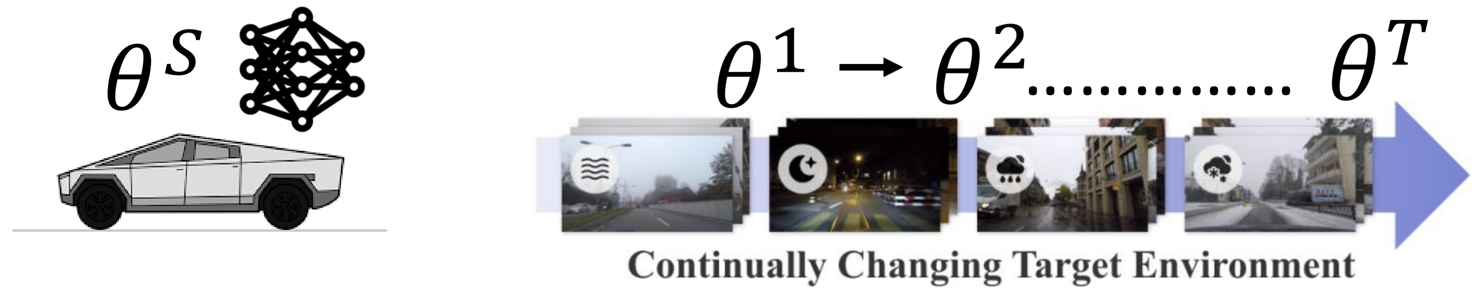
# PALM: Pushing Adaptive Learning Rate Mechanisms for Continual Test-Time Adaptation

Sarthak Kumar Maharana, Baoming Zhang, and Yunhui Guo

ERIK JONSSON
SCHOOL OF ENGINEERING AND COMPUTER SCIENCE

## Overview

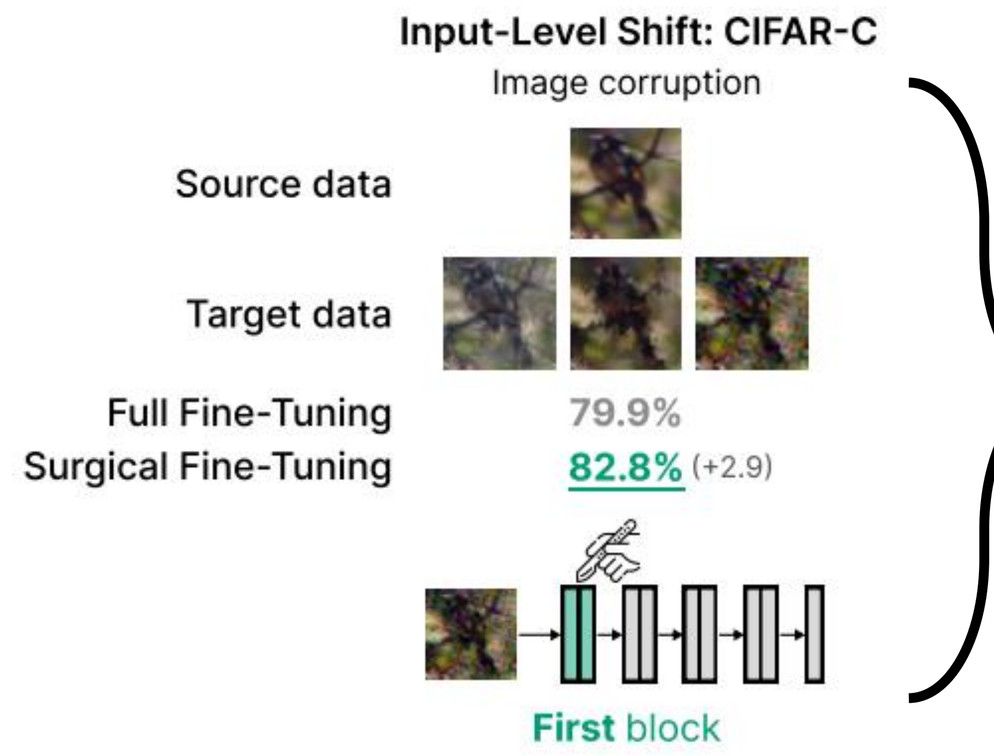
- CTTA involves **continual** source model adaptation to distributional shifts, at **test-time**, w/ no access to the source data.




