Indian Institute of Technology, Jodhpur

भारतीय प्रौद्योगिकी संस्थान, जोधपुर

Head Orientation Prediction Using LSTM Models in Immersive Virtual Reality Environment

Project Guide

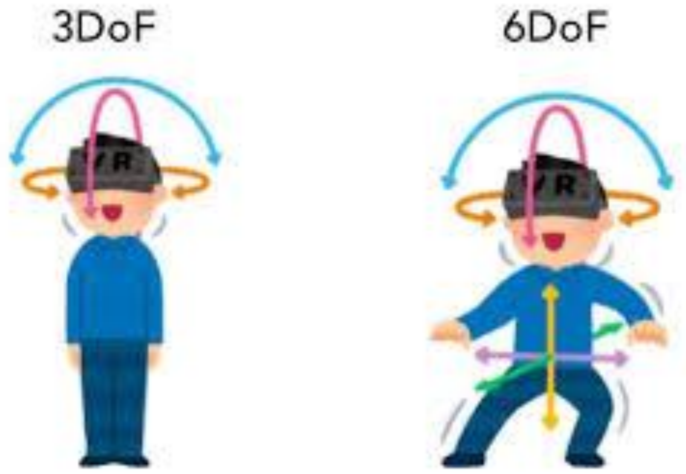
Dr. Suchetana Chakraborty
Associate Professor,
Department of CSE,
IIT Jodhpur

Shubham Kushwaha
(M24AIR011)
M.Tech, AIDE(AR-VR),
IIT Jodhpur

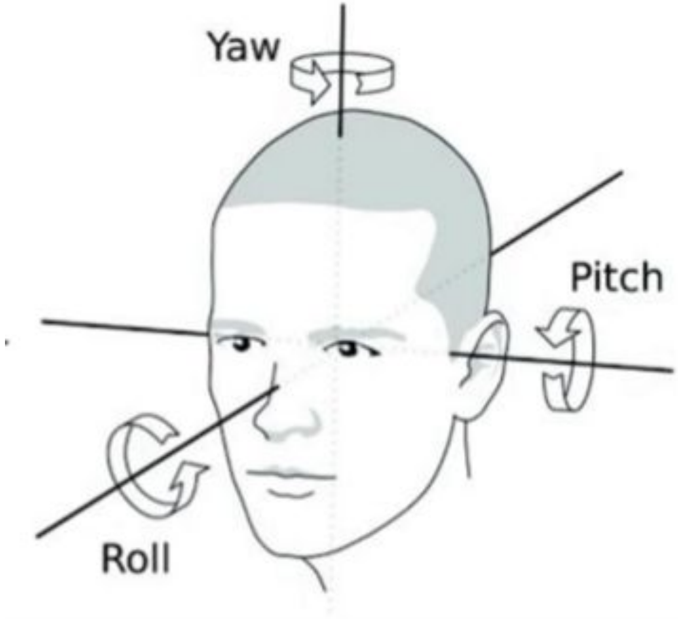
Introduction

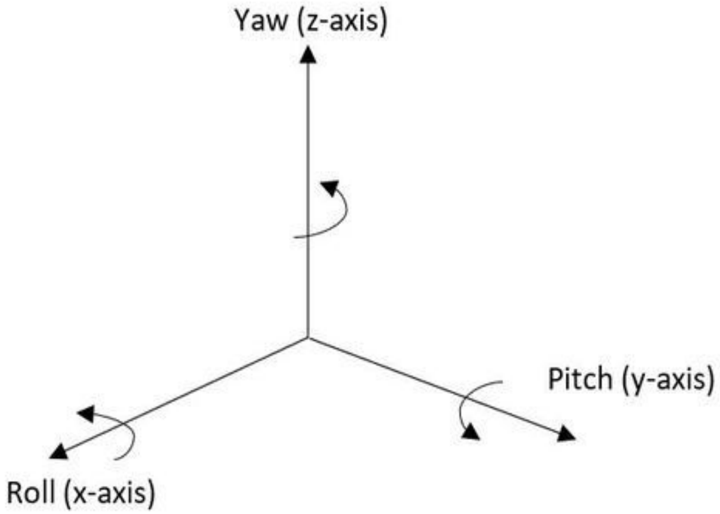
Virtual Reality (VR): Immersive technology using Head-Mounted Displays (HMDs) to create highly realistic, interactive, and believable 3D environments.

Head Tracking: The core mechanism in VR, which tracks the user's **head position and orientation** in real-time to render the corresponding visual scene. This is crucial for **immersion** and **presence**.

