From Clutter to Clarity: Visual Recognition through Foveated Object Centric Learning (FocL)

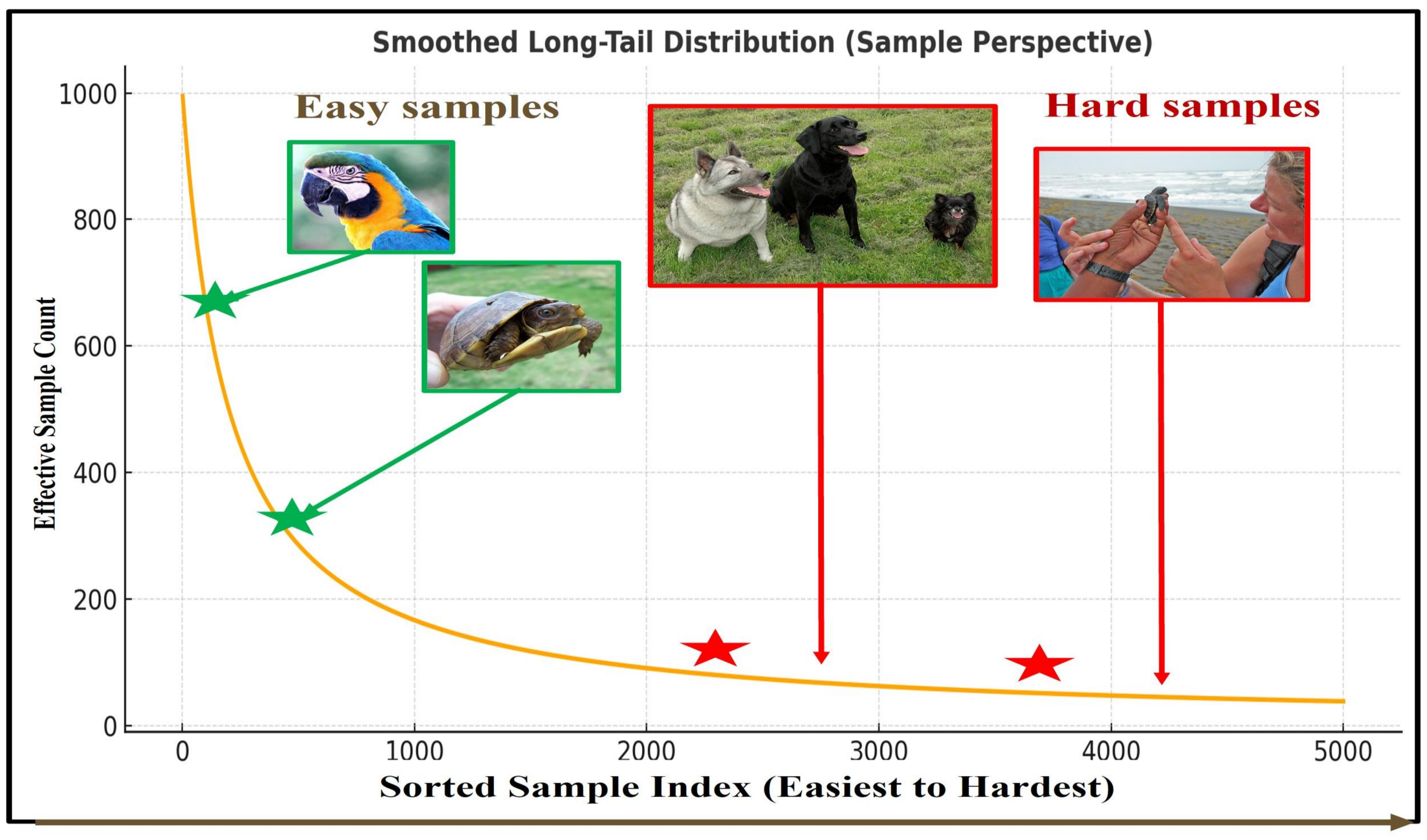
Amitangshu Mukherjee, Deepak **Ravikumar**, Kaushik Roy

