UCF
CENTER FOR RESEARCH IN COMPUTER VISION

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Microsoft Research

A Large-scale Robustness Analysis of Video Action Recognition Models

WED-PM-007


Madeline C. Schiappa[1]


Naman Biyani[3]


Prudvi Kamtam[1]


Shruti Vyas[1]


Hamid Palangi[2]


Vibhav Vineet[2]

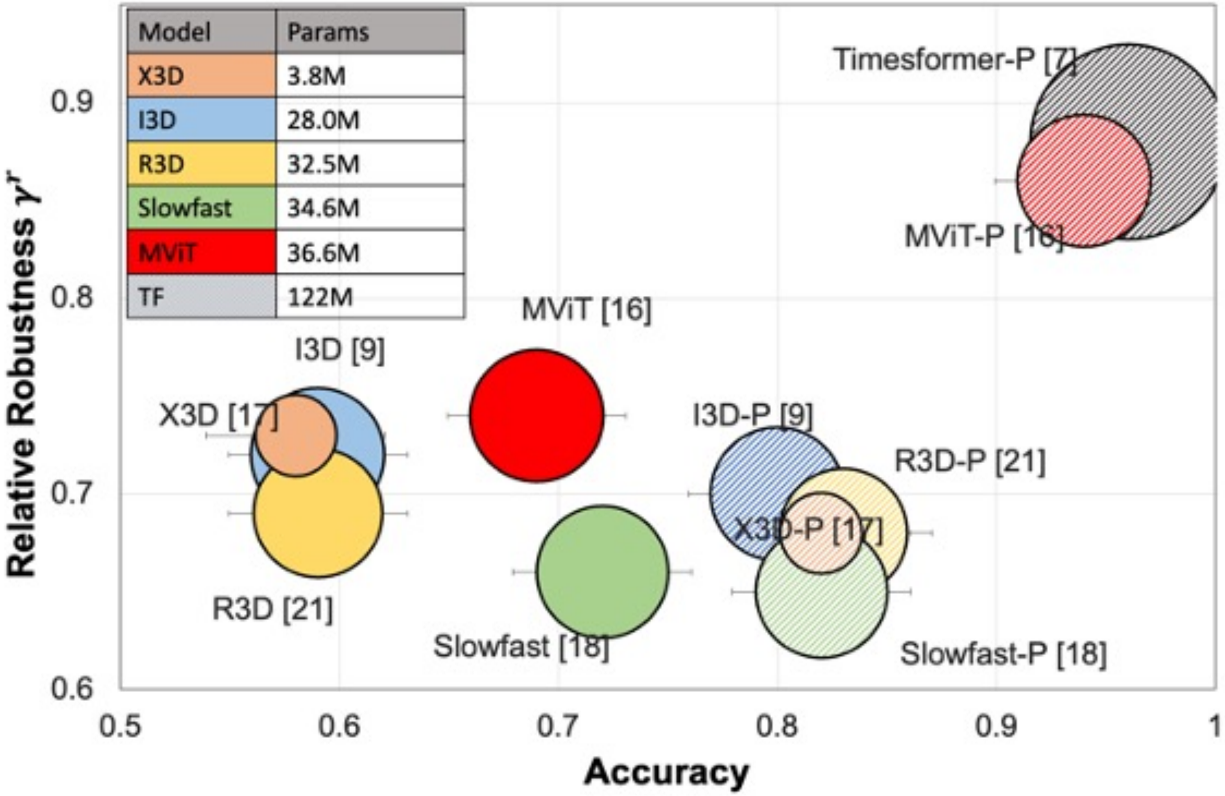

Yogesh S. Rawat[1]

University of Central Florida[1], Microsoft Research[2], IIT Kanpur, India[3]

Summary

- Benchmark action-recognition models on corrupted video
  - UCF101-P, HMDB51-P Kinetics-P, SSv2-P, UCF101-DS
  - 90 Perturbations: Noise, Camera, Compression, Temporal, Blur
- Findings:
  - **Pre-trained** typically more robust than scratch
  - Robust to time for most datasets, but **not robust when reversible actions** possible.
  - **Transformer**-based typically **more robust**
  - **CNN-based** models typically more robust than transformer-based models when **trained on corruptions**

