Back to the Source: Diffusion-Driven Adaptation to Test-Time Corruption

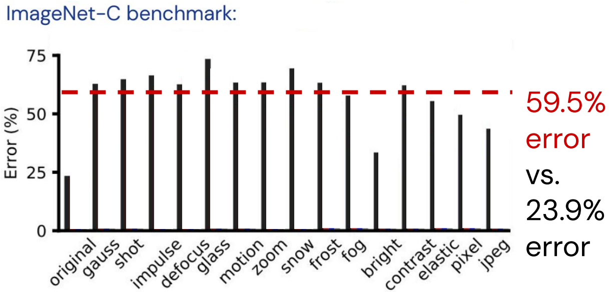
Jin Gao*, Jialing Zhang*,
Xihui Liu, Trevor Darrell,
Evan Shelhamer*, Dequan Wang*

SJTU BAIR DeepMind

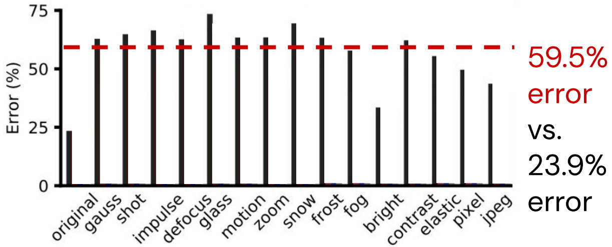
Poster #339 on Day 2: WED-AM-339
Demo #11 on Day 1: Tue. June 20, 2023

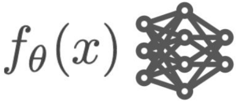
Corruption Happens, and Corruption Hurts

Corruption Happens, and Corruption Hurts



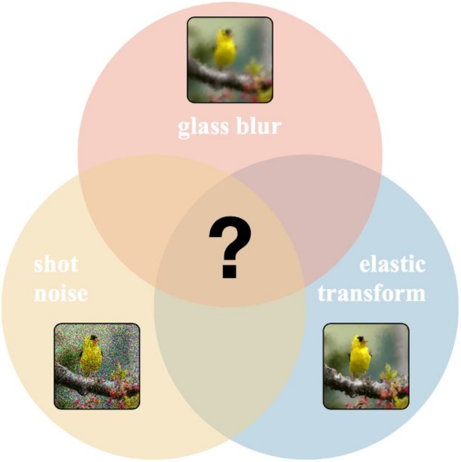
Corruption Happens, and Corruption Hurts





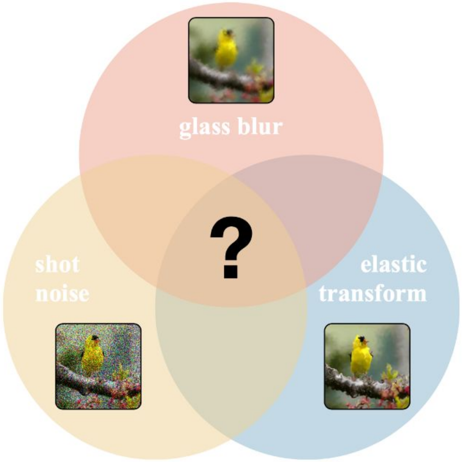
testing data varies in many ways
yet our models stay the same and
fail to generalize

Test-Time Adaptation can Counter Corruptions



How to adapt to
multiple shifts?

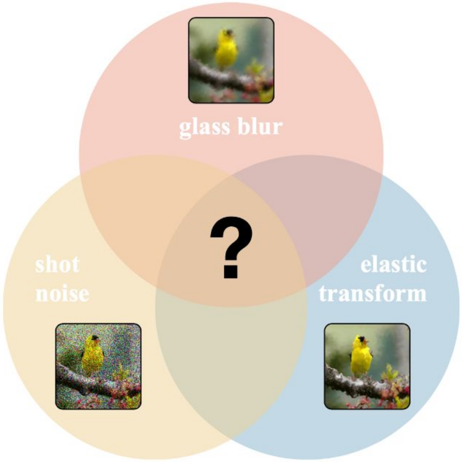
Test-Time Adaptation can Counter Corruptions



How to adapt to multiple shifts?