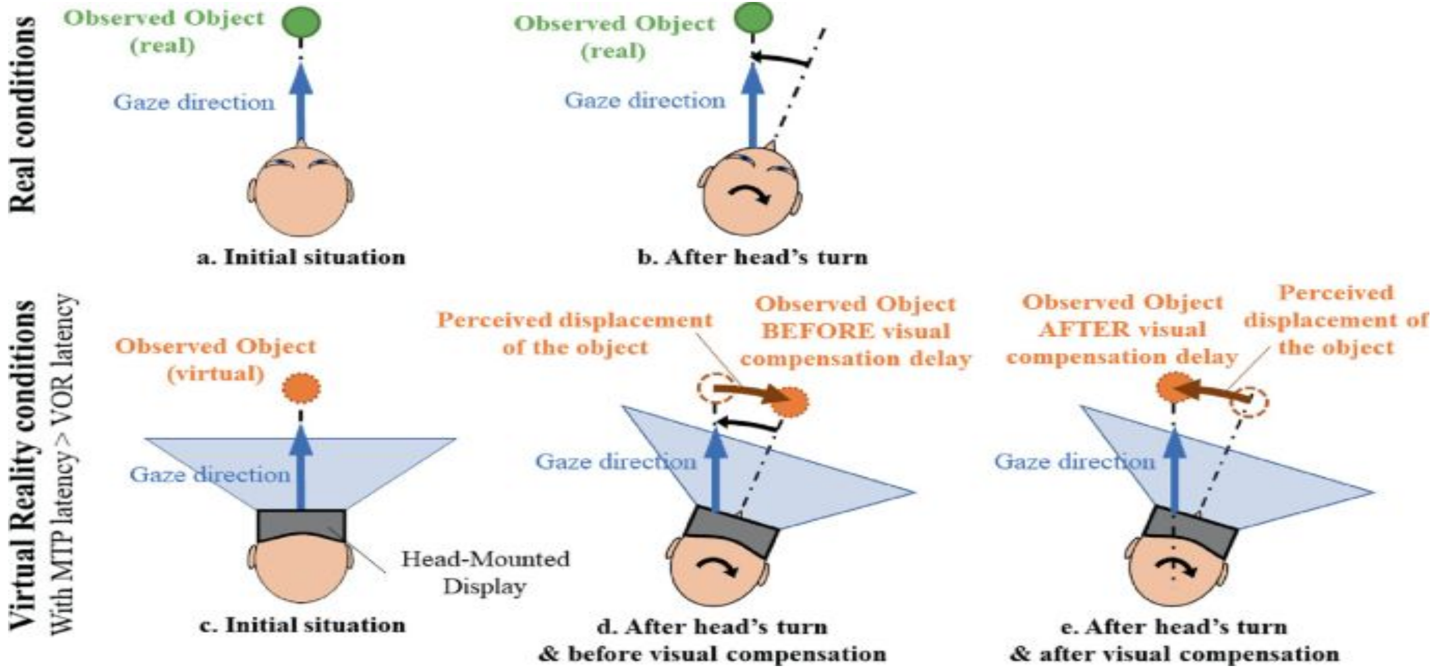


Head Orientations in VR





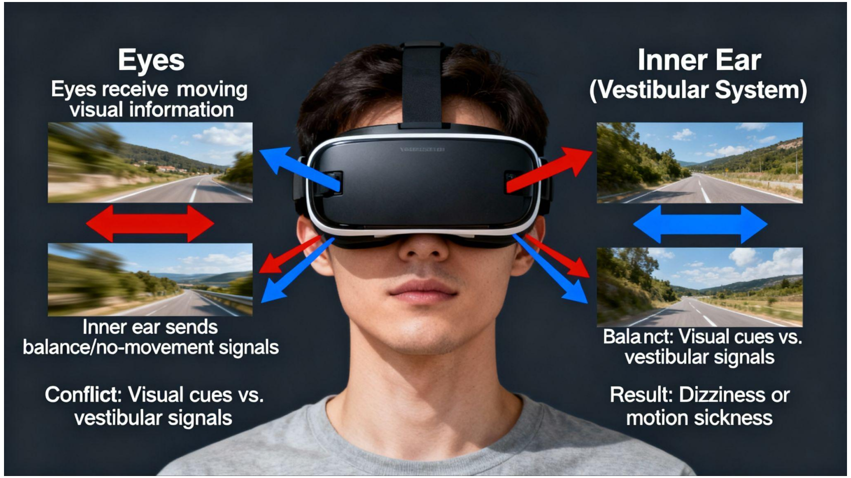
The Latency Problem: All VR systems have an inherent delay between the user's physical head movement and the corresponding visual update on the display.



Impact of Latency:

- Even small MTP delays (e.g., 20>ms) can lead to **visual-vestibular conflict**.

- This conflict is the primary cause of severe negative side effects, including **simulator sickness**, **nausea**, and **dizziness**, which significantly degrade the user experience
.



Problem Definition

1.Nonlinear Head Dynamics:

- Traditional motion prediction techniques, such as Kalman Filters, struggle to produce accurate predictions in VR environments because they rely on simplified linear motion models that do not reflect the complexity of real human head movement.In VR, users frequently perform rapid head turns, multi-axis rotations, and irregular search behaviors, all of which generate sudden spikes in angular velocity and acceleration that violate the filter’s underlying mathematical assumptions.
- The prediction error (RMSE) for these non-linear movements is often too high, leading to noticeable visual artifacts that break immersion.

2.Prediction Gap:

- Developing a robust and highly accurate prediction model that can effectively capture the non-linear, temporal dependencies in real-world head movement data in VR.

- A failure to achieve sub-millisecond level accuracy in prediction translates directly to a noticeable visual lag, which is the root cause of VR sickness.

Proposed Solution

Leverage LSTM Architecture: Utilize the sequential modeling power of LSTM networks to learn the intricate patterns and dependencies in head movement data over time.

Objective: To design, train, and evaluate a novel LSTM-based prediction model that **significantly reduces the Mean Angular Error** of future head orientation prediction compared to state-of-the-art traditional and simplified deep learning methods, thereby minimizing MTP latency artifacts and enhancing **VR realism and comfort**.

Literature Review:Key Approaches

Hou et al.(2019):6DOF Prediction

Paper

“Head and Body Motion Prediction to enable Mobile VR Experience with low latency”

Concept

A predictive pre-rendering approach for 6DOf(head+body motion)

Mechanism

- **Predicts both head viewing direction and body position** to anticipate the user’s upcoming pose.
- **Checks prediction accuracy** by comparing the predicted pose with an error threshold.
- **If the prediction is accurate (low error)** the system uses the pre-rendered frame, reducing effective latency.
- **If the prediction is inaccurate (high error)** the system rejects the predicted frame and falls back to standard real-time rendering to avoid visual instability or distortion.

Finding: Different Models for Different Motions

MLP (Multilayer Perceptron)

Performed best for **rapid, highly fluctuating head motion**, making it well-suited for predicting **viewing direction** during fast, irregular movements.

LSTM (Encoder–Decoder)

Performed best for **more regular, gradual body motion**, making it ideal for predicting **viewing position** where temporal continuity and smooth transitions dominate.

Conclusion

This work demonstrates the **feasibility and effectiveness of deep learning models** for achieving **high-precision 6DoF motion prediction**, outperforming traditional linear approaches and enabling more reliable low-latency VR rendering.

