CRCV REU - Week 9

EJIRO ADAMS UBINI
UNIVERSITY OF MARYLAND, BALTIMORE COUNTY
PROJECT LEAD: DR. YOGESH RAWAT

Project Overview

- Project Team: Dr. Yogesh Rawat
- Real-Time Detection of School Bullying Using CCTV
- The project focuses on using deep learning and video activity recognition models to detect bullying behaviors in real time from school surveillance footage.

Problem

- Existing surveillance systems in schools are passive and rely on human monitoring, which is inefficient and prone to oversight.
- There is a lack of robust, real-time systems that can detect aggressive or bullying behaviors using visual input from CCTV cameras.
- Moreover, there is limited publicly available data tailored to the school bullying context, especially in natural school environments.

Contributions

- 12-class curated dataset
- Clips from 2 datasets, movies & social media
- Normal + mixed-label bullying clips
- Fine-tuned action recognition model
- Baseline: 43% top-1 accuracy

Dataset Curation Progress

- Existing datasets curated
  - UCF-Crime
  - RWF-2000
- Clips collected online
  - YouTube
  - Facebook
  - X
  - Instagram

Dataset collected

n/o	Classes	Number of videos
1	forceful choking	90
2	dragging a person	41
3	elbow strike	12
4	hair pulling	48
5	headbutt attack	8
6	hitting a person with an object	247
7	aggressive holding	288
8	kicking a person	296
9	punching a person	706
10	forceful shoving	430
11	slapping a person	119
	multi label	347
	normal	449
	TOTAL	**3081**