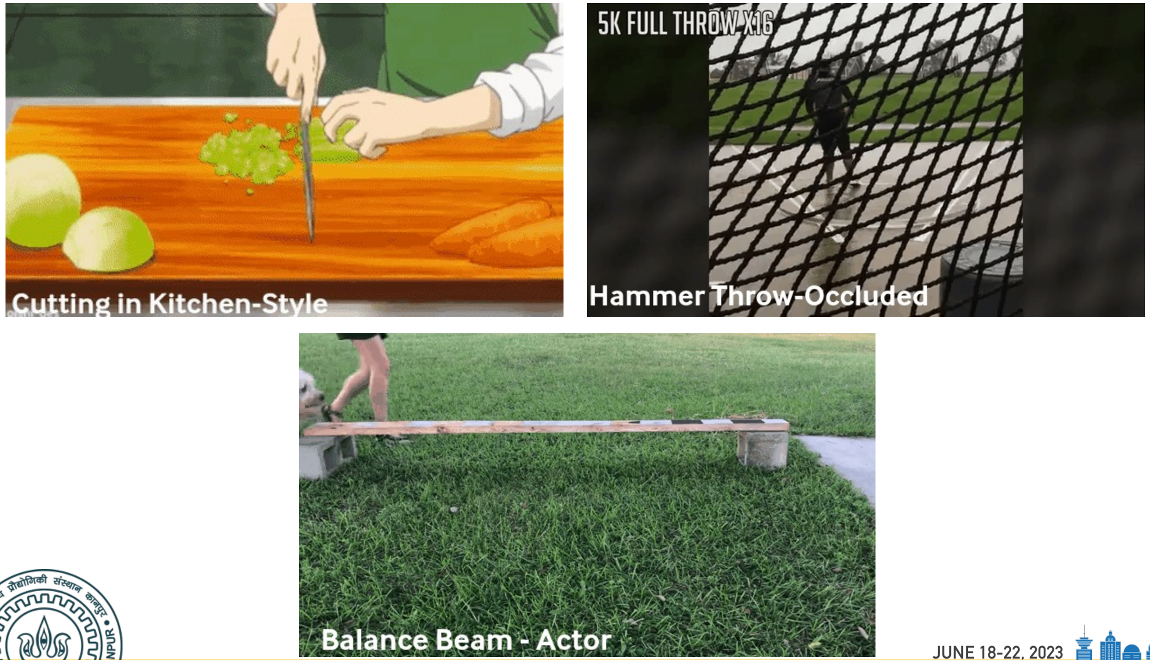
UCF101-P



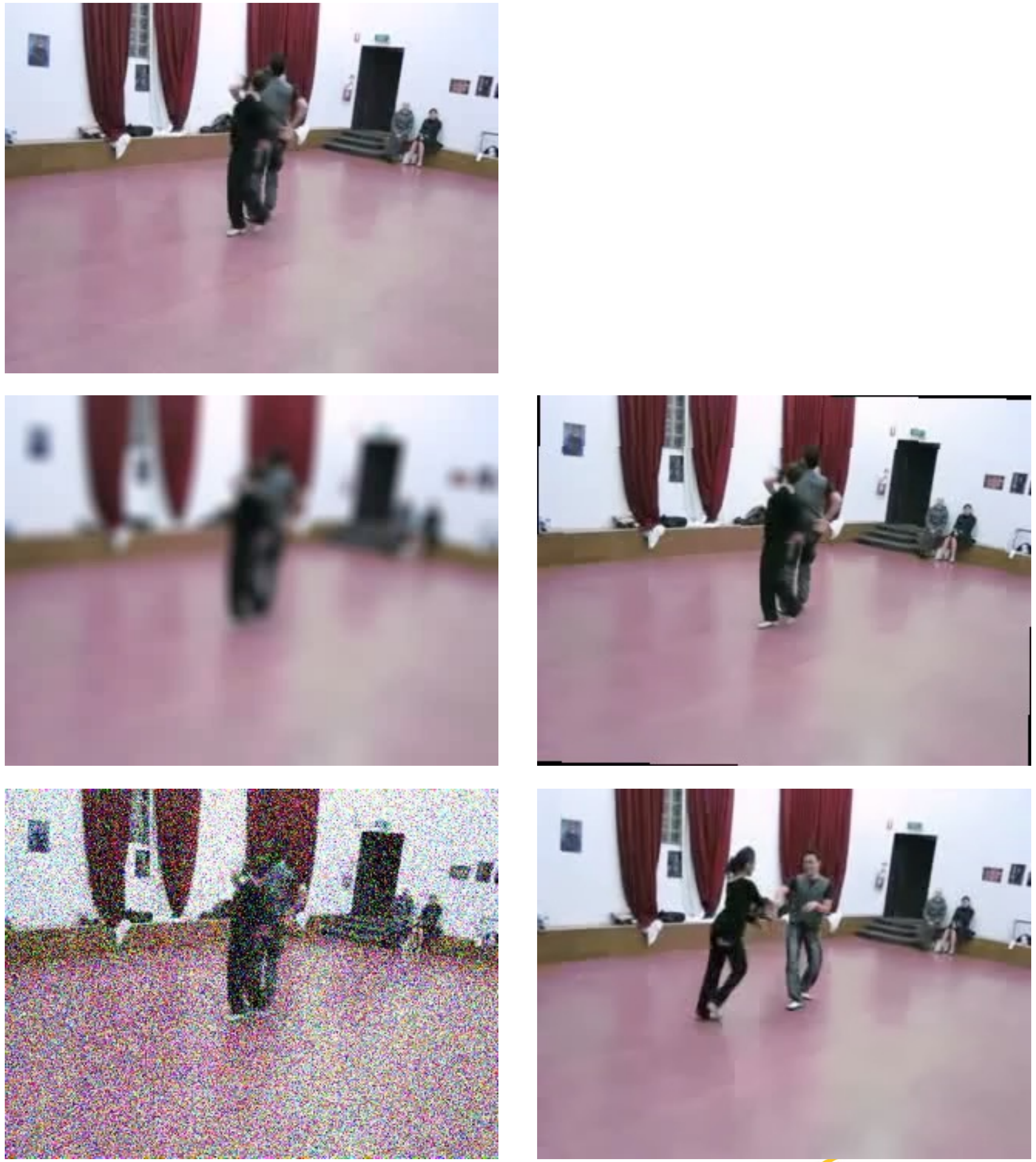
Motivation

"How is performance impacted by a natural distribution shift?"