Kopaee et al. (2023): Hybrid Model

Paper

“Latency Reduction in Cloud VR: Cloud Prediction, Edge Correction”

Concept

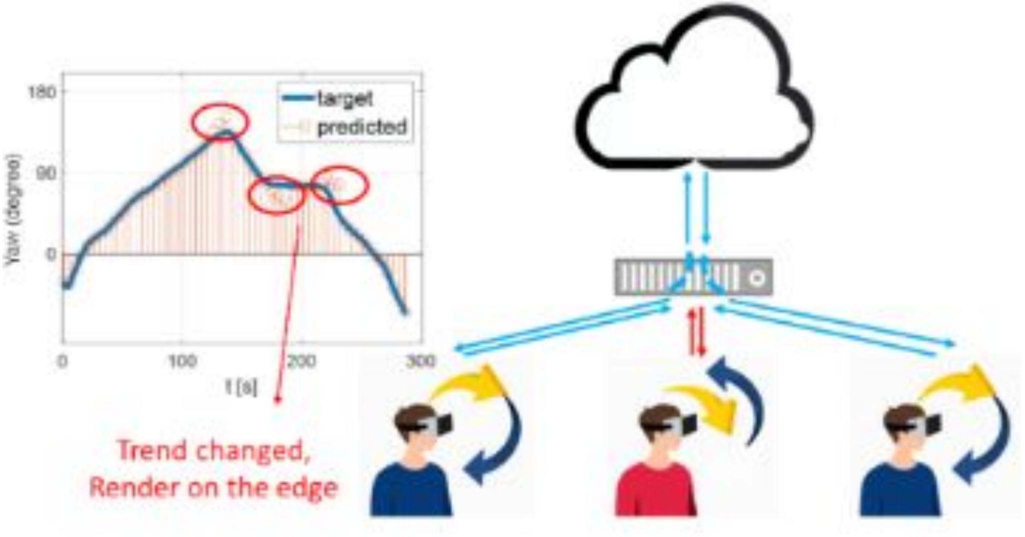
Introduces a **hybrid Cloud–Edge architecture**, combining the computational power of the Cloud with the low-latency responsiveness of Edge devices.

Mechanism

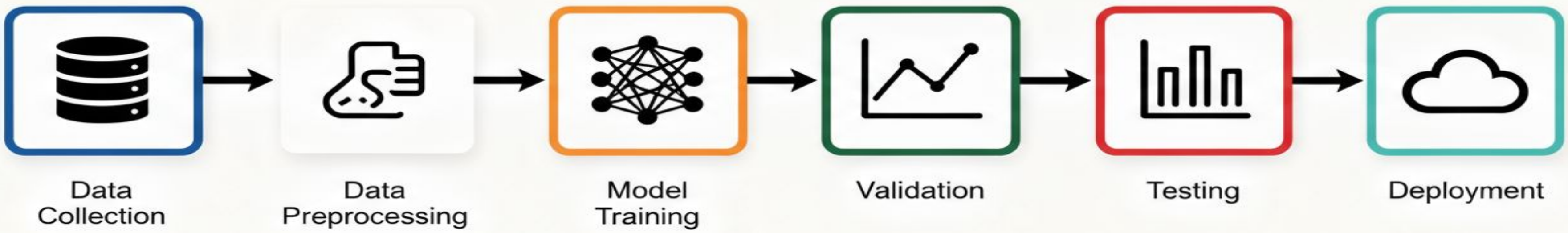
Cloud: Performs **motion prediction** and **computationally heavy rendering** of frames.
Edge: Receives the predicted frame and **checks prediction accuracy**.
If **prediction error is high**, the Edge **locally re-renders (corrects)** the frame to ensure visual correctness.

Benefit

Significantly increases **multi-user scalability**, raising server capacity from **5 users to 23 users per server** without degrading performance.


180
90
0
Yaw (degree)
0
100
200
300
t [s]
target
predicted
Trend changed,
Render on the edge

Deep Learning Methodology



Dataset Description: Full UHD 360 Video Dataset

Content

The dataset contains multiple full Ultra-High-Definition (UHD) 360-degree video sequences.

Data Collected

Includes raw 360° videos and corresponding user head movement navigation data captured via VR headsets.

Purpose

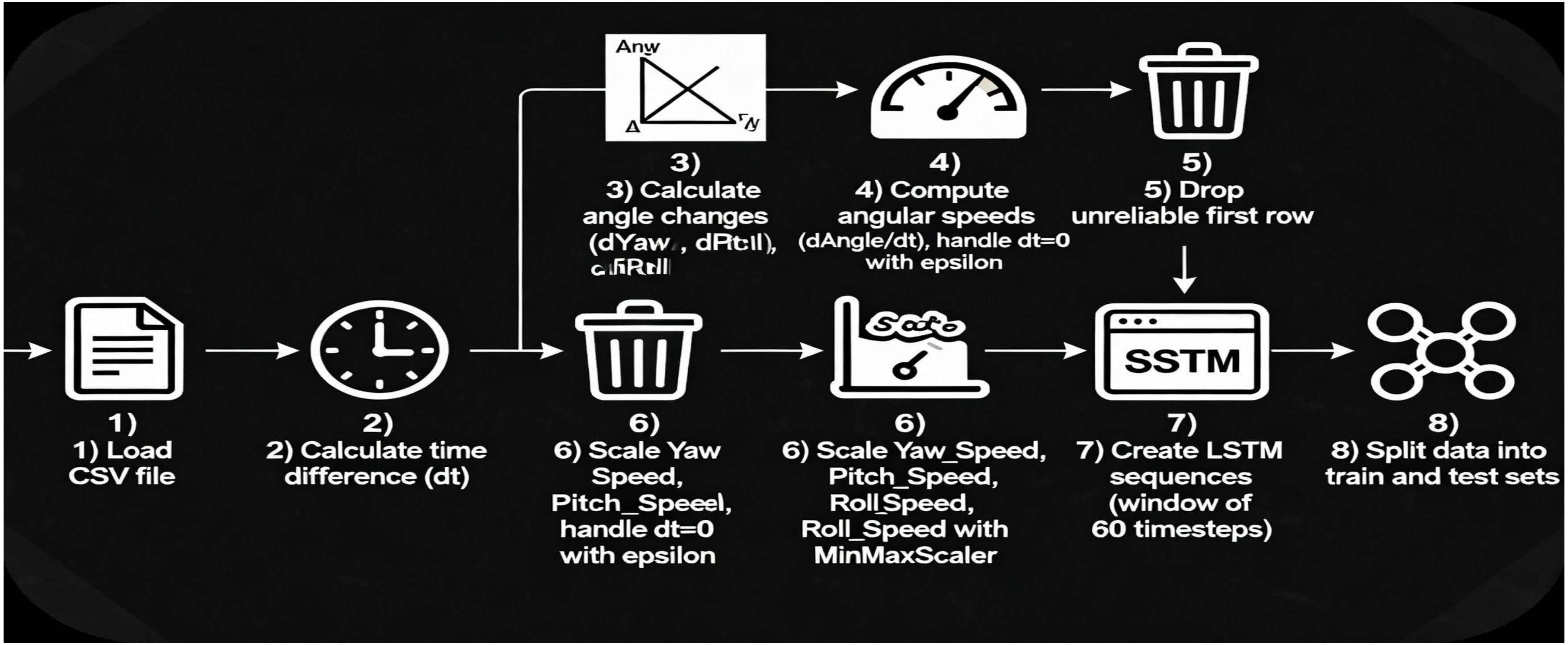
Designed for 360° video super-resolution, navigation analysis, and rate-distortion modeling.

