Concatenation of Attention Enhanced Spatial and Temporal Features for Violence Detection from Videos

Main Project Presentation: Final Review

Guided by:

Dr. Renjith Thomas, HOD AD

Presented by:

Batch 4

Ajay T Shaju, SJC20AD004
Emil Saj Abraham, SJC20AD028
Justin Thomas Jo, SJC20AD046
Vishnuprasad KG, SJC20AD063

Outline

- Introduction
- Literature Survey
- Problem Statement
- Objectives
- Data Collection
- Materials and Methods
- Model Architecture
- Results and Discussion
- Conclusion and Future Scope
- Conference Details and Project Competition
- References

Introduction

- **Violence** have been on the rise globally.

- **Stampede** at **CUSAT Campus**, Kochi, Kerala on 25th November 2023
  - Auditorium at max capacity.
  - No crowd monitoring.
  - Lack of safety measures.

- **AI to prevent violence**
  - Real-time detection.
  - Early warning.

Introduction

Figure 1: CUSAT Incident NEWS Report

Introduction

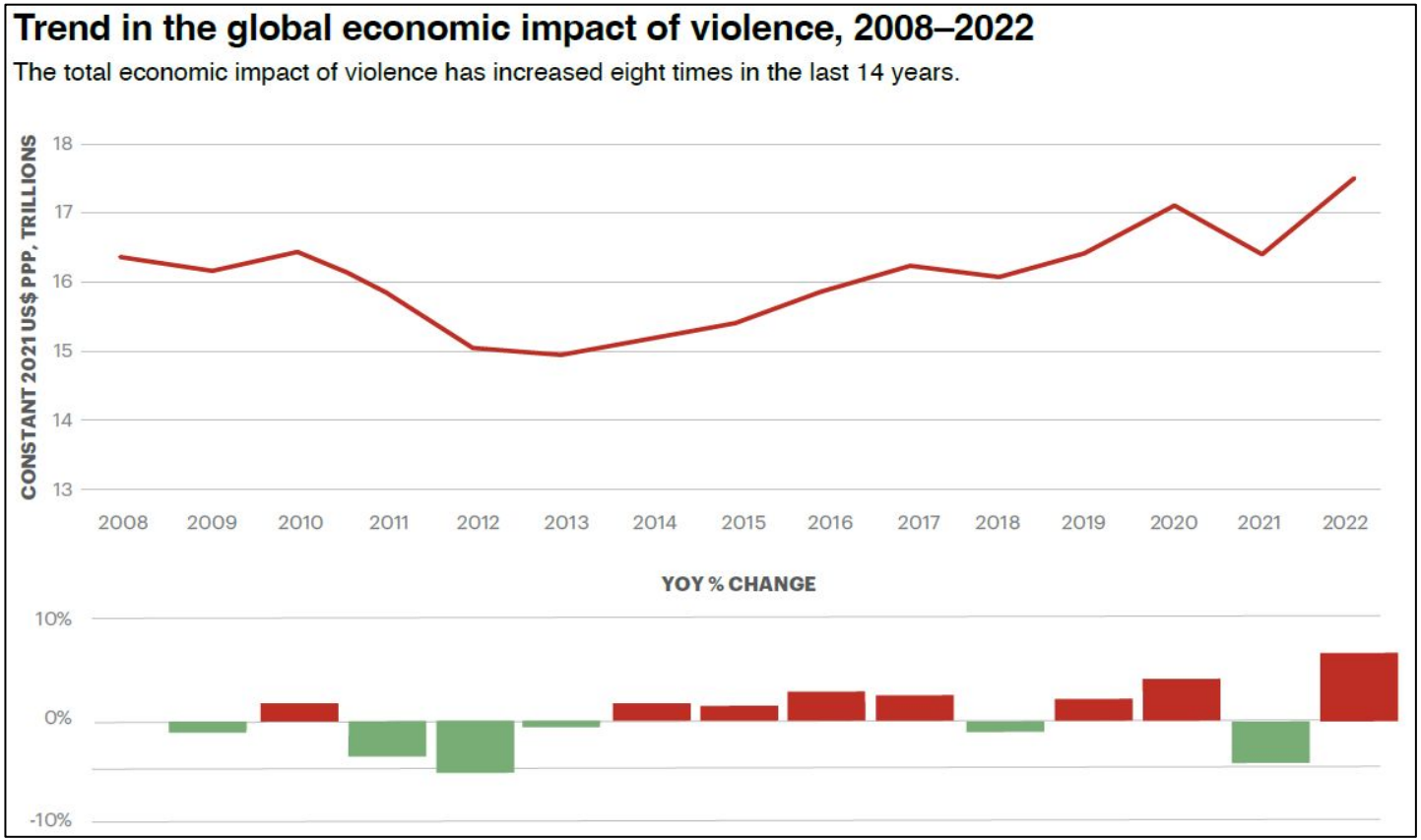


Figure 2: Economic Impact of Violence

Introduction

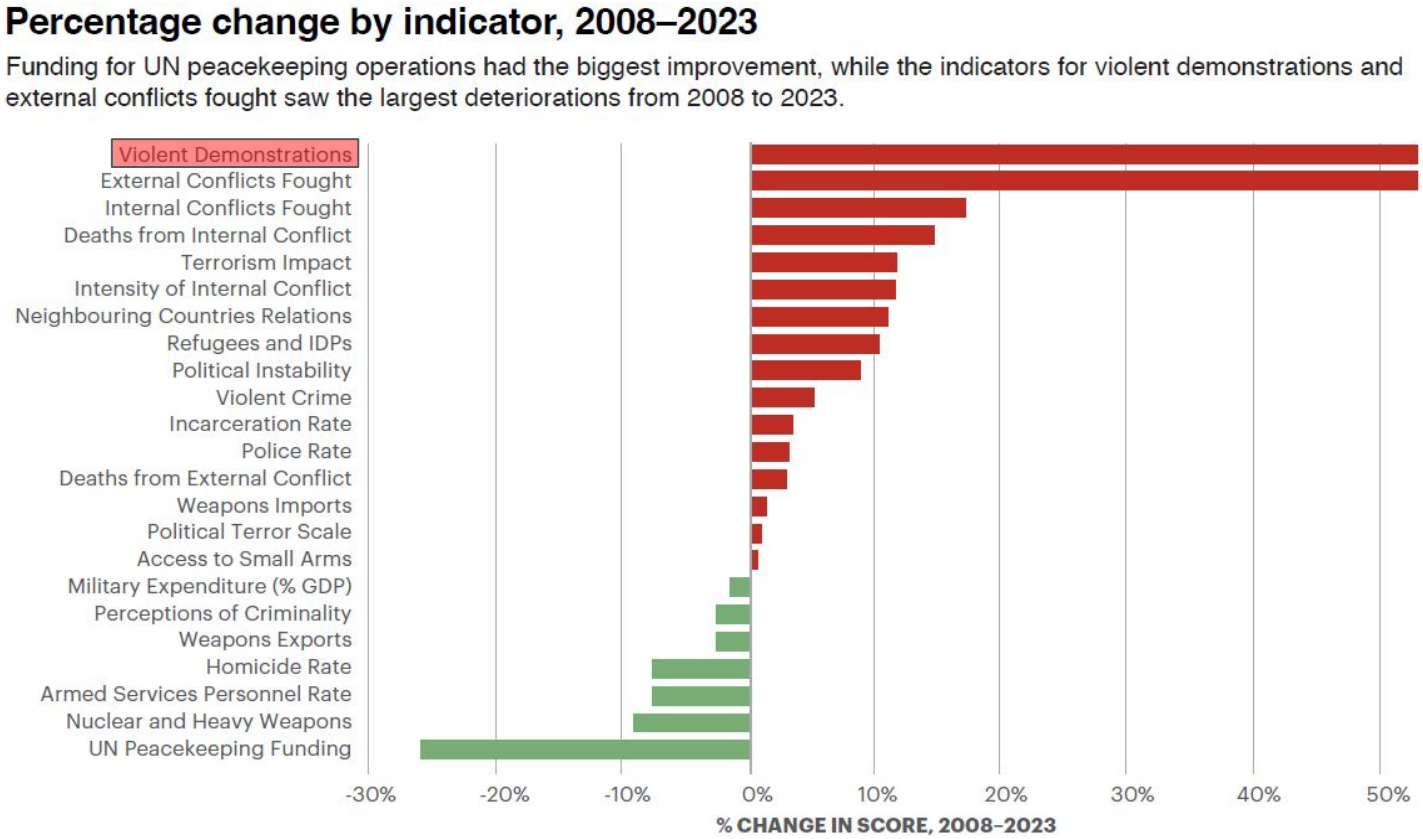


Figure 3: Percentage change in violence from 2008 - 23

Figure 4: Pictures of Violence

Literature Review

[1] M. Safaa M. Shubber, Z. Tariq M. Al-Ta'i, **"A review on video violence detection approaches"**, International Journal of Nonlinear Analysis and Applications, Vol. 13, No. 2, pp: 1117–1130, 2022.

