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

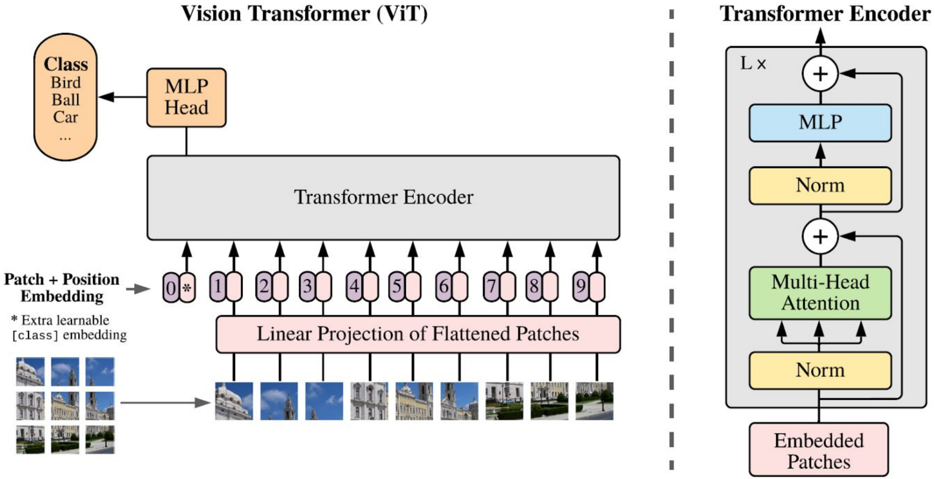
NeurIPS 2025

Table of Contents

An introduction to SSMs

The rise of ViTs

- ViTs treat an image as a sequence of patches
- We call them visual tokens
- Transformer encoder block process patches and enable the model to build global representation of the image



A Superpower: Global Attention

- Biggest strength of ViTs is self-attention
- Each patch can “look at” every other patch in the image to understand global context

A Curse for Inference

- Biggest strength of ViTs is self-attention
- Each patch can “look at” every other patch in the image to understand global context

Comes at a cost! $O\left(n^2\right)$

Can Convolutions Help?

- CNNs excel in modeling local patterns through strong inductive biases
- Struggle with global context (long-range dependencies)

Can Convolutions Help?

- CNNs excel in modeling local patterns through strong inductive biases
- Struggle with global context (long-range dependencies)

SSMs

What is State Space?

- Is a way to mathematically represent a problem by defining a system’s possible states

What is State Space?

- Is a way to mathematically represent a problem by defining a system’s possible states

What is State Space?

- Is a way to mathematically represent a problem by defining a system's possible states
- Maze navigation → state space shows:
    - where you are (current state),
    - where you can go next (possible future states),
    - and what changes take you to the next state (going right or left).

What is a State Space Model?

- SSMs are models used to describe these state representations and make predictions of what their next state could be depending on some input.

$$
\begin{aligned}
h'(t) &= Ah(t) + Bx(t) \\
y(t) &= Ch(t) + Dx(t)
\end{aligned}
$$

What is a State Space Model?

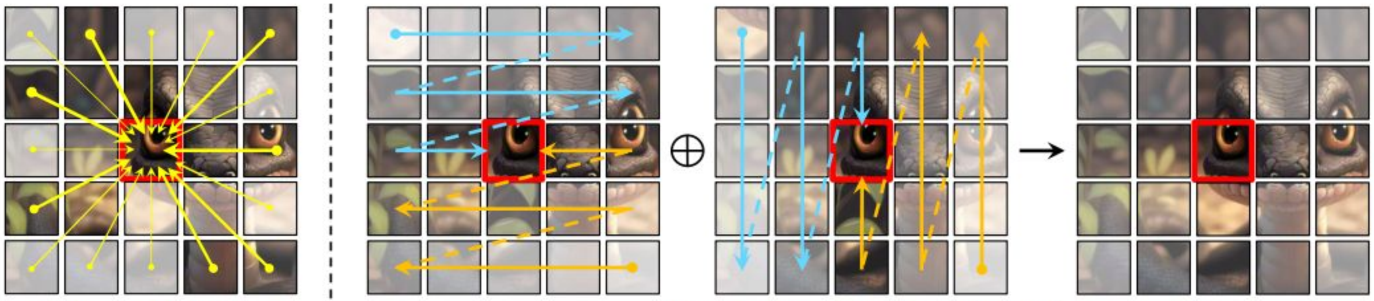
- The state equation describes:
  - How state changes (through A)
  - How the input influences the state (through B)

$$
\begin{aligned}
h'(t) &= Ah(t) + Bx(t) \\
y(t) &= Ch(t) + Dx(t)
\end{aligned}
$$

