# MECTA: Memory-Economic Continual Test-Time Model Adaptation

**Junyuan Hong**[1*], Lingjuan Lyu[2], Jiayu Zhou[1], Michael Spranger[2]

[1]Michigan State University, [2]Sony AI

*Work done during internship at Sony AI

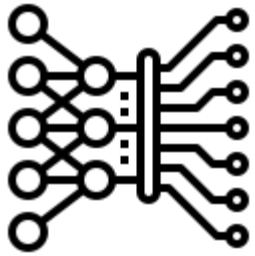
ICLR International Conference On Learning Representations

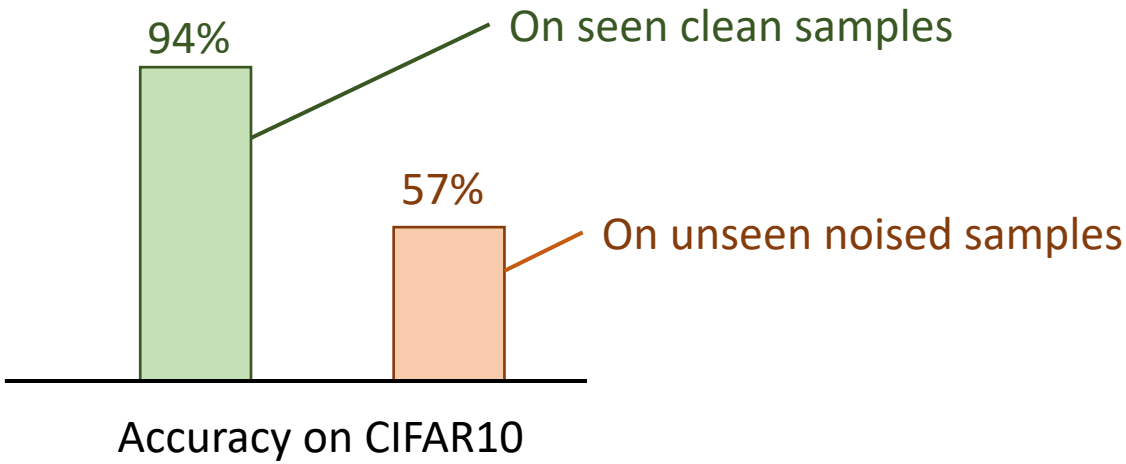
