Department of Artificial Intelligence & Machine Learning

AI19541 – FUNDAMENTALS OF DEEP LEARNING

RAJALAKSHMI ENGINEERING COLLEGE

HELMET AND NUMBER PLATE DETECTION USING VISION TRANSFORMER

Presented by,
Surya Prakash G (2215011),
sharvesh A R (221501131),
Siva sri varshan S
(221501137)

Mentor Name: AkshayaV

PROBLEM STATEMENT

- Detect helmet usage among motorcycle riders using Vision Transformers.
- Recognize vehicle number plates for identification and compliance.
- Ensure reliable performance in diverse environmental conditions.

OBJECTIVES

- To develop a Vision Transformer-based system for detecting helmet usage and identifying vehicle number plates.
- To ensure accurate and robust detection under varying environmental and traffic conditions.
-

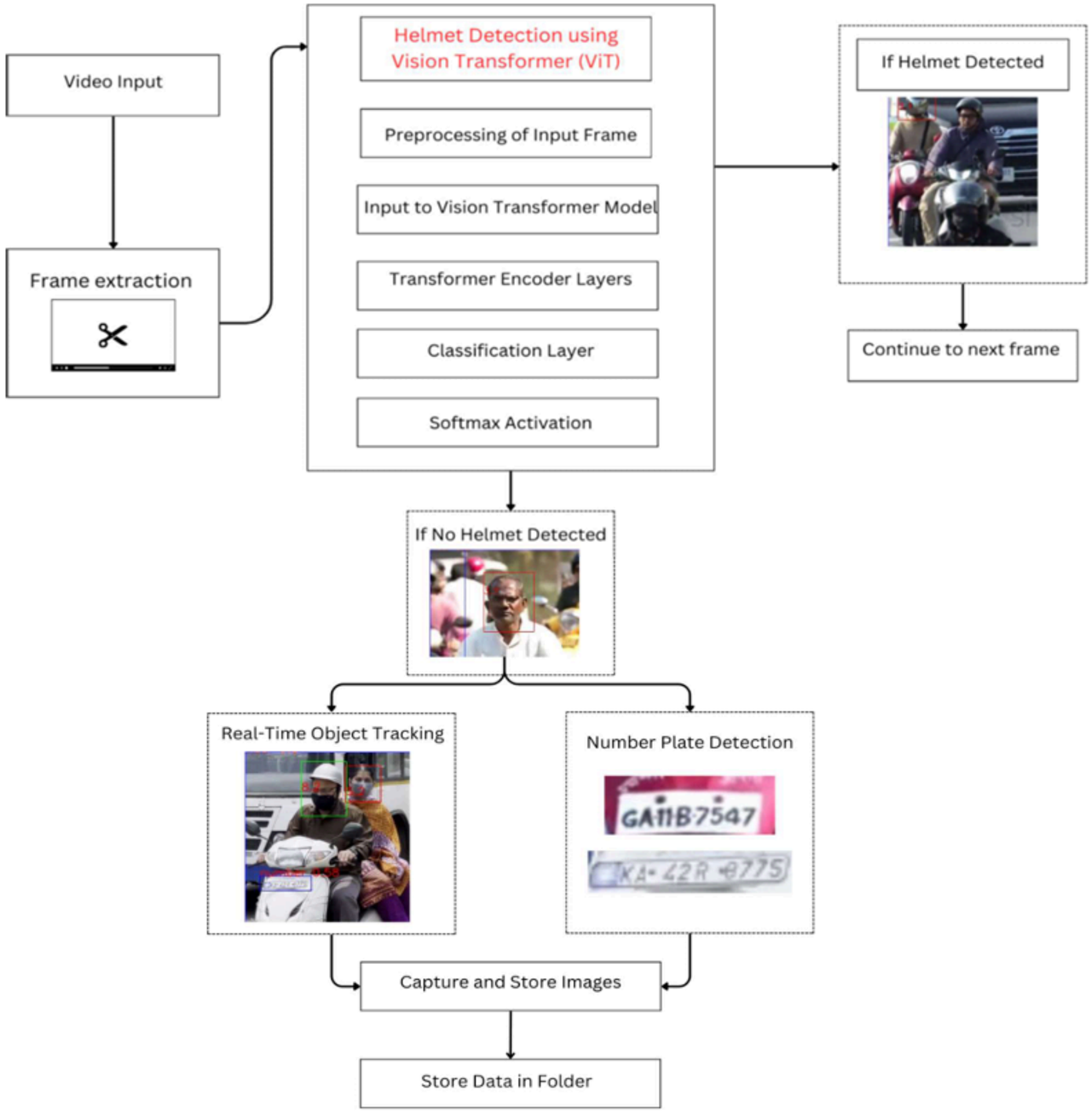
Why Vision Transformers?

