TRUST: Test-Time Refinement using Uncertainty-Guided SSM Traverses

Sahar Dastani*, Ali Bahri*, Gustavo Adolfo Vargas Hakim, Moslem Yazdanpanah, Mehrdad Noori, David Osoweichi, Samuel Barbeau, Ismail Ben Ayed, Herve Lombaert, Christian Desrosiers
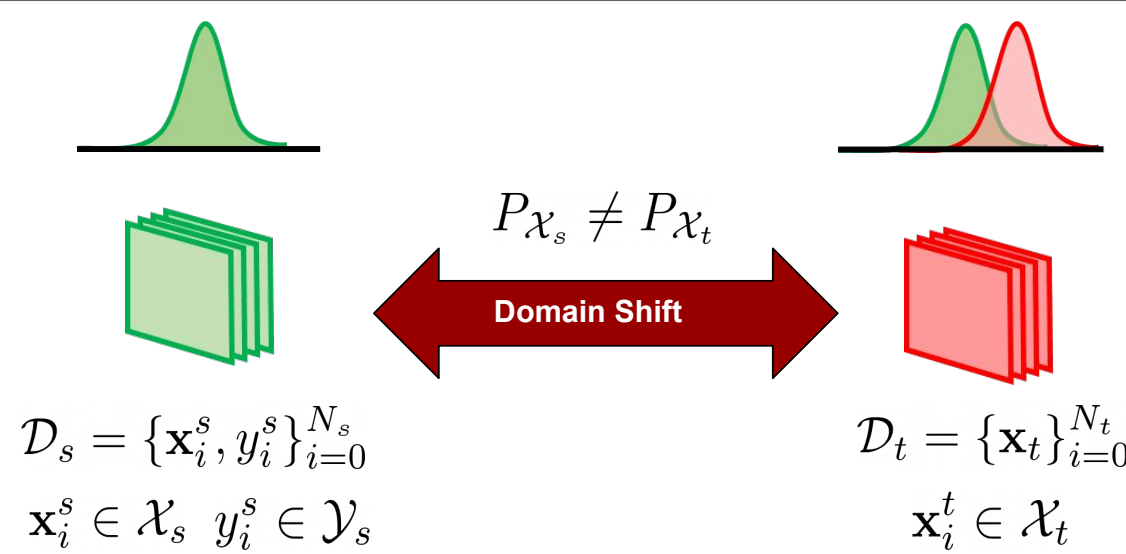
*equal contribution

TLDR

We propose **TRUST**, a TTA framework for VMamba vision state-space models under distribution shift. TRUST perturbs VMamba's four directional scans via multiple traversal permutations, adapts the model using pseudo-labels, and averages weights from the most confident traversals to obtain a single robust model. This significantly improves robustness over prior TTA methods on benchmarks such as CIFAR-C and ImageNet-C with moderate computational overhead.

Background

- Test-Time Adaptation (TTA): Adapting a pretrained model at test-time to a target dataset without accessing the source data.

$P_{\mathcal{X}_s} \neq P_{\mathcal{X}_t}$ — Domain Shift

$\mathcal{D}_s = \{\mathbf{x}_i^s, y_i^s\}_{i=0}^{N_s}$
$\mathbf{x}_i^s \in \mathcal{X}_s \; y_i^s \in \mathcal{Y}_s$

$\mathcal{D}_t = \{\mathbf{x}_t\}_{i=0}^{N_t}$
$\mathbf{x}_i^t \in \mathcal{X}_t$

- VMamba extends 1D state-space models to visual data by introducing the SS2D module, which traverses image patches in one of four predefined causal directions:

$$h(t) = Ah(t-1) + Bx(t), \quad y(t) = Ch(t)$$
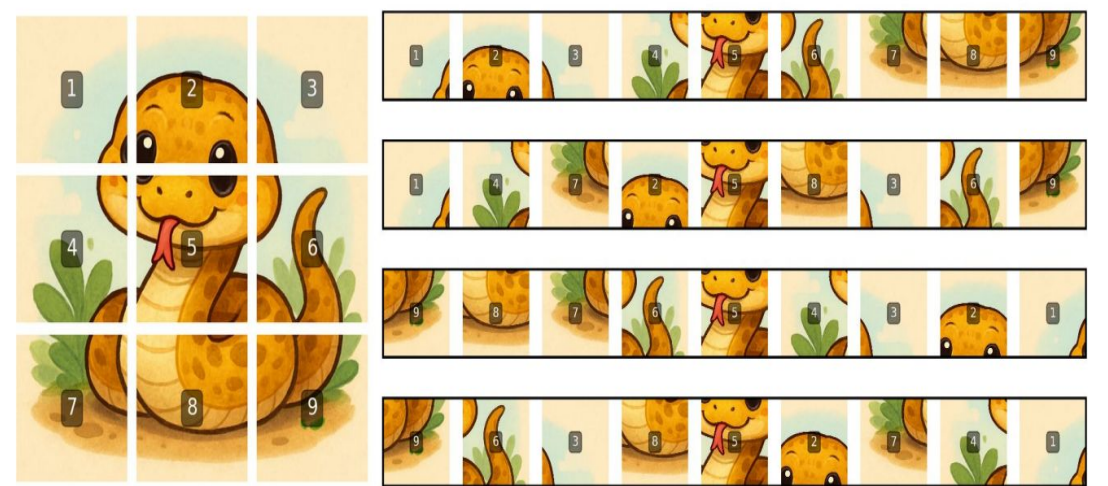
Motivations

① SSM perform poorly under distribution shifts:

- VMamba's directional processing introduces a strong inductive bias by aligning internal representations with fixed traversal paths.
- The hidden states of VMamba store historical context over the traversal sequence.

② Different traversal permutations in VMamba yield predictions with different entropy, reflecting differences in their reliability and performance.

③ Re-ordering prevents domain-specific artifacts from always entering the recurrent update at the same time step. Because the hidden state accumulates over time, an artifact injected early can affect many later updates, so its impact strongly depends on when it appears in the sequence.

$$h^{(1)}(t_\varepsilon) = f\left(h^{(1)}(t_\varepsilon - 1), \, x_{t_\varepsilon} + \varepsilon\right)$$