Sony AI

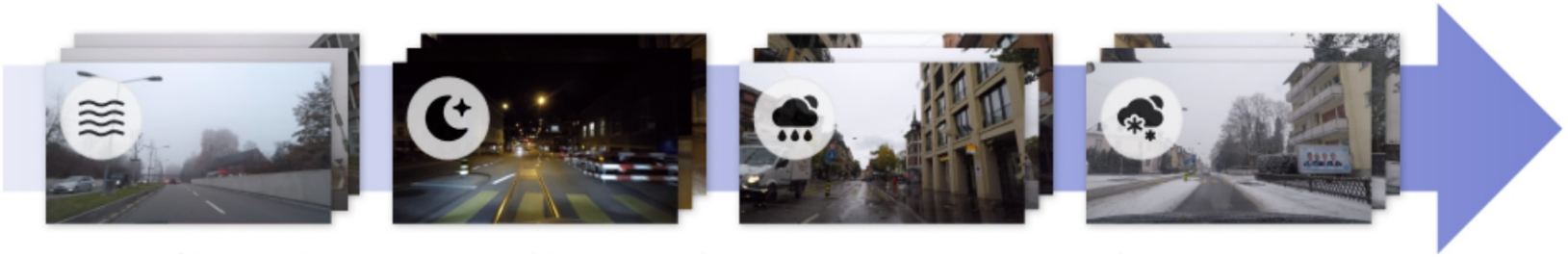
SONY

# Continually Changing Environments

Of-the-shelf pre-trained model

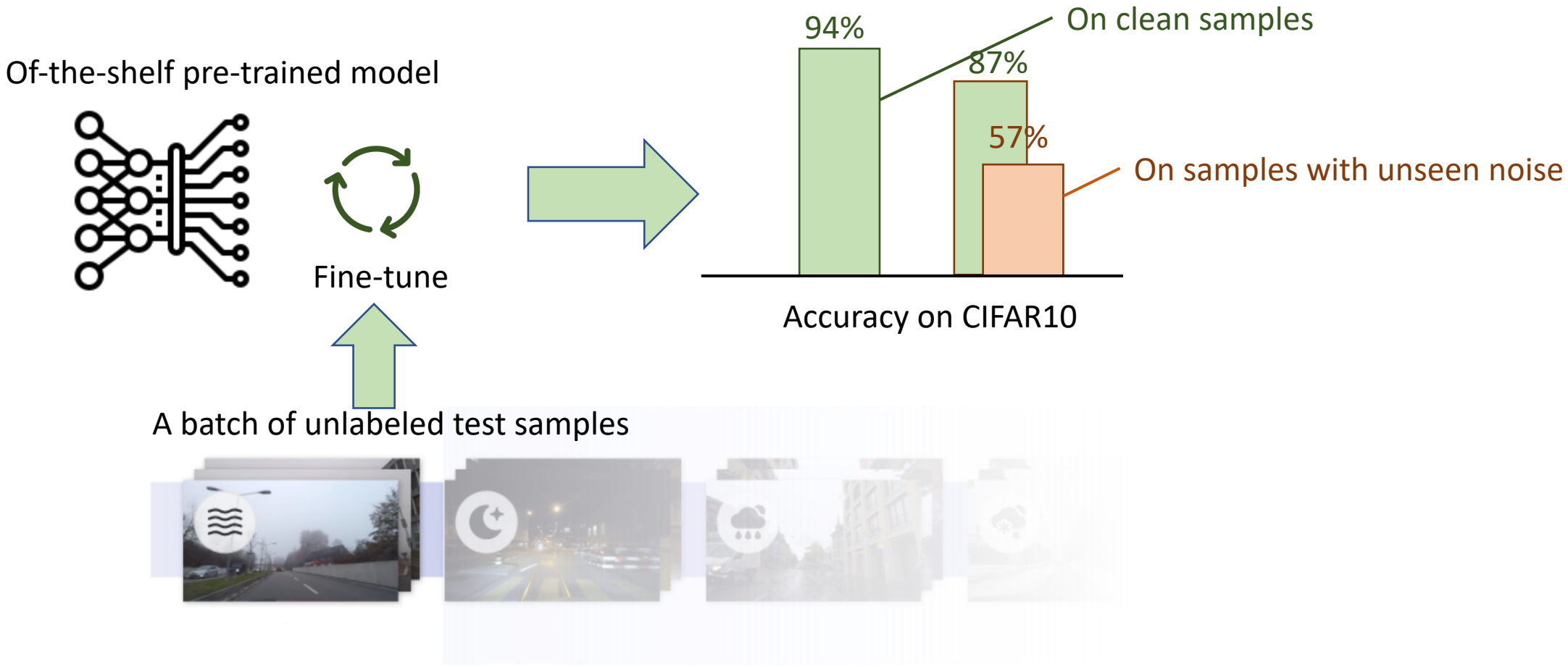




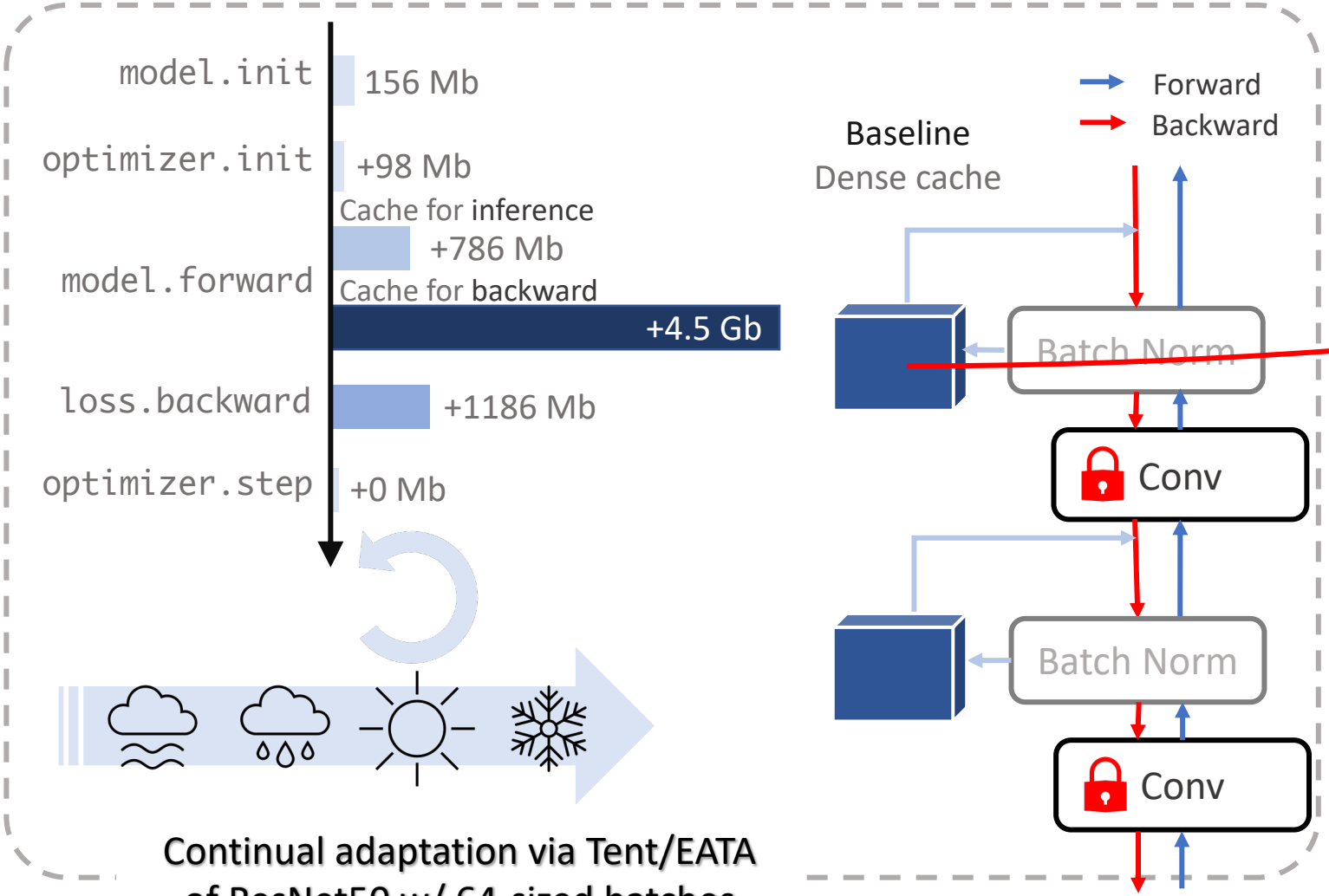
Wang, Q., Fink, O., Van Gool, L., & Dai, D. (2022). Continual test-time domain adaptation. CVPR.

# Continual Test-time Adaptation (CTA)




Wang, Q., Fink, O., Van Gool, L., & Dai, D. (2022). Continual test-time domain adaptation. CVPR.

# High memory consumption of CTA




Continual adaptation via Tent/EATA of ResNet50 w/ 64-sized batches




Batch-norm

$$z^l_{n,i,j,k} = \frac{x^l_{n,i,j,k} - \mu^l_i}{\sqrt{\sigma^2_i + \epsilon_0}} \qquad a^l_{n,i,j,k} = \gamma^l_i z^l_{n,i,j,k} + b^l_i$$

