University Institute of Information Technology
PMAS-Arid Agriculture University, Rawalpindi

Project Proposal
For
Deep Learning Based Vehicle Re-Identification based on Surveillance System

Submitted By:

Muhammad Ahsan Tayyab
22-Arid-816

Zaheer Ahmed
22-Arid-738

Supervised By:

Dr Noureen

Submission Date (11-Sep-2025)

Version 1.0.0

- **<u>Please select the appropriate category of your project</u>**

- [x] A - Desktop Application/Information System
- [x] B - Problem Solving and AI
- [ ] C - Web Application/Web Application based Information System
- [ ] D - Simulation & Modeling
- [ ] E - Smartphone Application
- [ ] F - Smartphone Game
- [x] G - Image Processing
- [ ] H - Networks
- [ ] I - Other: ____________________________

Group Members

1. Student Name: Muhammad Ahsan Tayyab
   Registration No: 22-Arid-816
   Class: BSCS
   Section: C
   Shift: Morning
   Email: ahsantayyab50999@gmail.com
   WhatsApp No: +92 319 5798625

2. Student Name: Zaheer Ahmed
   Registration No: 22-Arid-738
   Class: BSCS
   Section: C
   Shift: Morning
   Email: dev.zaheer.ahmad@gmail.com
   WhatsApp No: +92 347 5177267

TABLE OF CONTENTS

Contents

1. Introduction

With increasing urbanization, traffic jams, and rising crime, cities are turning to surveillance systems to ensure security and enforce law. Traditional surveillance technology is usually reactive and bounded — following vehicles through a series of cameras, under different viewpoints, lighting, and occlusions is still a serious challenge. Deep learning (DL) technologies have made outstanding progress in feature extraction, representation learning, and cross-camera matching. The goal of this project is to develop and deploy an efficient deep learning framework for vehicle re-identification in multi-camera surveillance settings. The system will also contain modules for detecting suspicious activities (e.g., multiple appearances in prohibited areas), thus facilitating more anticipatory law-enforcement and traffic monitoring.

2. Objective

1. Create a deep learning-based vehicle detection and re-identification pipeline operating across multiple camera views.
2. Make it robust against occlusion, varying lighting, varying angles, and cars with similar appearance (same make, color, etc.).
3. Embody anomaly / suspicious pattern detection from trajectories across cameras (e.g., a car entering restricted areas, standing still, etc.).
4. Test the system in real-time or near‑real-time environments with benchmark datasets (e.g VeRi-776, VehicleID) and, if accessible, local surveillance data.
5. Compare various architectures (transformers, CNN, attention mechanisms, part-based feature learning) on the basis of efficacy and computational expense.

3. Literature Review

A Comprehensive Survey on Deep-Learning-based Vehicle Re-Identification – Models, Data Sets and Challenges — overview of the field, shortcomings, datasets, evaluation procedures.

A Strong Baseline for Vehicle Re-Identification (CVPRW 2021) — set baseline models, losses, performance measures.

Strength in Diversity: Multi-Branch Representation Learning for Vehicle Re-Identification — several branches that merge diverse features in a bid to enhance discriminative ability.

Vehicle Re-Identification in Aerial Images and Videos: Dataset and Approach — discusses issues unique to aerial perspectives (greater viewpoint variance, resolution problems).

Learning Part-Based Features for Vehicle Re-Identification with Global Context (Applied Sciences, 2025) by Nath & Mitra. This publication suggests obtaining part-based features with global context in mind to enhance Re-ID performance against viewpoint and occlusion changes.

LKA-ReID: Vehicle Re-Identification with Large Kernel Attention (2024/2025) — proposes an attention mechanism in the form of large kernels, spatial & channel attention (hybrid channel attention) to enhance feature extraction for distinguishing similar vehicles.

