ICCV23
PARIS

The many faces of reliability of visual perception for autonomous driving

Trends and perspectives

Andrei Bursuc

valeo.ai

Growing number of complex datasets

- Several datasets to assess specific use-cases:
  - Weather: ACDC, Cityscapes Rainy / Foggy, Dark Zurich, Raincouver
  - Distribution shift: Cityscapes-C/OC, SHIFT (S), WildDash, WildDash2, RoboBEV
  - OOD: Fishyscapes, SegmentMeIfYouCan, StreetHazards (S), BDD-Anomaly

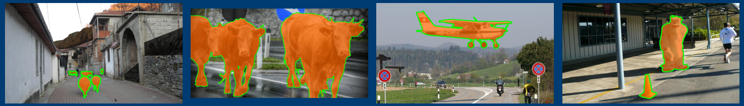
- Datasets with different sources of error in the same conditions: MUAD (S)
- BRAVO challenge: unified reliability bench
- Neural closed-loop simulators from real-data: Vista, Vista 2, UniSim
- New datasets for classification: ImageNet-C, ImageNet-R, ImageNet-A, ImageNet-O, ImageNetV2
- Similar trend of other sensors, e.g., Lidar

(S): synthetic data

Out-of-Context Cityscapes

Cityscapes-C

ACDC



SegmentMeIfYouCan

Reluctance to focus on multiple KPIs

We’re SoTA!

Method	Accuracy
Foo	***
Bar	***
Foo+Bar	***
Ours	**SoTA**

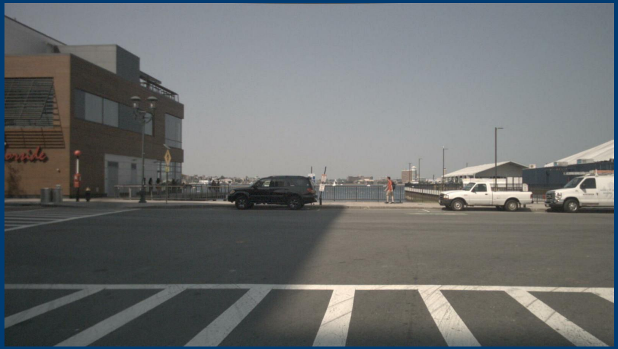
We’re SoTA … most of the time.
Here are our limitations.

Method	Metric 1	Metric 2	Metric 3	Metric 4
Foo	***	***	***	***
Bar	***	***	***	**SoTA**
Foo+Bar	***	**SoTA**	***	***
Ours	**SoTA**	***	**SoTA**	***

- The academic peer-review systems seems to reward bold numbers
- Potential solution: aggregate scores across conditions and metrics, e.g., nuScenes Detection Score (NDS)

Opportunity to adapt foundation generative models for testing

Editing to get edge cases (for Robustness evaluation or validation)



Original image

“People and kids are crossing the street”

“Trash is littering the street”

Spatially localized edits

Original image

“A baby crossing the street”

“An old lady fell in the middle of the road”

Image credit: Matthieu Cord et al.

Opportunity to adapt foundation generative models for testing

Create training or validation data for domain generalization

Original image

“In Paris”

“In India”

Annotations in the original domain are still valid

Original image

“Heavy fog, zero visibility”

“Heavy snow”

Image credit: Matthieu Cord et al.

Opportunity to adapt foundation generative model for training data

Create training data for OOD and corner-cases

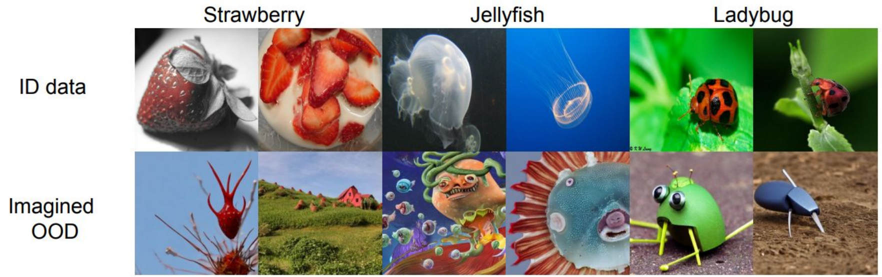


Figure 1: **Top**: Original ID training data in IMAGENET [13]. **Bottom**: Samples generated by our method DREAM-OOD, which deviate from the ID data.