NEURAL INFORMATION PROCESSING SYSTEMS

CoCoSys CENTER FOR THE CO-DESIGN OF COGNITIVE SYSTEMS

PURDUE UNIVERSITY

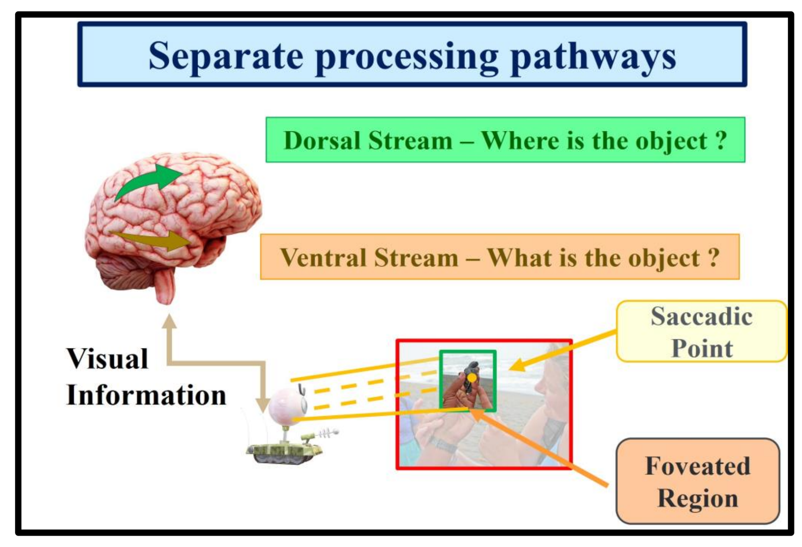
1. Motivation



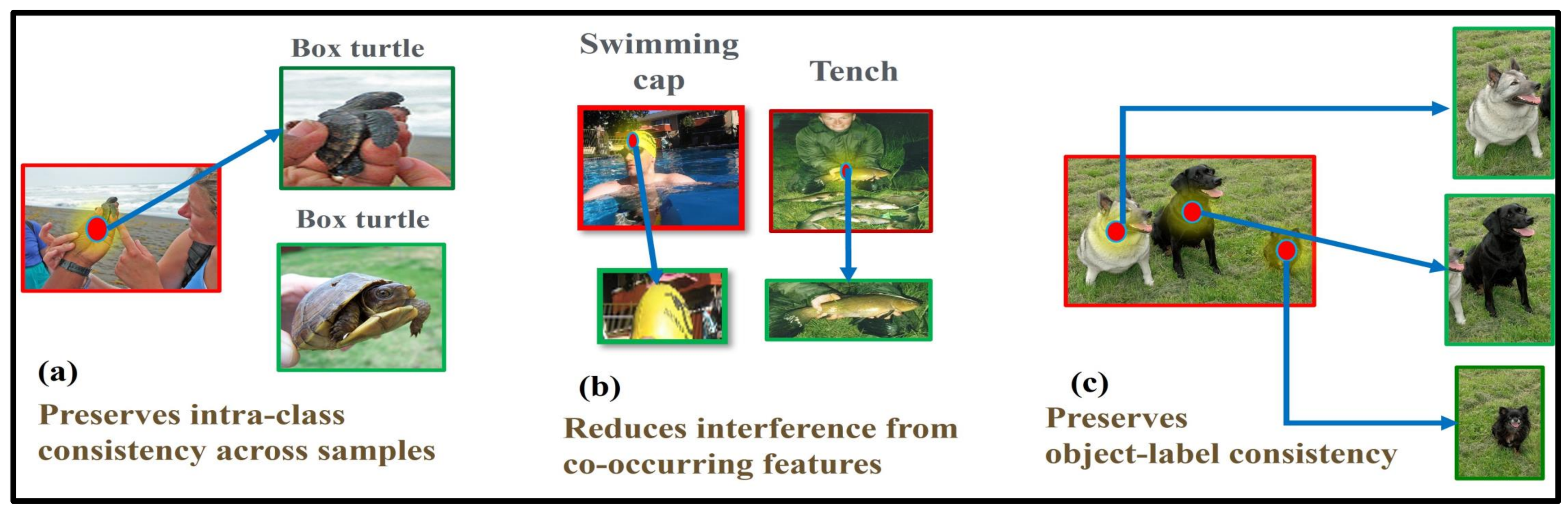
Easy (a) Measured by a proxy Hard

Hard samples cluster in the long tail region

