ICCV23
PARIS

paper & code

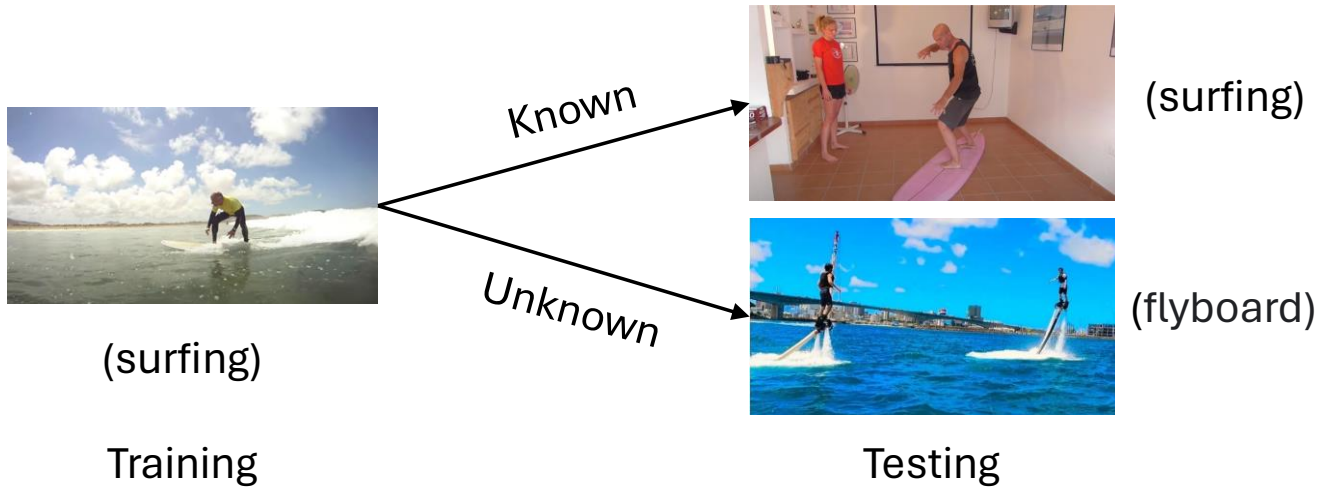
SOAR: Scene-debiasing Open-set Action Recognition

Yuanhao Zhai[1], Ziyi Liu[2], Zhenyu Wu[2], Yi Wu[2], Chunluan Zhou[2],
David Doermann[1], Junsong Yuan[1], Gang Hua[2]