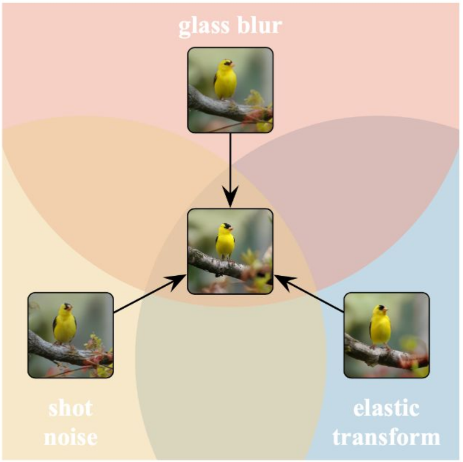
Update model for each shift

Test-Time Adaptation can Counter Corruptions



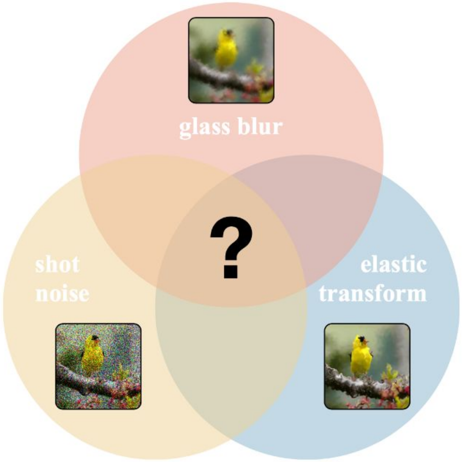
How to adapt to multiple shifts?

Update model for each shift



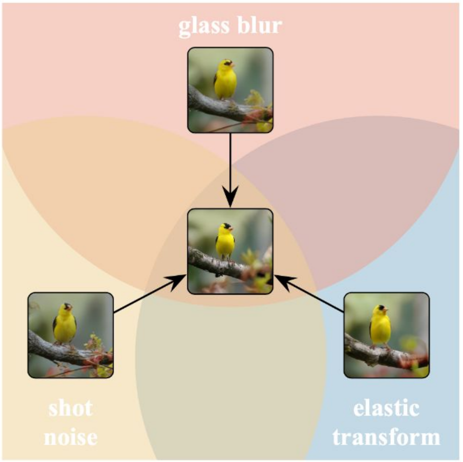
Update input for all shifts

Test-Time Adaptation can Counter Corruptions



How to adapt to multiple shifts?

Update model for each shift



Update input for all shifts

how do model and input updates compare?

Tricky Data Regimes for Test-Time Adaptation

Too Little Data

large batch

vs.

small batch

Tricky Data Regimes for Test-Time Adaptation

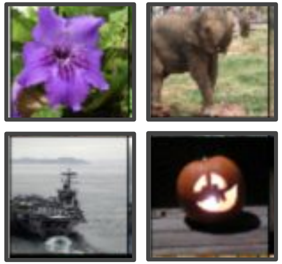
Too Little Data

large batch

vs.

small batch

Dependent Data



independent

vs.

dependent

Tricky Data Regimes for Test-Time Adaptation

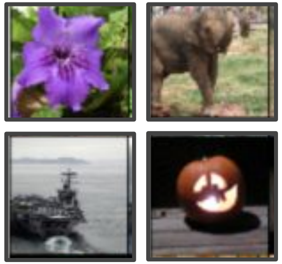
Too Little Data

large batch

vs.

small batch

Dependent Data



independent

vs.

dependent

Mixed Data

same/single shift

vs.

mixed/multiple shifts

Tricky Data Regimes for Test-Time Adaptation

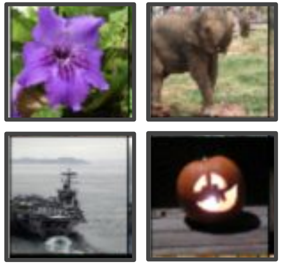
Too Little Data

large batch

vs.

