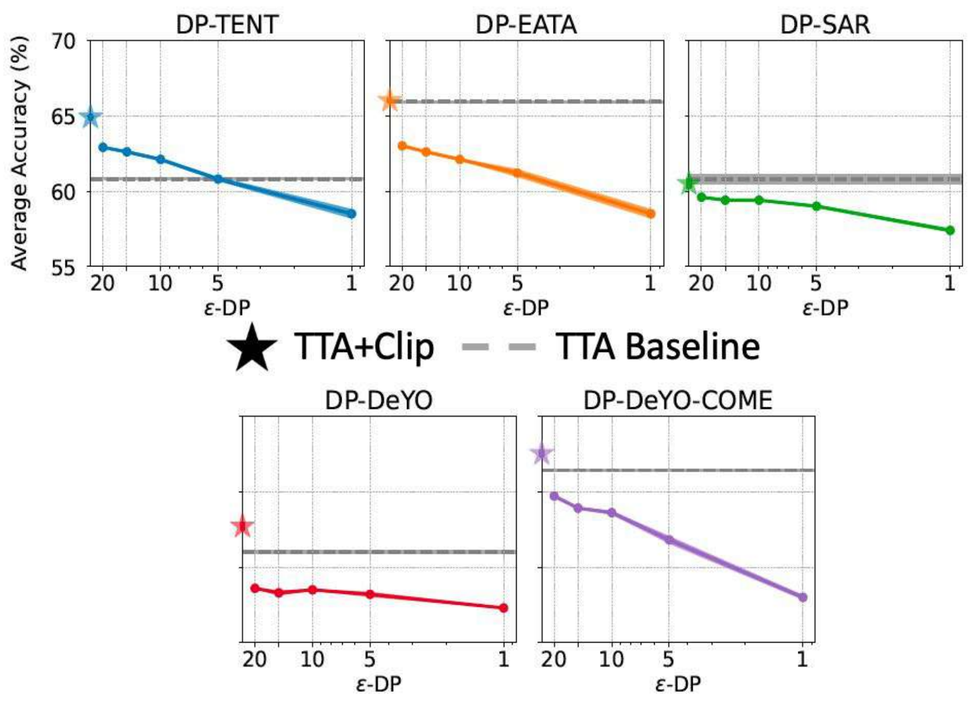
Privacy budgets: $\epsilon \in \{1, 5, 10, 15, 20\}$, $\delta = 10^{-6}$

Batch size: 64; per-sample clipping implemented with Opacus

5 PRIVACY–ACCURACY TRADE-OFF

Average top-1 accuracy (%) on ImageNet-C (severity 5)

under different privacy budgets ($\delta = 10^{-6}$)



Key Findings

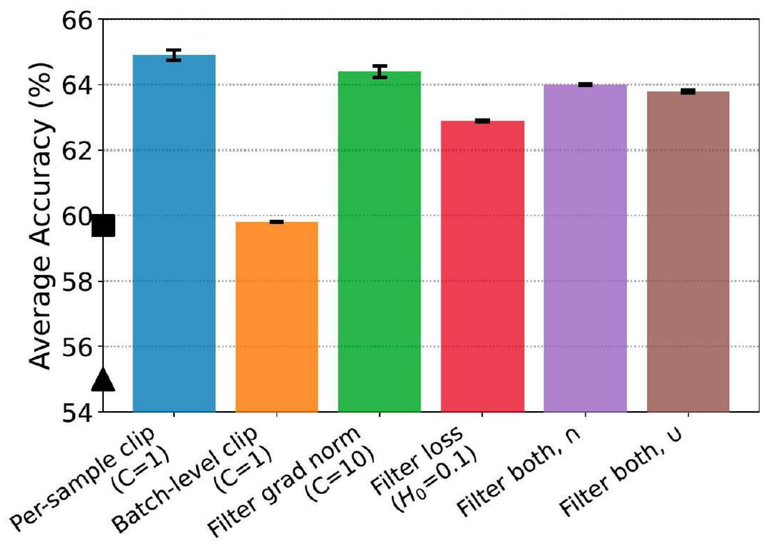
- Moderate privacy can still preserve strong adaptation performance, and some DP-TTA methods remain competitive with their non-private counterparts.
- DP-Tent is the clearest example: it achieves 62.9%, 62.6%, and 62.2% at ε = 20, 15, and 10, outperforming the non-private baseline of 60.8%.
- For other methods, the non-private versions remain stronger at ε = 20, but the gaps are modest: 2.9 points for EATA, 1.2 for SAR, 2.4 for DeYO, and 1.7 for DeYO-COME.
- Per-sample clipping consistently improves TTA, enhancing robustness and stability, and can even yield state-of-the-art continual accuracy.