[1]University at Buffalo [2]Wormpex AI Research

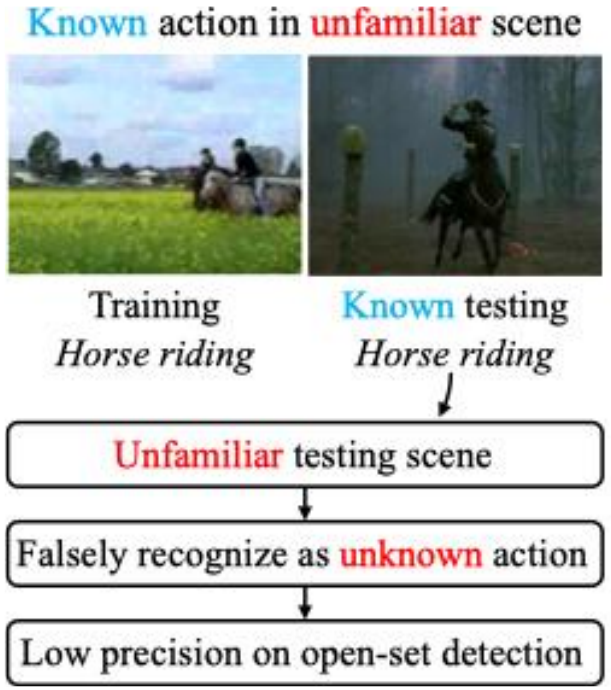
Open-set action recognition

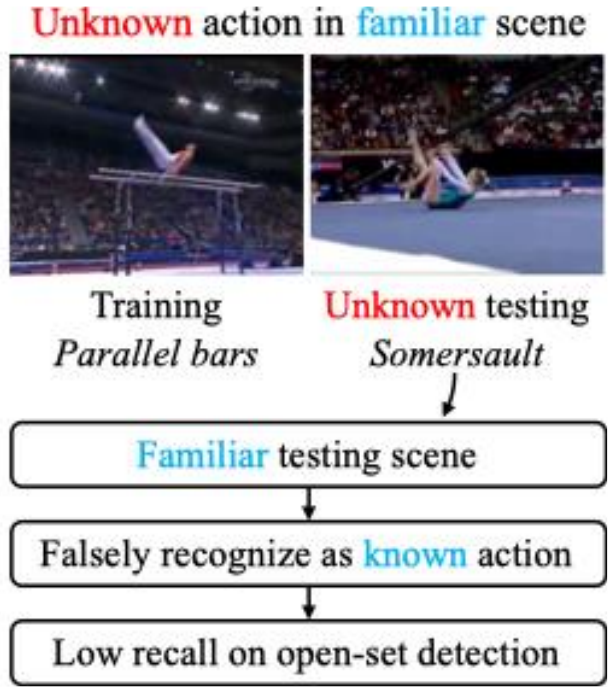
- Open-set action recognition (OSAR): two objectives
  1. Identify unknown new actions from the known actions (binary classification)
  2. Classify known actions (C-way classification)



Scene bias in OSAR

- Existing datasets exhibit heavy scene bias, e.g., linear probing a scene classifier yields ~55% accuracy on UCF101
- Two typical scenarios that OSAR methods may fail



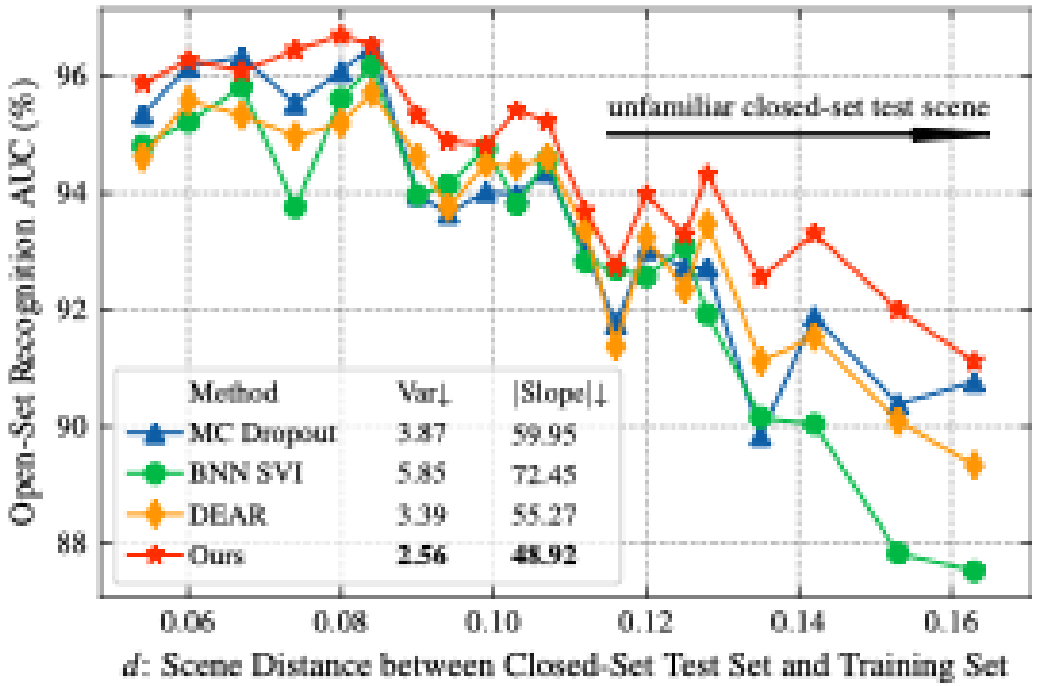


How OSAR methods are affected by the scene?