### What are the issues of current adaptive learning-rate CTTA approaches?

- The learning rate importance of parameters is selected based on the Fisher Information Matrix (FIM) i.e., a dependency on the pseudo-labels.
- Direct accumulation of FIM is ignorant of the imbalance caused with time.

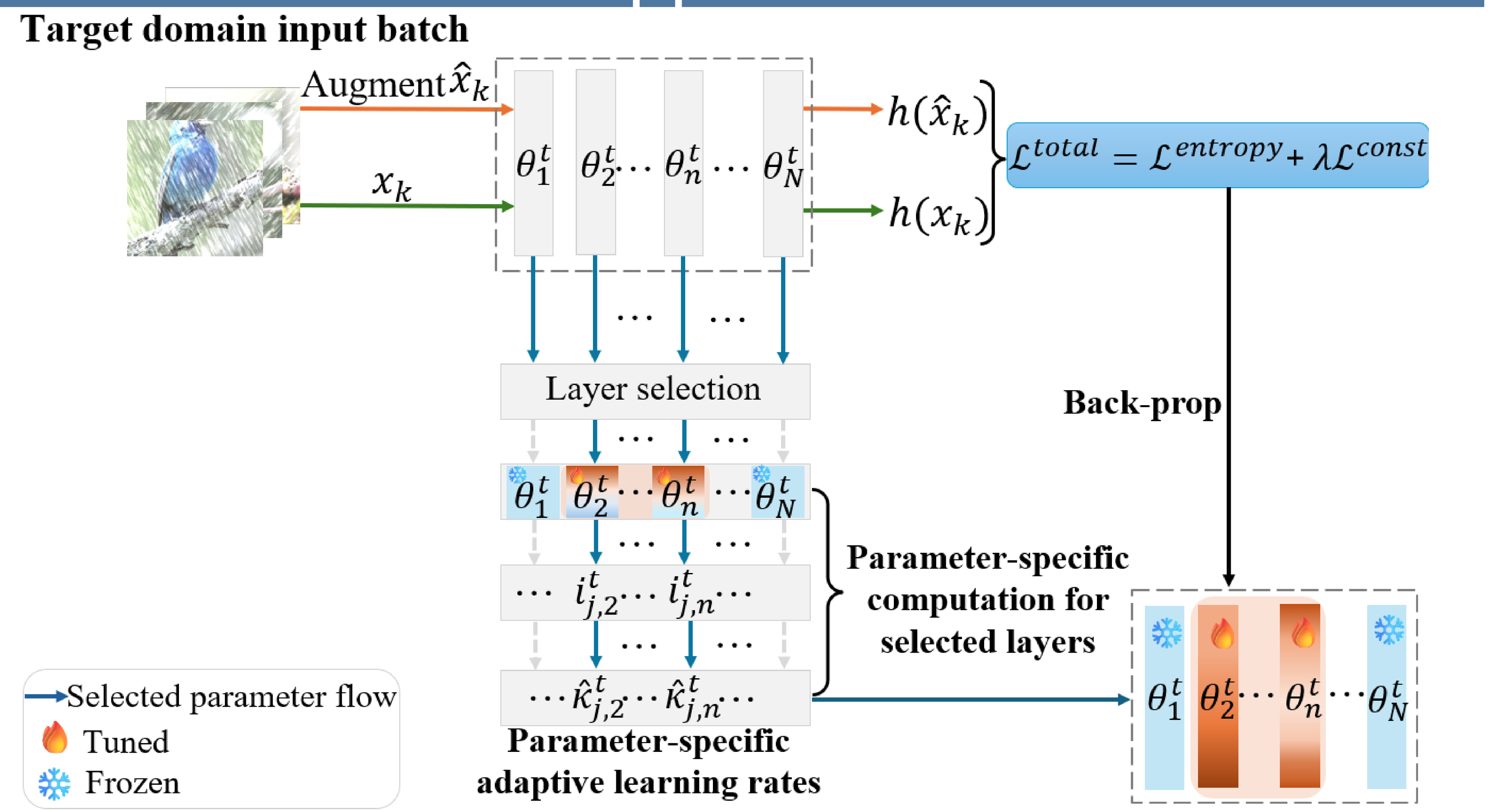
## Motivation




At test-time, it **cannot be heuristically** decided which subset of layers are suited for adaptation to rapid changes in domains.