4. Benchmarking

Feature / Aspect	Strong Baseline (CVPRW 2021) [2]	Multi-Branch Representation Learning [3]	Aerial Images & Videos Approach [4]	IBNT-Net (2025) [5]	CLIP-SENet (2025)	Proposed System
Global + Local Features	✔	✔	Partial	✔	✔	✔
Multi-branch Network	✔	✔	✔	✔	Partial	✔
Attention Mechanism / Transformer	✘ / limited	Limited	Some?	✔	✔	✔
Semantic Attribute Awareness	✘ / limited	Some attribute branches	Maybe orientation / viewpoint, but limited semantically	Limited semantic attributes	Strong	✔
Trajectory	✘	✘	✘	✘	✘	✔
Efficiency (Parameters / Compute)	Baseline; okay	More complex	heavy	Designed to reduce parameters (group conv etc.)	Overhead due to CLIP etc.	Targeted for reasonable computational cost (for surveillance deployment)
Suitable for Surveillance Camera Network	✔ / baseline	✔ / research	Specialized (aerial)	Yes	Yes	Yes, aimed for deployment in realistic surveillance networks
Dataset Diversity Tested	Standard datasets	Standard plus variations	Aerial datasets + ground	VeRi-776, VehicleID	VeRi, VehicleID, VeRi-Wild	Use standard + possible collected local data

5. Problem Statement

Vehicle Re-Identification (Re-ID) alone is insufficient in surveillance; it must include multi-camera tracking to build trajectories and analyze vehicle movement. Major challenges include inter-class similarity and intra-class variation under different views or occlusions. Limited compute, bandwidth, and low-resolution cameras also restrict real-world deployment. Hence, efficient yet accurate models are needed for practical systems.

Most existing research solves only partial issues, ignoring semantic attributes and real-world domain gaps. This project aims to create a robust Re-ID system effective occlusion variations. It will also focus on efficiency for real surveillance conditions. Through this, I aim to learn deep models, attention mechanisms, metric learning, and domain adaptation.

6. Problem Solution

This proposed approach effectively serves as a **solution to the identified problems** in vehicle Re-Identification and tracking under real surveillance conditions. The integration of **appearance feature extraction** (CNN + attention/transformer) with **semantic attribute extraction** directly addresses the challenges of inter-class similarity and intra-class variation by combining detailed visual cues with high-level semantic information.

The **multi-camera multi-object tracking framework** solves the issue of fragmented monitoring by linking vehicle identities across multiple cameras using spatial and temporal constraints. **Multi-task learning** enhances the model's robustness by simultaneously optimizing for classification, metric learning, and attribute recognition, ensuring viewpoint and occlusion invariance.

To tackle the **resource limitation problem**, the use of lightweight components such as group convolutions, pruning, and efficient attention mechanisms ensures that the system can run smoothly on edge devices or typical surveillance servers. Finally, evaluating the model on **benchmark and locally collected datasets** ensures both academic reliability and real-world applicability, making this approach a comprehensive and practical solution to the identified surveillance challenges.

7. Advantages/Benefits of proposed system

- Better discriminability via combining appearance features with semantic attribute features.
- Good accuracy while being computationally feasible for deployment in surveillance systems.
- Extended generalization: ability to perform well even when test domain is different (new camera, new location etc.).
- Contribution to knowledge in vehicle Re-ID with possibly annotated local data or domain-specific evaluation.
- Potential benefits for traffic monitoring, law enforcement, etc., in smart city contexts.

8. Scope

This project will cover:

- Implementation of a deep learning model combining CNN + attention/transformer + part-aware branches + semantic attribute fusion.
- Use of standard benchmark datasets (VeRi-776, VehicleID, VeRi-Wild etc.) plus possibly collecting a small local dataset from CCTV surveillance to test realistic conditions.
- Experiments covering variations: viewpoint, illumination, occlusion, weather.
- Evaluation metrics: mAP, Rank-1, Rank-5, possibly newer metrics (e.g. mean positive sample occupancy etc.).