- We design an empirical experiment to quantitative analyze
  - Known action in unfamiliar scene
  - Unknown action in familiar scene
- Setup:
  1. Scene feature extraction for videos
  2. Scene similarity computation between training and testing videos
  3. Divide testing videos into subsets according to the scene similarity to the training set
  4. Measure OSAR performance on each subset

How OSAR methods are affected by the scene? (cont'd)

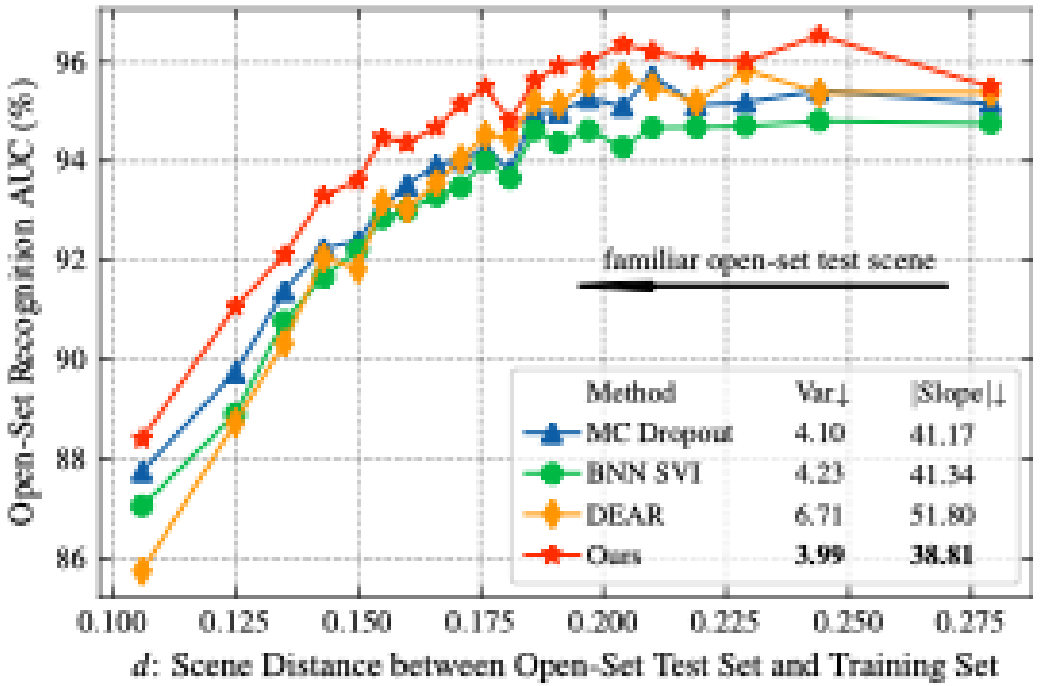
Known action in unfamiliar scene



- As the scene from known actions gets more unfamiliar, the OSAR performances decreases
- Our method is the least affected by the scene

How OSAR methods are affected by the scene? (cont'd)