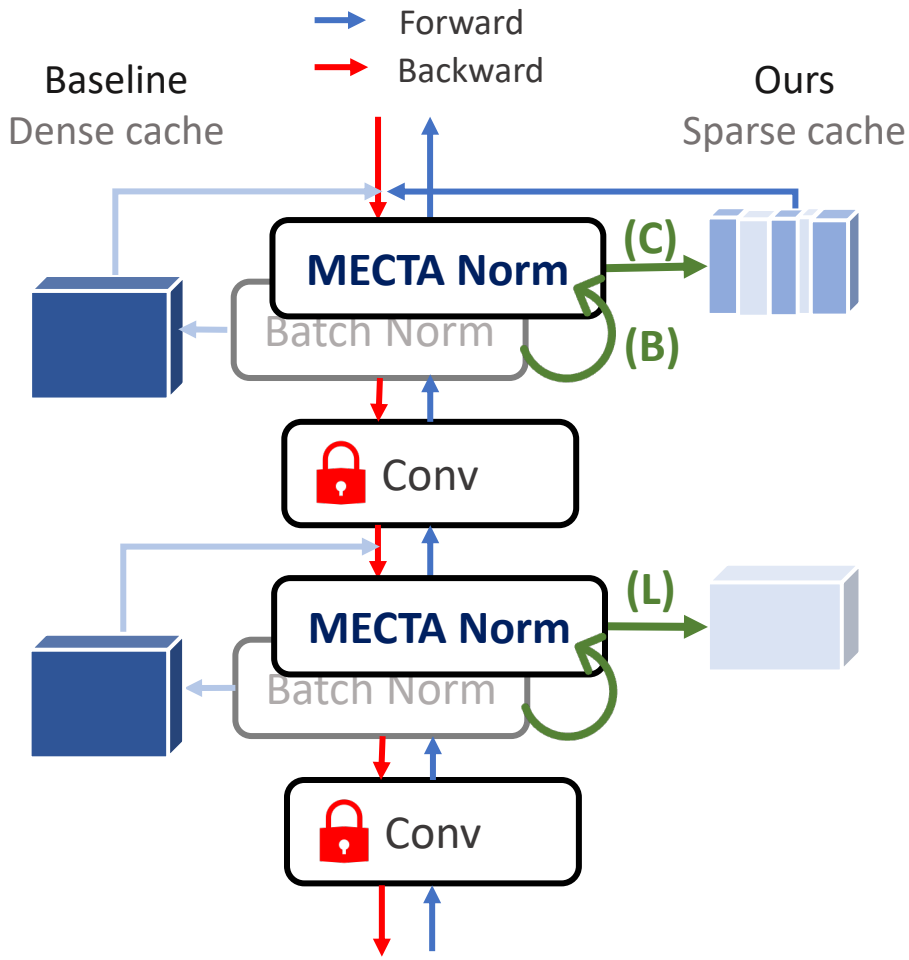
$$\sum_{n=1}^{B} \frac{\partial \ell_n}{\partial \gamma^l_i} = \sum_{n=1}^{B} \sum_{j=1}^{W} \sum_{k=1}^{H} \frac{\partial \ell_n}{\partial a^l_{i,j,k}} z^l_{n,i,j,k}.$$

Traditional CTA cannot fit into low-memory devices.

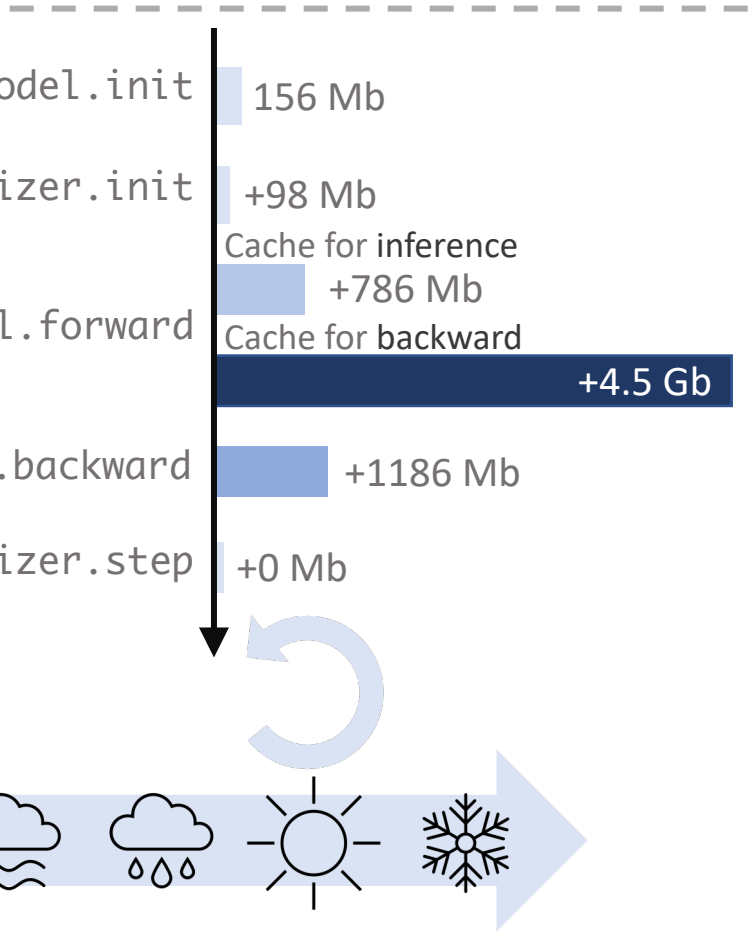
- **(B)** Require large batch size for statistic estimation.
- **(L&C)** The cache tensor $z$ scales by number of layers and channels.

