Real-Time Detection of School Bullying Using CCTV

Adams Ubini[1], Yogesh Rawat[2]

[1]University of Maryland, Baltimore County, [2]University of Central Florida

UCF Center for Research in Computer Vision

UMBC

Motivation

- School bullying often goes unreported
- CCTV is underused, no real-time analytics
- AI advances enable bullying detection
- No public dataset for school bullying

Problem Statement

- Passive surveillance in schools
- Manual monitoring required
- Models not trained on school bullying data

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

Classes	Number of Videos
Choking a person	90
Dragging a person	41
Elbow strike	12
Hair grab	48
Headbutt attack	8
Hit with an object	247
Aggressive holding	288
Kicking a person	296
Punching a person	706
Shoving a person	430
Slapping a person	119
Multi label	347
Non-bullying	449
TOTAL	3081

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

Samples from Dataset



Per-Class Accuracy of Action Recognition Model



Acknowledgement

The work is funded by NSF CNS-2349386 under Research Experiences for Undergraduates (REU) Program

Misclassification Patterns

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

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.