Methodology
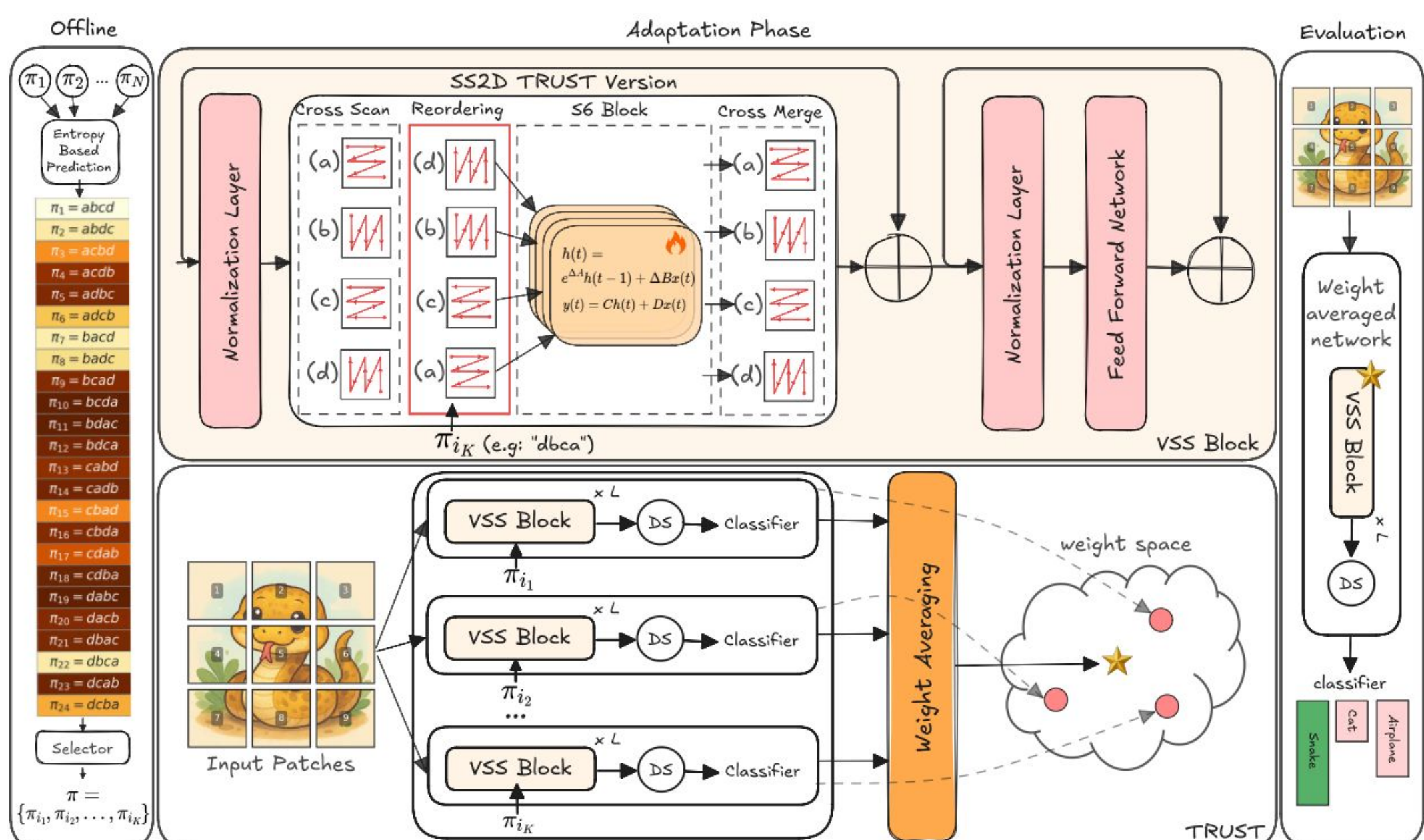
① Offline Phase

For each traversal permutation $\mathcal{P}_{\text{selected}} = \{\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_K}\}$, we compute the output Shannon entropy, rank all permutations, and select the K with the lowest entropy.

② Adaptation Phase

For each selected traversal, we process the input, take the highest-probability class as a pseudo-label $\hat{y}_k$, and update the Mamba parameters by minimizing cross-entropy with this pseudo-label over the target batch.

$$\hat{y}_k = \arg\max_{c \in \{1,\ldots,C\}} \left[p(X; \pi_{i_k})\right]_c, \quad \boldsymbol{\theta}_k = \arg\min_{\boldsymbol{\theta}} -\frac{1}{|\mathcal{B}|} \sum_{X \in \mathcal{B}} \log\left[p(X; \pi_{i_k})\right]_{\hat{y}_k}$$