2. Active Vision decouples the "where" from the "what"

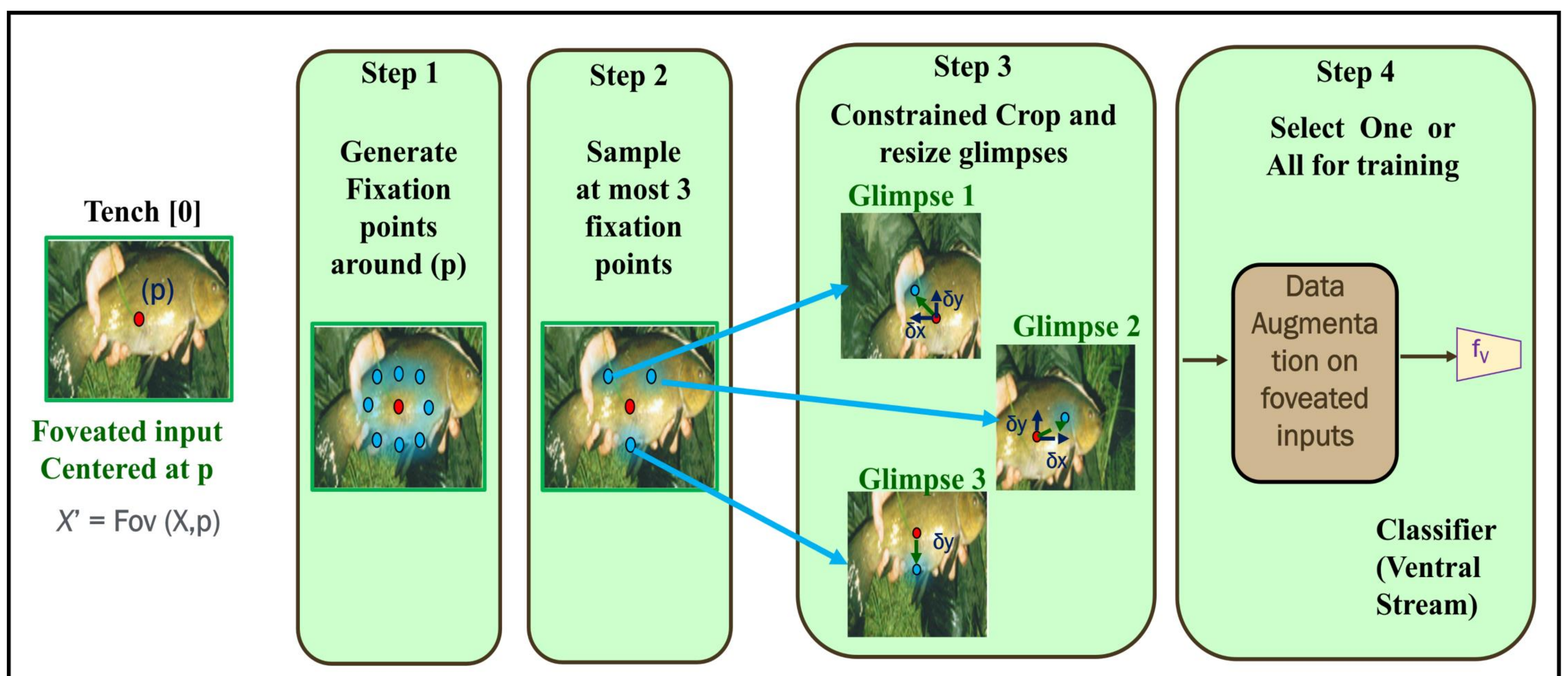


- Active Vision processes visual information through two pathways
  - Dorsal ("where" stream) - localization
  - Ventral ("what" stream) - recognition
- Saccades fixate on objects, and foveation extracts clean object-centric glimpses