★ Selected quantitative results

- Per-sample gradient clipping improves adaptation across most corruption types, raising the average gain across TTA methods from **0.1%** to **4.1%**.
- DeYO-COME + clipping achieves state-of-the-art continual performance, reaching **67.5%** average accuracy.
- Per-sample clipping rescales any gradient with norm larger than C to norm C, preserving its direction while reducing its magnitude.
- Batch-level clipping is much less effective: accuracy stays around **57–59%** across thresholds C, even when the fraction of clipped batches ranges from **84%** to **1.5%**.

6 PER-SAMPLE CLIPPING IMPROVES TTA

Per-sample Clip is better than Other Operations

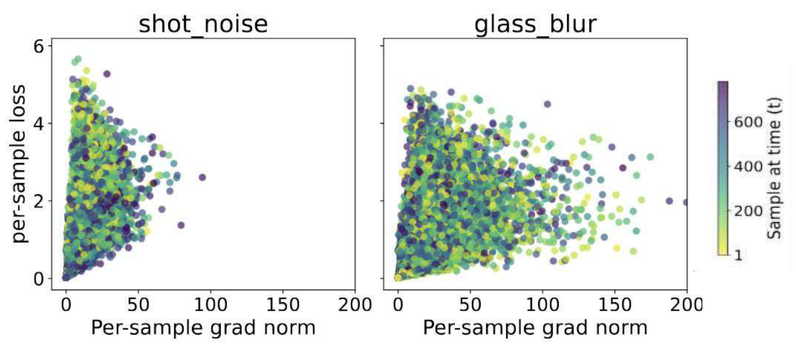


▲ Source model ■ Tent baseline

Per-sample clipping **controls per-sample influence** to model updates, batch-level clipping do not.

Controlling per-sample influence (**clipping**) **is more efficient than** removing difficult samples (**filtering**).

Gradient Norm and Loss Measure Different Aspect of Difficult Sample



Per-sample gradient norm and loss are **not strongly aligned**.

They can be **complementary indicators** of difficult samples.

7 EFFICIENCY AND TAKEAWAYS

DP adds only modest overhead: roughly 20–30 ms per batch.