Modules of the Project:

1. **Data Preparation & Augmentation Module**
   - Collect & preprocess benchmark datasets; optionally collect local CCTV data.
   - Perform augmentations: changing lighting, blur, occlusion simulation, weather effects.
   - Annotate semantic attributes (color, type, model, etc.) if needed.

2. **Feature Extraction Backbone Module**
   - Build CNN backbone, possibly ResNet or ResNet-IBN, for extracting base features.
   - Incorporate local/part branches and transformer / self-attention modules for global context.

3. **Semantic Attribute Module**
   - Extract attributes either via pretrained vision-language model or via attribute classification branch.
   - Fuse semantic features with appearance features.

4. **Loss / Metric Learning Module**
   - Implement classification loss, metric loss (triplet, contrastive), and maybe additional losses to enforce viewpoint invariance.
   - Multi-task learning combining multiple losses.

5. **Evaluation & Testing Module**
   - Evaluate standard datasets and local test data.
   - Measure mAP, Rank-1, etc., possibly newer/evaluation metrics.
   - Ablation studies (e.g., effect of semantic attributes, effect of attention vs no attention, etc.)

6. **Optimization & Efficiency Module**

- Optimize network size, inference speed, memory usage.
- Use techniques like group convolution, pruning, perhaps model quantization.

9. Software Methodology

We plan to use **Agile development with iterative prototyping**, for the following reasons:

It allows rapid prototyping of different architectures (baseline, attention, semantic fusion etc.), and quick feedback from experiments.
Flexibility to adjust depending on performance, e.g. change module designs or dataset augmentation. Helps manage tasks among team members effectively.

10. Tools and Technologies

Tools And Technologies	**Tools**	**Version**
	VS Code	1.104
	PyTorch	2.7.0
	Git/Github	Latest
	GPU(s)	2013
	OpenCV, custom scripts	Latest
	Standard Re-ID metrics code (mAP, Rank-k), plus possibly custom metric implementation	Latest
	Technology	**Version**
	Python	3.8 or higher
	TensorFlow etc	Latest
	ResNet50, ResNet-IBN, Vision Transformer modules, CLIP image encoder etc	Latest

11. Concepts

Deep convolutional neural networks, transformer/self-attention architectures for computer vision. Metric learning methods: triplet loss, contrastive loss; designing multi-task losses.Semantic attribute extraction and fusion (vision-language models, attribute classification).Spatio-temporal modeling for cross-camera association (trajectory prediction, constraints based on camera layout). Model optimization and efficiency (parameter reduction, group conv, pruning, inference speed).

12. Intended Users

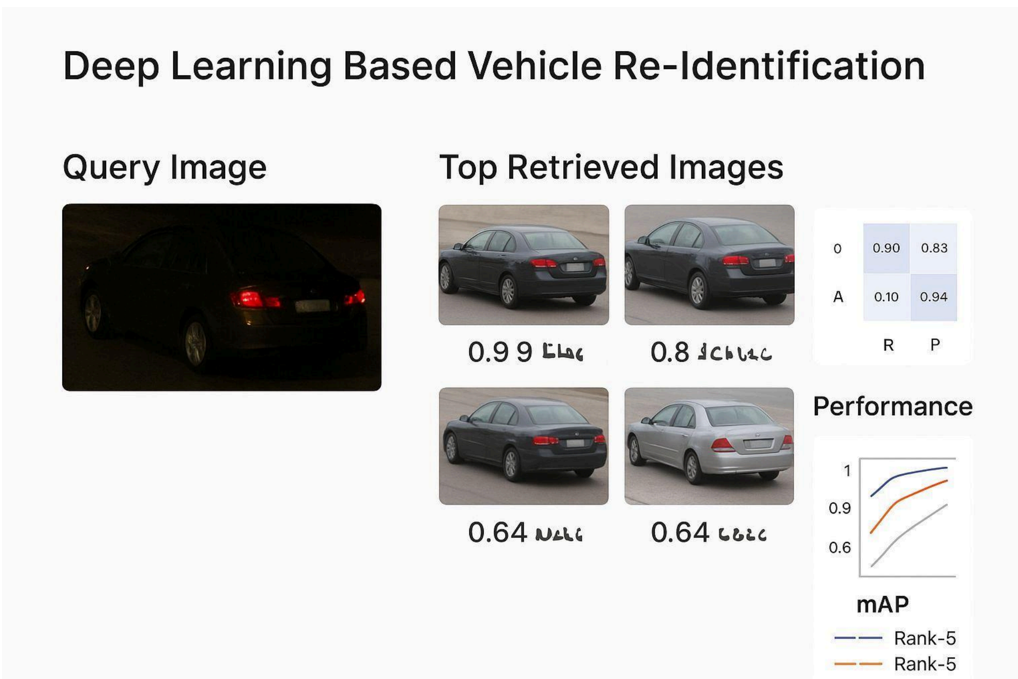
Surveillance / security agencies wanting to track vehicles across multiple cameras.
Traffic management authorities for monitoring, path/flow analysis, speed estimation.
Smart city systems that need to detect anomalies or cross-camera movement.
Researchers developing Re-ID + tracking systems in computer vision.