Dream-OOD: generate useful images for OOD skills

Robusta: generate segmentation data with unseen layouts, corruptions and OOD

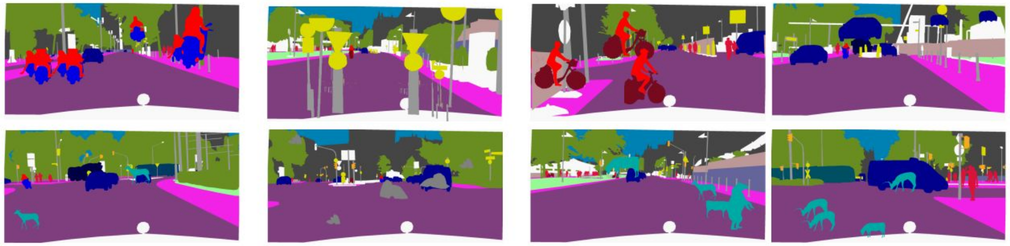
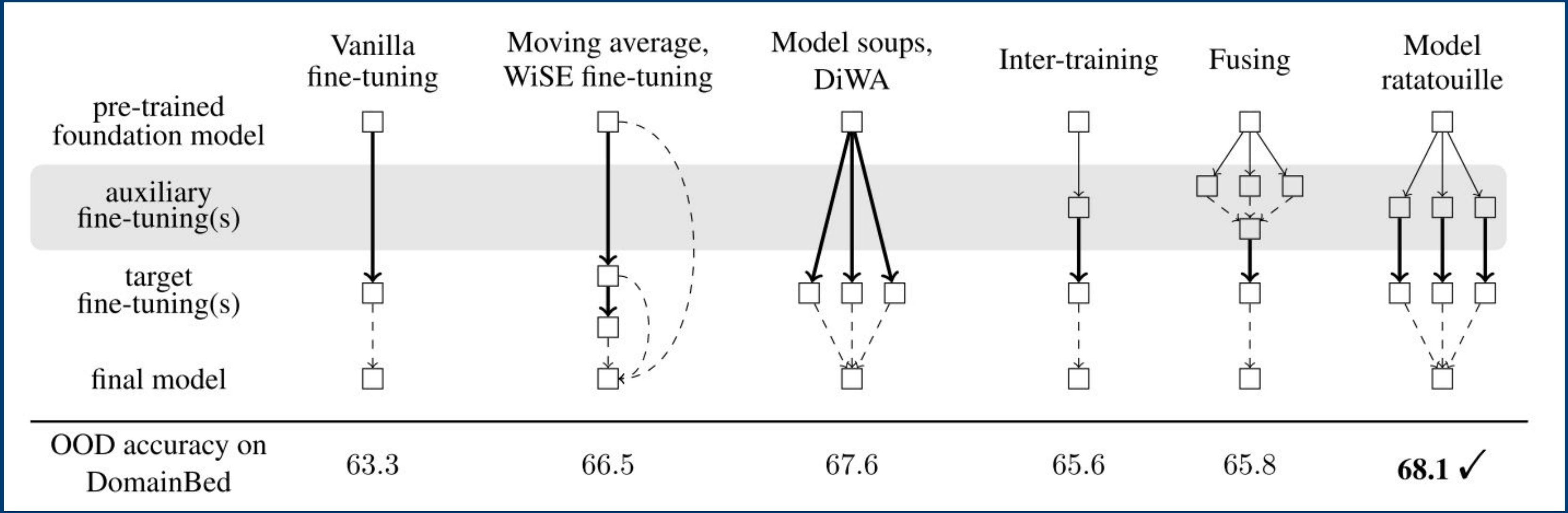


Figure 4. **Perturbed label maps.** Different label maps from Corrupted-Cityscapes (top) and Outlier-Cityscapes (bottom).