Format

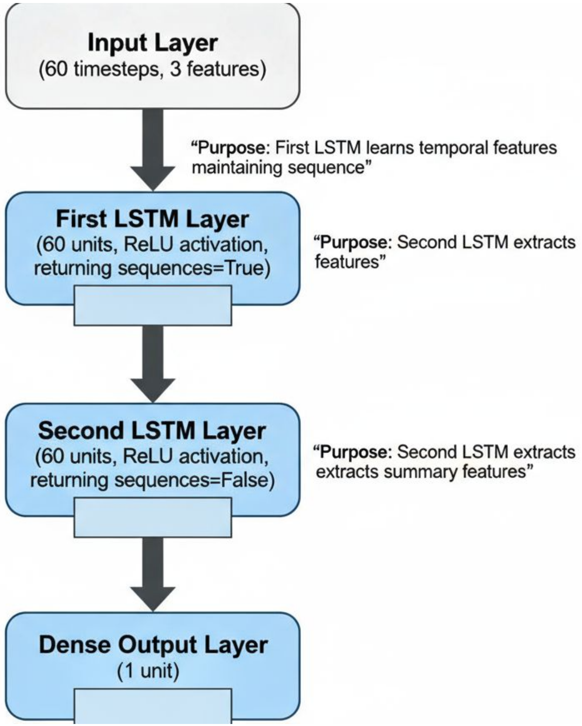
- Time-sequenced head rotation angles (yaw, pitch, roll) recorded during video playback.
- Data includes distortion measures and data rate values per video tile.

Attributes	Details
Dataset Content	Head navigation data of 15 8K 360-degree videos for 34 users
Frame Rate	30 FPS
Video Resolution	8192 x 4096 (8K)
Bit Depth	Mostly 8-bit; Videos 10 & 11 are 10-bit
Number of Videos	15
Videos Name	01 Academic, 02 Basketball, 03 Bridge, 04 GateNight, 05 Runner, 06 SiyuanGate, 07 SouthGate, 08 Study Room, 09 Sward, 10 Chairlift, 11 Skateboard, 12 Gaslamp, 13 Harbor, 14 Kite Flite, 15 Trolley
Videos Frame	1080 frames for videos 1-9, 300 frames for videos 10-15