# Memory-Efficient Adaptation by MECTA

- (Reduce B) Adaptive and online statistic estimation on dynamic distributions for accurate statistics on small batch sizes.
- (Reduce C) Channel-sparse gradients via stochastically-pruned caches.
- (Dynamic L) Cache and train layers on demand.

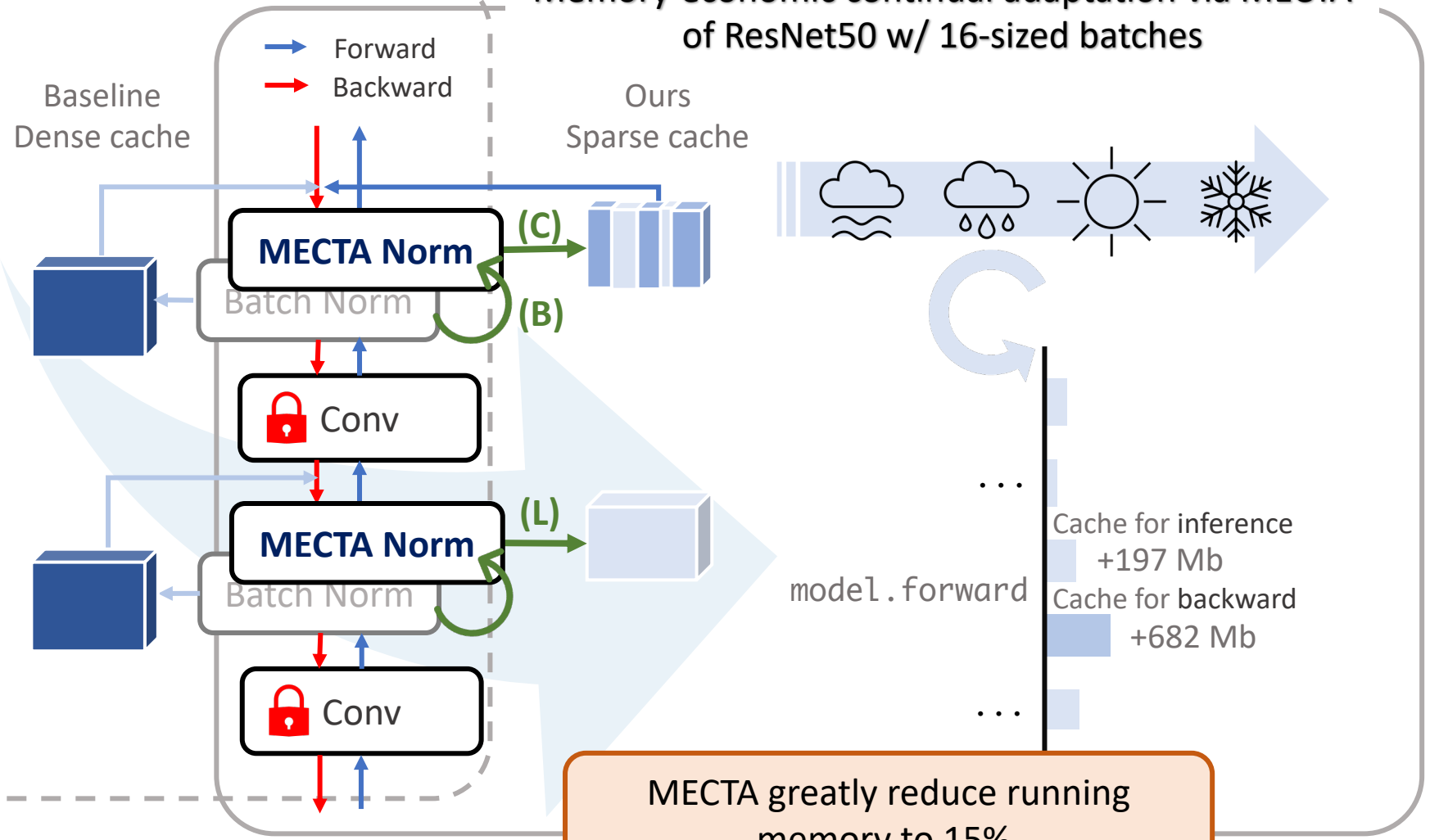


# Memory-Efficient Adaptation by MECTA




Continual adaptation via Tent/EATA of ResNet50 w/ 64-sized batches

Memory-economic continual adaptation via MECTA of ResNet50 w/ 16-sized batches




MECTA greatly reduce running memory to 15%.