Curated



Simulated

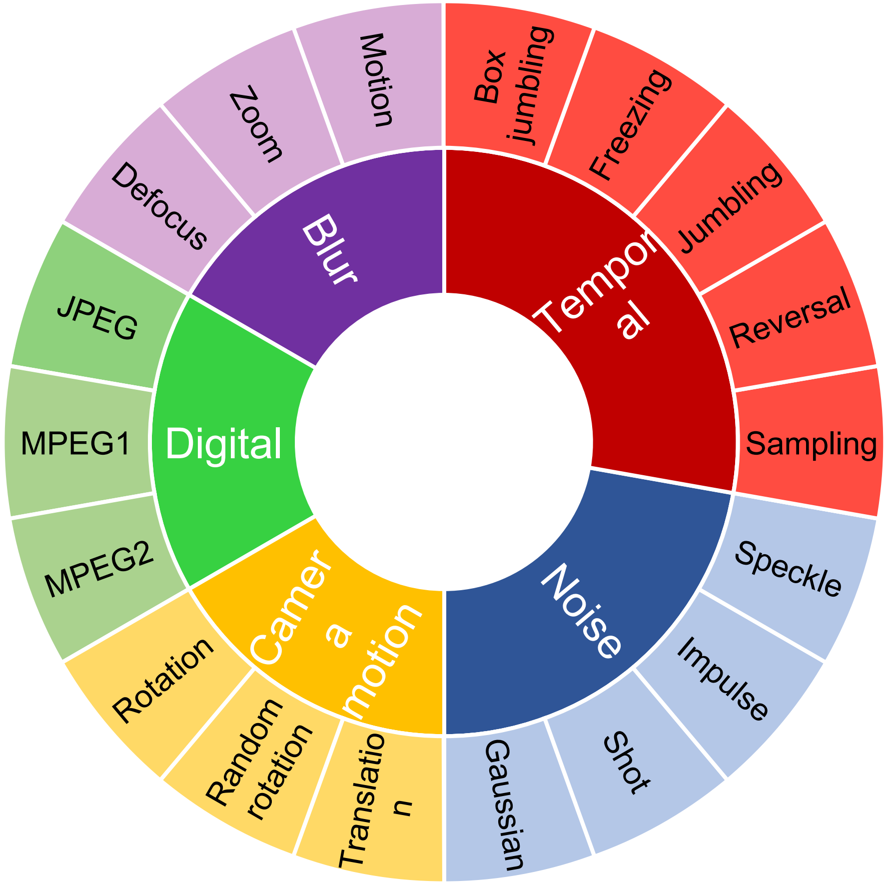


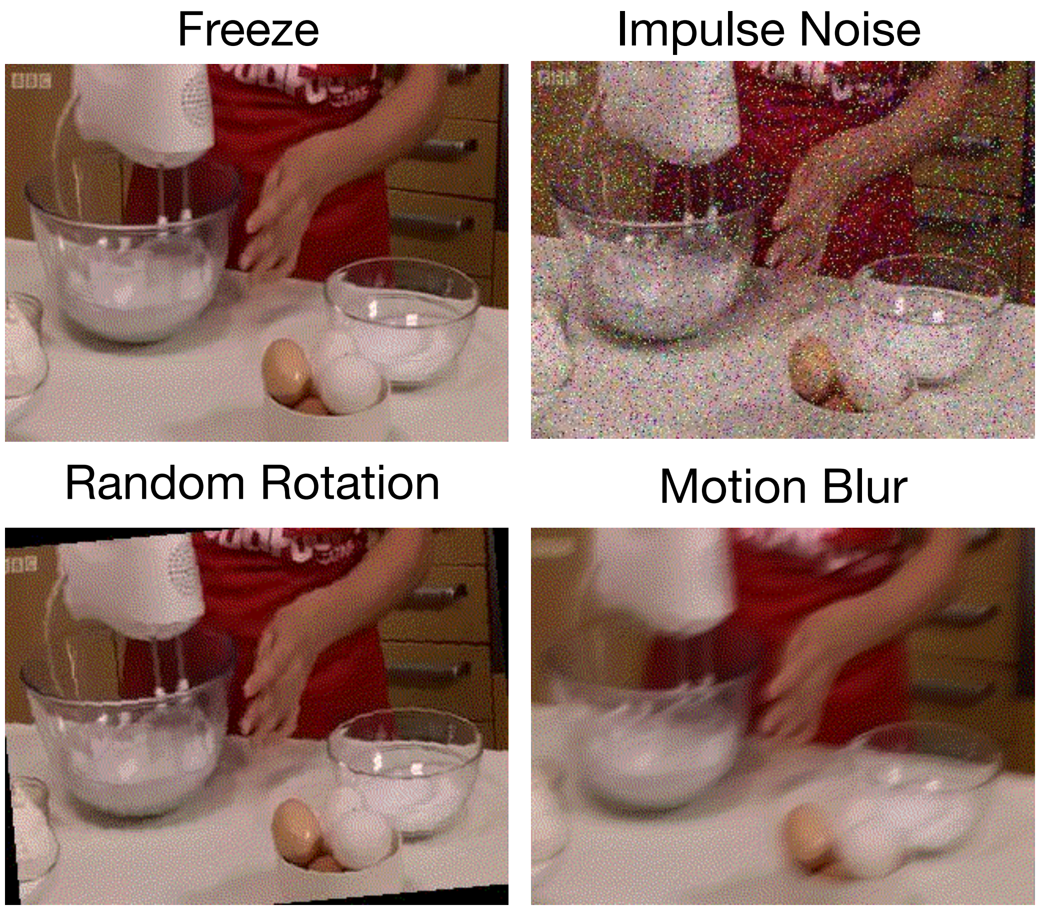
Microsoft Research

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

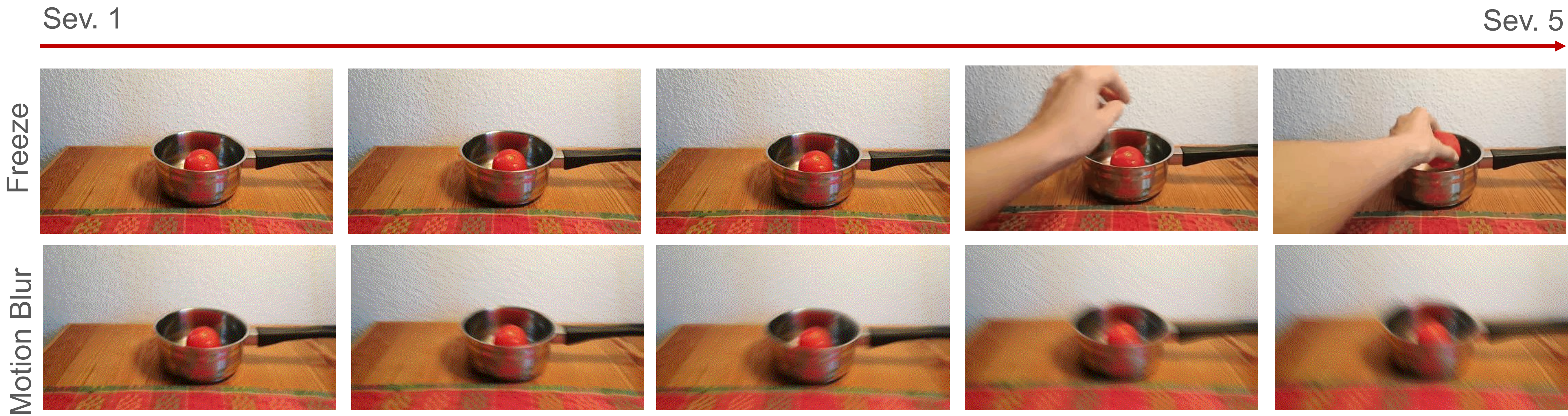
UCF Center for Research In Computer Vision

Perturbations





Severities



Images from SSv2 dataset.

Datasets

- **UCF101-P**
  - 101 action classes.
  - ~350K videos
- **HMDB51-P**
  - 51 action classes