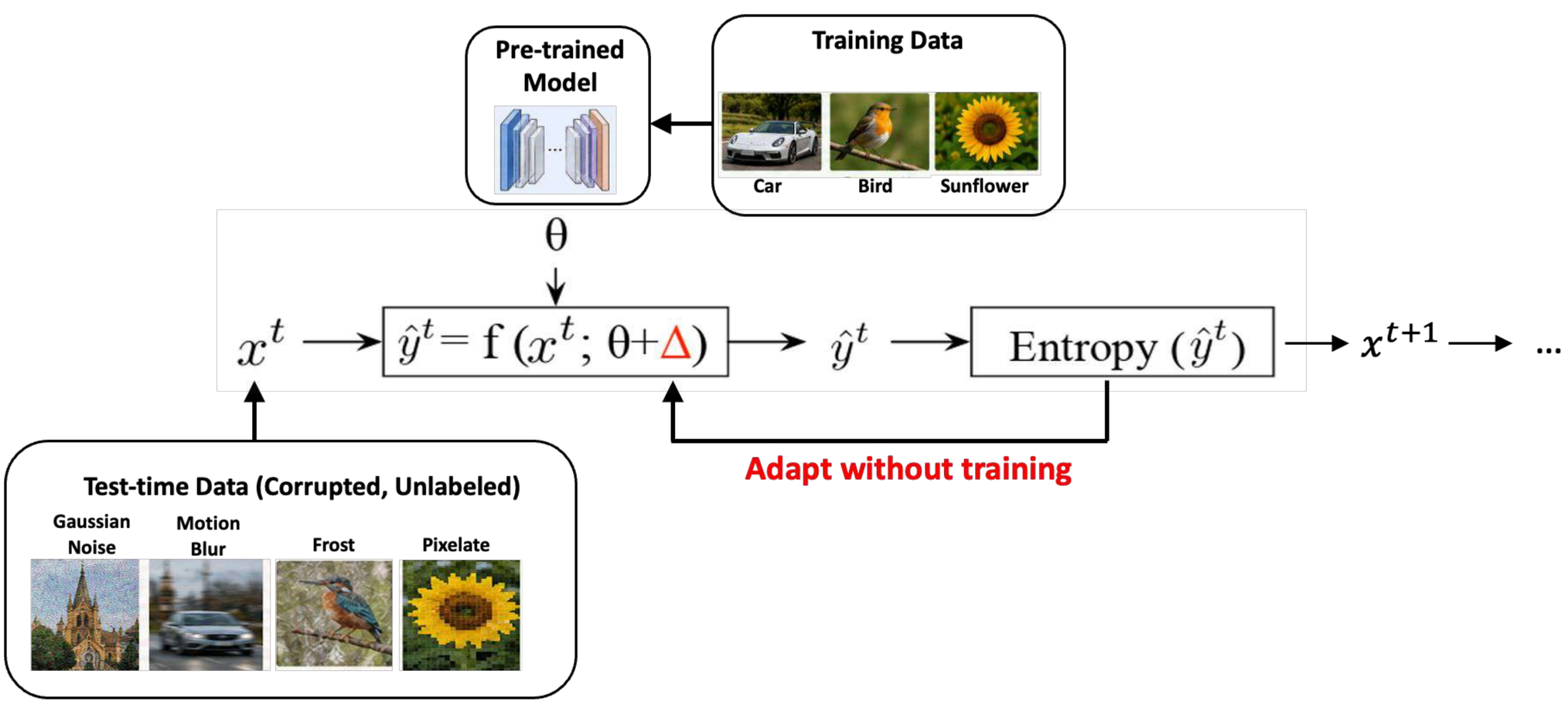
The slowest DP variant takes only 1.28x the original runtime.