M. Hariat et al., Learning to Generate Training Datasets for Robust Semantic Segmentation, WACV 2024
X. Du et al., Dream the Impossible: Outlier Imagination with Diffusion Models, NeurIPS 2023

Opportunity to adapt foundation generative model for perception

Repurpose foundation models for improved generalization with computational efficiency



Different model soups variants

A. Ramé et al., Model Ratatouille: Recycling Diverse Models for Out-of-Distribution Generalization, ICML 2023

Reliability of vision-language models

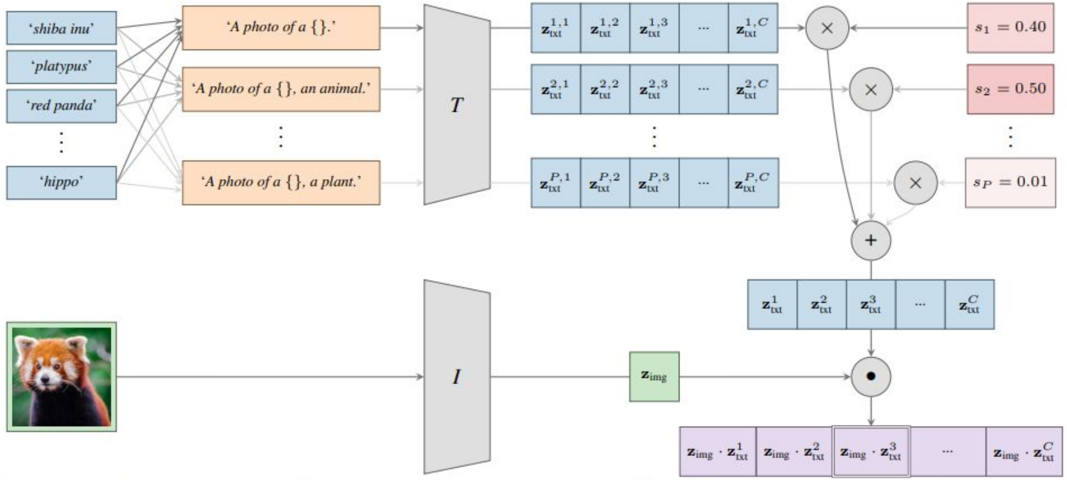
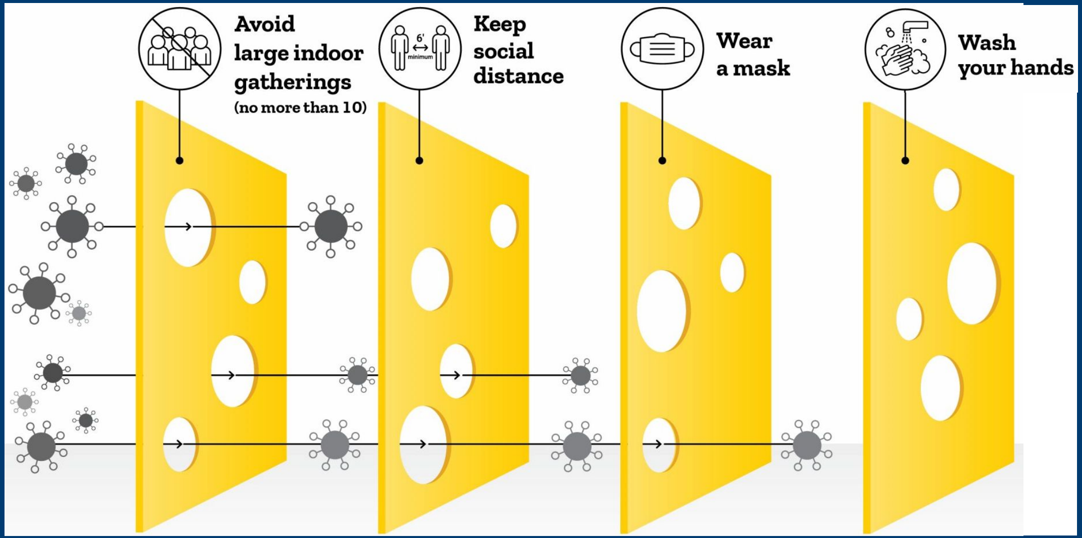