- **Machine learning + Deep learning**
- CNN: Reliable results.
  LSTM: Gradient extinction, time dimension information
- **Drawback:** Supervised learning; large no. of training samples, expensive hardware

Literature Review

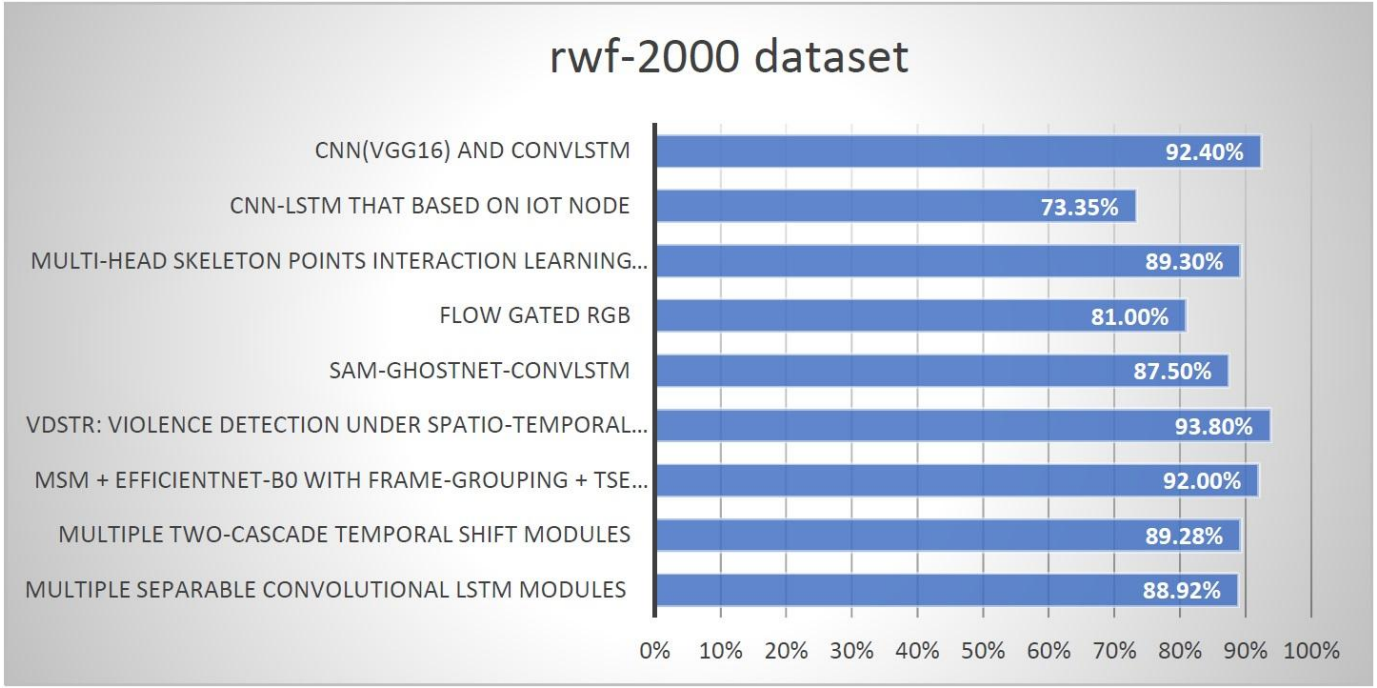


Figure 5: Violent Behavior Detection accuracy results on RWF-2000 dataset

Literature Review

[2] S. Lomlen, **“Impact of Artificial Intelligence in enhancing National Security”**, Artificial Intelligence Studies, Vol. 1, pp: 1-7, Feb 2024.

- Exponential growth of AI, enhance national security and safety.
- Challenges of AI in defense and military like situations.
- **Drawback:** Lack of robustness, safety of implementation, ethical concerns - decision making capability.

Literature Review

[3] H.Gupta, Syed T. Ali, **“Violence Detection using Deep Learning Techniques”,** International Conference on Emerging Techniques in Computational Intelligence (ICETCI), pp: 121-124, Oct 2022.

- **LSTM and BiLSTM** - **sequence modeling** and **capturing temporal dependencies** in data.
- **Drawback: Biases/imbalances**, computationally expensive, significant resources required.

Literature Review

Figure 6: Contents of Violent Flow dataset used in the proposed work

Literature Review

[4] K. Aarthy, A. Alice Nithya, **"Crowd Violence Detection in Videos Using Deep Learning Architecture"**, IEEE 2nd Mysore Sub Section International Conference (MysuruCon), 1-6, Dec 2022.

- Research done on **Hockey dataset**.
- **Keyframe extraction technique** - remove duplicate consecutive frames from the input video; reduces the **training data** and **computational cost.**
- **Drawback:** Not easy to generalize to new and diverse datasets; suboptimal hyperparameter tuning - average model performance.

Literature Review

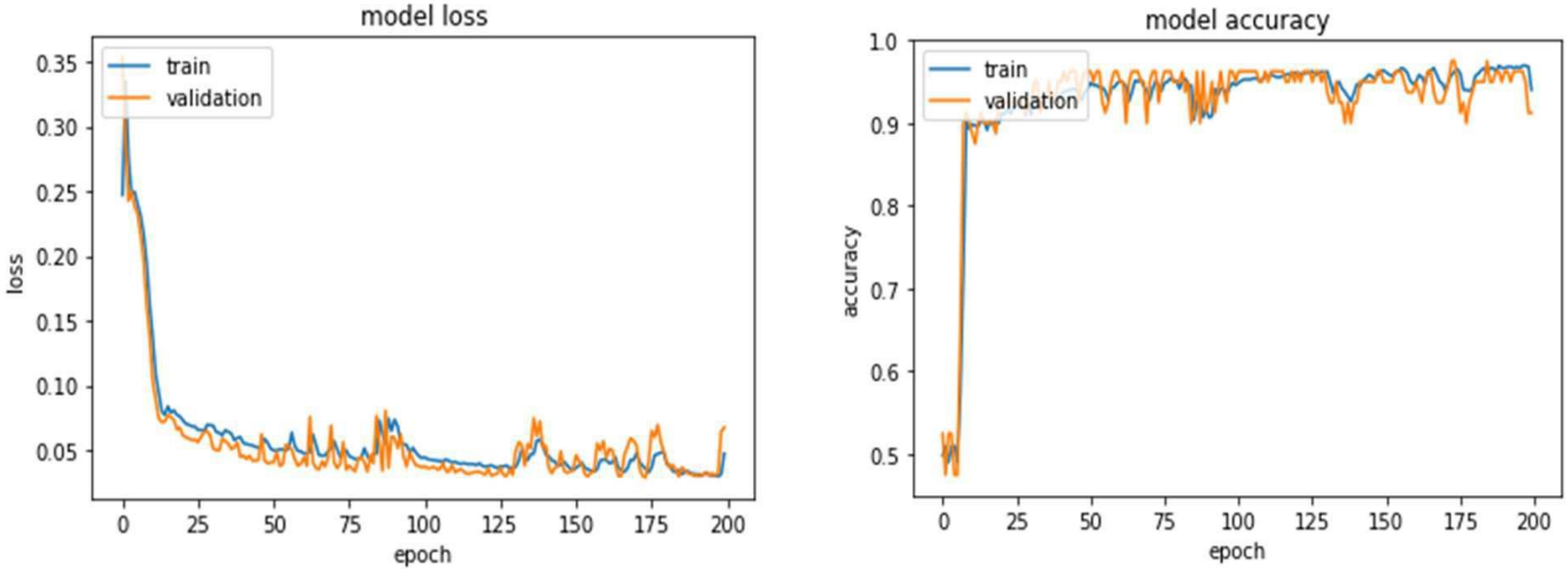


Figure 7: Model loss and model accuracy on Hockey dataset

Problem Statement

- **Modern models are too big for edge devices**.
- **Limited resources and storage**
- **High hardware demands**; **lengthy training times**.
- **Overfitting**