small batch

Dependent Data



independent

vs.

dependent

Mixed Data

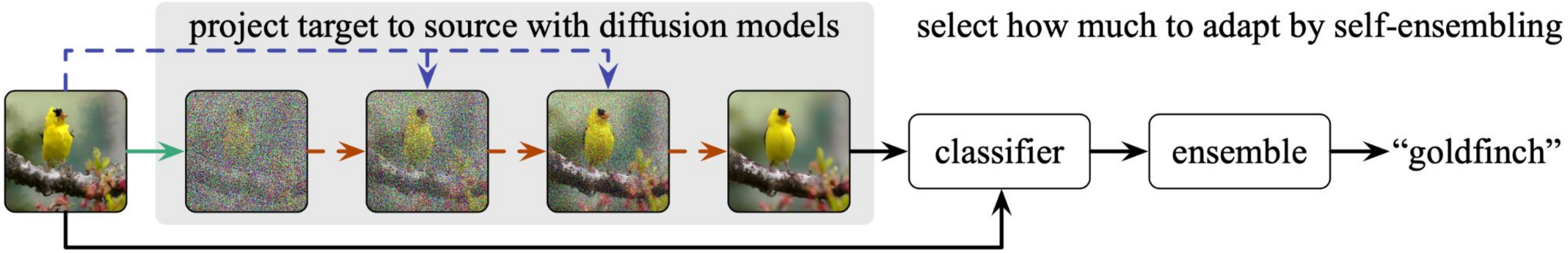
same/single shift

vs.

mixed/multiple shifts

Input updates can help!

Input Updates by Diffusion-Driven Adaptation



DDA projects target inputs back to the source domain.
Adapting the input enables direct use of the source classifier without model adaptation.

- **add noise** by perturbing the corrupted input with gaussian noise
- **iteratively update** by applying the reverse process of the diffusion model
- **refine** by guiding updates to match image structure

For reliability, we ensemble predictions with and without input adaptation weighted by confidence.

Input Updates Improve on Model Updates for Episodic Adaptation

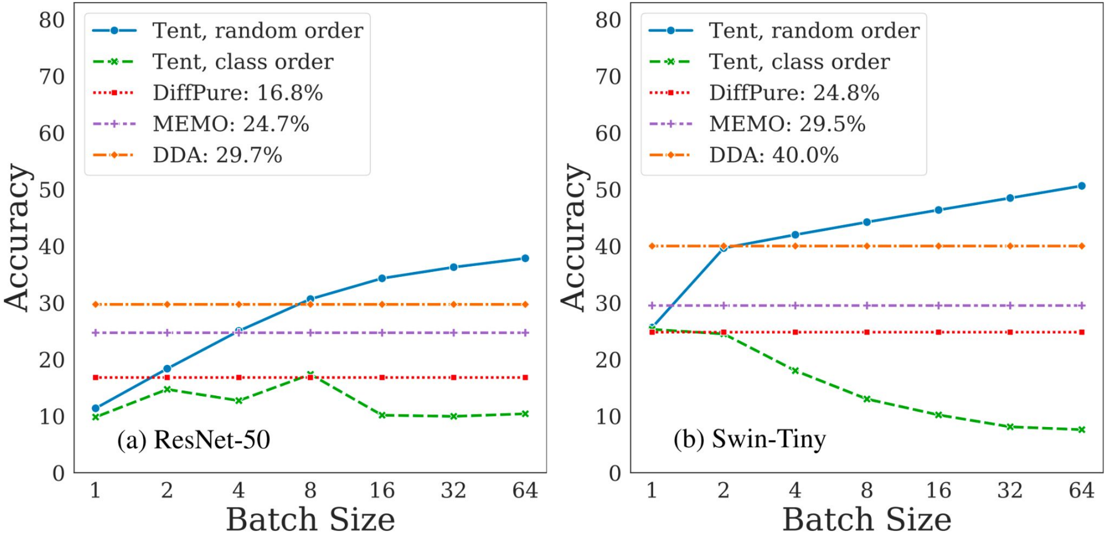
DDA **input updates improve** on MEMO model updates.
MEMO can hurt without tuning to the architecture, because model optimization can be sensitive to its hyperparameters.