  - ~140K videos
- **Kinetics400-P**
  - 400 action classes
  - ~1.6M videos
- **SSv2-P**
  - 174 action classes
  - ~2.2M videos

Jumble UCF101-P

JPEG Compression HMDB51-P

Defocus Blur Kinetics400-P

Freeze SSv2-P

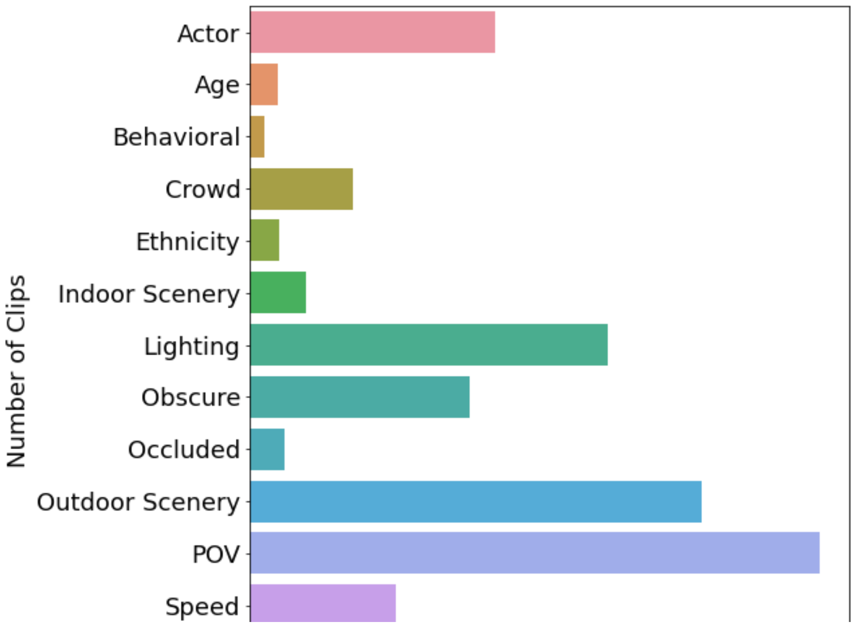
UCF101-DS

- Real distribution shifts
  - 4,708 clips
  - 47 action classes from UCF101
- Keywords for query
  - e.g. “bike riding+fog”
- Higher level categories