Objectives

- **Make model that**
  - **Less prone to overfitting.**
  - **Shallow network with additional mechanisms for classification.**
  - **Fit into the memory of an edge devices like surveillance cameras.**
  - **Trains in hardware with a reasonable specifications.**

Data Collection

Hockey Fight Dataset

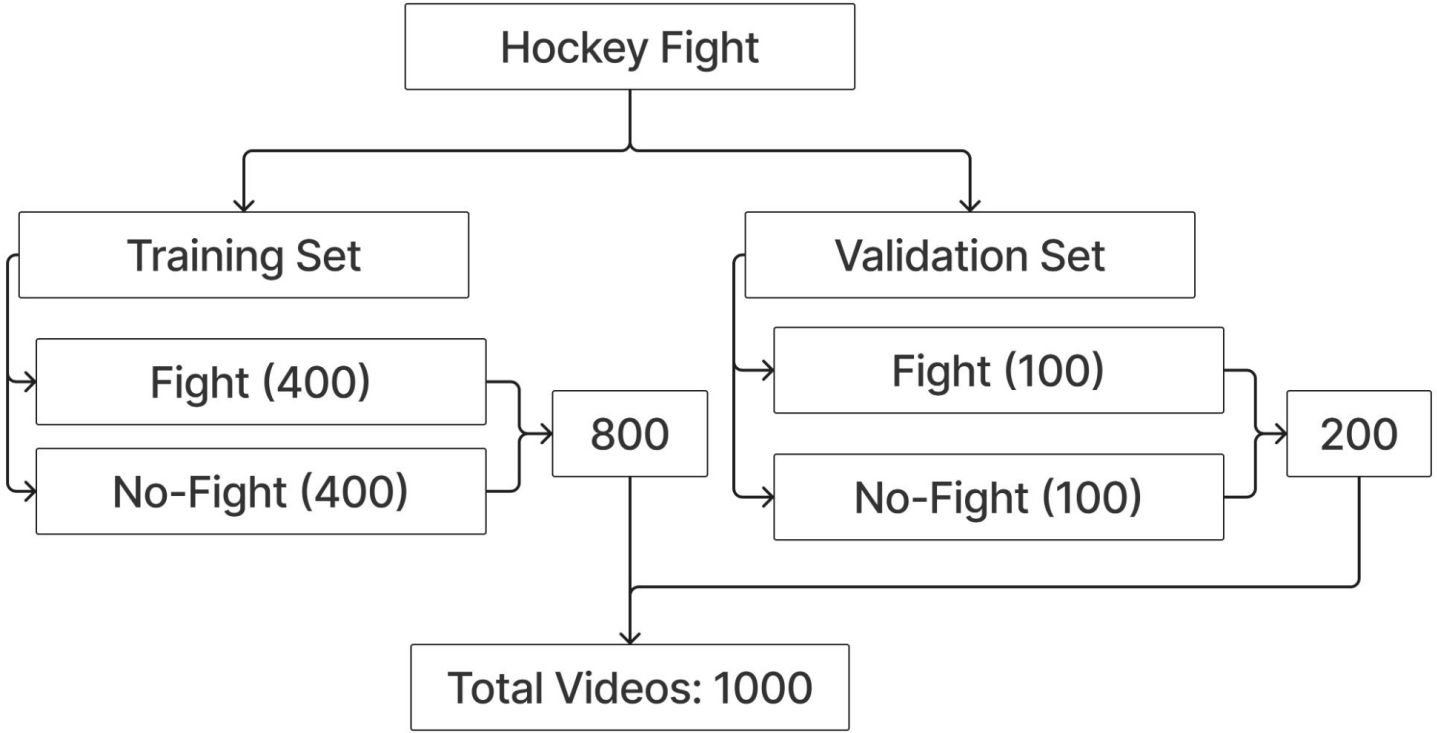


Figure 8: Hockey Fight Dataset Arrangement

Data Collection

```
├──train
│    ├──Fight
│    │       fi5_xvid.avi
│    │
│    └──No_Fight
│            no150_xvid.avi
│
└──val
     ├──fight
     │       fi29_xvid.avi
     │
     └──no_fight
             no303_xvid.avi
```

Figure 9: Directory structure of Dataset

Hockey Fight Dataset

Figure 10: Fight videos

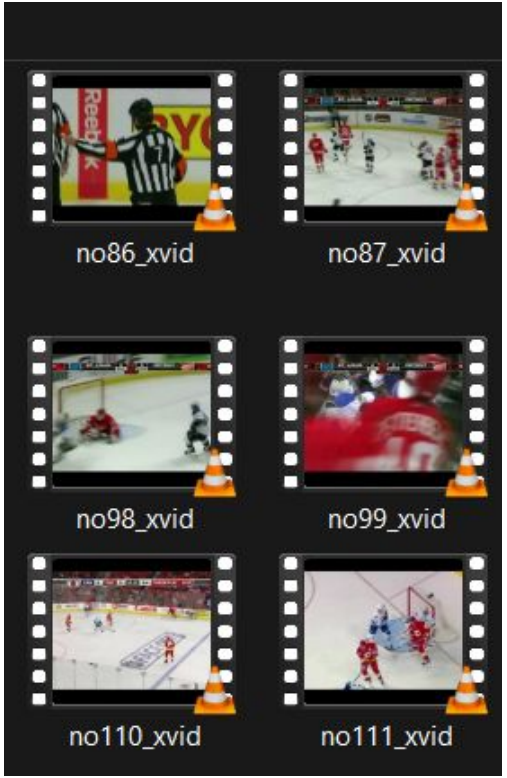


Figure 11: No-Fight videos

Data Collection

Hockey Dataset

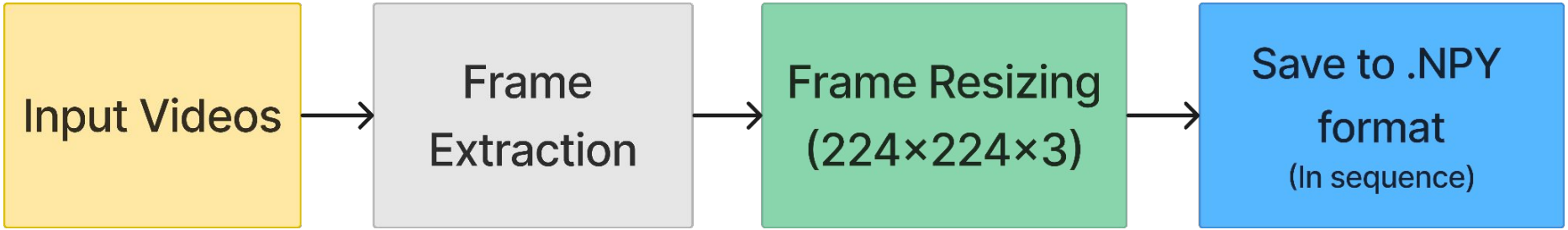
Advantages

- Compact Dataset with **1000 videos**
- **Clarity** of motion and actions
- Clear background **separation**
- **Uniform Size** (360x228)