Model	IN Acc.	ImageNet-C Accuracy Source-Only	MEMO	DiffPure	DDA
ResNet-50	76.6	18.7	24.7	16.8	**29.7**
Swin-T	81.2	33.1	29.5	24.8	**40.0**
ConvNeXt-T	82.1	39.3	37.8	28.8	**44.2**
Swin-B	83.4	40.5	37.0	28.9	**44.5**
ConvNeXt-B	83.9	45.6	45.8	32.7	**49.4**

DDA is architecture agnostic—although model adaptation can apply to different architectures, it may need tuning.

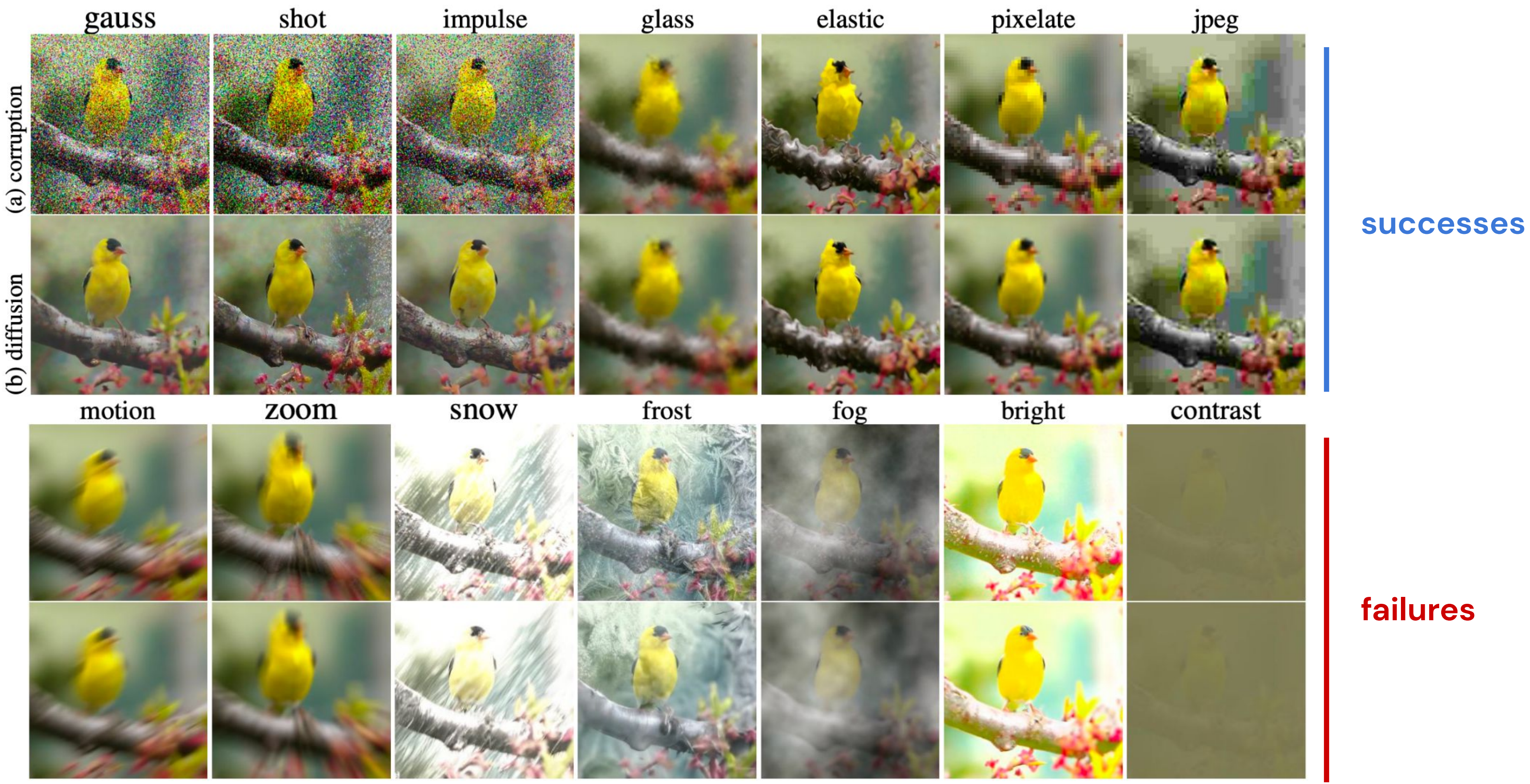
Input Updates are Insensitive to Batch Size and Order