13. Mockups

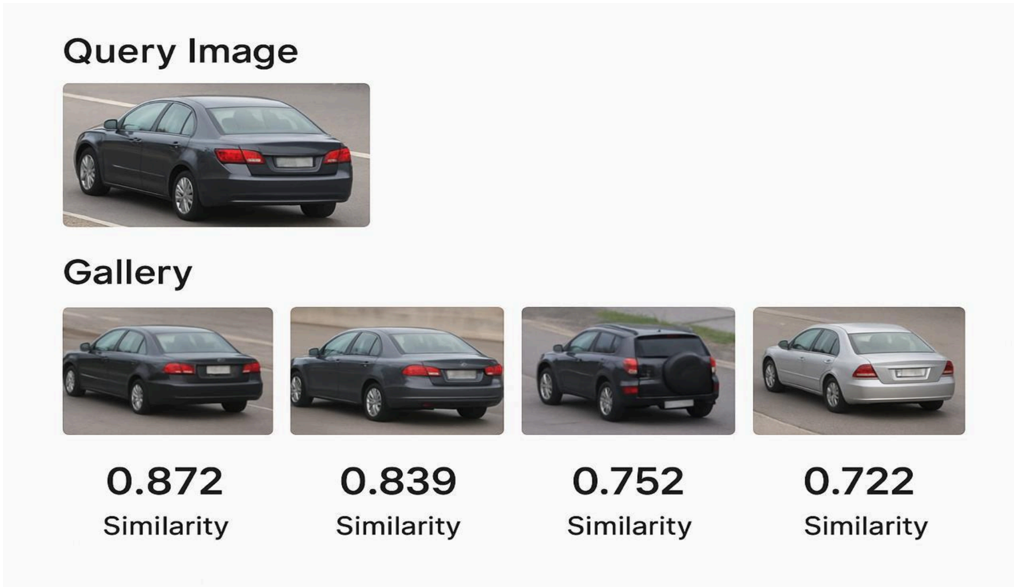
Since this is mostly a deep learning / backend system, mockups could include:

1. Comparative view: query image, and top-5 retrieved images, with similarity scores.
2. Interface for uploading local CCTV images and seeing performance under various conditions (lighting, occlusion).