SSM for Vision

SSM in Vision → VMamba

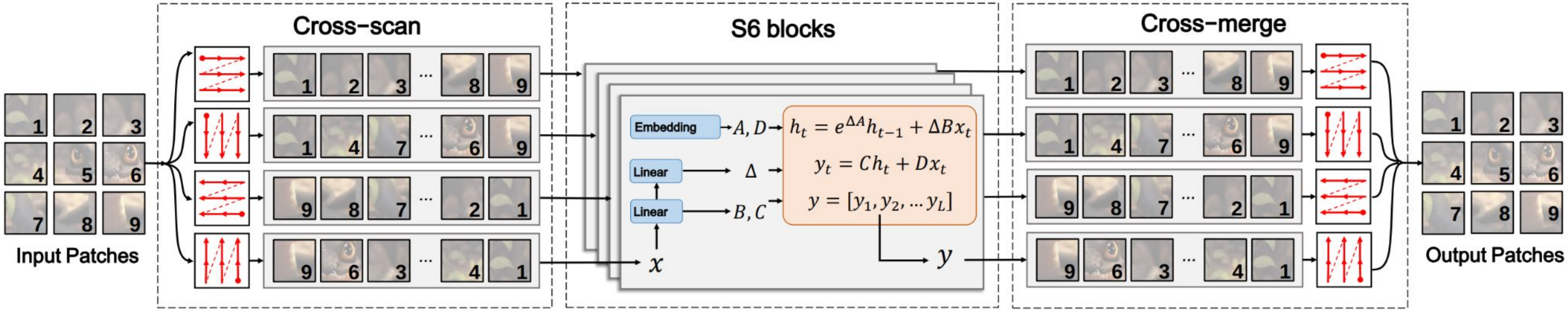
- 2D structured state space model (SS2D) to process images recursively and sequentially



(a) Self-Attention (b) 2D-Selective-Scan (SS2D)

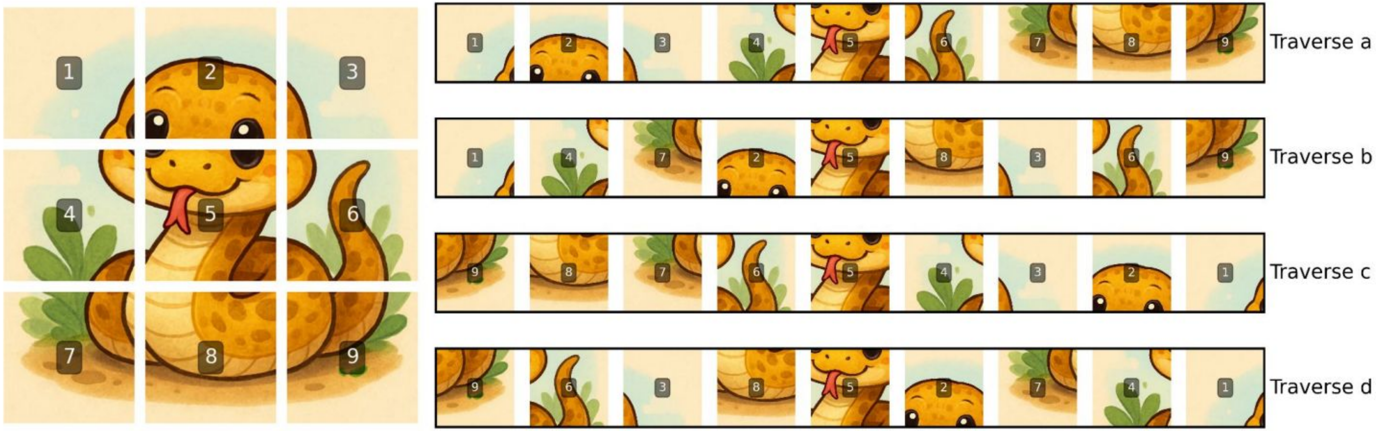
SSM in Vision → VMamba

- 2D structured state space model (SS2D) to process images recursively and sequentially



SSM in Vision → VMamba

- Cross Scan:



SSM under Distribution Shifts

SSM Performance under Distribution Shifts

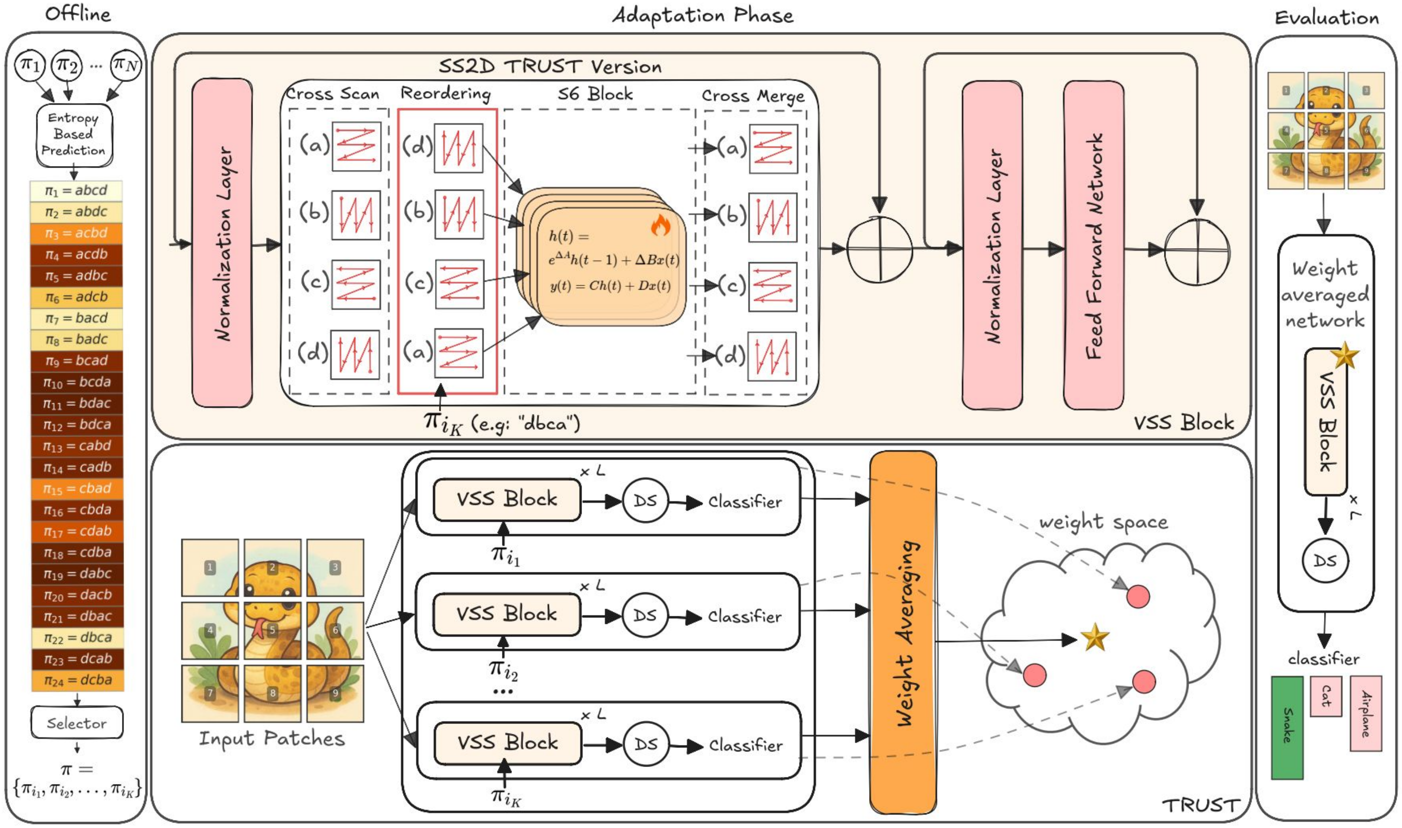
- Directional processing introduces a strong inductive bias by aligning internal representations with fixed traversal paths
- The hidden states of VMamba store historical context over the traversal sequence

TRUST

Solution: TRUST

- The first TTA approach specifically designed for Mamba-based vision models
- Our approach takes advantage of the internal traversal dynamics of VMamba
- Introduces a novel weighted averaging strategy to promote robustness
- Validation on seven standard benchmarks (SOTA among them)

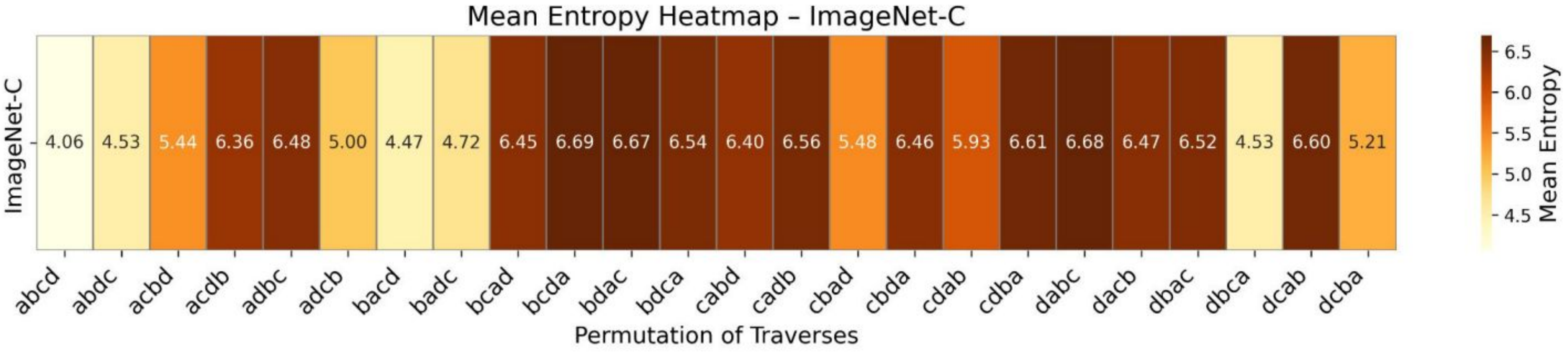
How TRUST works?