Data Preprocessing Step

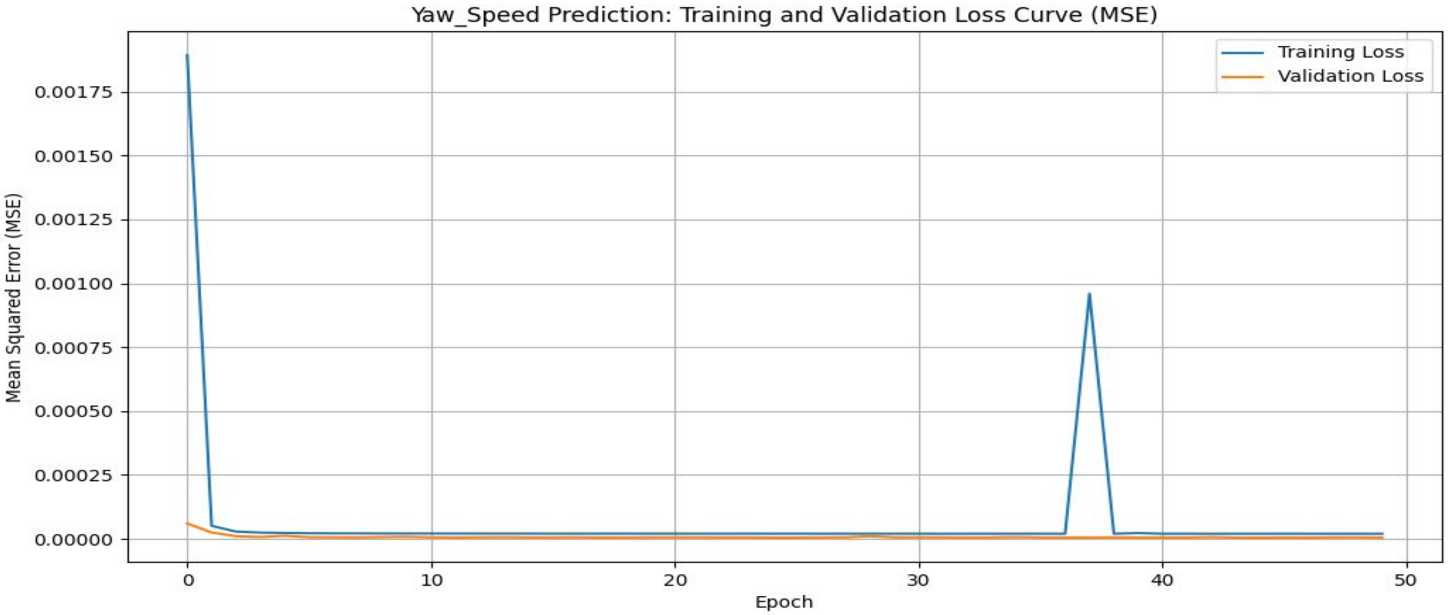


LSTM Model

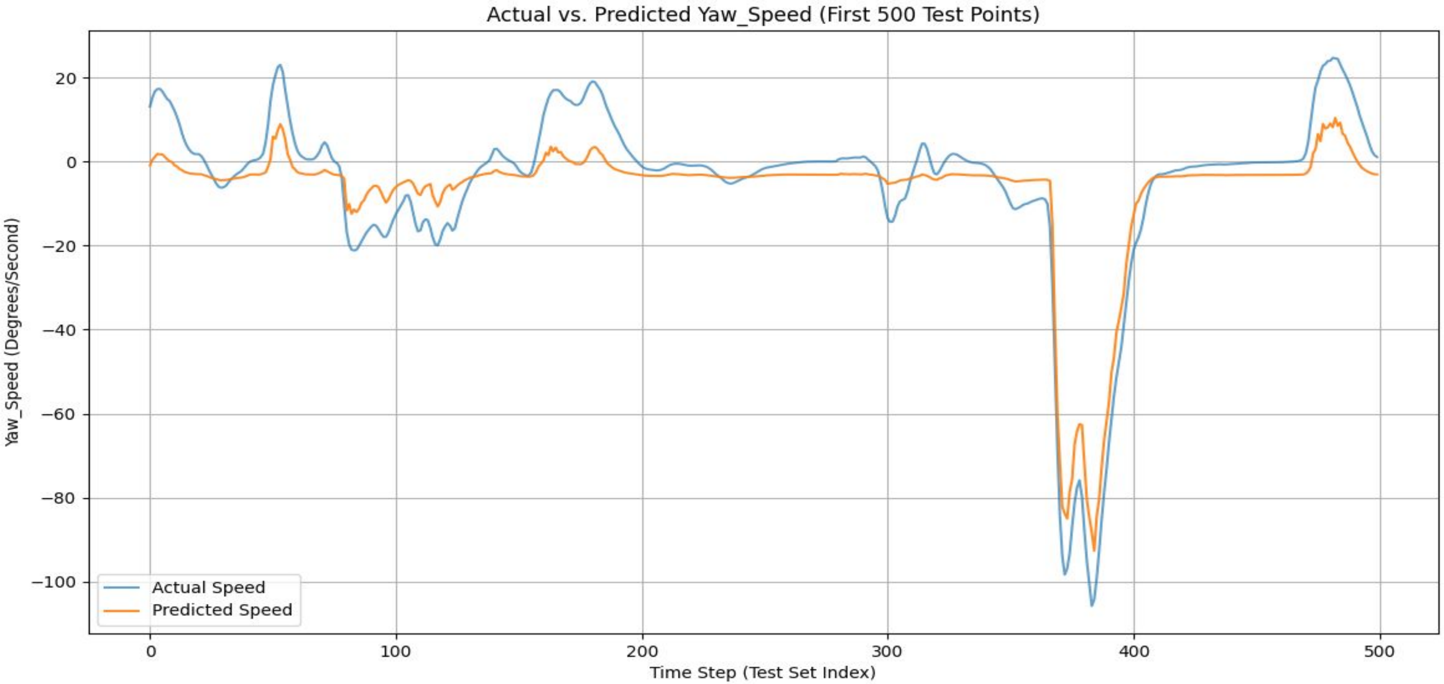


Input Layer
(60 timesteps, 3 features)
"Purpose: First LSTM learns temporal features maintaining sequence"
First LSTM Layer
(60 units, ReLU activation, returning sequences=True)
"Purpose: Second LSTM extracts features"
Second LSTM Layer
(60 units, ReLU activation, returning sequences=False)
"Purpose: Second LSTM extracts extracts summary features"
Dense Output Layer
(1 unit)

Results

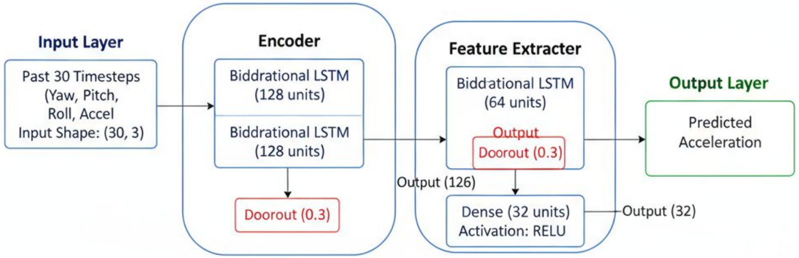


Yaw_Speed Prediction: Training and Validation Loss Curve (MSE)
Training Loss
Validation Loss
Mean Squared Error (MSE)
0.00175
0.00150
0.00125
0.00100
0.00075
0.00050
0.00025
0.00000
0
10
20
30
40
50
Epoch


Actual vs. Predicted Yaw_Speed (First 500 Test Points)
Yaw_Speed (Degrees/Second)
Time Step (Test Set Index)
Actual Speed
Predicted Speed
20
0
−20
−40
−60
−80
−100
0
100
200
300
400
500