Disadvantages

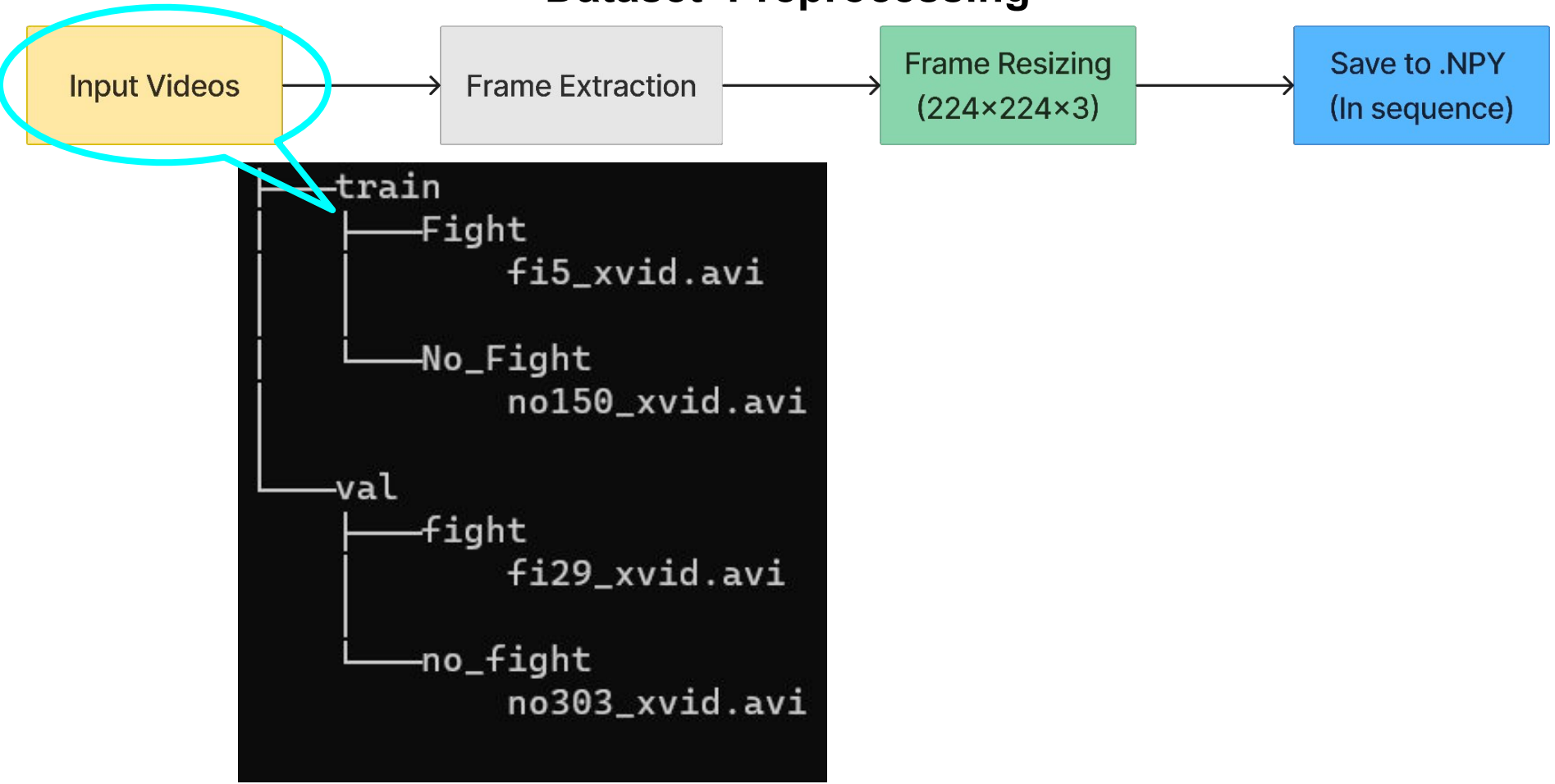
- **Lack of diversity** in the types of events captured
- **Variable no. of participants**,harder for algorithm to detect
- **Ethical considerations** on the content used **without proper context**

Dataset Preprocessing

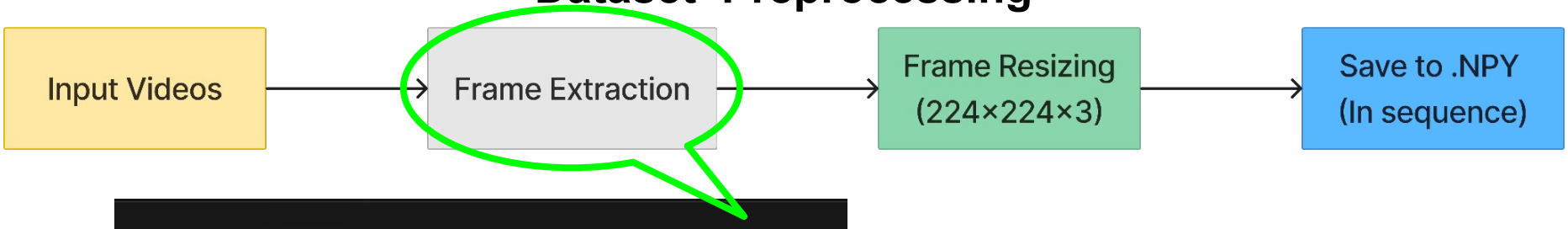


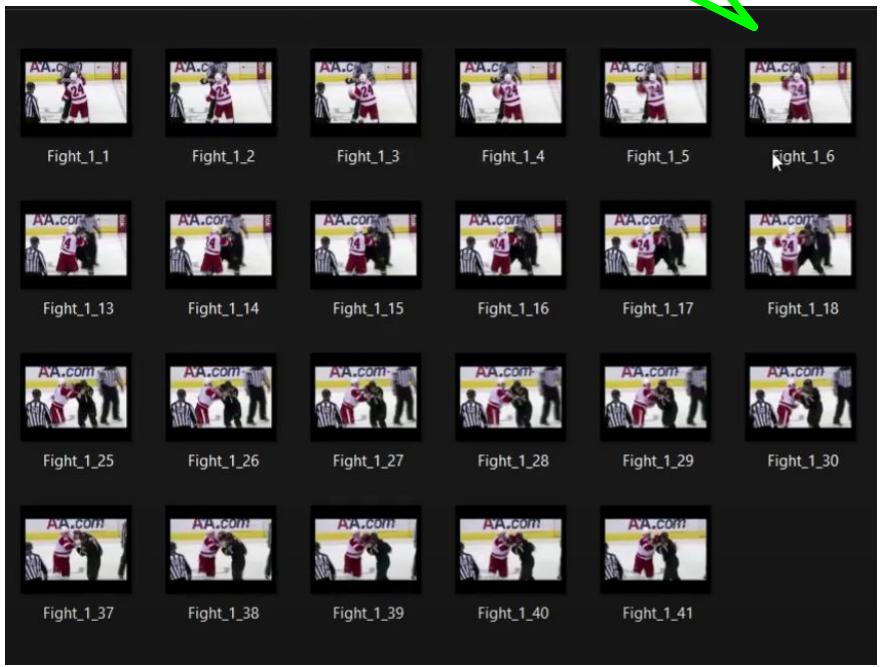
Block Diagram 1: Dataset Preprocessing steps

Dataset Preprocessing



Dataset Preprocessing

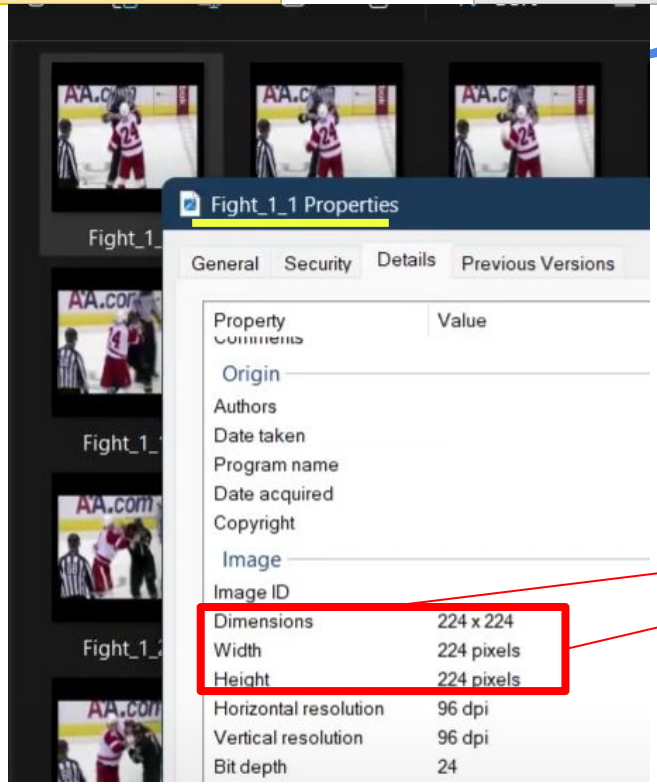




Frame Extraction: Videos can't be directly processed, it has to be converted to frames in sequence.