Class: Balance Beam **Shift**: Actor | **Class:** Frisbee Catch **Shift:** Obscure | **Class:** Basketball Dunk **Shift**: Point-of-View | **Class:** Marching Band **Shift:** Lighting

UCF101-DS

- Real distribution shifts
  - 4,708 clips
  - 47 action classes from UCF101
- Keywords for query
  - e.g. “bike riding+fog”
- Higher level categories



Category	Distribution Shift
Actor	[animal, costume, toy]
Age	[kids, old_person]
Behavioral	[caught_on_cam!, prank, reaction, scary]
Crowd	[crowd]
Ethnicity	[african, asian, black, indian_brown]
Indoor Scenery	[at_home, at_the_club, at_the_gym, indoor, indoors, in_court, in_garage, mirror]
Lighting	[low_light, at_late_night, at_night, dark, low_light_conditions]
Obscure	[unsual, unusual]
Occluded	[obstructed, obstructed_view]
Outdoor Scenery	[at_the_beach, desert, in_backyard, in_garden, in_the_fields, on_the_road, outdoors, outside, underwater]
POV	[camera_angle, camera_angles, go_pro, on_TV, pov, pov_at_night, shaky, tutorial, upside_down]
Speed	[alow_mo, fastest, slowmotion, slow_mo]
Style	[animated, animation, filter, text_on_screen, vintage]
Weather	[fog, in_rain, muddy, rain, snow]

Metrics and Evaluation

Relative Robustness

- Measures relative drop in accuracy

$$\gamma^r_{p,s} = 1 - (A^f_c - A^f_{p,s})/A^f_c$$

Absolute Robustness

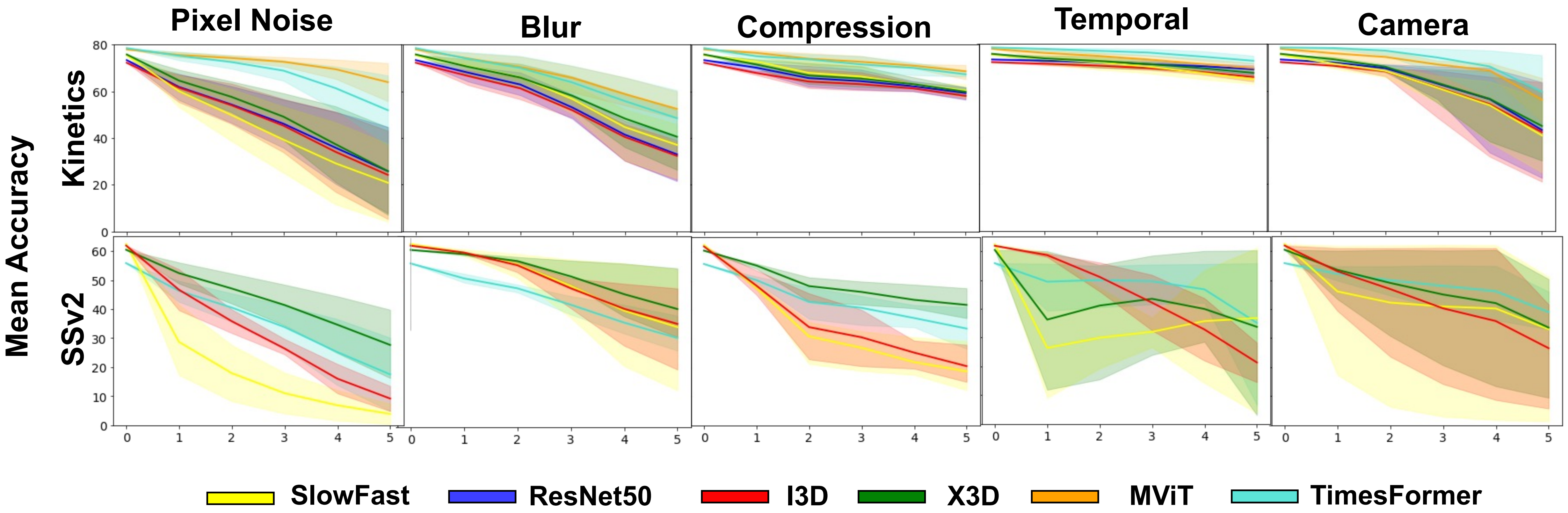
- Measures absolute drop in accuracy

$$\gamma^a_{p,s} = 1 - (A^f_c - A^f_{p,s})/100$$

- p: Perturbation
- s: Severity
- A^f_c: Accuracy on clean video
- $A^f_{p,s}$: Accuracy on perturbed (p) video at severity s

