ALBAR: Adversarial Learning approach to mitigate Biases in Action Recognition

Joseph Fioresi, Ishan Rajendrakumar Dave, Mubarak Shah
University of Central Florida
https://joefioresi718.github.io/ALBAR_webpage/

UCF Center for Research in Computer Vision

Webpage/Code:

ICLR

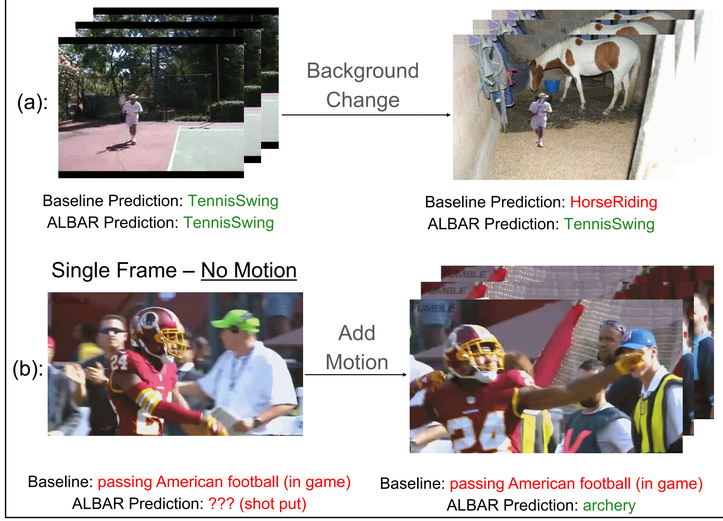
Bias in Action Recognition

- Video action recognition models improperly exploit both background & foreground cues over motion information.
- Below figure: (a) background change → misclassification, (b) adding motion does not overcome foreground bias.

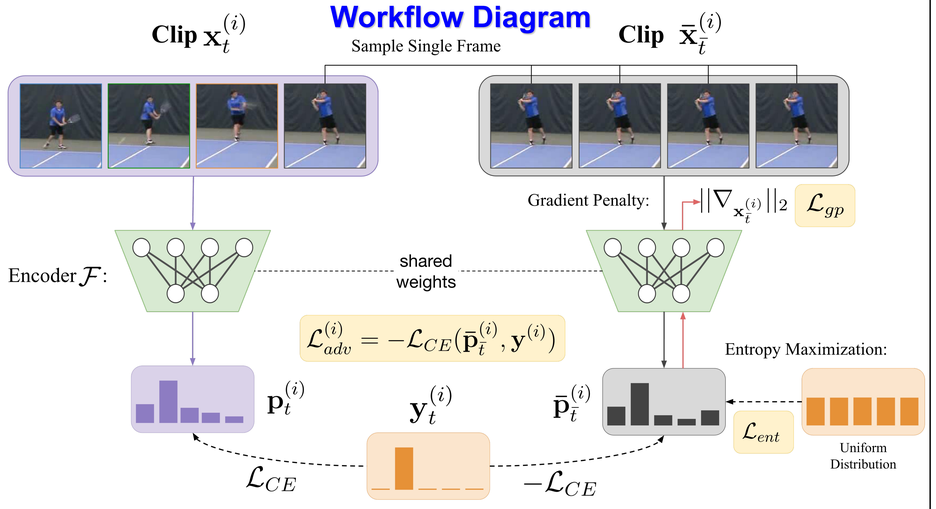
Contributions

- We propose a novel adversarial setup that uses a *static* (no motion) clip passed through the *same* encoder model.
- Designed supplemental static input entropy maximization and gradient penalty objectives to stabilize training.
- ALBAR achieves SOTA combined debiasing scores on 3 major AR datasets: UCF101, HMDB51, and Kinetics400.
- Propose a background leakage fix for UCF101 protocol.

Background/Foreground Bias



Workflow Diagram



Overall Loss: $\mathcal{L}^{(i)} = \mathcal{L}_{CE}^{(i)} + \omega_{adv} * \mathcal{L}_{adv}^{(i)} + \omega_{ent} * \mathcal{L}_{ent}^{(i)} + \omega_{gp} * \mathcal{L}_{gp}^{(i)}$

Loss Descriptions

Baseline – Cross-Entropy: $\mathcal{L}_{CE}^{(i)} = -\sum_{c=1}^{N_C} \mathbf{y}_c^{(i)} \log \mathbf{p}_c^{(i)}$

- Standard action classification, susceptible to bias

Adversarial Loss – Static Inputs: $\mathcal{L}_{adv}^{(i)} = -\mathcal{L}_{CE}(\bar{\mathbf{p}}_{\bar{t}}^{(i)}, \mathbf{y}^{(i)})$

- Negative cross-entropy of static clip prediction
- Leads to class flipping trivial solution on its own

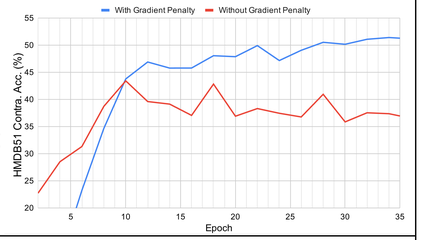
Static Entropy Maximization: $\mathcal{L}_{\text{ent}}^{(i)} = \sum_{c=1}^{N_C} \bar{\mathbf{p}}_{\bar{t},c}^{(i)} \log(\bar{\mathbf{p}}_{\bar{t},c}^{(i)})$

- Encourages uncertainty in static clip prediction
- Mitigates adversarial trivial solution

Static Gradient Penalty:

- Static inputs: drastic weight updates
- Regularize gradients, reduce impact

$$\mathcal{L}_{gp}^{(i)} = ||\nabla_{\bar{\mathbf{x}}_{\bar{t}}^{(i)}} \mathcal{F}(\bar{\mathbf{x}}_{\bar{t}}^{(i)})||_2$$



Comparison With Prior Works (HMDB51)

Debiasing Method	Original	OOD			
		SCUBA (↑)	SCUFO (↓)	Confl-FG (↑)	Contra. Acc.
None	73.92	43.93	20.46	36.58	27.84
Mixup [1] ICLR'18	74.58	43.10	21.17	36.62	26.09
VideoMix [2] arXiv'20	73.31	39.39	20.44	32.68	23.13
SDN [3] NeurIPS'19	74.66	40.02	20.22	34.87	22.88
BE [4] CVPR'21	74.31	43.56	19.96	35.99	27.84
ActorCutMix [5] CVIU'23	74.05	46.79	22.07	36.97	28.12
FAME [6] CVPR'22	73.79	51.40	26.92	39.61	29.66
StillMix [7] ICCV'23	**74.82**	51.81	13.39	47.38	40.28
Ours	73.20	**53.2** ↑21.1%	**0.42** ↓98.0%	**49.84** ↑36.3%	**53.02** ↑90.5%

Downstream Task Performance

Method	Action Recognition HMDB51 (OOD)	Anomaly Detection UCF_Crime	Temporal Action Localization THUMOS14
Baseline	27.84	82.39	54.89
Ours	**53.02**	**84.91**	**55.20**

UCF101 SCUBA/SCUFO Protocol Fix