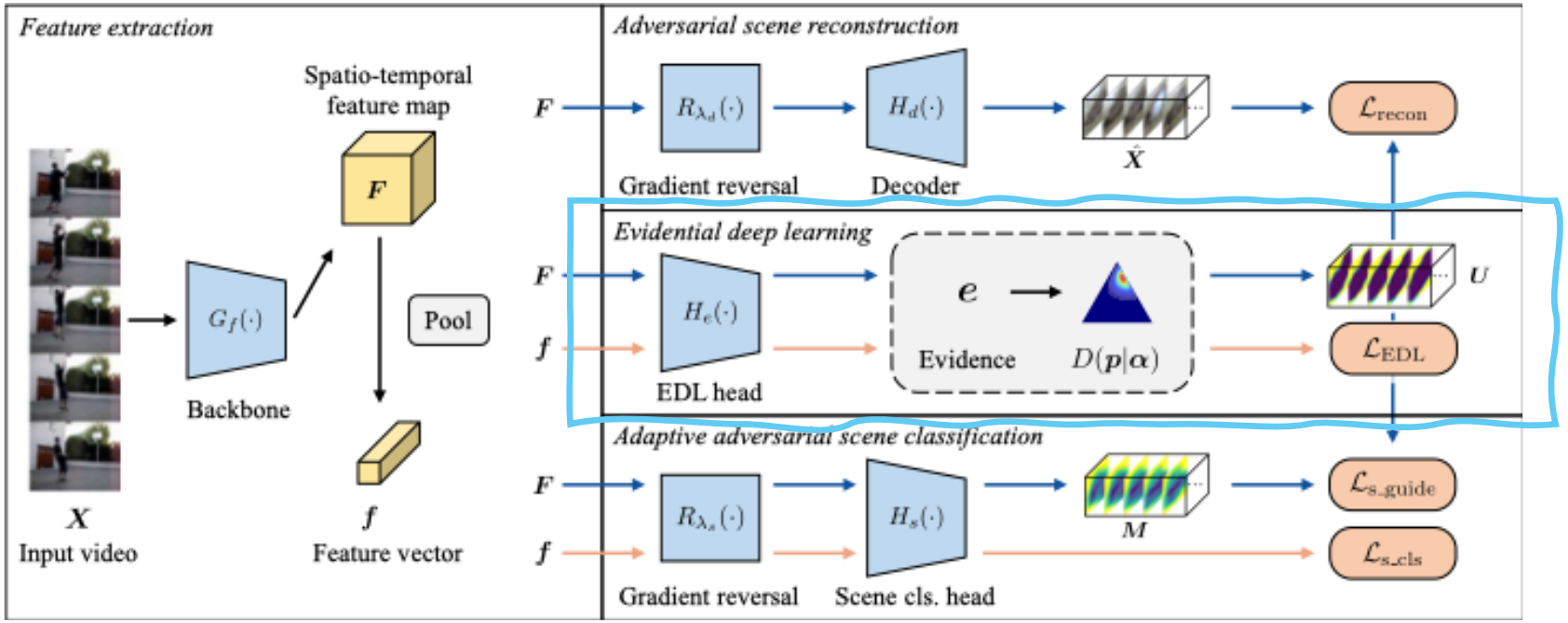
Unknown action in familiar scene



- As the scene from unknown actions gets more familiar, the OSAR performances decreases
- Our method is the also least affected by the scene

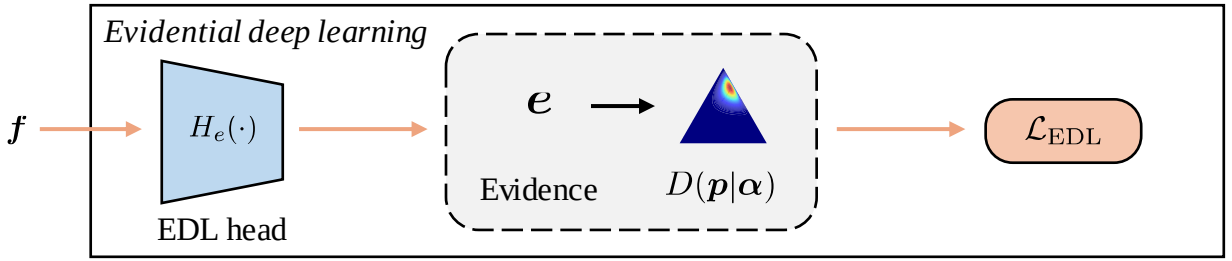
Scene-debiasing Open-set Action Recognition (SOAR)

- Evidential deep learning baseline
- Adversarial scene reconstruction
- Adaptive adversarial scene classification



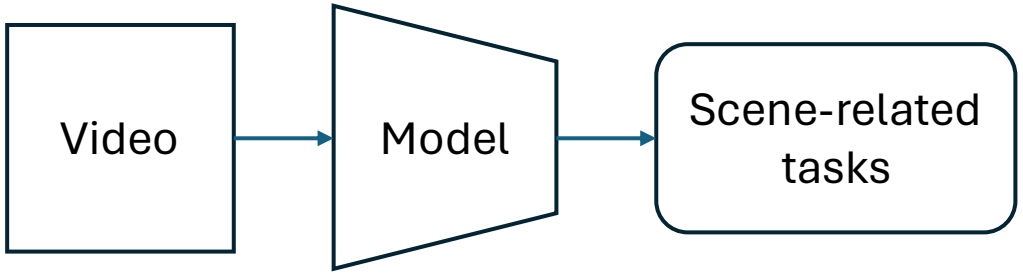
Evidential deep learning (EDL) baseline

- EDL predicts an uncertainty to measure the likelihood that the sample is unknown
  1. Evidence is predicted for each class
  2. A Dirichlet distribution is built based on the evidence
  3. Uncertainty is deterministically given via EDL