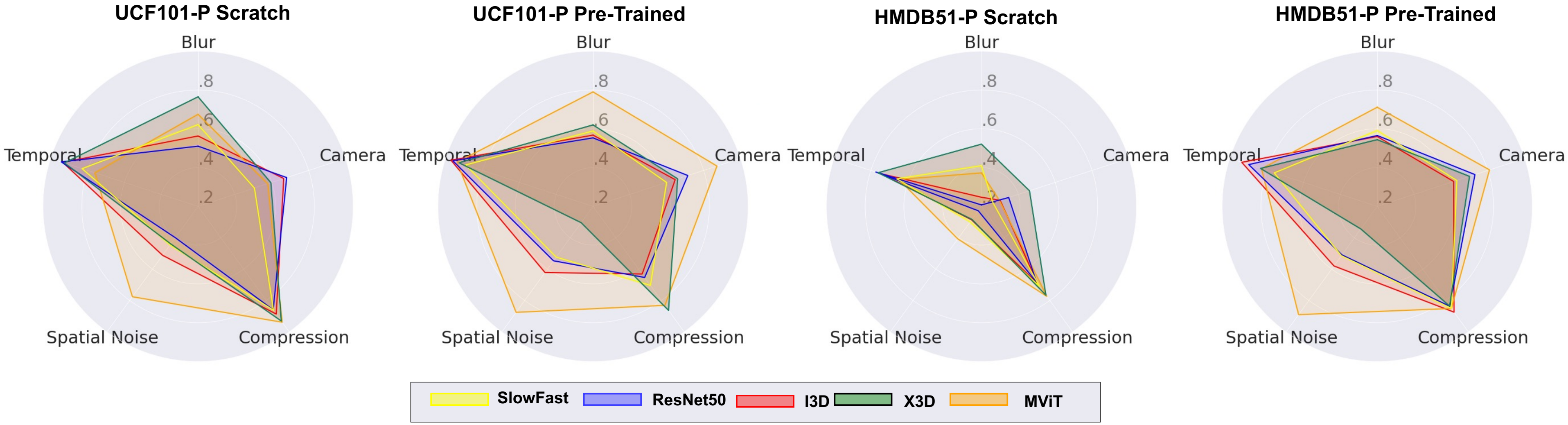
Findings

Overall Results



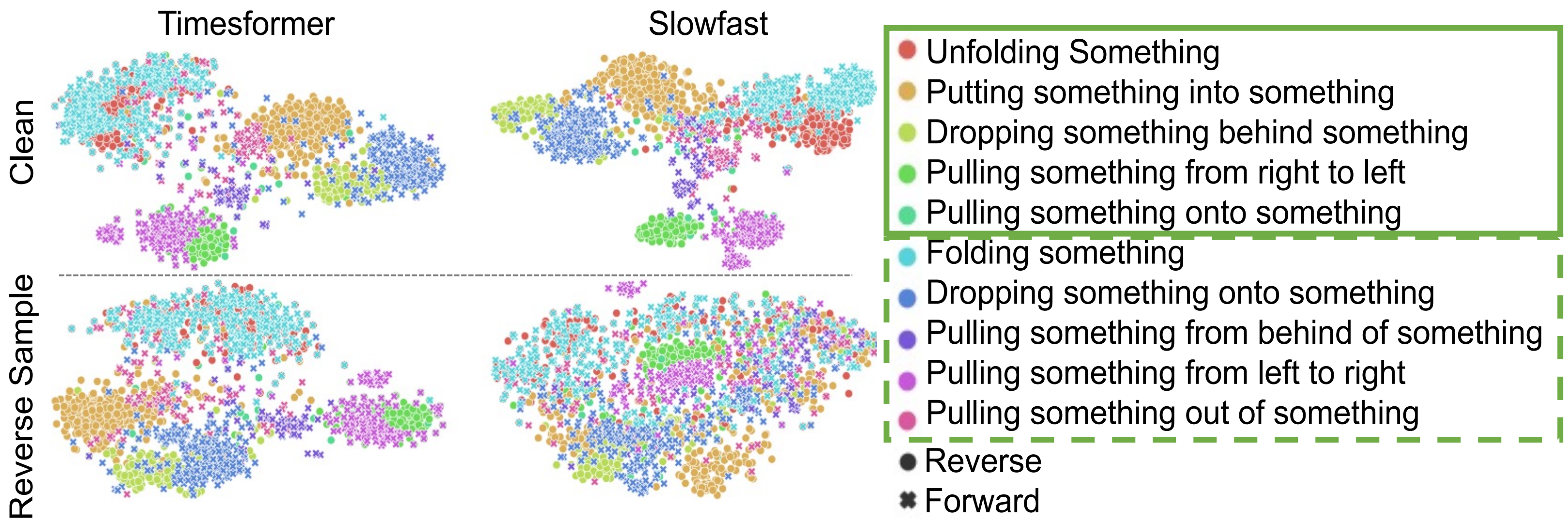
Pre-Training

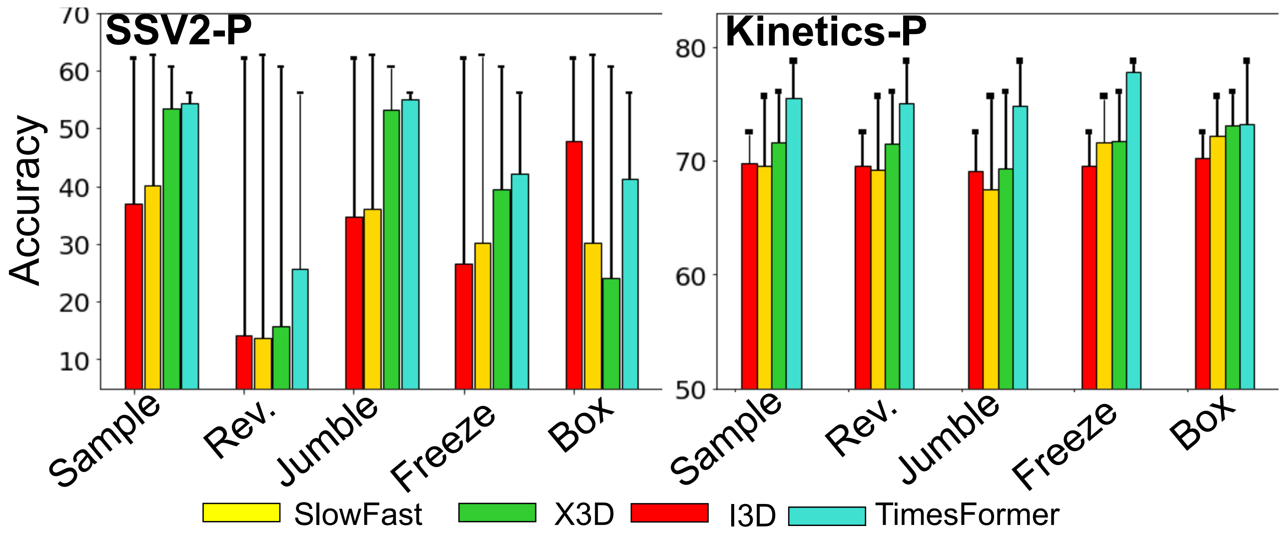
- Pre-training generally improves robustness



Temporal

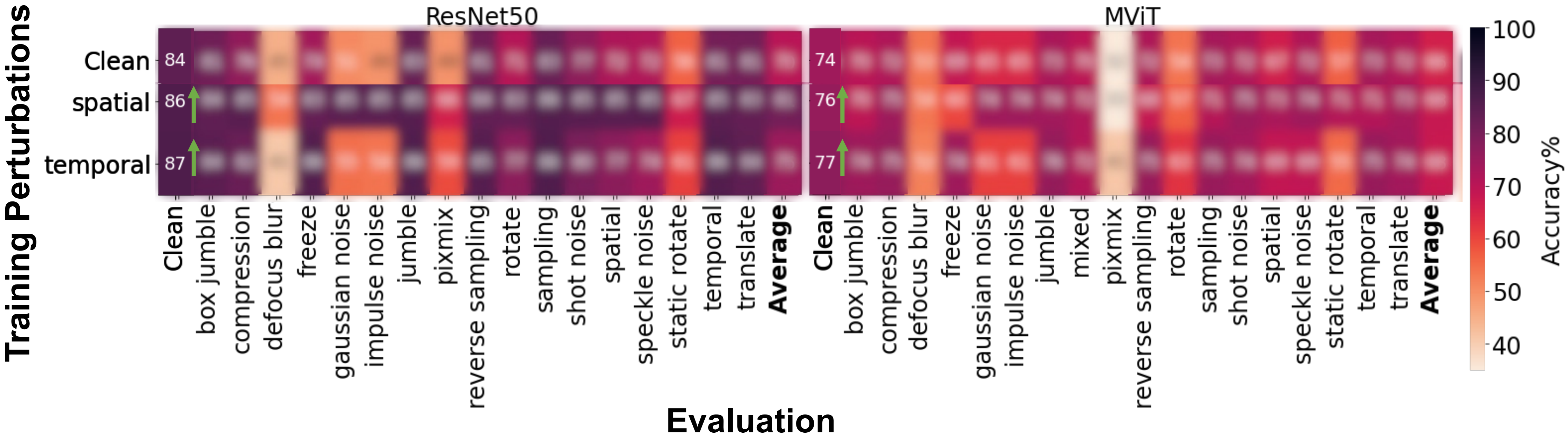