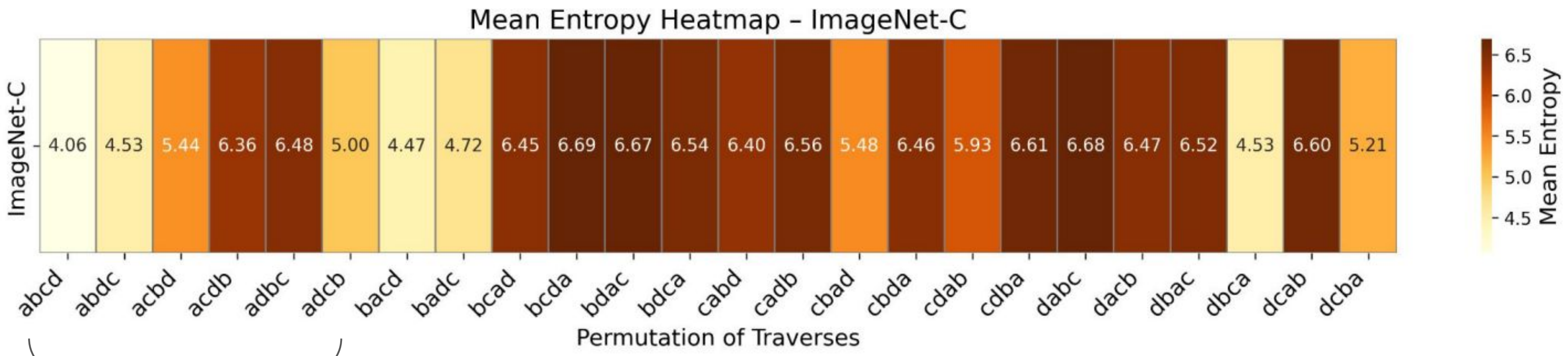
ÉTS

Mila

Offline

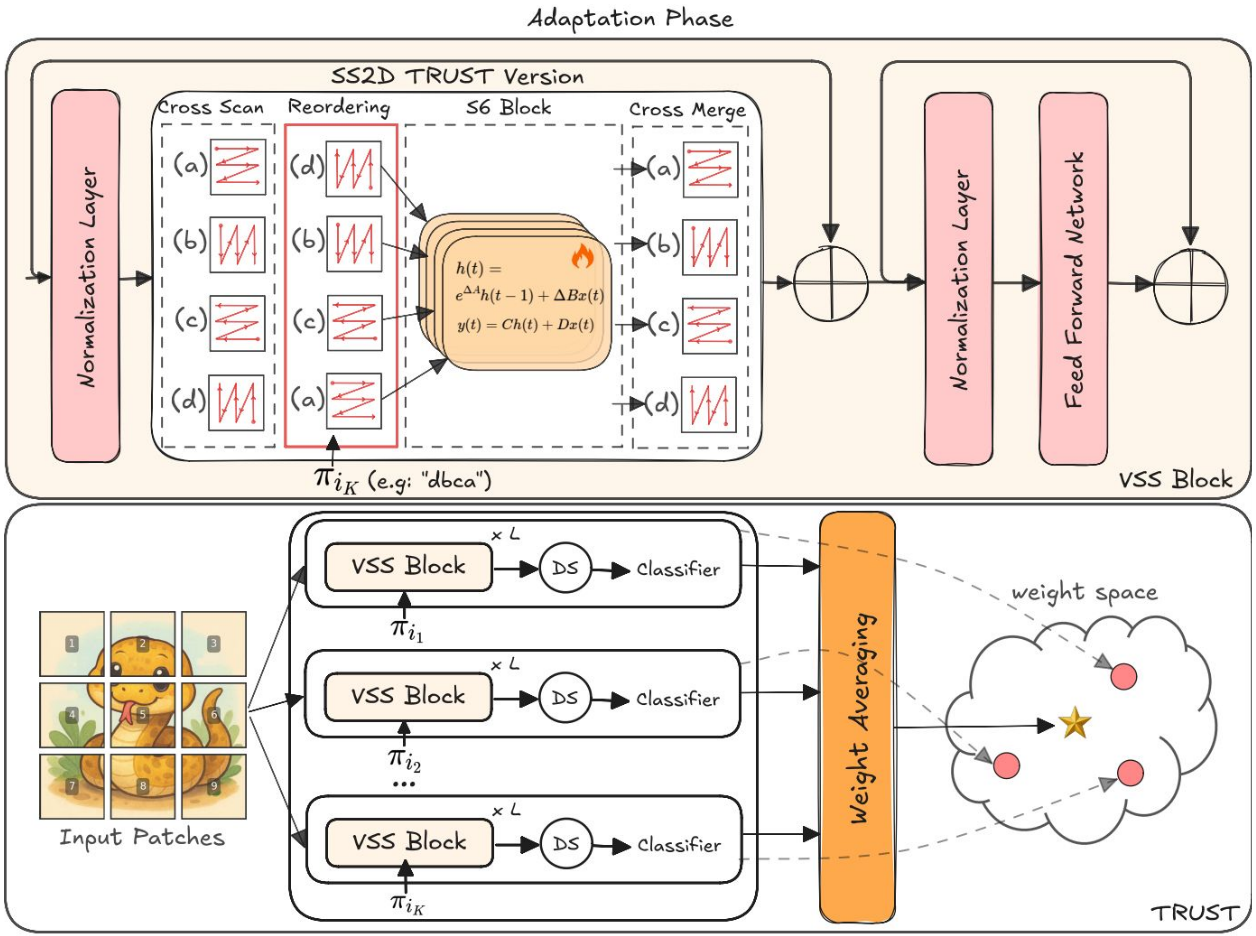

Mean Entropy Heatmap – ImageNet-C
ImageNet-C
4.06
4.53
5.44
6.36
6.48
5.00
4.47
4.72
6.45
6.69
6.67
6.54
6.40
6.56
5.48
6.46
5.93
6.61
6.68
6.47
6.52
4.53
6.60
5.21
abcd
abdc
acbd
acdb
adbc
adcb
bacd
badc
bcad
bcda
bdac
bdca
cabd
cadb
cbad
cbda
cdab
cdba
dabc
dacb
dbac
dbca
dcab
dcba
Permutation of Traverses
6.5
6.0
5.5
5.0
4.5
Mean Entropy

Offline



Top-K

Adaptation


Adaptation Phase
SS2D TRUST Version
Normalization Layer
Cross Scan
Reordering
S6 Block
Cross Merge
(a)
(b)
(c)
(d)
$h(t) = e^{\Delta A}h(t-1) + \Delta Bx(t)$
$y(t) = Ch(t) + Dx(t)$
π_{i_K} (e.g: "dbca")
Normalization Layer
Feed Forward Network
VSS Block
Input Patches
VSS Block
DS
Classifier
× L
π_{i_1}
π_{i_2}
π_{i_K}
Weight Averaging
weight space
TRUST

Adaptation

- Processes the same image through multiple directional permutations
- Enable VMamba to exploit complementary causal views of the input