Foveated vision reduces variability and improves recognition consistency

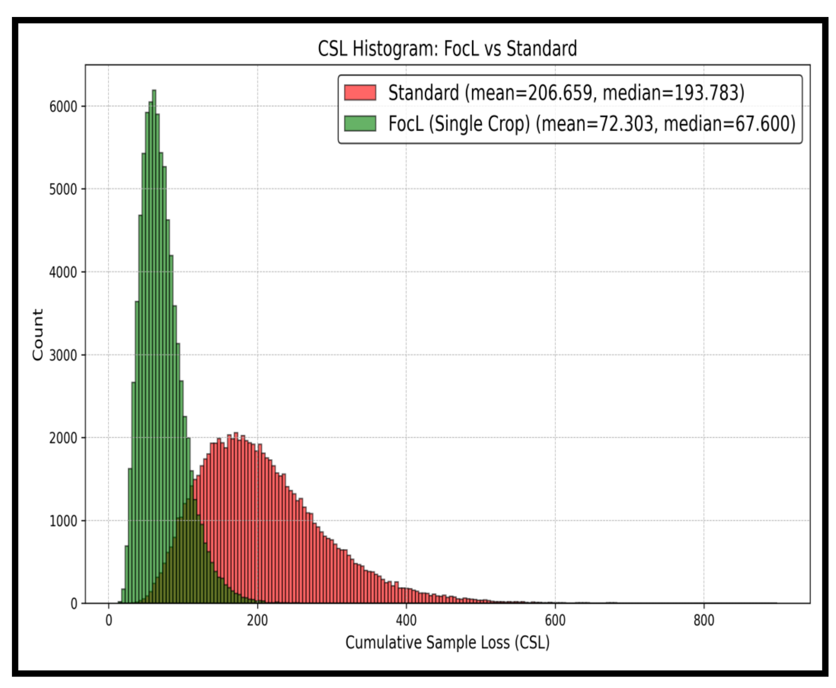
3. Methodology



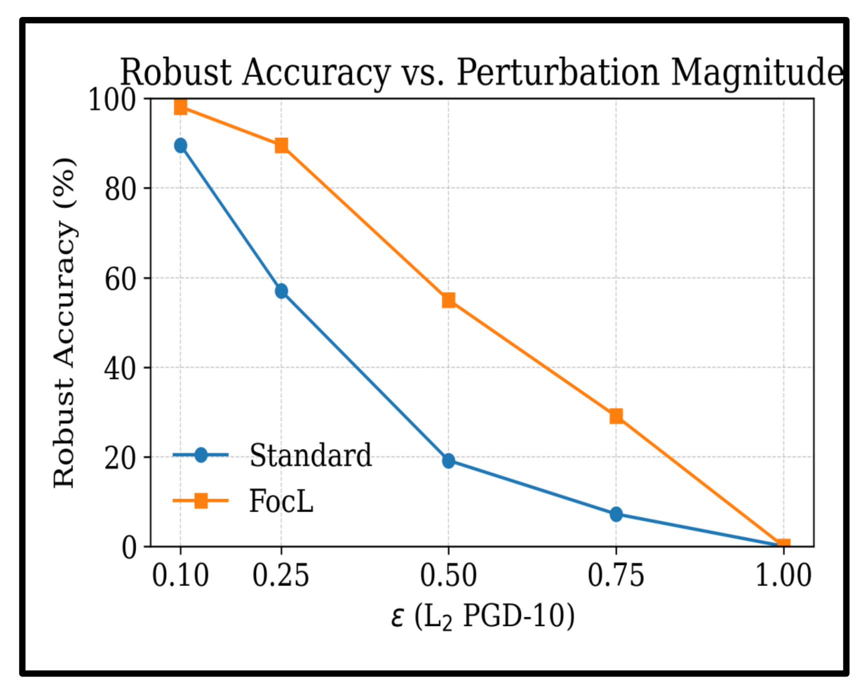
FocL : Multi-glimpse emulating small saccadic shifts