DDA is episodic so batching & ordering do not matter.
For Tent, not so! Online model updates are sensitive.

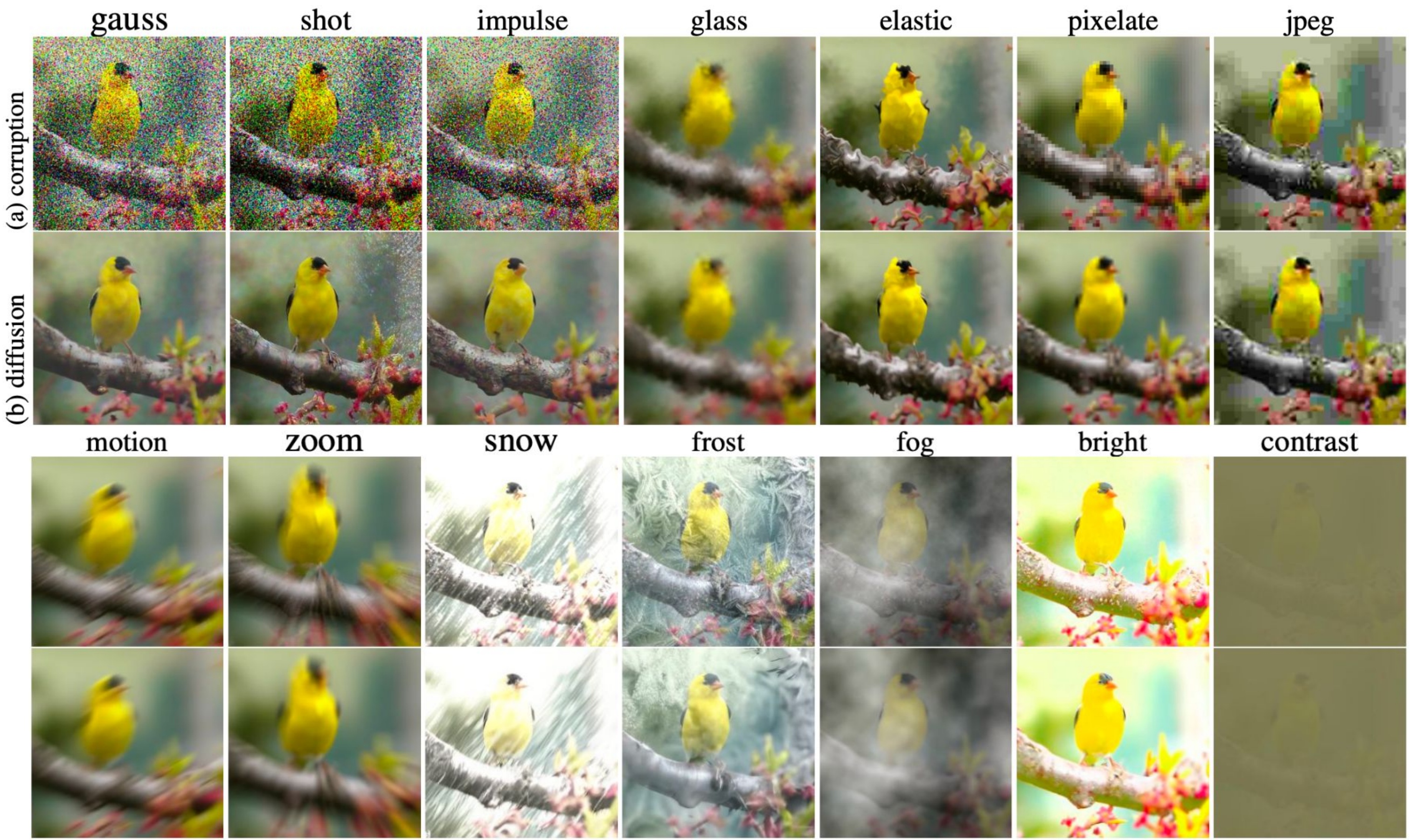


DDA can adapt on inputs one-by-one in whatever order.