- These distinct trajectories expose the model to both global consistency and local variation, which helps find a flatter minima

Global consistency
(identical token set)

Local variation
(different h(t) evolution)

How it helps in practice? (hidden state updates)

- Default traversal π_1

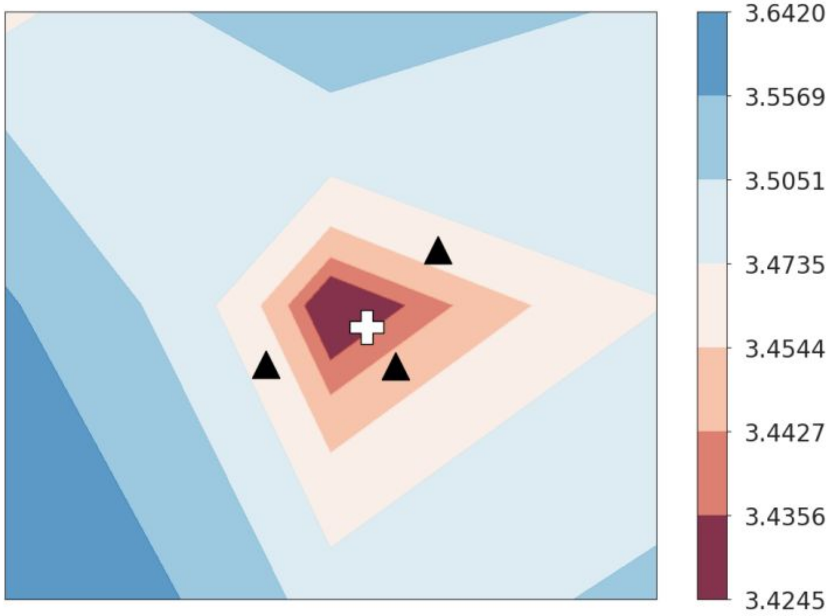
$$\mathbf{h}^{(1)}(t_\varepsilon) = f\left(\mathbf{h}^{(1)}(t_\varepsilon - 1),\, \mathbf{x}_{t_\varepsilon} + \varepsilon\right)$$

Time step

Corrupted patch

Weight Averaging

- Each triangle: model adapted via a different traversal permutations
- Flat minima in the loss landscape



Evaluation

- We tested the performance with weighted averaged network and reported the results

Datasets

- CIFAR10-C
- CIFAR100-C
- ImageNet-C
- ImageNet-Sketch
- ImageNet-V2
- ImageNet-R
- PACS

Results (Classification)

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)	41.8 (↑0.6)	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)

Results (Classification)