Dataset Preprocessing

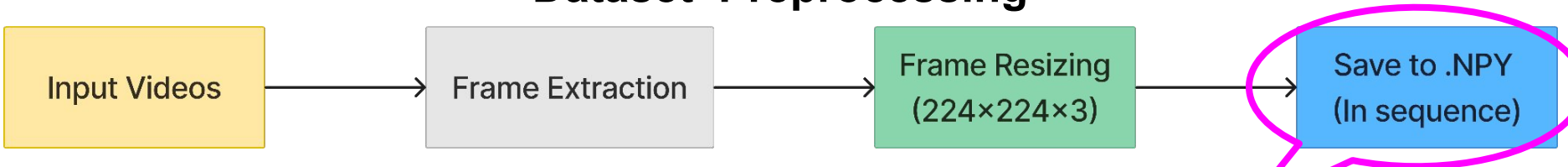
Input Videos → Frame Extraction → Frame Resizing (224×224×3) → Save to .NPY (In sequence)



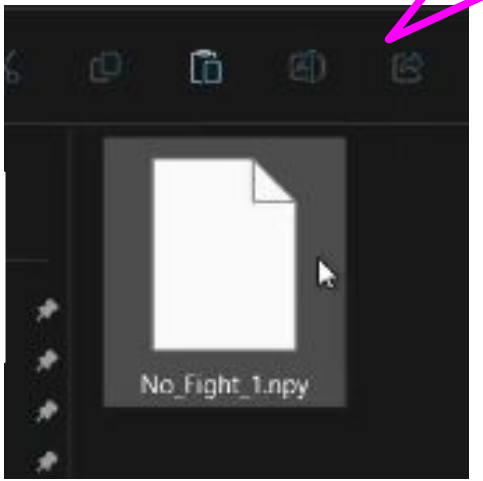
Frame Resizing: Models can't work with dynamic resolutions, resizing creates uniform and standard resolution across dataset.

Dimensions	224 x 224
Width	224 pixels
Height	224 pixels

Dataset Preprocessing



Npy File Conversion: Better handling–less complexity, memory mapping.

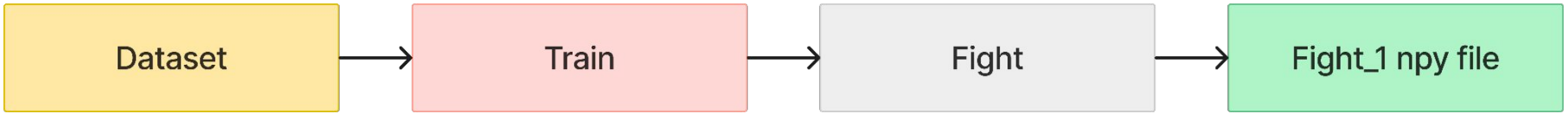



```
Saved D:/Hockey fight/Resized\train\Fight\Fight_1.npy with shape (41, 224, 224, 3)
Deleted folder D:/Hockey fight/Resized\train\Fight\Fight_1
```

Storing and searching in normal files:

Block Diagram 2: Normal files search

Storing and searching in npy files:



Block Diagram:3 Npy files search

Information about a single npy file(filename: Hockey/Fight_10.npy):
{'data_type': dtype('float32'), 'shape': (41, 224, 224, 3), 'number_of_frames': 41}

Figure 12: Visualization of frames stored in a npy file

Materials and Methods

Materials:

- **Primary Laptop: Acer Nitro AN515-57**
  - Intel Core i7-11th Gen CPU | 24GB RAM | 1TB SSD ROM | **Nvidia Geforce RTX 3070 GPU (5,888 CUDA Cores) 8GB VRAM**
- **Secondary Laptop: Asus ROG Strix G15**
  - Intel Core i7-10th Gen CPU | 16GB RAM | 1TB SSD ROM | **Nvidia GeForce GTX 1660Ti GPU (1,536 CUDA Cores) 6GB VRAM**
- **Coding Platform:** VS Code, **Python** 3.10.11, **TensorFlow 2.10** GPU Accelerated with CUDA and cuDNN Integration*.

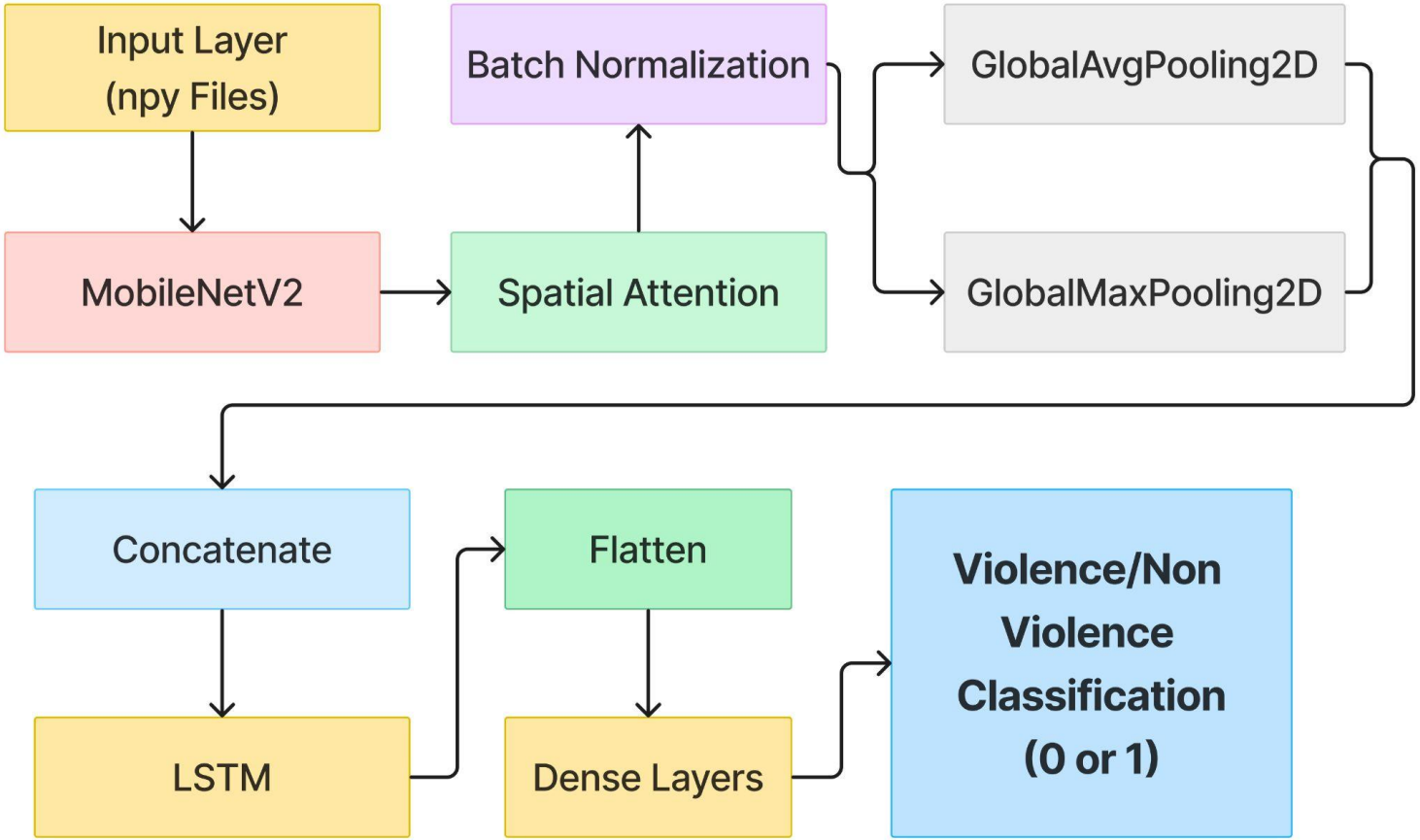
** TensorFlow do not support GPU acceleration its versions greater than 2.10(Windows)*