Scene-debiasing via adversarial learning

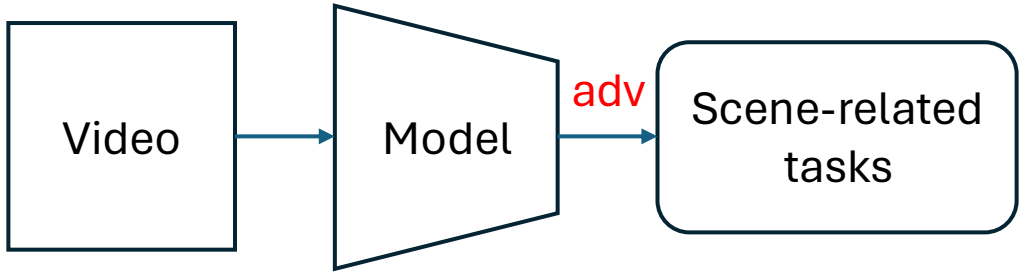
- Model learns scene information if it is trained towards scene-related tasks.



learns scene info

e.g., scene classification

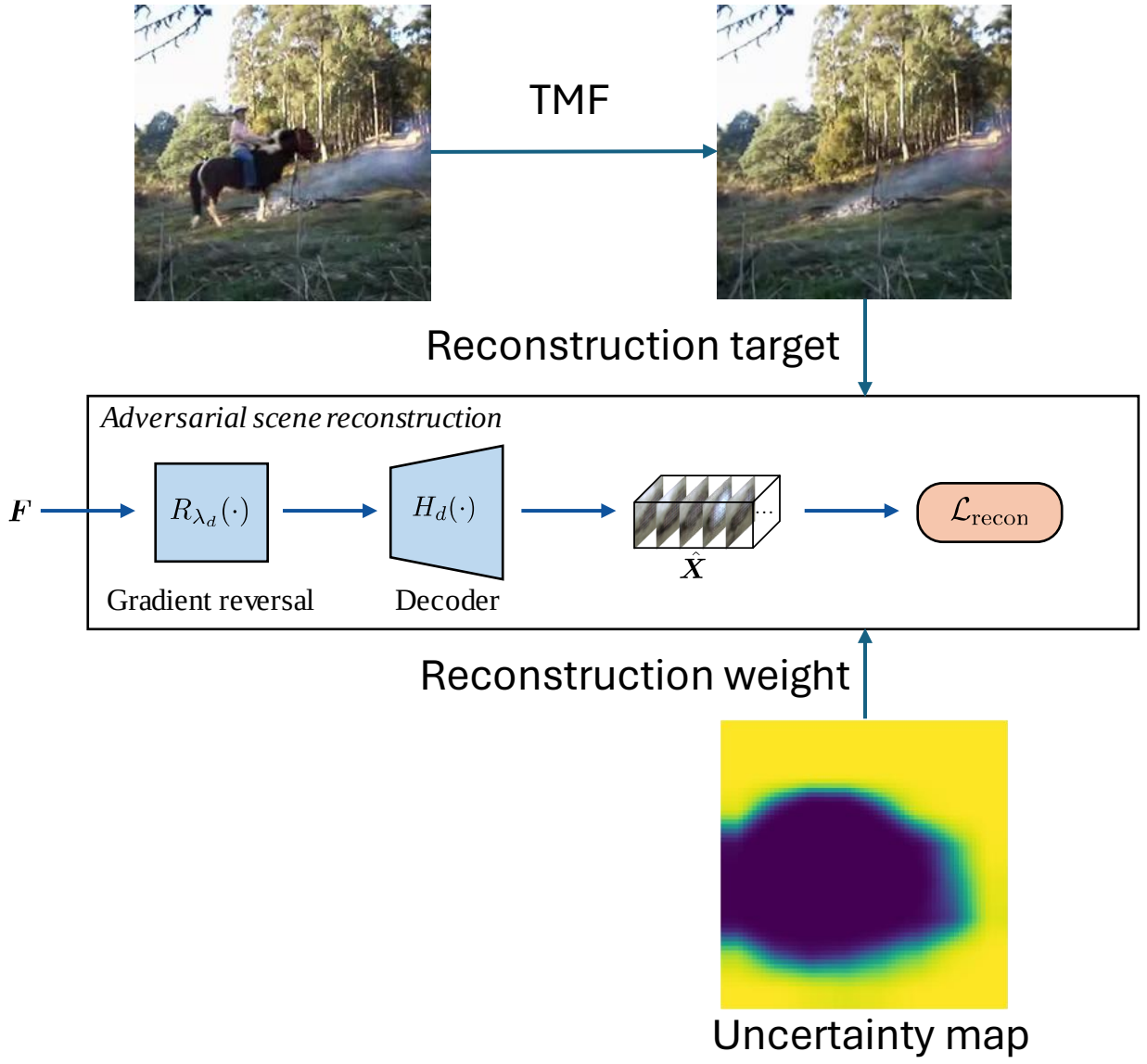
- Model learns to remove scene information if it is adversarially trained towards scene-related tasks.