4. Does FocL reduce memorization ?

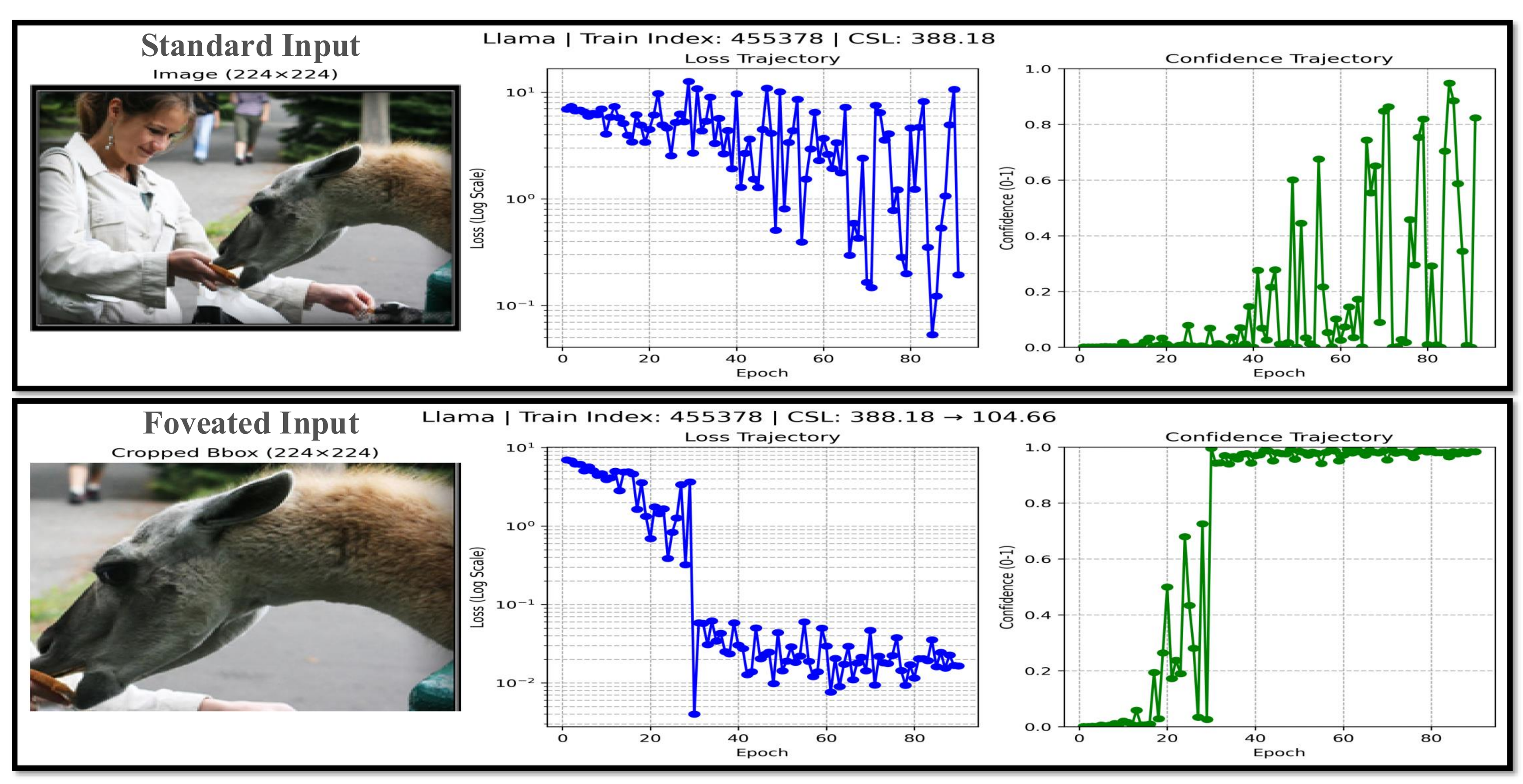
CSL shift from tail to mode



Higher Adversarial Resistance



FocL reduces sample-level difficulty



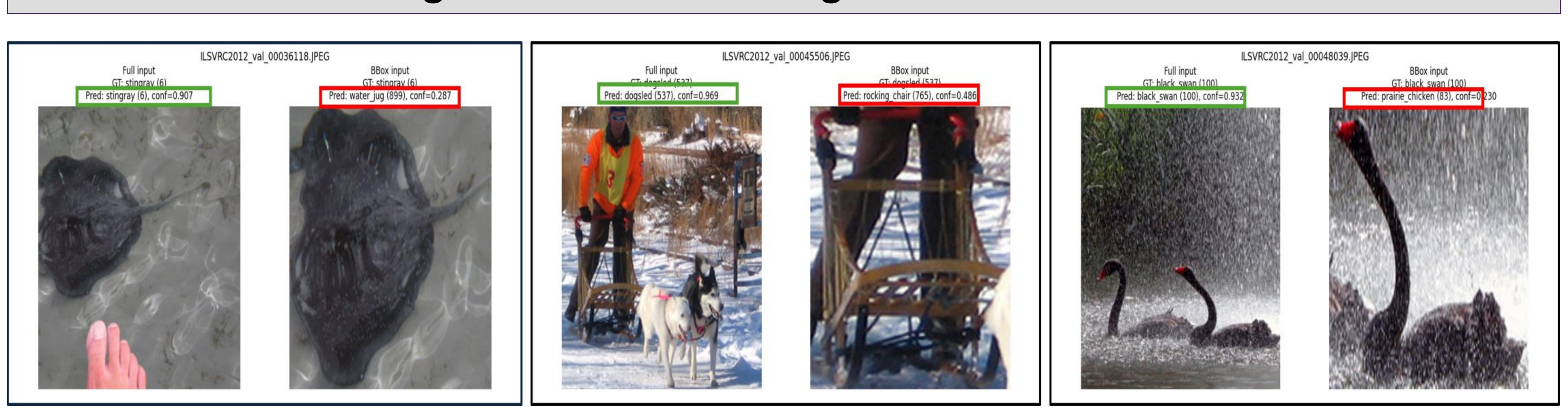
FocL reduces memorization on background heavy, top 1% FZ samples

5. Does FocL enhance generalization ?

Boosts Any Top-1 by 8% and 10% on 2K samples from ImageNet V1 and V2

Dataset	Standard only	SAM + Standard	SAM + FocL
ImageNet-V1	60.90 ± 1.10	68.08 ± 0.97	**75.82** ± 0.23
ImageNet-V2	49.77 ± 0.94	57.50 ± 1.92	**65.80** ± 0.55