Materials and Methods

Methods:

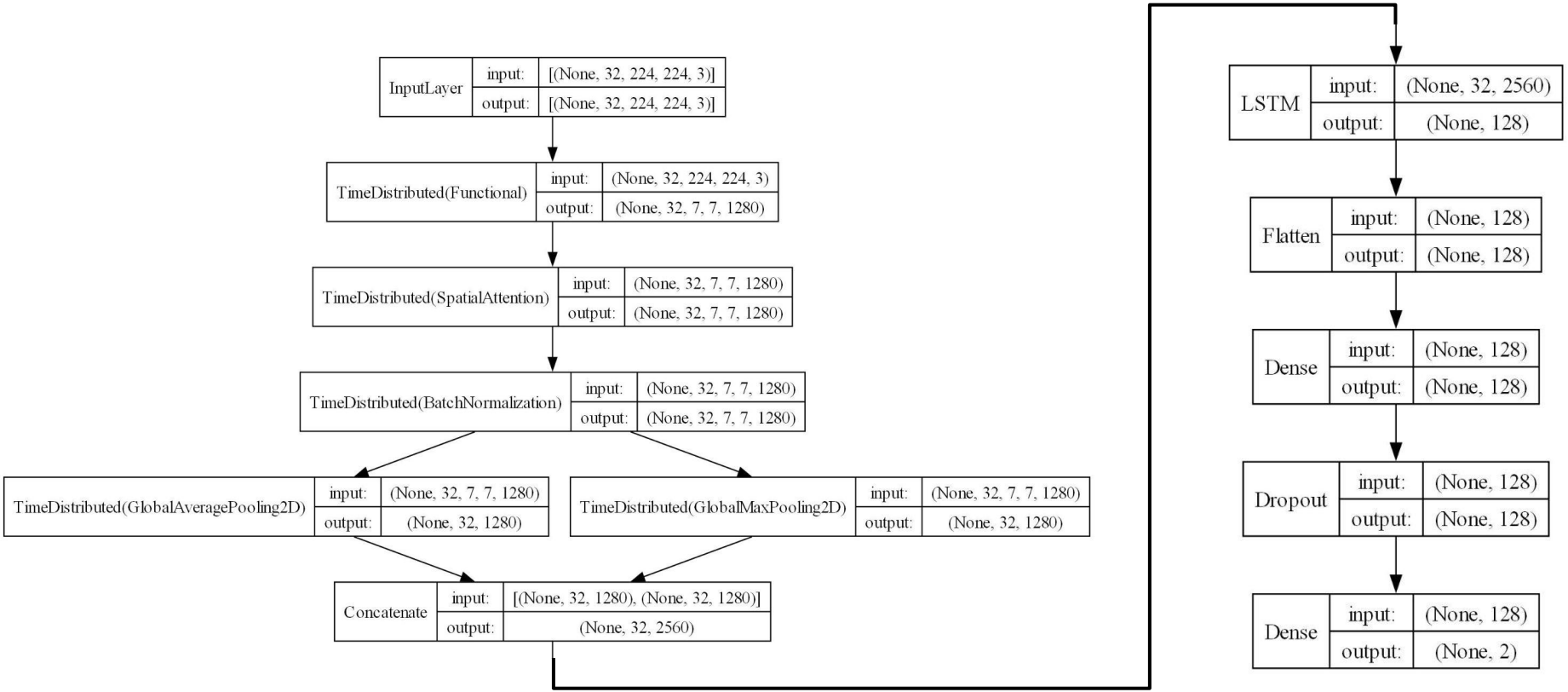
- Dataset Loading: **Data Generator** and **Memory Mapping**
- **Incremental GPU Memory Growth**
- Tackling Overfitting:
  - **Early Stopping** – monitor: 'validation loss'
  - **Reduce LR on Plateau** – monitor: 'validation loss'
  - **Dropout Layers** – 50% of total neurons
- **Model Checkpointing** – Resume from a broken training

Model Architecture



Block Diagram 4: Proposed Model Architecture

Model Architecture



Block Diagram 5: Model Plotted using Keras Model Plotting Utility

Results and Discussion: **Results**

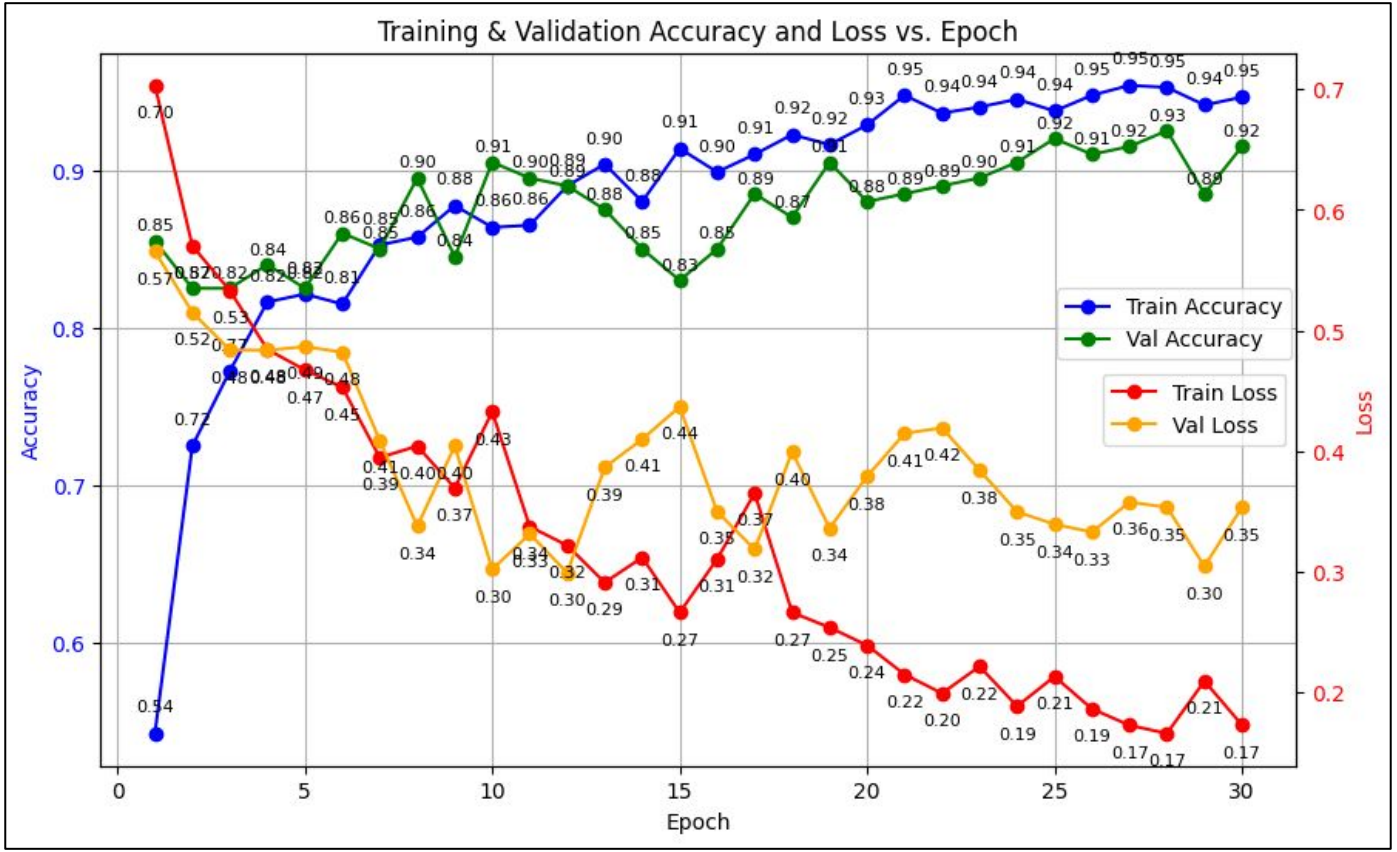


Figure 13: Accuracy and Loss vs Epoch Graph (Hockey Dataset)