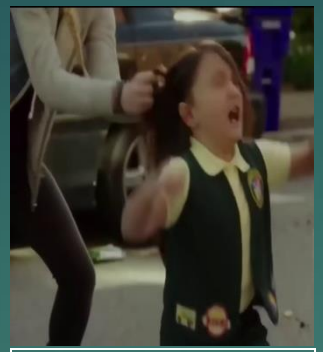
Samples from Dataset



Choke



Drag



Hair grab



Hit with object



Punch



Shove



Elbow strike



Kick



Headbutt



Hold



Slap



Non-bullying



Elbow strike



Kick



Choke



Hold & Kick



Hold & Punch



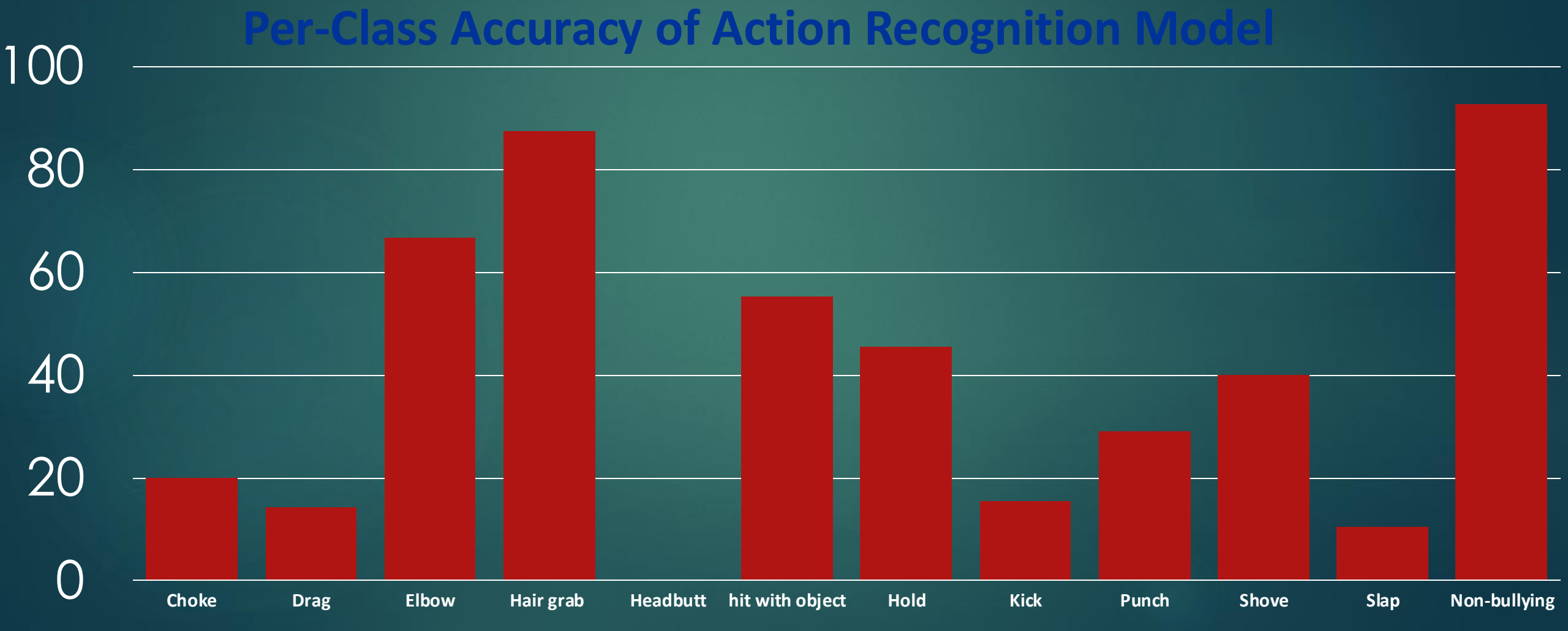
Slap

T2L video recognition model zero shot test (100 frames)

n/o	Classes	Accuracy
1	forceful choking	Top1: 5.555555555555555, Top5: 31.11111111111111
2	dragging a person	Top1: 0.0, Top5: 7.317073170731707
3	elbow strike	Top1: 0.0, Top5: 8.333333333333332
4	hair pulling	Top1: 2.083333333333333, Top5: 91.66666666666666
5	headbutt attack	Top1: 0.0, Top5: 100.0
6	hitting a person with an object	Top1: 0.404858299595141 7, Top5: 46.558704453441294
7	aggressive holding	Top1: 0.6944444444444444, Top5: 5.555555555555555
8	kicking a person	Top1: 2.364864864864865, Top5: 59.797297297297305
9	**punching a person**	**Top1: 41.21813031161473, Top5: 99.0084985835694**
10	forceful shoving	Top1: 0.23255813953488372, Top5: 20.46511627906977
11	slapping a person	Top1: 1.680672268907563, Top5: 68.90756302521008

T2L Test Results after training (Sampled 25 frames, 20 epochs)

- ~43% overall accuracy



Misclassification Patterns

- Choking a person
- Dragging a person
- Hair grab
- Hit with an object
- Punching a person
- Shoving a person
- Elbow strike
- Kicking a person
- Headbutt attack
- Aggressive holding
- Slapping a person
- Non-bullying



Analysis

- Trained for 20 epochs on the dataset
- Sampled 25 frames per video uniformly
- Hard recognizing subtle action differences
- Non-bullying scenes are mostly accurate
- Dataset imbalance contributes to bias

Conclusion

- ~43% overall accuracy
- >90% accuracy for non-bullying scenes
- Strong on distinct actions like hair grab
- weak on subtle or overlapping motions
- smarter sampling may boost accuracy

Future Work

- Expand and balance dataset
- Add distinct bullying actions
- Annotate dataset for higher-quality labels
- Explore open-vocabulary action recognition
- Benchmark more SOTA models

References

- [1] Ahmad, Shahzad, Sukalpa Chanda, and Yogesh S. Rawat. "T2L: Efficient Zero-Shot Action Recognition with Temporal Token Learning." *Transactions on Machine Learning Research*.
- [2] Cheng, Ming, Kunjing Cai, and Ming Li. "RWF-2000: An open large scale video database for violence detection." *2020 25th International conference on pattern recognition (ICPR)*. IEEE, 2021.
- [3] Sultani, Waqas, Chen Chen, and Mubarak Shah. "Real-world anomaly detection in surveillance videos." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.