removes scene info

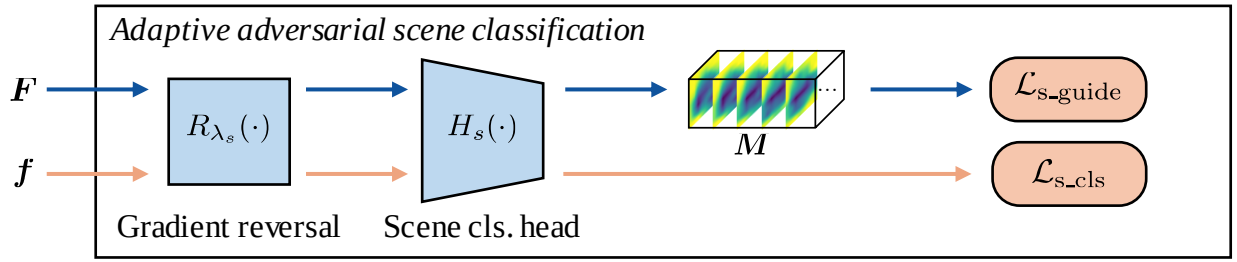
Adversarial scene reconstruction

- Adversarially reconstruct the background
- Background estimation via temporal median filter (TMF)
  - Works well if background is static
- Uncertainty-weighted reconstruction to focus on the background regions
  - Helps on dynamic background



Adaptive adversarial scene classification

- Given input video, adversarially classify its scene
  - With a spatial uncertainty guidance to apply more weight on the scene regions

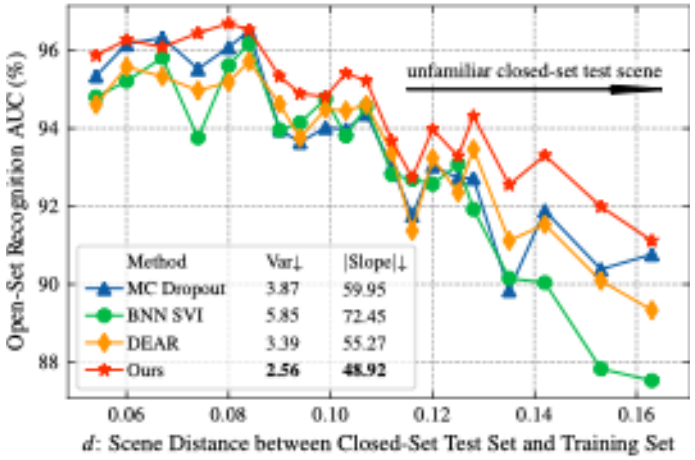


Experiments – comparison with SOTA

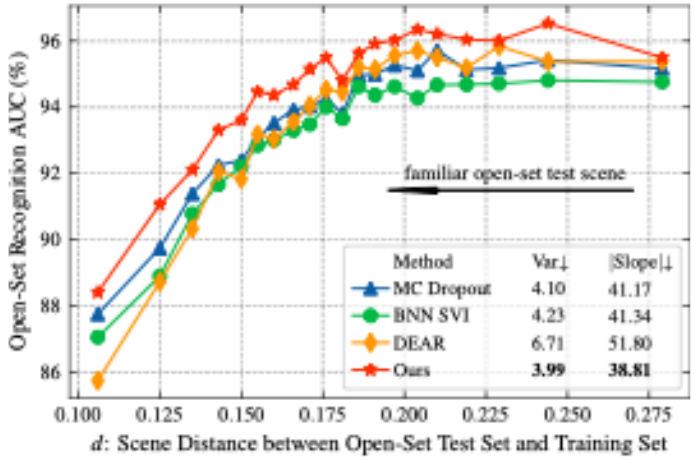
- Best OSAR and closed-set action recognition performances