Results and Discussion

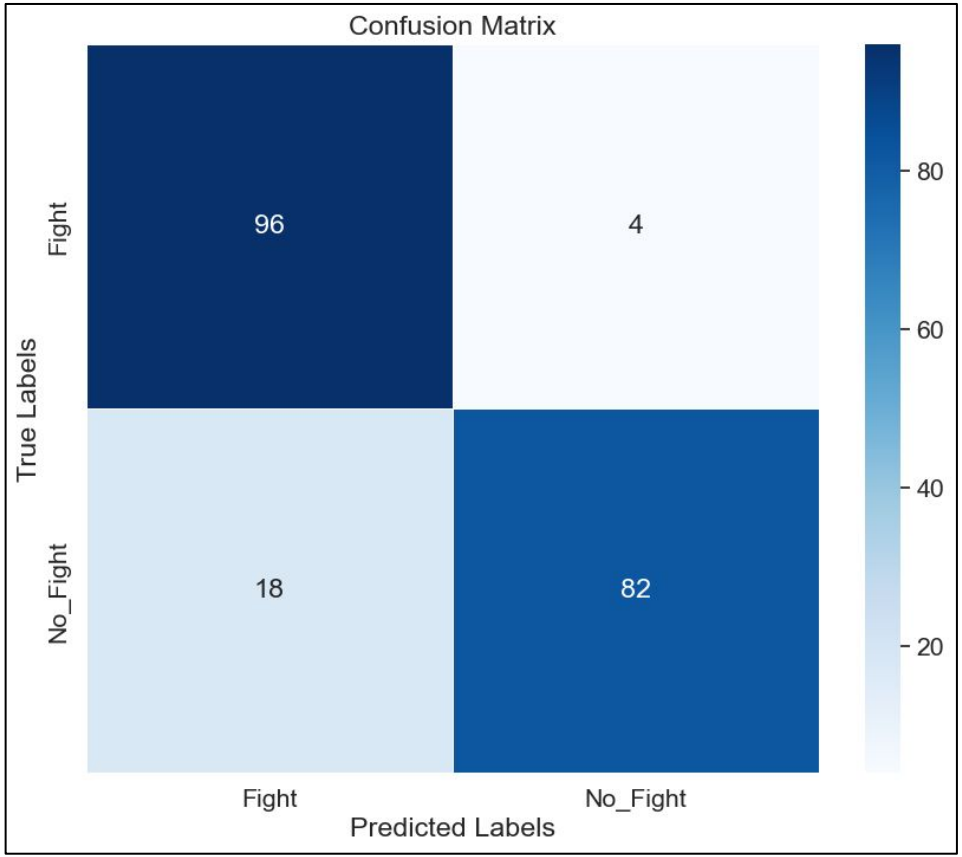


Figure 14: Confusion Matrix

Metric	Value(%)	
	Fight	**No Fight**
Precision	89.00	82
Recall	89.00	95.3
F1-Score	88.95	88.2

Table1: Other Metrics

Results and Discussion

Original Image

Attention Map

Figure 15: Correct Attention Map

Results and Discussion

Original Image

Attention Map

Figure 16: Correct Attention Map

Conclusion

- **Overfitting**: a challenge in violence detection even with larger models.
  - **minimizing overfitting** and **good accuracy.**
- **Real-world data**: **poor imaging quality** and **low lighting**.
- **Existing models**: impractical for **edge devices(ex: camera)**
  - Reason: size and complexity.
  - Proposed model prioritizes deployment on **edge devices**.

Future Scope

- **Auditory Integration:** Video + Audio = Wider applicability
- **Contextual Factors:** Video + Situation Understanding
- **Predictive Classifications:** Proactive Detection
- **Diverse Dataset Training:** Robust Model

Conference Details

- **Conference Name:** 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI 2024)
- **Publication:** Institute of Electrical and Electronics Engineers (IEEE)
- **Conference Dates:** 18th-20th November, 2024
- **Conference Website:** https://www.icdici.com
- **Date of Submission:** 2nd May, 2024
- **Notification of Acceptance:** 15th August 2024
- **Acceptance Intimation:** 12th September, 2024
- The selected paper immediately after the conference presentation will be submitted to IEEEXplore Digital Library.