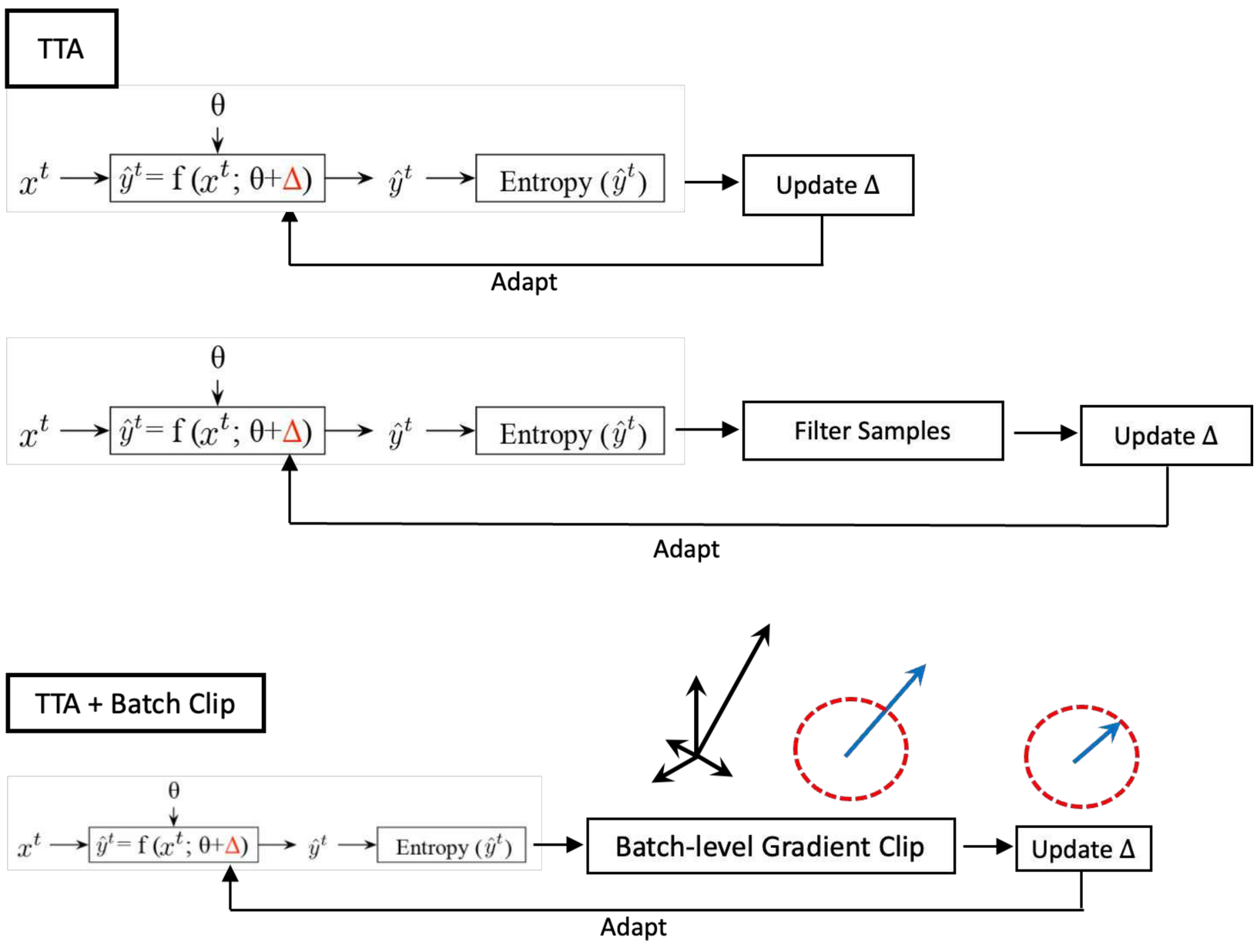
This is the first study of private test-time adaptation.