Diffusing Inputs to Defuse Corruptions



Diffusing Inputs to Defuse Corruptions



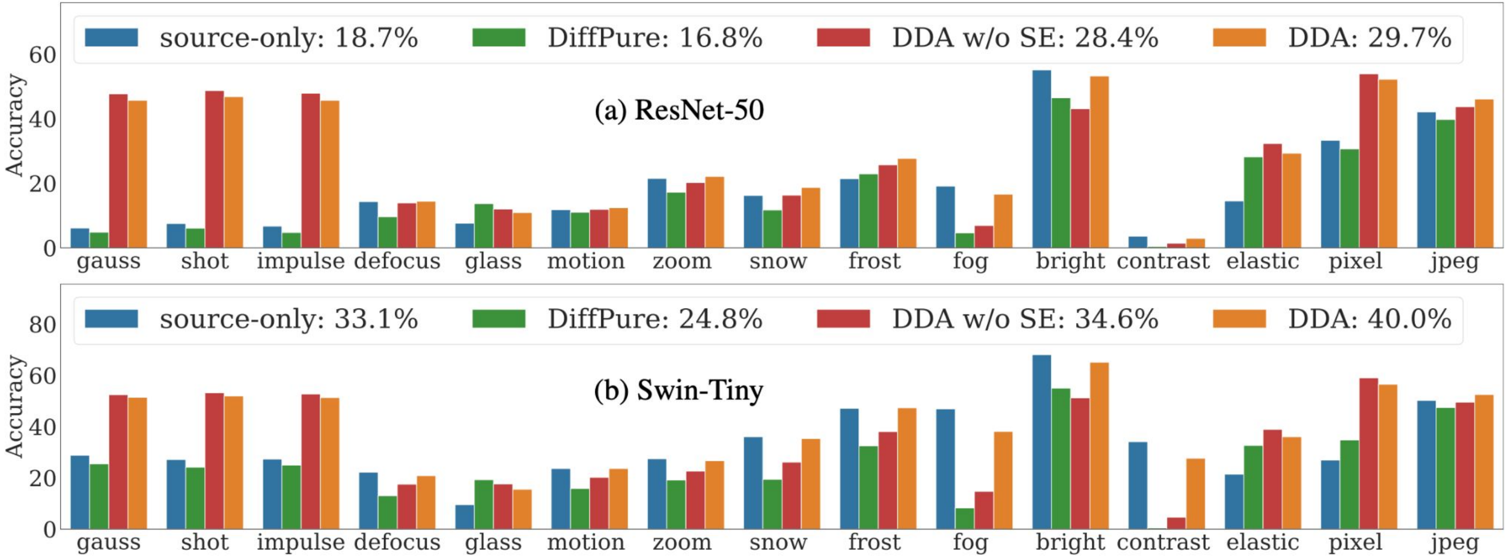
successes

failures

why we need self-ensemble to choose how much to adapt

DDA Helps across Corruption Types

DDA reliably helps due to self-ensembling and refinement.
Diffusion alone can hurt when the projection fails.



Is It Time To Update our Opinion of Model Updates?

While input updates shine in certain settings

small batches, ordered data, and mixed corruptions

model and input updates are complementary

More adaptation is needed to make the best of both kinds of updates—take the next steps with us!

Come check out the poster and see our

- poster #399 on day 2: WED-AM session (WED-AM-339)
- demo #11 on day 1: Tue. June 20, 2023
- arxiv https://arxiv.org/abs/2207.03442
- code https://github.com/shiyegao/DDA

Thanks!