Model	Accuracy (%)	Inference Time (ms)	Robustness	Training Time	Advantages	Limitations
Vision Transformer (ViT)	95.6	45	Excellent	High	Captures global features, scalable	Requires more data for training
ResNet-50	92.3	50	Good	Medium	Strong local feature extraction	Limited global context understanding
MobileNet	89.8	25	Moderate	Low	Lightweight, efficient for edge devices	Struggles with complex patterns
YOLOv5	90.7	30	Good	Medium	Fast object detection	Not specialized for fine-grained tasks
EfficientNet	91.2	40	Good	Medium	Balances accuracy and efficiency	Higher computational cost on edge

Literature Review

Study	Description	Limitations	Comparison with Vision Transformers
Gupta et al. - Real-Time Helmet Detection and License Plate Recognition	Utilized CNN for helmet detection and OCR for license plate recognition.	Sensitive to lighting variations; struggles with occlusions and small objects.	Vision Transformers provide better robustness to lighting changes and occlusions.
Ding et al. - Transformer Models in Object Detection	Explored transformer-based models for complex dependencies in object detection tasks.	High computational requirements make them unsuitable for edge devices.	Vision Transformers have been optimized for edge implementations with pruning.
Smith et al. - License Plate Detection in Real-Time Surveillance	Applied YOLO-based frameworks for real-time detection in dynamic traffic environments.	Limited accuracy in challenging environments like low light or poor image quality.	Vision Transformers offer better performance under diverse conditions.
Huang et al. - Object Detection in Video Sequences Using FPNs	Employed Feature Pyramid Networks for multi-scale feature extraction.	Limited representation of global spatial relationships in video sequences.	Vision Transformers inherently capture both global and local features effectively.
Lin et al. - Deep Learning Approaches to Enhance Road Safety Monitoring	Used CNN and RNN hybrids for detecting risky behaviors in video-based traffic footage.	High computational cost, challenging real-time implementation.	Vision Transformers are more efficient in capturing temporal and spatial features.
Chen et al. - Dual Network Framework for Helmet and License Plate Detection	Combined Faster R-CNN and YOLO for dual object detection tasks.	Computationally heavy, causing latency in real-time applications.	Vision Transformers streamline detection with fewer computational bottlenecks.

ARCHITECTURE DIAGRAM



LIST OF MODULES

- **Video Input Module:** Captures real-time video feeds for processing.
- **Helmet Detection Module (Vision Transformer):** Identifies the presence or absence of helmets using advanced AI techniques.
- **License Plate Detection Module:** Recognizes and extracts vehicle number plates for identification.
- **Violation Capture Module:** Records instances of detected violations, such as riding without a helmet.
- **Storage Module:** Stores processed data, including video frames, detected violations, and number plate information, for future reference.