- Robust on all datasets but SSv2
- Some SSv2 frames in reverse, become a different action
  - CNN-based model confused more





Training on Corruptions

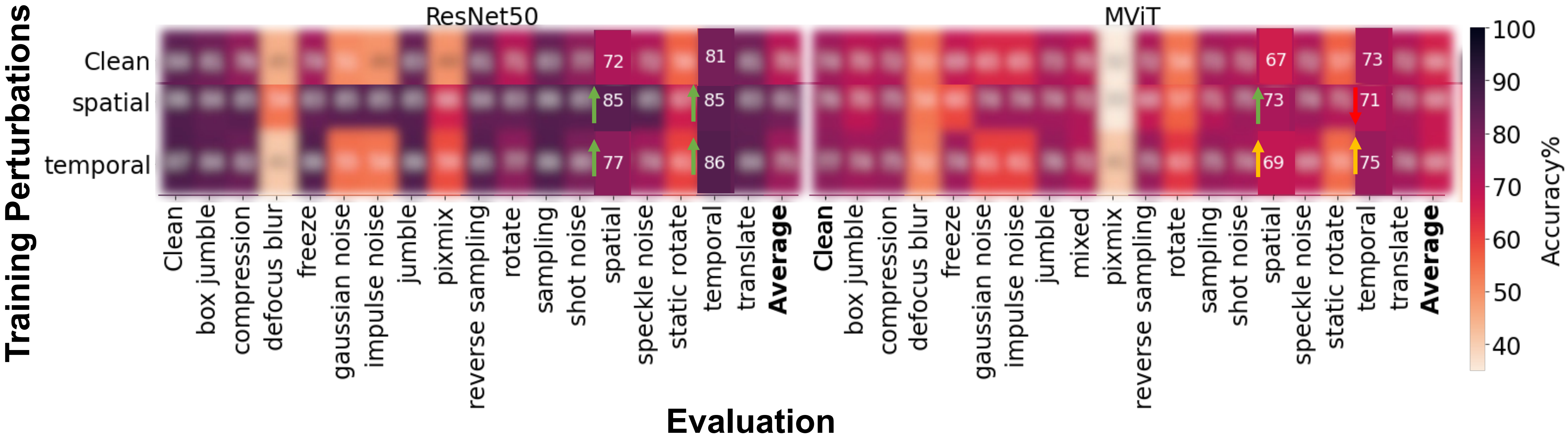
- Train foundational models on corruptions