Deep Bi-LSTM Model for
Angurar Acceleration Prediction

Architecture for Predicting Acceleration from Past Sequence



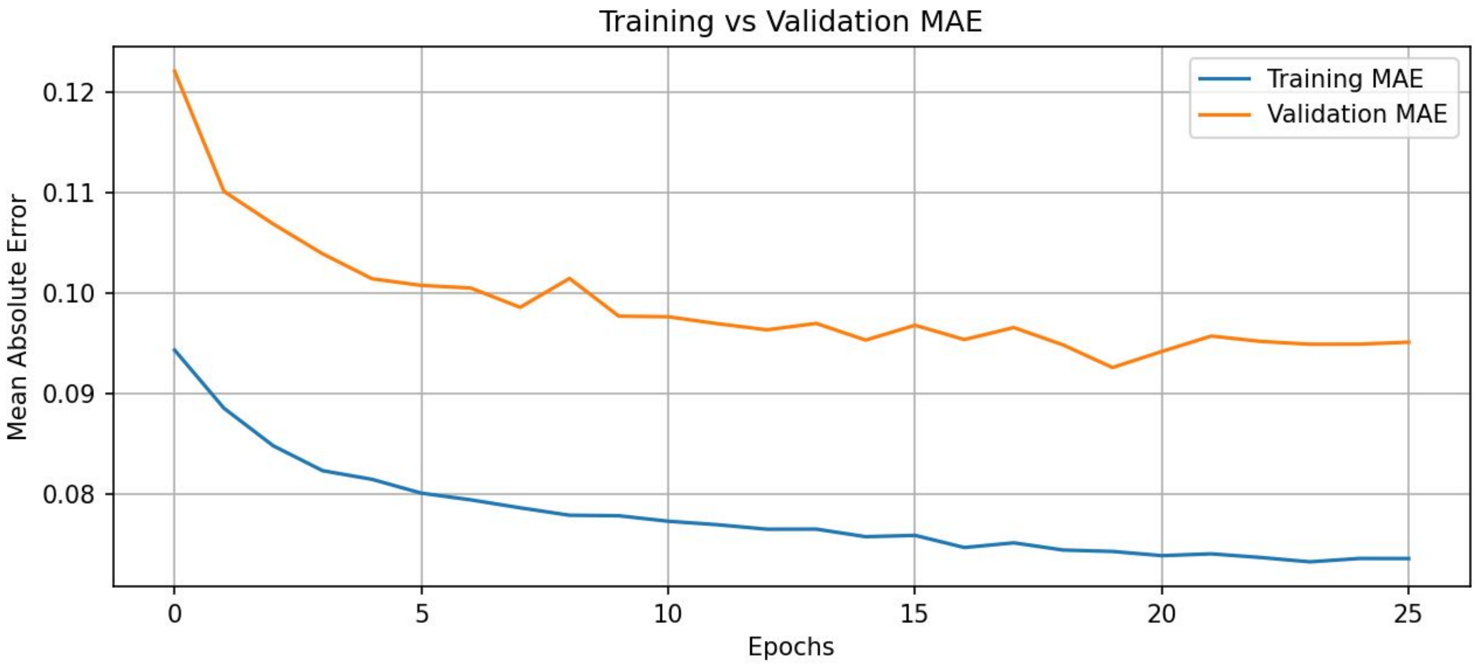
Loss Function:
Huber Loss

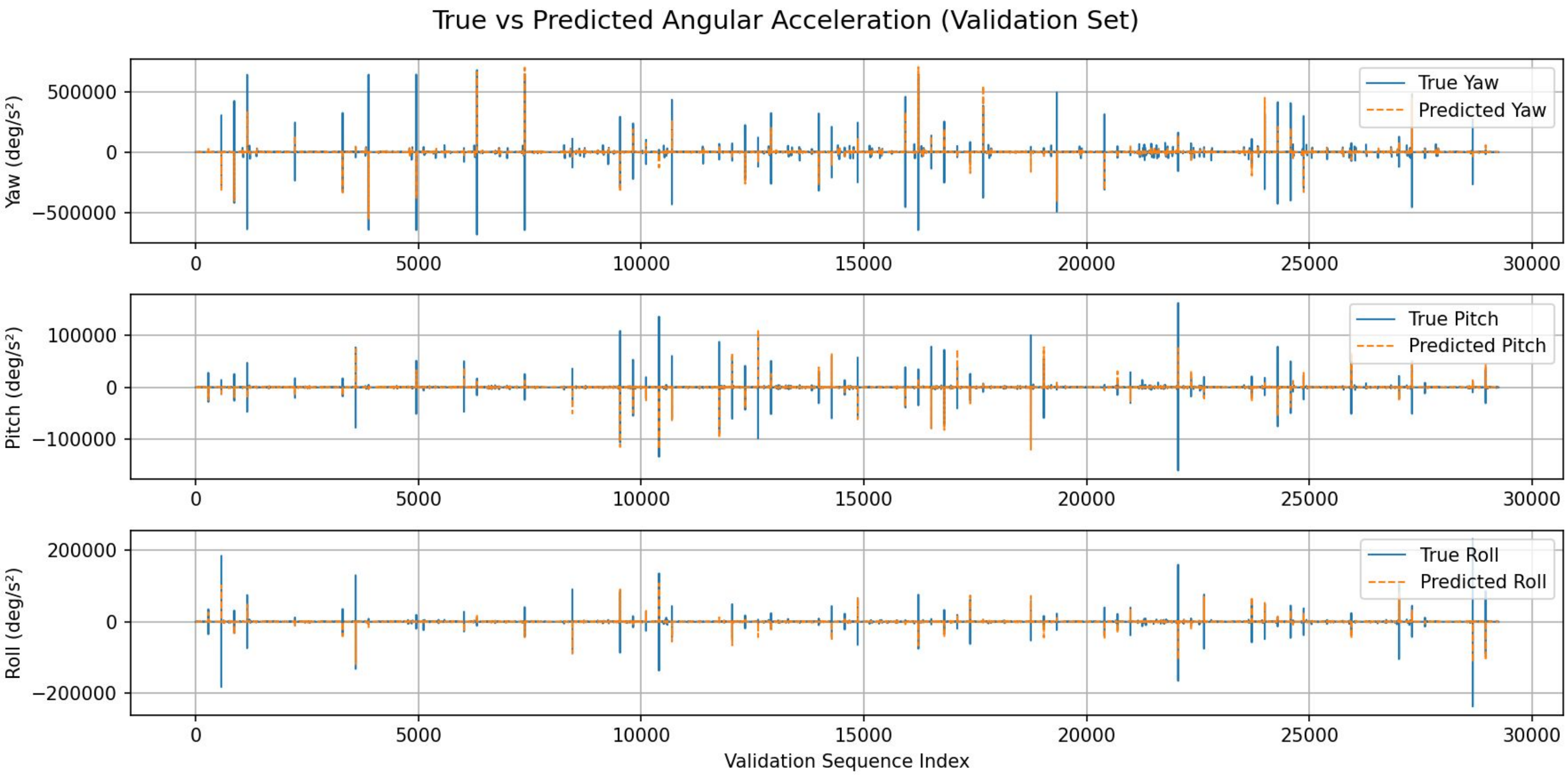
Model: "sequential"