	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
CIFAR10-C	Source only	46.8	48.4	45.0	73.5	52.6	73.0	78.7	71.8	75.8	77.3	85.7	69.6	63.7	67.9	59.0	65.9
	ETA [21]	46.7	48.3	44.8	73.5	52.6	73.0	78.6	75.7	71.4	77.2	85.7	69.6	63.7	67.9	59.0	65.8 (↓0.1)
	LAME [31]	46.7	48.3	44.8	73.5	52.6	73.0	78.6	71.8	75.8	77.2	85.7	69.6	63.7	67.9	59.0	65.9
	SAR [23]	47.7	49.5	46.2	74.3	53.4	73.8	79.1	72.5	76.5	78.0	86.1	70.8	64.5	68.9	60.0	66.8 (↑0.9)
	SHOT [19]	47.8	49.7	46.3	74.3	53.7	74.0	79.3	72.6	76.6	78.1	86.3	70.7	64.5	68.9	59.9	66.8 (↑0.9)
	TENT [20]	47.3	49.2	45.8	74.2	53.1	73.7	79.1	72.2	76.3	77.9	86.1	70.4	64.3	68.7	59.6	66.5 (↑0.6)
	TRUST naive	58.9	61.8	62.0	79.8	60.9	79.1	82.6	80.5	81.8	83.6	88.8	81.8	70.3	75.1	66.0	74.2 (↑8.3)
	TRUST	**63.1**	**67.8**	**70.3**	**81.0**	**64.5**	**81.4**	**85.0**	**83.2**	**85.4**	**85.8**	**90.1**	**85.7**	**72.1**	**79.1**	**68.6**	**77.5 (↑11.6)**
CIFAR100-C	Source only	21.0	22.1	18.3	50.6	27.7	51.0	56.2	45.3	50.6	52.4	65.3	43.2	39.0	41.7	33.4	41.2
	ETA [21]	21.2	22.3	18.7	50.8	27.8	51.2	56.3	45.4	50.8	52.7	65.5	43.5	39.2	42.0	33.7	41.4 (↑0.2)
	LAME [31]	21.0	22.1	18.3	50.6	27.7	51.0	56.2	45.3	50.6	52.5	65.4	43.2	39.0	41.7	33.4	41.2
	SAR [23]	21.9	22.8	19.3	51.1	28.2	51.5	56.7	46.4	51.4	53.1	65.8	44.2	39.9	42.8	34.2	41.9 (↑0.7)
	SHOT [19]	21.9	22.9	19.1	51.4	28.3	51.8	56.8	46.3	51.4	53.3	66.0	44.2	39.8	42.7	34.3	42.0 (↑0.8)
	TENT [20]	21.6	22.6	18.9	51.1	28.2	51.5	56.7	46.2	51.1	53.1	65.8	44.0	39.7	42.5	34.0	41.8 (↑0.6)
	TRUST naive	32.1	32.8	34.1	56.9	35.4	57.2	61.6	54.6	57.8	60.1	69.6	55.6	46.6	50.8	41.0	49.8 (↑8.6)
	TRUST	**37.8**	**38.9**	**42.3**	**60.9**	**36.6**	**60.8**	**65.4**	**59.0**	**62.2**	**64.5**	**71.7**	**63.1**	**50.3**	**56.6**	**44.9**	**54.3 (↑13.1)**
ImageNet-C	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
	ETA [21]	26.4	28.4	27.2	23.5	24.6	37.2	45.1	50.8	51.0	58.8	70.6	29.1	47.7	46.9	45.0	40.8 (↑2.1)
	LAME [31]	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	38.8 (↑0.1)
	SAR [23]	26.5	29.2	28.0	24.5	25.3	37.4	45.1	51.0	51.7	59.1	70.5	31.5	48.2	48.6	46.3	41.5 (↑2.8)
	SHOT [19]	28.0	30.1	28.8	25.0	26.0	38.0	45.7	51.0	51.5	59.1	70.6	30.2	48.4	47.8	45.8	41.7 (↑3.0)
	TENT [20]	27.8	30.0	28.8	24.9	25.9	38.0	45.5	51.0	51.3	59.1	70.6	30.0	48.2	47.8	45.7	41.7 (↑3.0)
	TRUST naive	43.4	45.6	44.9	38.3	36.6	53.0	54.9	57.1	60.2	66.0	72.2	50.2	59.0	61.1	58.5	53.4 (↑14.7)
	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 (↑17.4)**

Results (Segmentation)

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

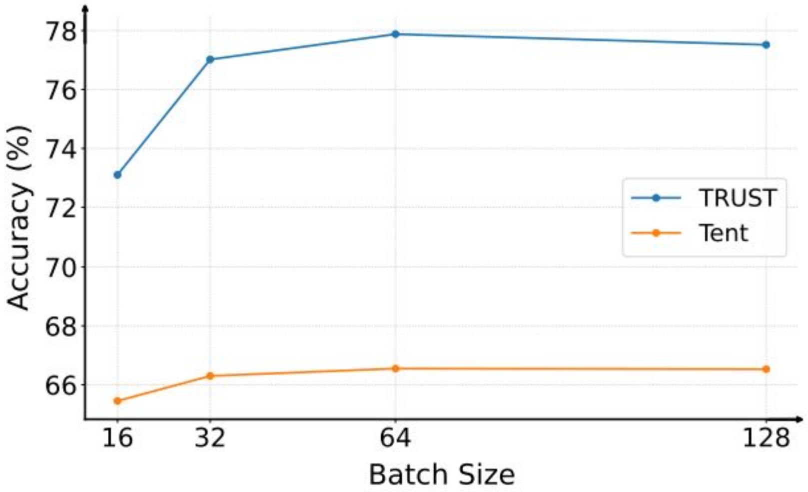


Figure 3: Accuracy comparison between TRUST and TENT across varying batch sizes on CIFAR10-C dataset.