Methods	UCF101 [63]+MiTv2 [45]				UCF101 [63]+HMDB51 [38]				Closed-set Accuracy
	AUC ↑	FAR@95 ↓	TPR@10 ↑	Open maF1 ↑	AUC ↑	FAR@95 ↓	TPR@10 ↑	Open maF1 ↑	
SoftMax	44.47	96.93	8.85	55.50 ± 0.45	44.34	97.91	3.66	73.13 ± 0.12	94.10
OpenMax [5]	63.96	45.89	3.78	66.21 ± 0.16	63.67	80.53	6.54	67.81 ± 0.12	56.54
MC Dropout [21]	93.66	25.43	85.72	68.12 ± 0.20	86.11	77.50	70.13	71.13 ± 0.15	94.13
BNN SVI [36]	93.16	25.88	79.36	67.96 ± 0.19	85.63	71.52	66.14	71.57 ± 0.17	93.89
DEAR [3]	93.52	29.53	84.03	75.12 ± 0.27	87.12	71.32	72.21	88.07 ± 0.20	93.97
SOAR (Ours)	**94.60**	**25.33**	**86.47**	$\mathbf{76.22 \pm 0.32}$	**88.10**	**69.57**	**72.75**	$\mathbf{89.55 \pm 0.22}$	**95.24**

- The least affected by the scene



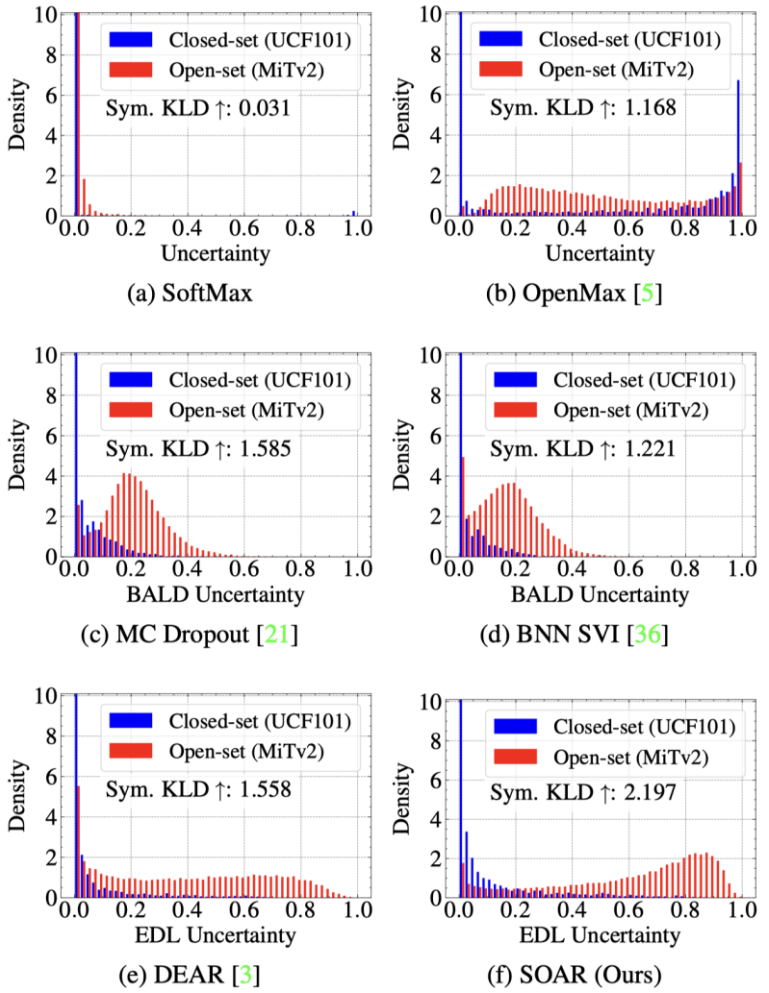
Known action in unfamiliar scene



Unknown action in familiar scene

Experiment – comparison with SOTA

- Best open-set/closed-set uncertainty separation (highest sym. KLD)



Contributions

- Design an empirical experiments to quantitatively analyze the scene bias
- Adversarial scene reconstruction & adaptive adversarial scene classification to reduce the learned scene information
- State-of-the-art OSAR performance, and effectively reduces scene bias

paper & code