We **automatically** select the required subset of layers for adaptation, based on the domain shift. The degree of adaptation of **selected** parameters is determined by their sensitivity to domain shift in a continual setup.

## Approach




**Selected layers are adaptively fine-tuned, based on the domain. Unselected layers are frozen to preserve the source knowledge.**

### Layer selection based on the model uncertainty:

- $L_1$ norm on the gradients computed from ***KL-divergence b/w softmax outputs (smoothed by a temp. T) and a uniform distribution***. Select layers below a threshold.

### Parameter sensitivity as a domain shift indicator:

- Sensitivity approximates error induced upon removal.

For the j[th] param of n[th] selected layer,

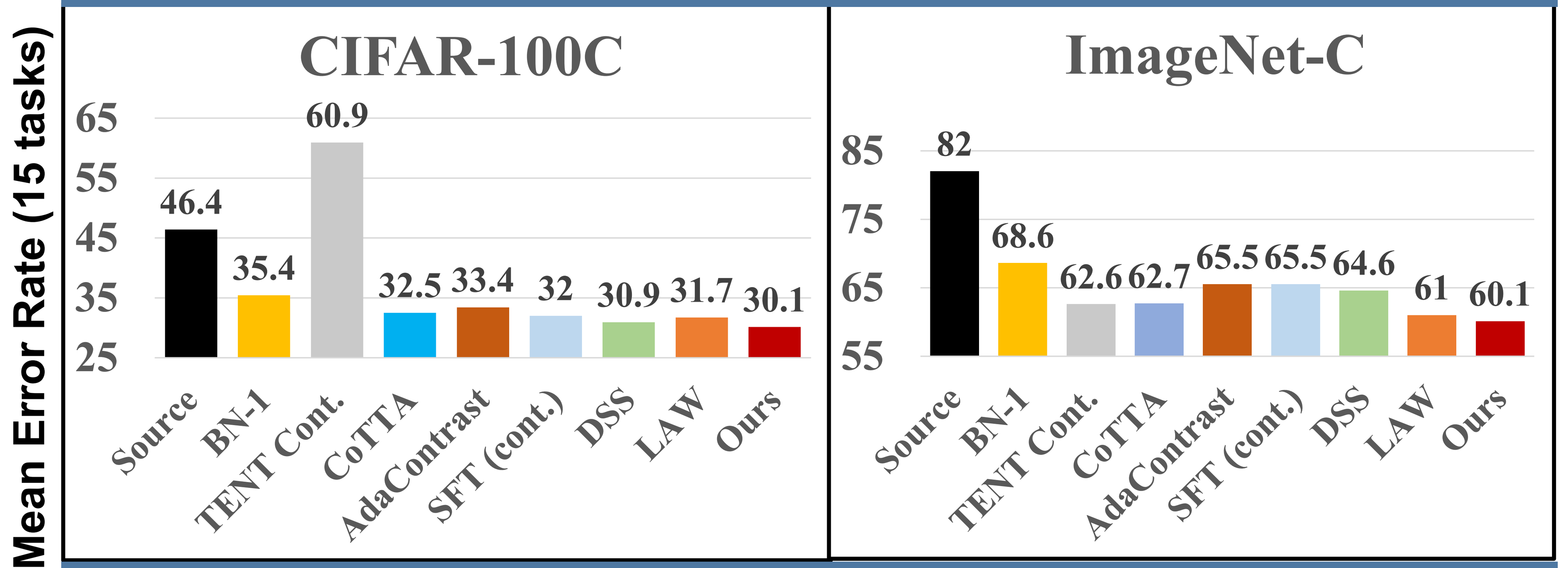
$$\underbrace{\hat{S}^t_{j,n} = \alpha S^t_{j,n} + (1-\alpha)\hat{S}^{(t-1)}_{j,n}}_{\text{"Domain-level" sensitivity}} \quad ; \quad \underbrace{\hat{D}^t_{j,n} = |S^t_{j,n} - \hat{S}^t_{j,n}|}_{\text{"Domain-level" uncertainty}}$$

$$\underbrace{i^t_{j,n} = \hat{D}^t_{j,n}/\hat{S}^t_{j,n}}_{\text{Learning rate importance}} \quad ; \quad \underbrace{\hat{\kappa}^t_{j,n} = \kappa * i^t_{j,n}}_{\text{Adjusted learning rate}}$$