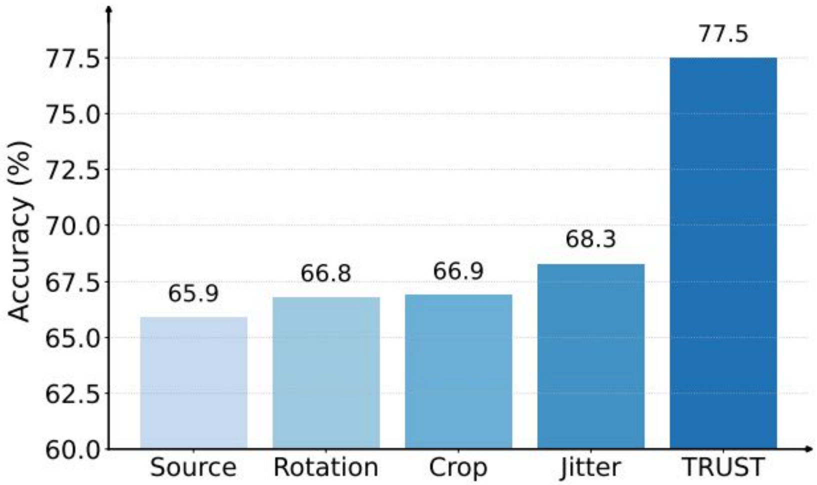


Figure 4: Performance comparison between standard augmentations and TRUST on CIFAR10-C dataset.

Ablation Study

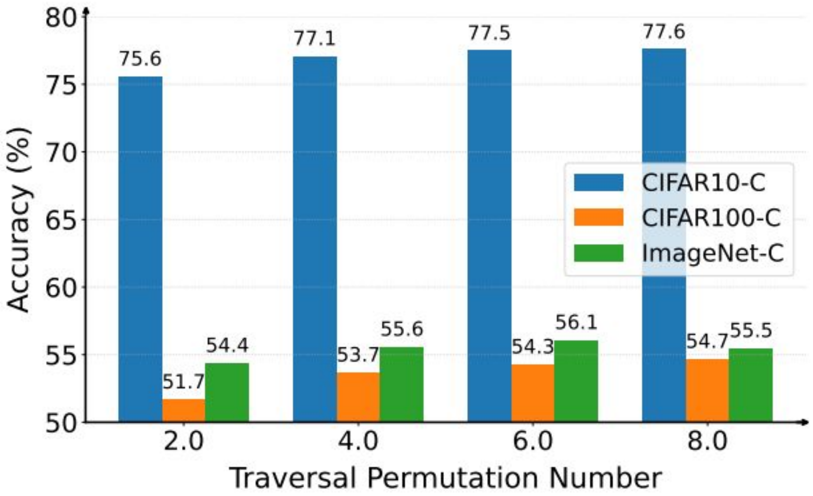


Figure 5: Effect of traversal permutation count on accuracy across three datasets.

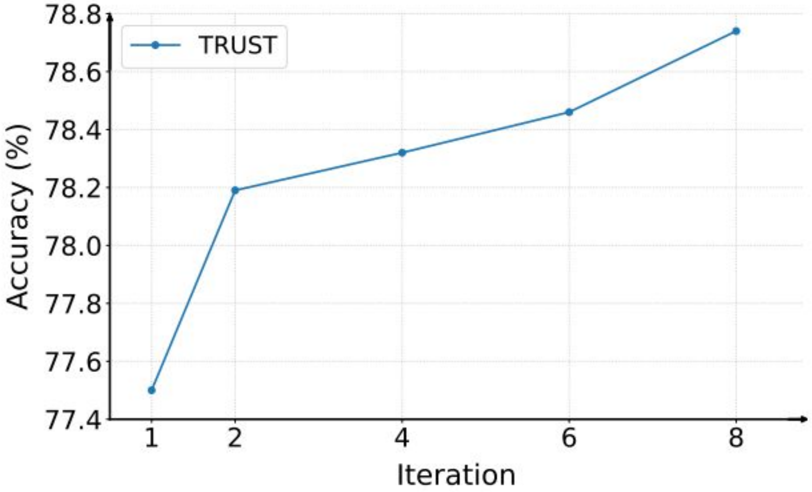


Figure 6: Model performance across adaptation iterations on CIFAR10-C dataset.