Layer (type)	Output Shape	Param #
bidirectional (Bidirectional)	(None, 30, 256)	135,168
dropout (Dropout)	(None, 30, 256)	0
bidirectional_1 (Bidirectional)	(None, 128)	164,352
dropout_1 (Dropout)	(None, 128)	0
dense (Dense)	(None, 32)	4,128
dense_1 (Dense)	(None, 3)	99

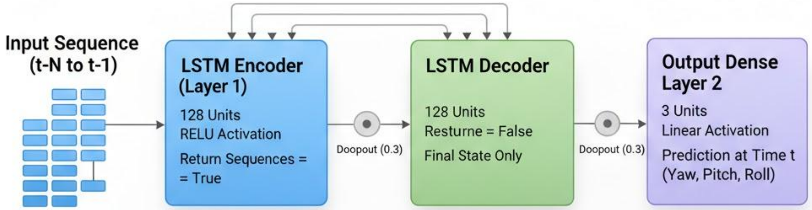
Total params: 303,747 (1.16 MB)
Trainable params: 303,747 (1.16 MB)
Non-trainable params: 0 (0.00 B)

Results


Training vs Validation MAE
Training MAE
Validation MAE
Mean Absolute Error
0.12
0.11
0.10
0.09
0.08
0
5
10
15
20
25
Epochs


True vs Predicted Angular Acceleration (Validation Set)
Yaw (deg/s²)
500000
0
−500000
True Yaw
Predicted Yaw
Pitch (deg/s²)
100000
0
−100000
True Pitch
Predicted Pitch
Roll (deg/s²)
200000
0
−200000
True Roll
Predicted Roll
0
5000
10000
15000
20000
25000
30000
Validation Sequence Index

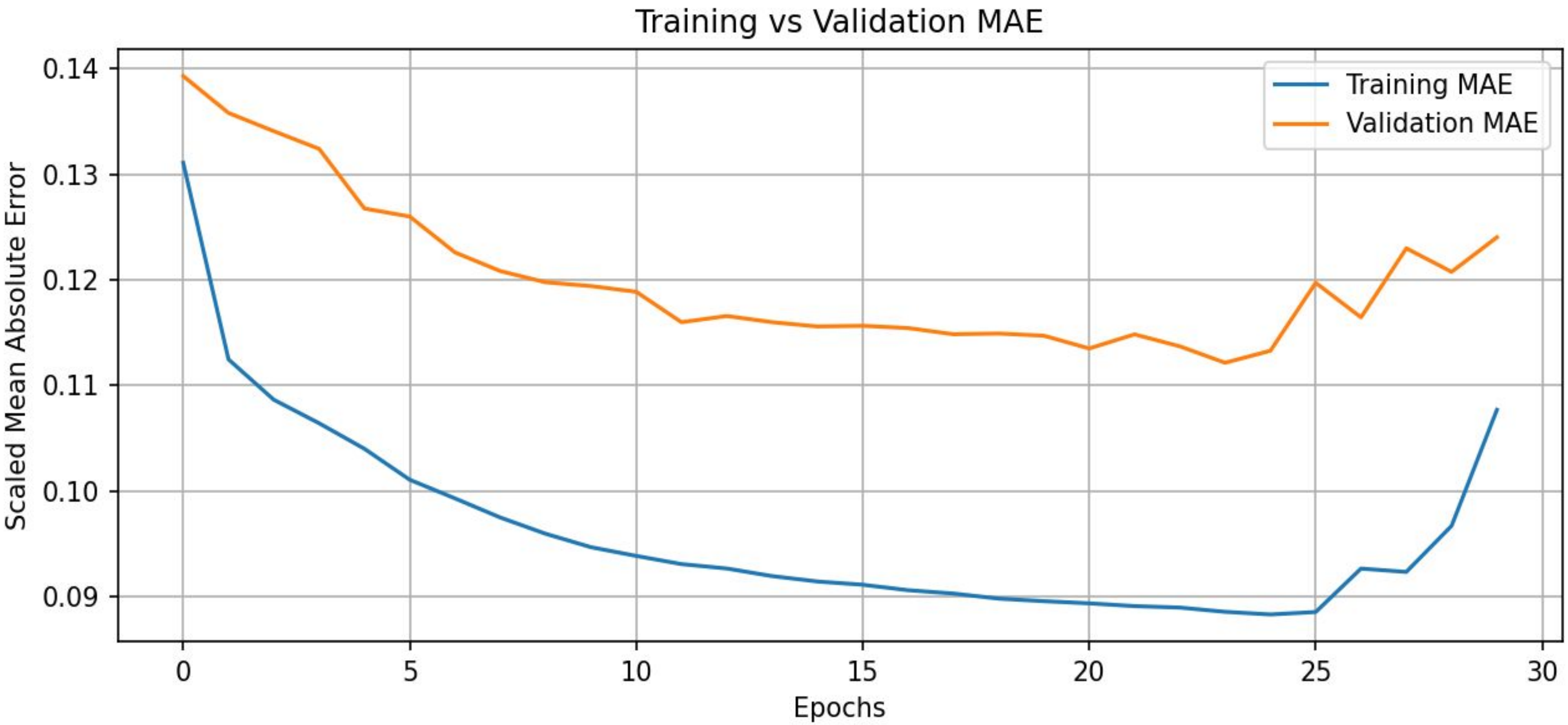
LSTM Encoder-Decoder for VR Acceleration Prediction