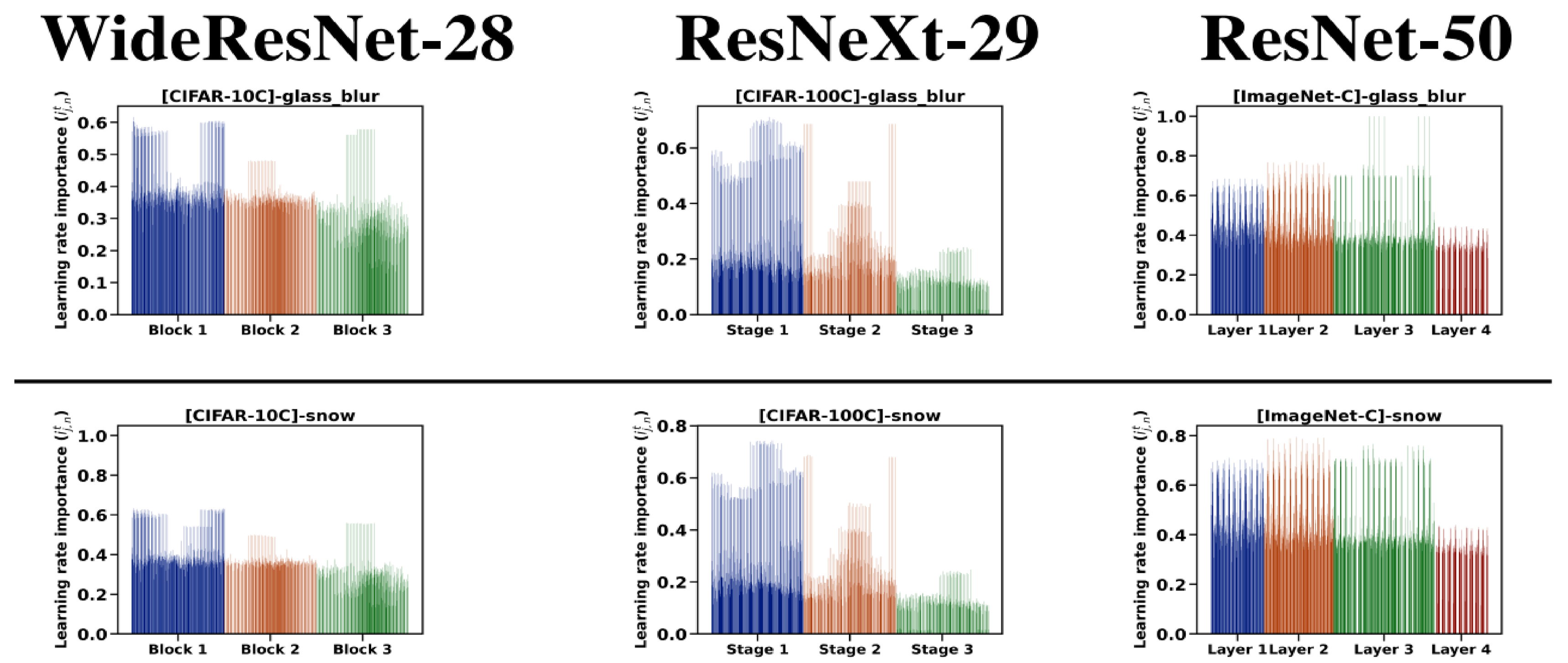
### Optimization:

- Entropy minimization + Consistency loss w/ augmentation.

## Results




## Visuals: Learning Rate Importance




PALM **automatically** selects and puts more weights on the initial layers - aligning with Surgical FT.

## Ablation Results

| Dataset | Pseudo-labels | Ours |
|---|---|---|
| CIFAR-10C | 17.9 | **15.5** |
| CIFAR-100C | 33.8 | **30.1** |
| ImageNet-C | 68.6 | **60.1** |

Layer selection based on the gradient of log-likelihood.

Large temps. capture uncertainty and push the output distribution closer to **u**.