Ablation Study

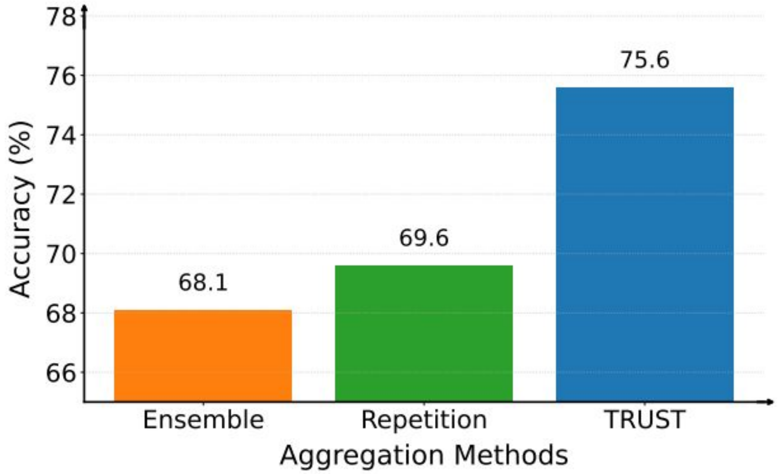


Figure 7: Accuracy comparison of different aggregation strategies on CIFAR10-C dataset.

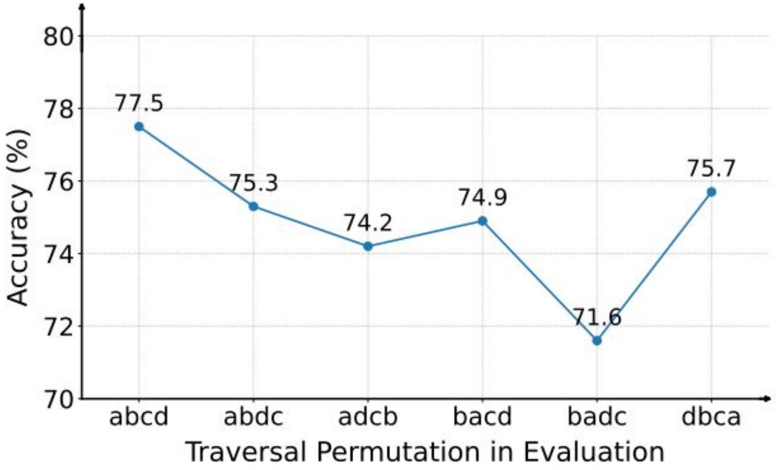


Figure 8: Impact of traversal permutation during evaluation on CIFAR10-C dataset.

Ablation Study

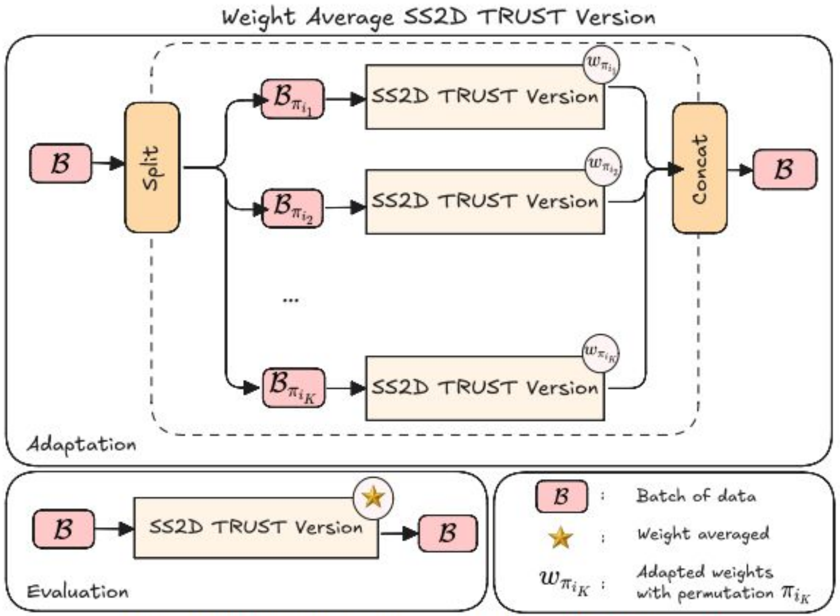


Figure 10: Detailed diagram of TRUST in Parallel mode.

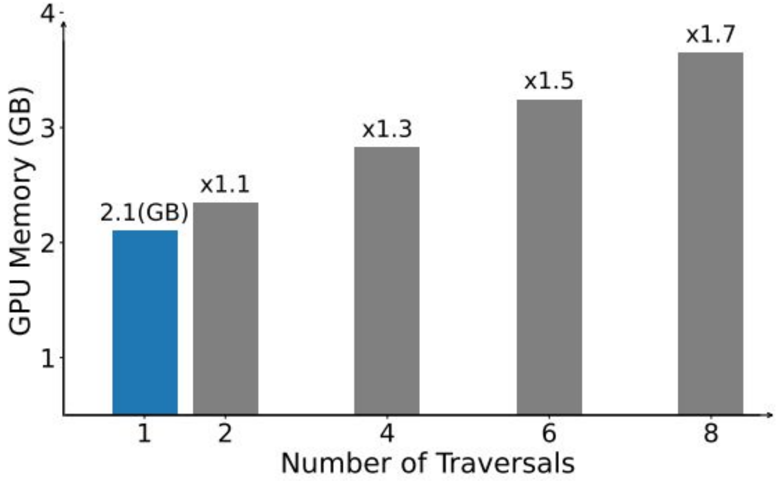


Figure 9: GPU memory usage across traversals.

Questions?