③ Evaluation Phase

We aggregate the adapted models from the top-K traversal permutations by averaging their weights, and at evaluation we use this averaged model with the default traversal path.
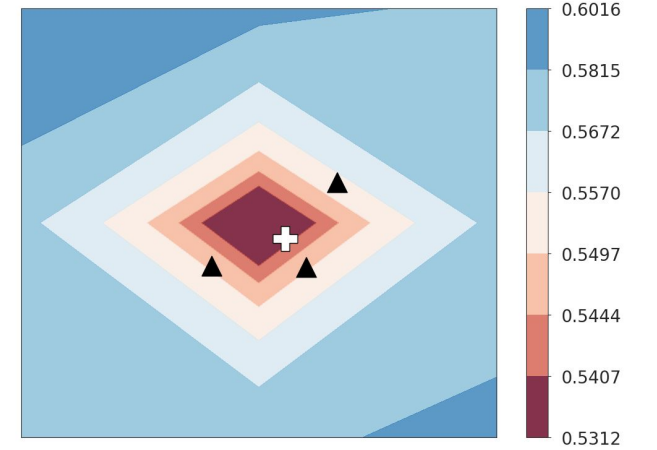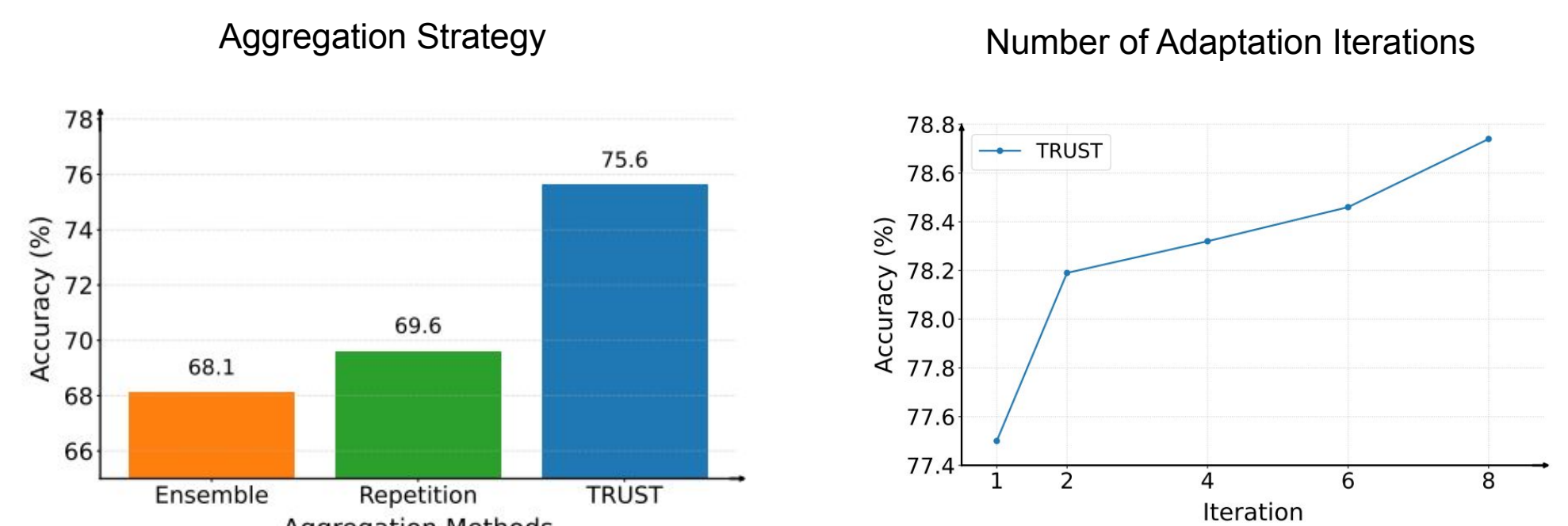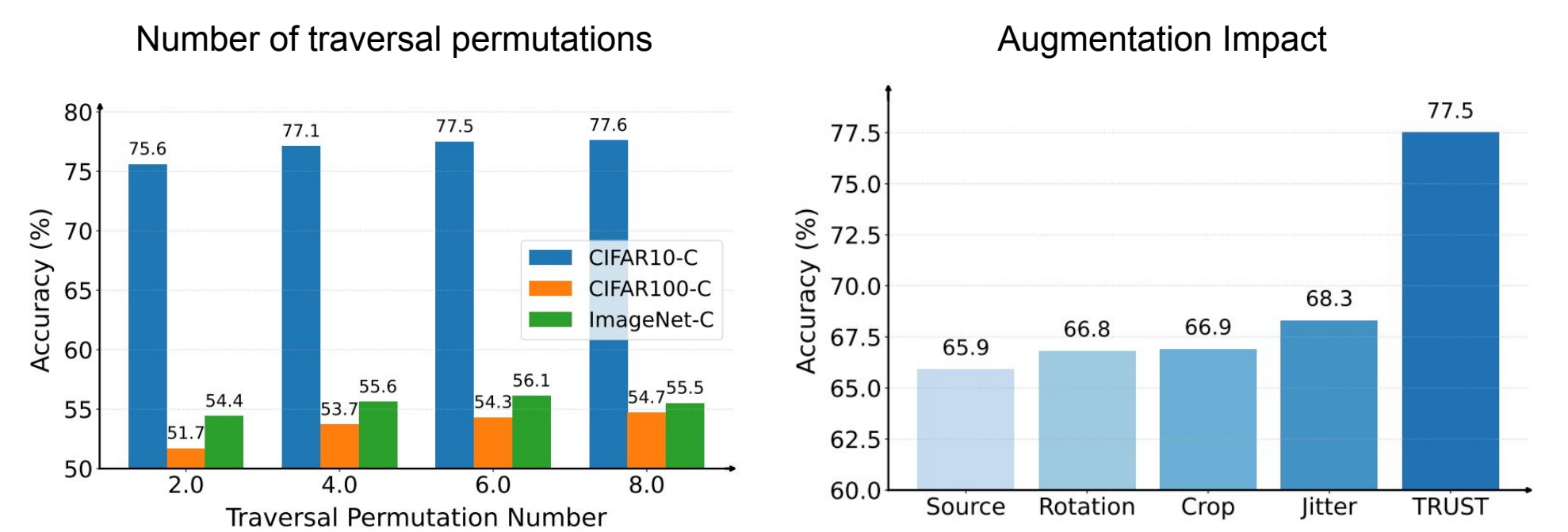