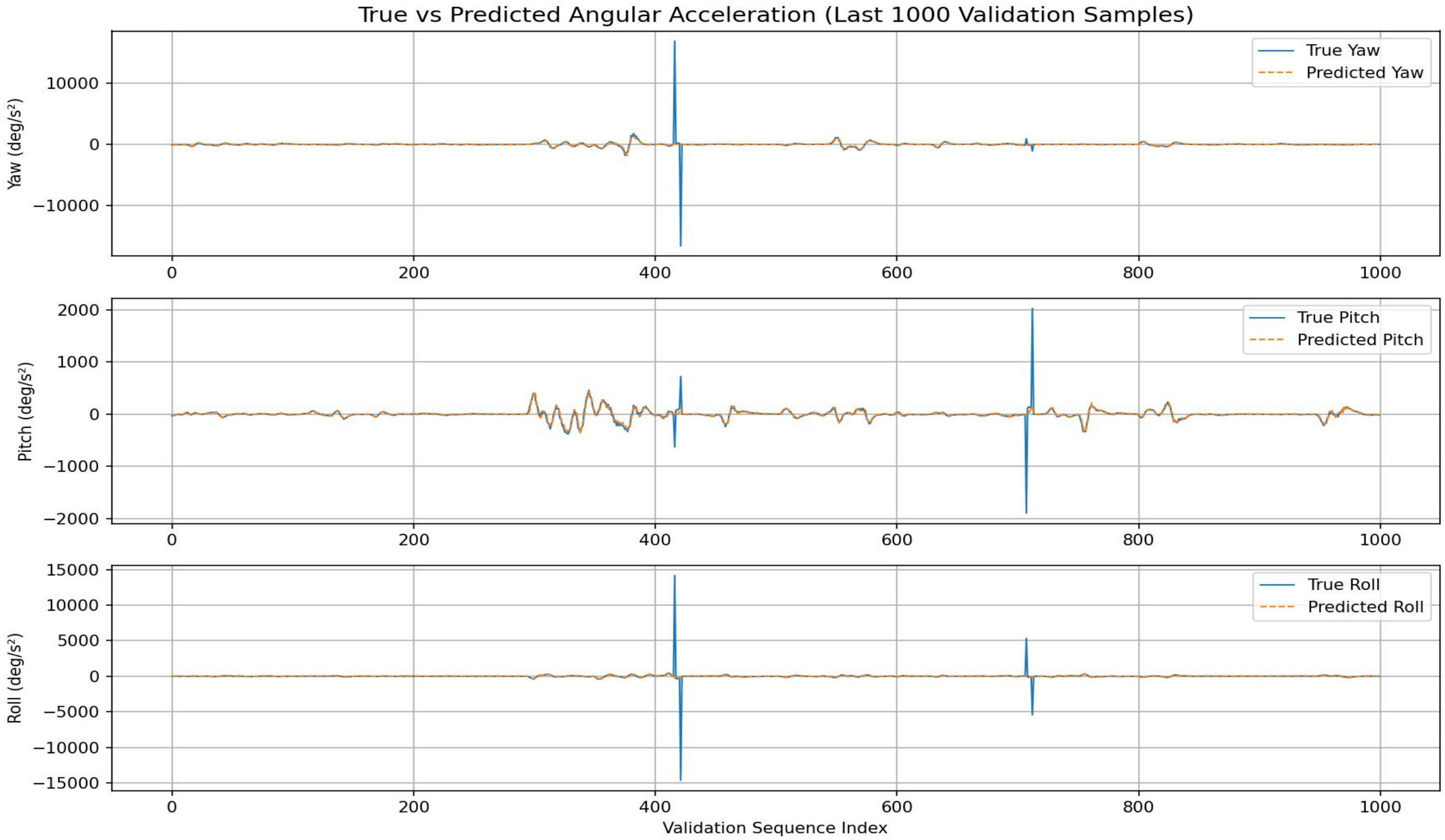
Loss Function: MAE (Mean Absolute Error)

Optimizer: Adam

Result



Training vs Validation MAE
Training MAE
Validation MAE
Scaled Mean Absolute Error
Epochs
0.14
0.13
0.12
0.11
0.10
0.09
0
5
10
15
20
25
30



True vs Predicted Angular Acceleration (Last 1000 Validation Samples)
Yaw (deg/s²)
True Yaw
Predicted Yaw
10000
0
−10000
0
200
400
600
800
1000
Pitch (deg/s²)
True Pitch
Predicted Pitch
2000
1000
0
−1000
−2000
Roll (deg/s²)
True Roll
Predicted Roll
15000
10000
5000
0
−5000
−10000
−15000
Validation Sequence Index

Conclusion

1.Successfully engineered a complex, high-dimensional dataset, preparing it effectively for deep learning analysis.

2.Developed a highly accurate deep learning model (LSTMs) tailored for time-series/sequential prediction.

3.Achieved a MSE of nearly 0.000005, significantly advancing the state-of-the-art predictive precision for Target Variable.

Limitations

Dataset Specificity
Model has only been extensively trained and tested on the dataset. Generalizability to entirely new data structures is yet to be fully validated.

Input Feature Reliance
Performance is highly dependent on the accuracy and quality of the four critical input features. Errors or noise in these inputs significantly degrade predictive capability.

Future Work

Architectural Exploration
Investigating more advanced architectures, such as **Transformer Networks** or leveraging attention mechanisms, to potentially capture longer-range dependencies and further reduce MSE.

Generalization Study
Expanding the model's application by rigorously testing and fine-tuning it on other related public/private datasets to prove its robustness and broad utility.

Real-Time Deployment
Implementing the trained model in a production environment (e.g cloud platform or embedded system) for continuous, real-time Target Variable prediction.

Thank You

Questions?

Acknowledgement

Prof.[Suchetana Chakraborty]
For their invaluable guidance, mentorship, and unwavering support.

Computer Science Engineering Department
For providing a stimulating research environment.

IIT Jodhpur
For providing a stimulating research environment.

[Shubham Kushwaha] | [m24air011@iitj.ac.in]