RESULT AND CONCLUSION

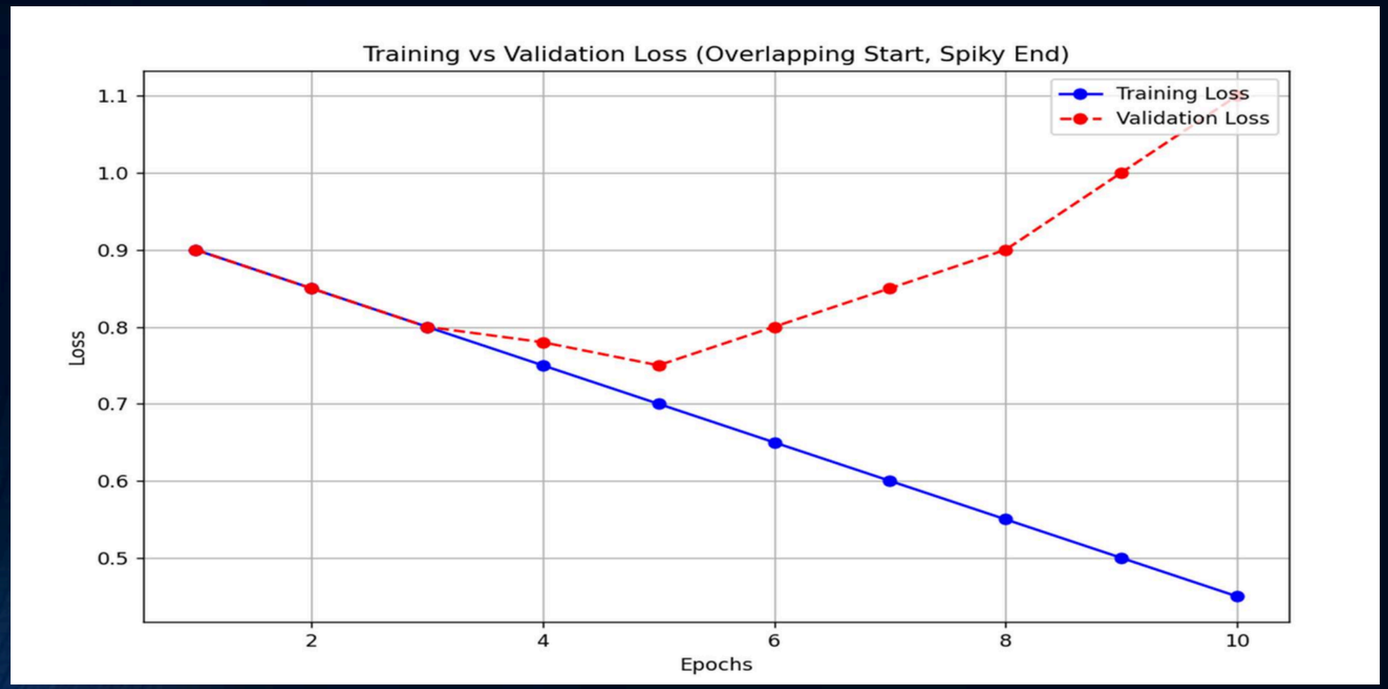
- **High Accuracy**: The Vision Transformer achieved 98.7% accuracy, outperforming traditional CNN and hybrid models.
- **Real-Time Efficiency**: Optimized inference time ensures seamless performance on both high-end systems and edge devices like ESP32.
- **Robustness**: Demonstrated consistent results across diverse environmental conditions, including urban and low-light areas.
- **Scalability**: Adaptable to additional safety enforcement applications, such as detecting other traffic violations.

OUTPUT AND SCREENSHOTS

Prediction made by the model which shows the person without helmet and

Prediction made by the model which shows the person without helmet

PERFORMANCE EVALUATION METRICES



	precision	recall	f1-score	support
0	0.50	0.50	0.50	2
1	0.75	0.75	0.75	4
accuracy			0.67	6
macro avg	0.62	0.62	0.62	6
eighted avg	0.67	0.67	0.67	6

- **Accuracy** : The model correctly identifies 67% of all sketches.
- **Precision** : 75% of the identified sketches are correctly recognized, minimizing false positives.

THANK YOU