Conference Details

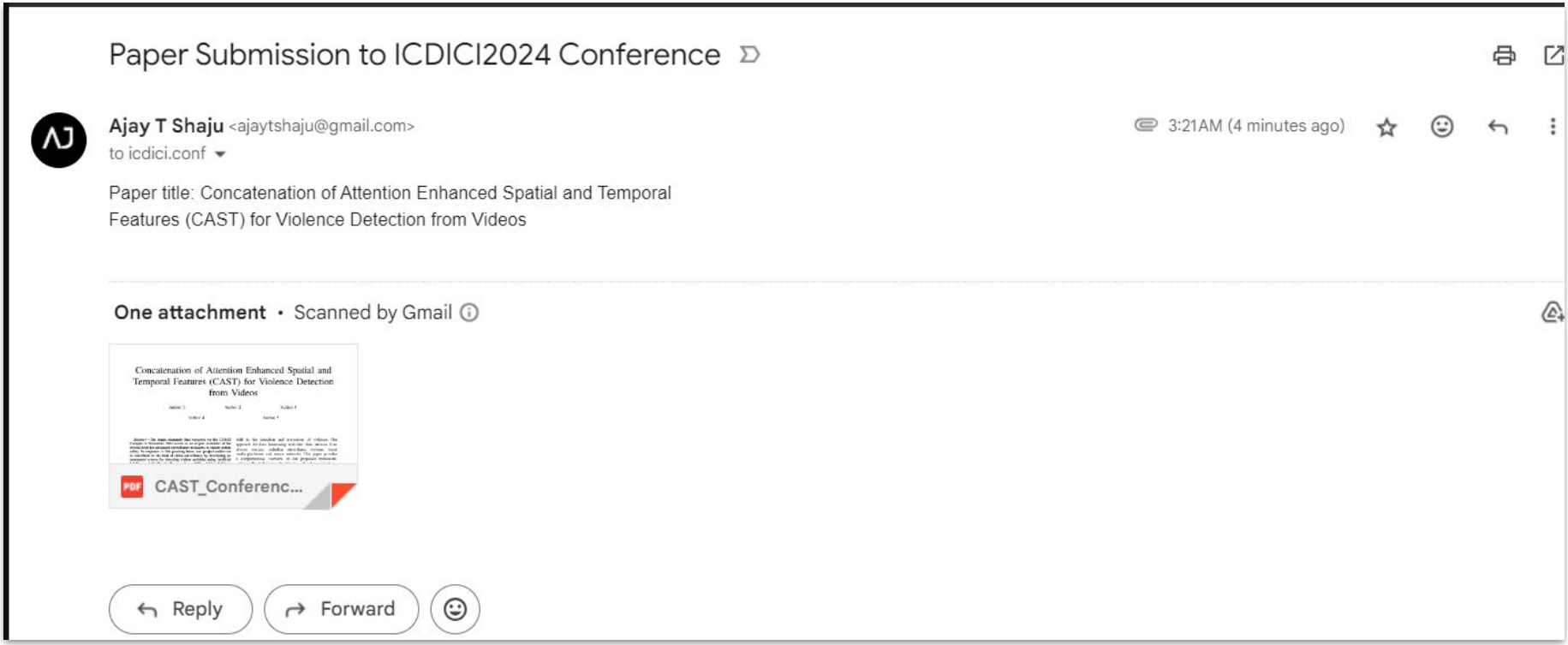


Figure 17: ICDICI 2024 Conference Submission Mail

Conference Details

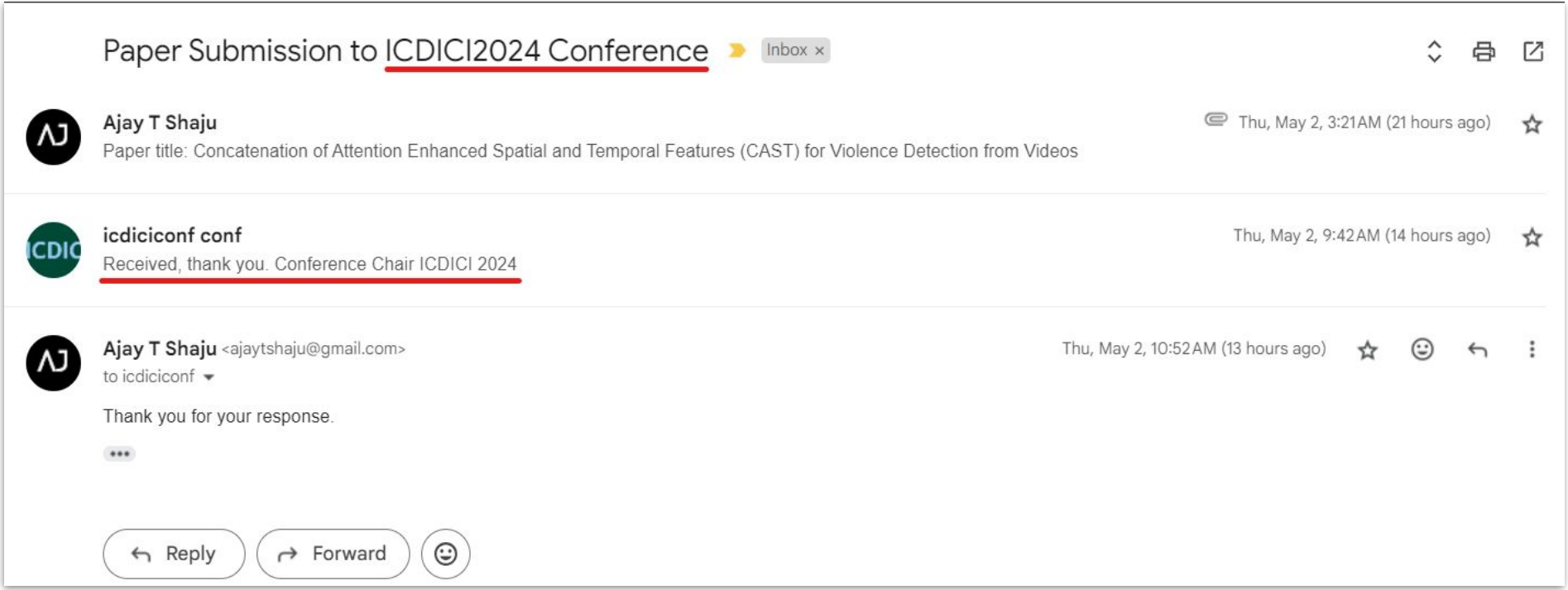


Figure 18: Acknowledgment of Receipt - ICDICI 2024 Conference Paper

Project Competition

Figure 19: Team at Carmel College Of Engineering, Alappuzha
From left: Ajay, Emil, Vishnu and Justin

References

[1] M. Safaa M. Shubber, Z. Tariq M. Al-Ta'i, **"A review on video violence detection approaches"**, International Journal of Nonlinear Analysis and Applications, Vol. 13, No. 2, pp: 1117–1130, 2022.
[2] S. Lomlen, **"Impact of Artificial Intelligence in enhancing National Security"**, Artificial Intelligence Studies, Vol. 1, pp: 1-7, Feb 2024.
[3] H. Gupta, S. T. Ali, **"Violence Detection using Deep Learning Techniques"**, International Conference on Emerging Techniques in Computational Intelligence (ICETCI), pp: 121–124, 2022.
[4] K. Aarthy, A. A. Nithya, **"Crowd Violence Detection in Videos Using Deep Learning Architecture"**, IEEE 2nd Mysore Sub Section International Conference (MysuruCon), pp: 1–6, 2022.
[5] M. Cheng, K. Cai, M. Li, **"RWF-2000: An Open Large Scale Video Database for Violence Detection"**, 25th International Conference on Pattern Recognition (ICPR), pp: 4183–4190, 2022.

[6] H. H. Nguyen, Q. Trung Le, V. Q. Nghiem, M. S. Hoang, D. A. Pham, **"A novel violence detection for drone surveillance system"**, International Conference on Communication, Circuits, and Systems (IC3S), May 2023.

[7] Y. Lyu, Y. Yang, **"Violence Detection Algorithm Based on Local Spatio–temporal Features and Optical Flow"**, International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration, pp: 307-311, 2015.

[8] A. N. Appavu, C. Nelson K. Babu, **"An Xception Model Based Real-time Violence Detection"**, IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET), April 2023.

[9] L. Sachan, P. Katiyar, Y. Kumbhawat, G. K. Rajput, T. Mehrotra, **"Comparative Analysis on Violence Detection Using Yolo and ResNet"**, 12th International Conference on System Modeling & Advancement in Research Trends (SMART), 2023.

[10] A. Chauhan, R. Gupta, **"Human Violence Detection Using LHOGF Algorithm and Deep Learning Model"**, 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2022.

[11] P. Sernani, N. Falcionelli, S. Tomassini, P. Contardo, A. F. Dragoni, **"Deep Learning for Automatic Violence Detection: Tests on the AIRTLab Dataset"**, IEEE Access, Vol. 9, pp: 160580–160595, Dec 2021.

[12] Deepak K., Vignesh L.K.P., Chandrakala S., **"Autocorrelation of gradients based violence detection in surveillance videos"**, ICT Express, Vol. 6, No. 3, pp: 155–159, July 2020.

[13] H. Su, X. Wang, T. Han, Z. Wang, Z. Zhao, P. Zhang, **"Research on a U-Net Bridge Crack Identification and Feature-Calculation Methods Based on a CBAM Attention Mechanism"**, Buildings, Vol. 12, No. 10, pp: 1-18, Sep 2022.

[14] K. Yun, J. Honorio, D. Chattopadhyay, Tamara L. Berg, D. Samaras, **"Two-person interaction detection using body-pose features and multiple instance learning"**, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp: 28–35, June 2012.

[15] T. Khalil, Javed I. Bangash, Abdul W. Khan, Saima A. Lashari, A. Khan, Dzati A. Ramli, **"Detection of Violence in Cartoon Videos Using Visual Features"**, 25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Procedia Computer Science, vol. 192, pp: 4962-4971, Oct 2021.

[16] S. Vosta, K-C. Yow, **“A CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras”**, Computing and Artificial Intelligence for Visual Data Analysis II, Applied Sciences, Vol. 12, pp: 1-15, Jan 2022.

[17] S. Manjula, K. Lakshmi, **“Human Abnormal Activity Pattern Analysis in Diverse Background Surveillance Videos Using SVM and ResNet50 Model”**, IoT and Analytics for Sensor Networks: Proceedings of ICWSNUCA 2021, pp: 47-60, Sep 2021.

[18] A. Traoré, Moulay A. Akhloufi, **“Violence Detection in Videos using Deep Recurrent and Convolutional Neural Networks”**, IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp: 154-159, Oct. 2020.

[19] Y. Su, G. Lin, J. Zhu, Q. Wu, **“Human Interaction Learning on 3D Skeleton Point Clouds for Video Violence Recognition”**, Proceedings of the 16th European Conference (ECCV) - Part IV, pp: 1-17, Aug 2020.

[20] Z. Islam, M. Rukonuzzaman, R. Ahmed, M. H. Kabir, M. Farazi, **“Efficient Two-Stream Network for Violence Detection Using Separable Convolutional LSTM”**, International Joint Conference on Neural Networks (IJCNN), pp: 1-8, July 2021.

[21] G. Garcia-Cobo, Juan C. SanMiguel, **“Human skeletons and change detection for efficient violence detection in surveillance videos”**, Computer Vision and Image Understanding, Vol. 233, pp: 1-11, Aug 2023.

[22] H. Mohammadi, E. Nazerfard, **“Video Violence Recognition and Localization Using a Semi-Supervised Hard Attention Model”**, Expert Systems with Applications, Vol. 212, pp: 1-33, Sep 2022.

[23] R. Hachiuma, F. Sato, T. Sekii, K. Minolta, **“Unified Keypoint-based Action Recognition Framework via Structured Keypoint Pooling”**, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp: 1-10, June 2023.

[24] Y. Qiao, W. Cui, T. Shi, **“LaM-2SRN: A Method Which Can Enhance Local Features and Detect Moving Objects for Action Recognition”**, IEEE Access, Vol. 8, pp: 192703-192712, Oct 2020.

Questions?

Thank You