Visualization : Background removal hurts generalization for standard classifiers



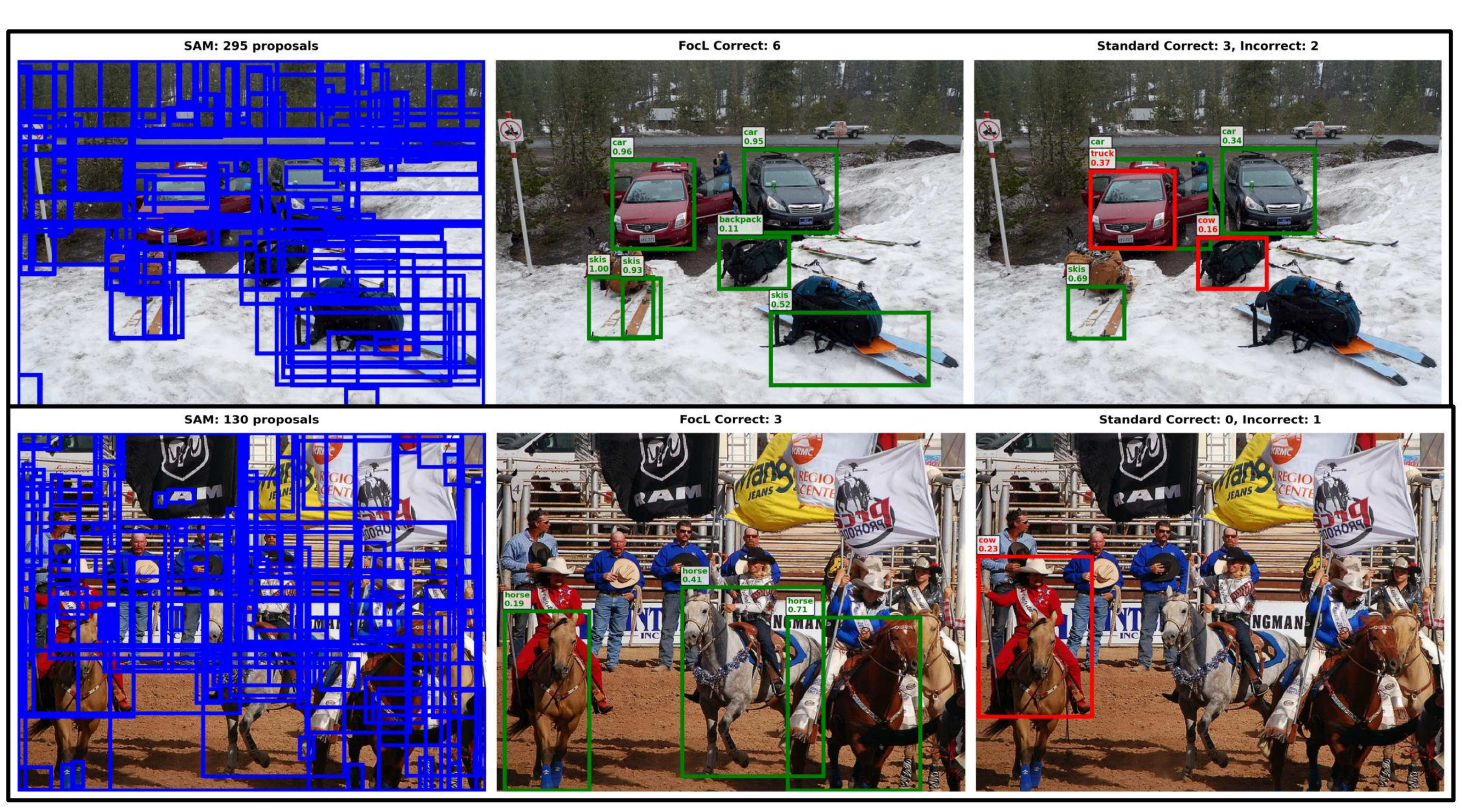
Improves mAP by 3-4 pp on COCO subset showing cross-domain transfer

(a) Max 20 proposals / image

IoU	SAM + Standard → SAM + FocL
0.3	24.39 → **28.22** (+3.83 pp / +16%)
0.5	13.10 → **16.22** (+3.12 pp / +24%)

(b) Max 300 proposals / image

IoU	SAM + Standard → SAM + FocL
0.3	28.19 → **31.76** (+3.57 pp / +12.7%)
0.5	14.29 → **17.24** (+2.95 pp / +20.6%)



6. Key Takeaways

- **FocL removes background reliance through object-first viewpoint variations.**
- **FocL cuts CSL by 62% overall and improves nearly all top FZ samples.**
- **FocL boosts cross domain generalization (ImageNet V2, COCO) from ImageNet-1K training alone.**

Acknowledgment

This work was supported in part by the Center for Co-design of Cognitive Systems (CoCoSys), one of the seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program, and in part by the Department of Energy (DoE).