Results

- ### Classification Results

Method	CIFAR10-C	CIFAR100-C	ImageNet-C	ImageNet-S	ImageNet-V2	ImageNet-R	PACS
Source only	65.9	41.2	38.7	31.4	62.2	31.3	66.7
ETA [21]	65.8 (↓0.1)	41.4 (↑0.2)	40.8 (↑2.1)	31.4	62.2	31.3	66.7
LAME [31]	65.9	41.2	38.8 (↑0.1)	31.4	62.2	31.3	66.7
SAR [23]	66.8 (↑0.9)	41.9 (↑0.7)	41.5 (↑2.8)	32.6 (↑1.2)	62.4 (↑0.2)	32.0 (↑0.7)	67.3 (↑0.6)
SHOT [19]	66.8 (↑0.9)	42.0 (↑0.8)	41.7 (↑3.0)	32.6 (↑1.2)	62.4 (↑0.2)	31.9 (↑0.6)	67.4 (↑0.7)
TENT [20]	66.5 (↑0.6)	41.7 (↑0.5)	41.7 (↑3.0)	32.5 (↑1.1)	62.3 (↑0.1)	31.9 (↑0.6)	67.4 (↑0.7)
TRUST naive	74.2 (↑8.3)	49.8 (↑8.6)	53.4 (↑14.7)	41.1 (↑9.7)	63.4 (↑1.2)	39.7 (↑8.4)	67.1 (↑0.4)
TRUST	77.5 (↑11.6)	54.3 (↑13.1)	56.1 (↑17.4)	41.5 (↑10.1)	64.0 (↑1.8)	44.3 (↑13.0)	69.9 (↑3.2)

- ### Segmentation Results

Dataset	Method	gaussian noise	shot noise	impulse noise	defocus blur	glass blur	motion blur	zoom blur	frost	snow	fog	brightness	contrast	elastic	pixelate	jpeg compression	Mean
V21	Source only	29.1	33.1	28.3	21.0	8.2	33.1	25.4	50.9	50.3	70.7	76.5	63.9	25.5	22.2	59.2	39.8
	Tent	33.0	35.7	32.0	22.3	14.7	38.2	25.3	46.5	49.0	60.2	63.9	66.2	38.5	28.8	43.9	39.9
	TRUST	38.8	42.0	38.7	29.8	22.6	45.1	29.8	50.5	53.5	63.4	66.4	68.5	45.1	37.7	48.6	45.4 (↑5.6)
P59	Source only	17.1	19.6	17.4	27.4	14.9	29.2	19.5	30.2	28.5	42.1	50.8	41.0	23.9	30.4	38.4	28.7
	Tent	17.6	18.9	17.8	22.2	15.9	27.5	17.9	26.9	30.0	36.7	41.9	42.9	25.8	28.2	28.3	26.6
	TRUST	24.4	27.4	25.4	24.6	21.2	30.1	19.8	29.8	32.8	39.2	42.4	43.2	31.5	36.1	31.6	30.6 (↑1.9)