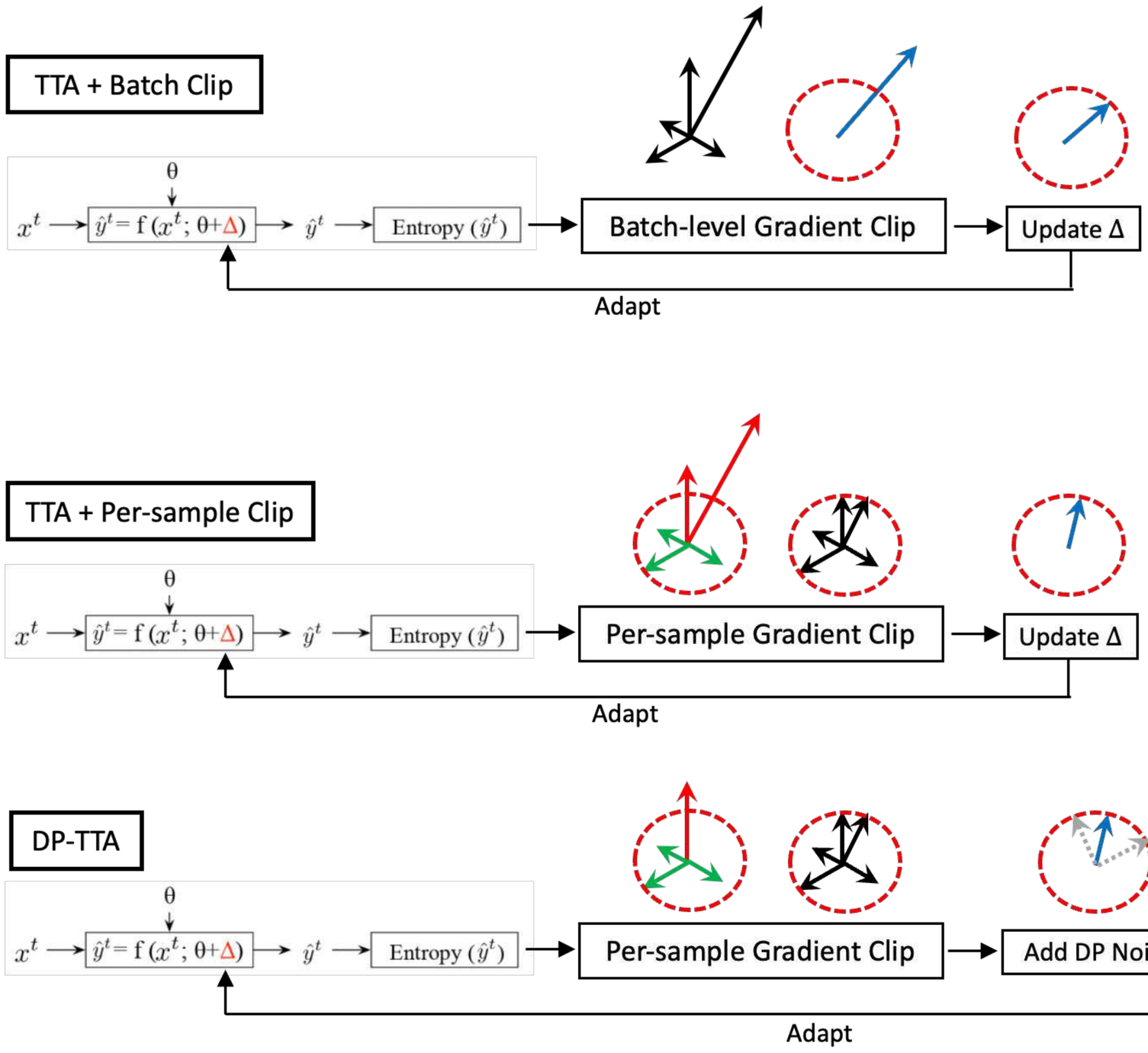
Per-sample clipping is not only a privacy tool; it is also a strong stability mechanism for TTA.

Private test-time updates remain practical and effective.

Key message: Private TTA is practical and effective, with modest overhead; per-sample clipping improves privacy and stability.


Pre-trained Model
Training Data
Car
Bird
Sunflower
θ
x^t
$\hat{y}^t = \mathrm{f}(x^t; \theta+\Delta)$
$\hat{y}^t$
Entropy ($\hat{y}^t$)
x^{t+1}
...
Adapt without training
Test-time Data (Corrupted, Unlabeled)
Gaussian Noise
Motion Blur
Frost
Pixelate


TTA
θ
x^t
$\hat{y}^t = \mathrm{f}(x^t; \theta+\Delta)$
$\hat{y}^t$
Entropy ($\hat{y}^t$)
Update Δ
Adapt
θ
x^t
$\hat{y}^t = \mathrm{f}(x^t; \theta+\Delta)$
$\hat{y}^t$
Entropy ($\hat{y}^t$)
Filter Samples
Update Δ
Adapt


TTA + Batch Clip
θ
x^t
$\hat{y}^t = \mathrm{f}(x^t; \theta+\Delta)$
$\hat{y}^t$
Entropy ($\hat{y}^t$)
Batch-level Gradient Clip
Update Δ
Adapt
TTA + Per-sample Clip
θ
x^t
$\hat{y}^t = \mathrm{f}(x^t; \theta+\Delta)$
$\hat{y}^t$
Entropy ($\hat{y}^t$)
Per-sample Gradient Clip
Update Δ
Adapt

DP-TTA
θ
x^t
$\hat{y}^t = \mathrm{f}(x^t; \theta+\Delta)$
$\hat{y}^t$
Entropy ($\hat{y}^t$)
Per-sample Gradient Clip
Add DP Noise
(Private) Update Δ
Adapt