Figure 1. Construction of a zero-shot classifier with zero-shot prompt ensembling (ZPE) for text-image models. Logits (□) are calculated by combining text (□) and image (□) representations. The final text representation is a weighted ensemble of representations corresponding to different prompts (□). Crucially, the ZPE scores (□) for weighting each prompt are calculated without access to any labeled training data, as described in Section 3 and Algorithm 2.

Zero-shot prompt ensembling

J. Urquhart Allingham et al., A Simple Zero-shot Prompt Weighting Technique to Improve Prompt Ensembling in Text-Image Models, ICML 2023

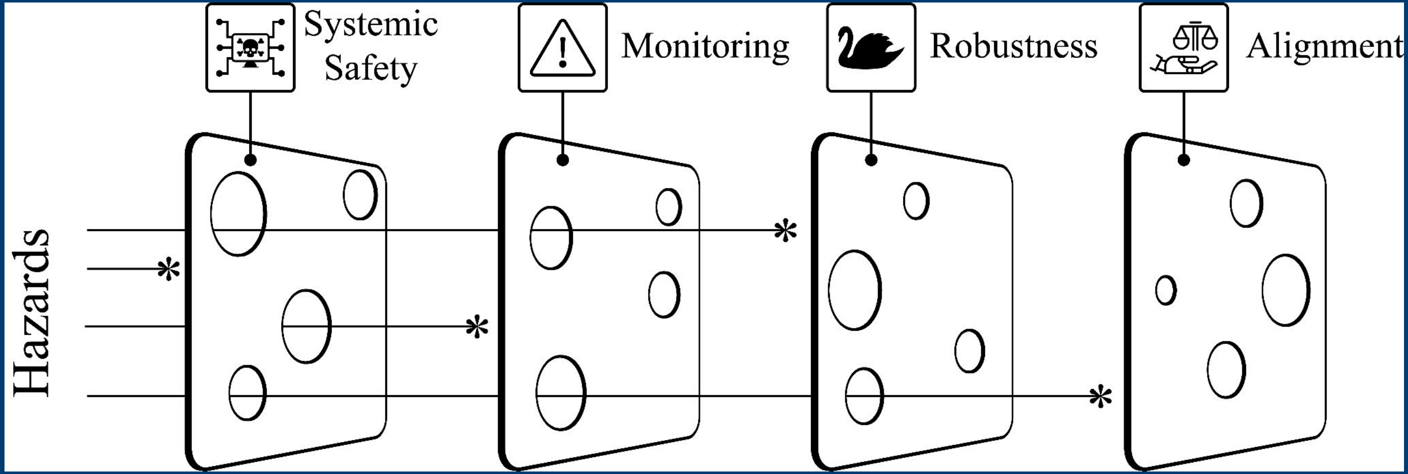
Swiss Cheese model



- Use multiple layers of robustness and safety barriers
- Pursuing multiple avenues creates multiple layers of protection that mitigate hazards and make ML systems safer

---
D. Hendrycks al., Unsolved Problems in ML Safety, arXiv 2021

Swiss Cheese model



- Use multiple layers of robustness and safety barriers
- Pursuing multiple avenues creates multiple layers of protection that mitigate hazards and make ML systems safer

---
D. Hendrycks al., Unsolved Problems in ML Safety, arXiv 2021

The end.