Ablation Study

Number of traversal permutations

Augmentation Impact

Aggregation Strategy

Number of Adaptation Iterations

Batch Norm Adaptation

|	Method	gaussian	shot	impulse	defocus	glass	motion	zoom	frost	snow	fog	brightness	contrast	elastic	pixelate	jpeg	Mean
	Source only	24.3	26.1	25.1	22.2	23.2	35.4	43.2	49.3	48.4	56.9	70.0	26.8	45.1	43.7	41.4	38.7
BN	TENT	27.8	30.0	28.8	24.9	25.9	38.0	45.5	51.0	51.3	59.0	70.6	30.0	48.2	47.8	45.7	41.7
	TRUST	32.8	35.1	34.0	26.8	28.5	40.8	47.7	52.9	53.8	61.1	71.3	34.1	50.7	51.8	50.2	44.8 (↑3.1)
SS2D	TENT	29.6	31.7	34.7	25.1	22.0	45.7	44.8	46.0	55.8	62.2	69.8	32.5	52.8	56.8	54.3	44.3
	TRUST	46.8	49.4	48.5	42.8	40.8	57.1	57.9	57.3	61.7	66.8	71.9	54.9	61.4	63.6	60.2	56.1 (↑11.8)

Parallel Implementation
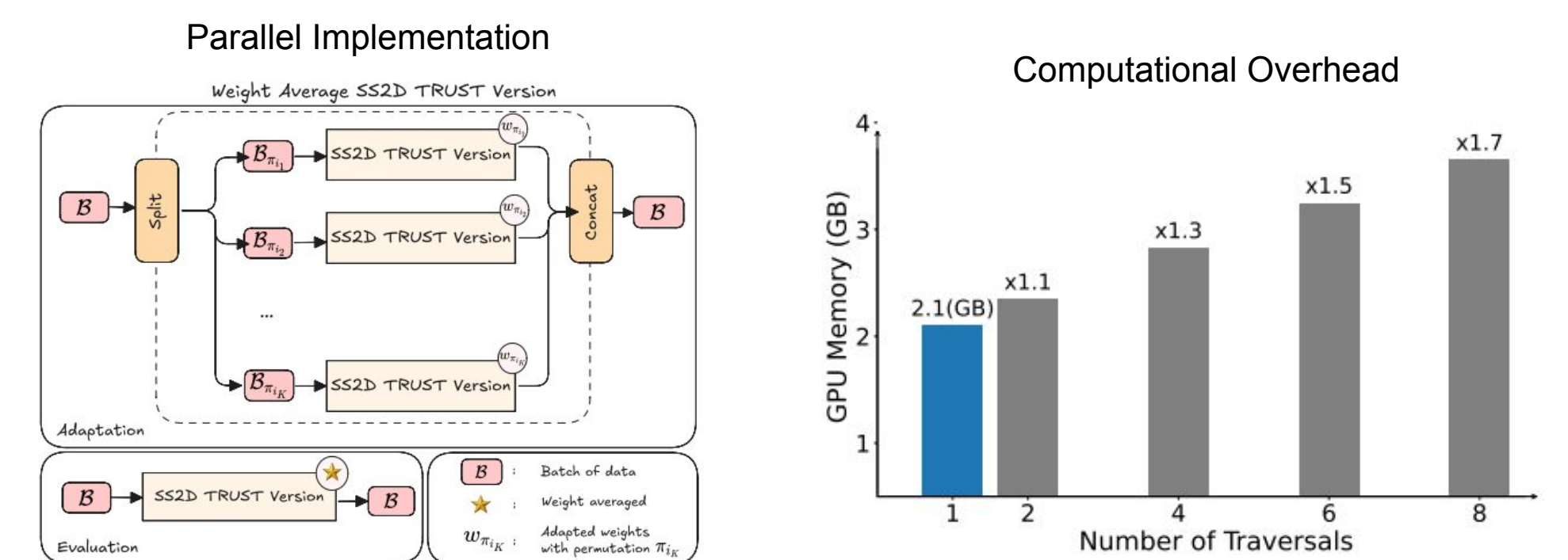
Figure 10: Detailed diagram of TRUST in Parallel mode.

Computational Overhead