# Benchmark with Constrained Cache

MECTA norm avoid forgetting of TENT.

| | | Noise | | | Blur | | | | Weather | | | | Digital | | | | | Acc. | Cache | | GFLOPs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Alg. | BS | Gauss. | Shot. | Impul. | Defoc. | Glass. | Motion | Zoom. | Snow | Frost | Fog | Bright. | Contr. | Elast. | Pixel. | JPEG | Orig. | Avg | Avg | Max | |
| | | | | | | | | | | | *ImageNet-C* | | | | | | | | | | |
| BN | 128 | 39.2 | 42.6 | 39.6 | 29.9 | 32.9 | 40.8 | 47.4 | 45.0 | 47.7 | 55.8 | 68.5 | 36.0 | 54.8 | 65.4 | 55.7 | 74.2 | 48.5 | 411 | 411 | 4.1 |
| TENT | 8 | 33.8 | 16.5 | 0.8 | 0.4 | 0.3 | 0.4 | 0.4 | 0.4 | 0.4 | 0.3 | 0.4 | 0.2 | 0.3 | 0.4 | 0.3 | 0.4 | 3.5 | 355 | 355 | 8.183 |
| +GC | 16 | 43.3 | 46.1 | 42.8 | 25.8 | 14.8 | 5.0 | 1.3 | 0.7 | 0.7 | 0.7 | 0.8 | 0.6 | 0.7 | 0.7 | 0.7 | 0.7 | 11.6 | 404 | 404 | 10.3 |
| +MECTA | 30 | **48.6** | **50.9** | 48.5 | 35.7 | **38.3** | 39.6 | **44.2** | **37.0** | **37.4** | **42.1** | **51.9** | **31.7** | **42.9** | **47.6** | **42.5** | **53.6** | **43.3** | 338 | 397 | 8.190 |
| EATA | 8 | 34.1 | 37.0 | 35.0 | 27.5 | 28.1 | 35.5 | 38.6 | 39.6 | 39.7 | 47.8 | 56.6 | 35.5 | 44.1 | 53.3 | 46.7 | 63.2 | 41.4 | 355 | 355 | 8.183 |
| +GC | 16 | 44.4 | 47.1 | 45.4 | 39.0 | 39.4 | 47.4 | 49.7 | 49.7 | 48.4 | 57.6 | 64.3 | 47.8 | 54.5 | 61.7 | 56.3 | 69.5 | 51.4 | 404 | 404 | 10.3 |
| +MECTA | 30 | **50.6** | **53.3** | **51.7** | **44.7** | **46.1** | **52.2** | **56.1** | **53.4** | **53.0** | **62.0** | **68.9** | **52.9** | **60.4** | **67.1** | **61.7** | **73.6** | **56.7** | 342 | 397 | 8.190 |

Better accuracy on all noise, reasonable computation load.

# MECTA: Memory-Economic Continual Test-Time Model Adaptation

Improve **on-device** machine learning memory efficiency on **changing** environments.

- **New Problem**: We initiate the study on the **memory efficiency** of continual test-time adaptation (CTA), revealing the substantial obstacle in practice.
- **New Method**: We propose a novel method with a simple **plug-in MECTA Norm** layer that improves the memory efficiency of different CTA methods.
- **Better Memory-Robustness Trade-off**: Our method maintains comparable performance to full back-propagation methods while significantly reducing the dynamic and maximal cache overheads.

paper

code

# MECTA: Memory-Economic Continual Test-Time Model Adaptation

**Junyuan Hong**[1*], Lingjuan Lyu[2], Jiayu Zhou[1], Michael Spranger[2]

[1]Michigan State University, [2]Sony AI

*Work done during internship at Sony AI

ICLR International Conference On Learning Representations

Sony AI

SONY