Video Game-Based Domain Generalization for Computer Vision

Kuan-Cheng Lee, SCAI
Advisors: Yiran Luo, Professor Yezhou Yang
Arizona State University

Motivation

The project is to provide a new challenge to Domain Generalization (DG) in Computer Vision, by curating a novel benchmark dataset made from video game scenes.

Requirements

In order train a DG model, we require the DG dataset must have:

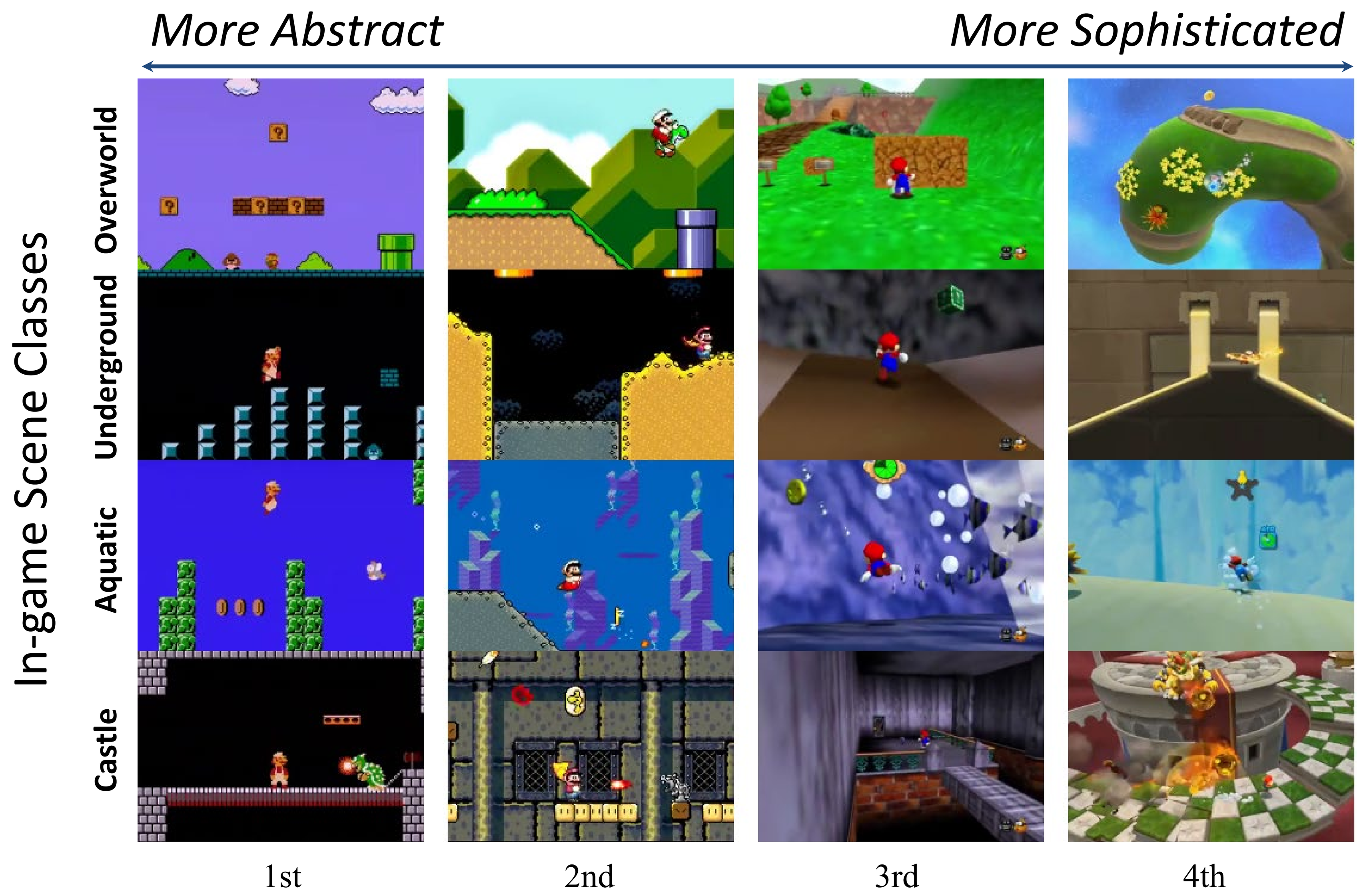
- Multiple collections (domains) of images **be in distinct image styles**, and
- Images in all domains **be classified using a shared group of class labels**.

Source of Data

We define **4 classes** of in-game scenes, and sample **10k images per domain**, for **4 domains** of Mario games rendered on evolving 4 generations of hardware:

- 1st: Super Mario Bros & Mario 3
- 2nd: Super Mario Bros World & All-Star
- 3rd: Mario 64 & Mario Maker 3DS
- 4th: Mario Galaxy & Mario Maker 2

Dataset Overview & Benchmarks



Domains (Generations) of Video Game Graphics

Leave-one-out Multi-Source Domain Generalization Benchmarks

Method	Training Domains → Test Domain				
	234 → 1	134 → 2	124 → 3	123 → 4	Avg.
Using ResNet-50 as backbone. 23M parameters.					
ERM	46.7	36.4	23.5	26.0	33.2
MIRO	50.9	42.6	24.7	28.9	36.8
Using CLIP-ViT-b32 as backbone. 86M parameters.					
ERM	43.6	33.0	22.2	29.0	32.0
MIRO	44.5	44.7	30.1	29.6	37.2

In these experiments, we hold out all images of one domain at a time as the unseen test set, e.g. 123 → 4 means we jointly use three domains of 1st + 2nd + 3rd for training a model, then test the model with data only from the 4th domain.

Evaluation & Takeaways

- We use DomainBed[1] to evaluate the performances by current state-of-the-art methods on DG (ERM[2], MIRO[3]) paired with different backbone models (ResNet-50[4] and CLIP-ViT-b32[5]) on our benchmark.
- We observe our benchmark presents a unique challenge, where generalizing to more sophisticated graphics is increasingly harder along the order of hardware generation.
- We also find the size of the model is irrelevant on our benchmark as all methods end up with close results in general.

References

1. Gulrajani, I., & Lopez-Paz, D. (2020). *In search of lost domain generalization.* arXiv:2007.01434.
2. Vapnik, V. (1991). *Principles of risk minimization for learning theory.* Advances in neural information processing systems, 4.
3. Cha, J., et al. (2022). *Domain generalization by mutual-information regularization with pre-trained models.* ECCV.
4. He, K., et al. (2016). *Deep residual learning for image recognition.* CVPR.
5. Radford, Alec, et al. (2021). *Learning transferable visual models from natural language supervision.* ICML.

FURI

Ira A. Fulton Schools of Engineering
Arizona State University