- Helps on original data

Training on Corruptions

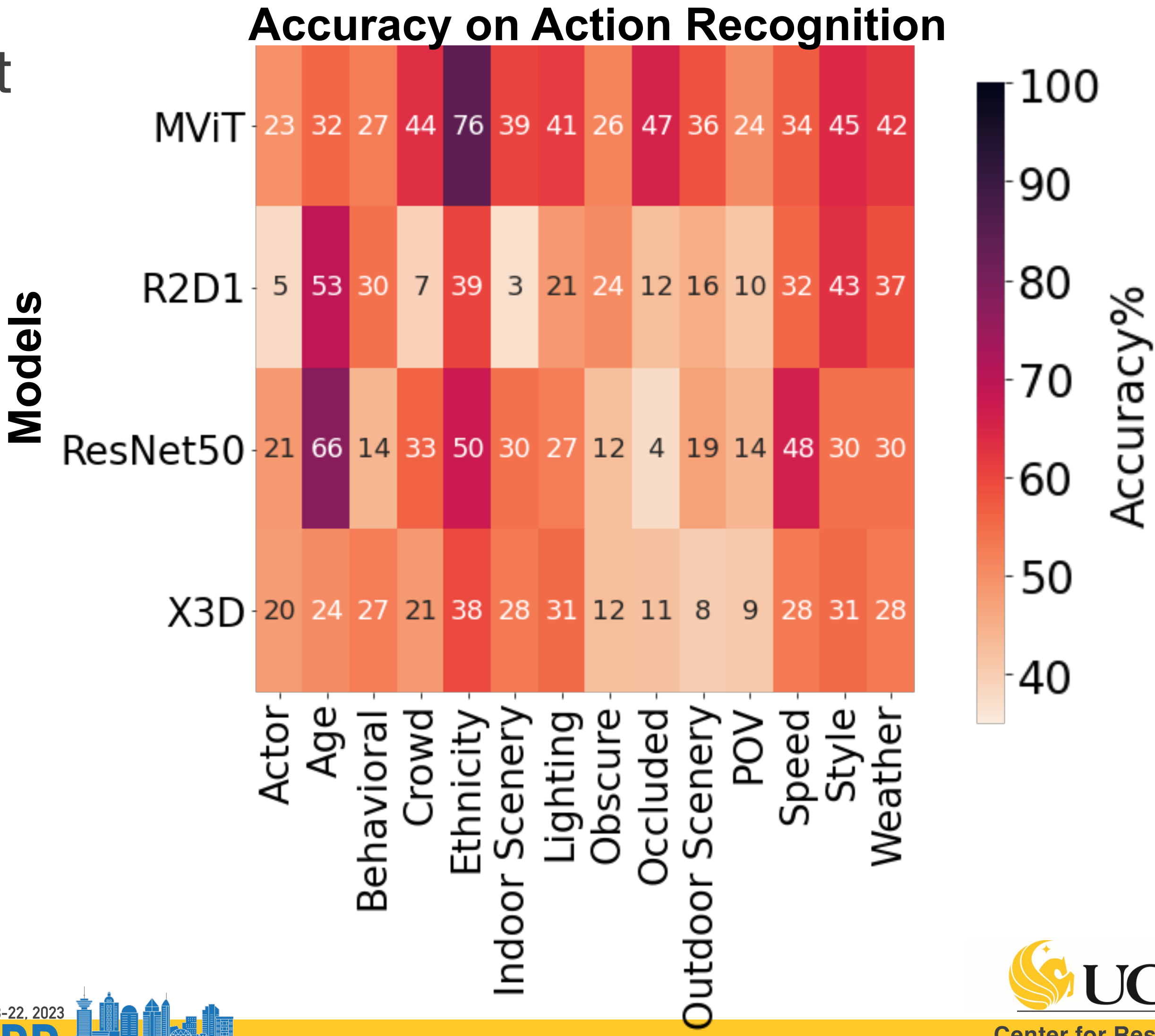
- Train foundational models on corruptions



- Helps on original data
- Helps more on CNN-based architecture

Real Distribution Shifts: UCF101-DS

- Certain architectures more robust



UCF
CENTER FOR RESEARCH IN COMPUTER VISION

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Microsoft Research

Thank You!

WED-PM-007

Paper

Website







Madeline C. Schiappa[1] *Naman Biyani*[3] *Prudvi Kamtam*[1] *Shruti Vyas*[1] *Hamid Palangi*[2] *Vibhav Vineet*[2] *Yogesh S. Rawat*[1]

University of Central Florida[1], *Microsoft Research*[2], *IIT Kanpur, India*[3]