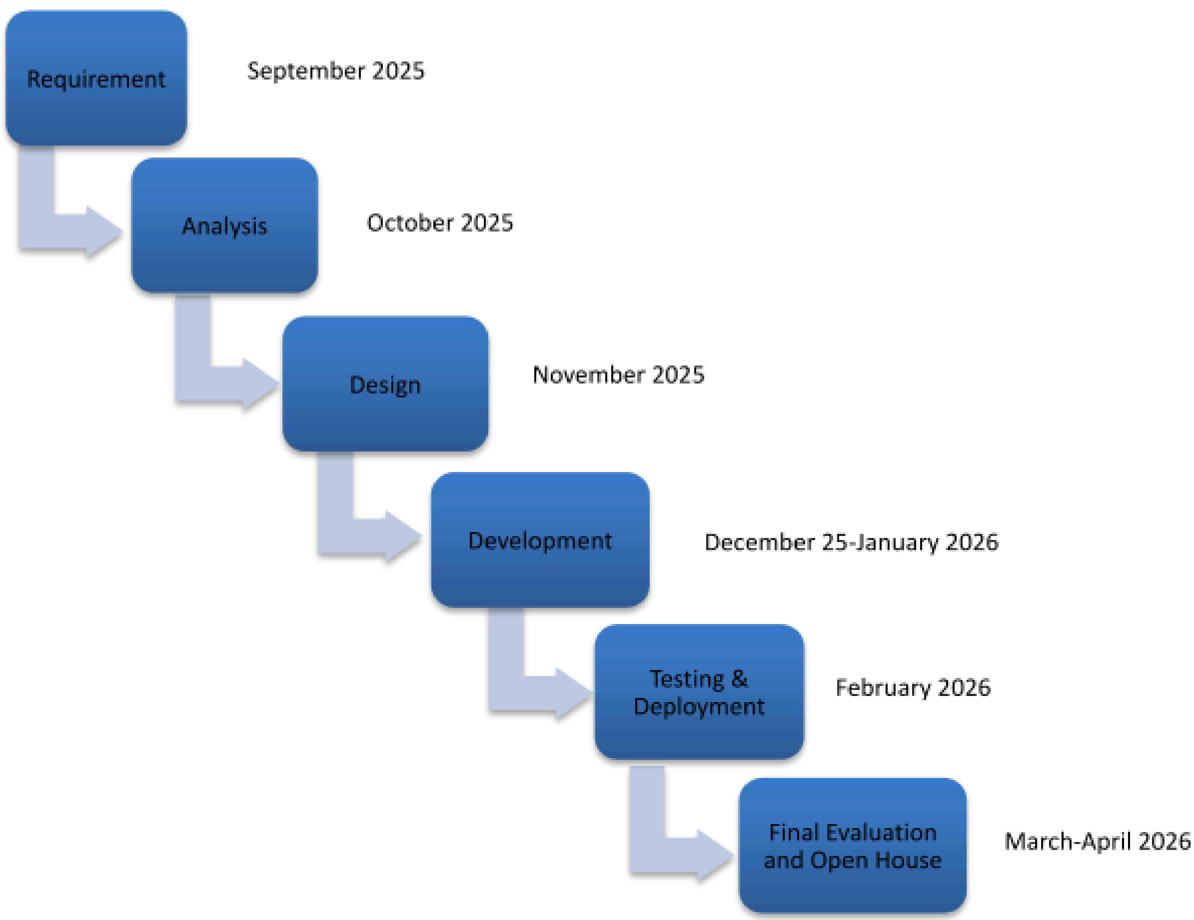
3. Performance visualization: confusion matrix, mAP / Rank-1 / Rank-5 graphs, possibly visualizing feature attention maps (which parts the model attends to).



4. A dashboard showing query image and retrieval results from gallery (matched vehicles).



14. Timeline



15. Conclusion

In our work, we aim to build a system that does not only re-identify vehicles across cameras, but also tracks their trajectories through multi-camera networks, combining strong Re-ID with tracking. By integrating semantic attributes, robust augmentation, and spatio-temporal trajectory association, we expect to achieve better performance in realistic surveillance conditions. This dual focus on both Re-ID and trajectory tracking will increase the practical value of our system, making it suitable for real-world surveillance, traffic analysis, and smart city applications.

16. References

1. Amiri, A. Kaya, and A. S. Keceli, “A Comprehensive Survey on Deep-Learning-based Vehicle Re-Identification: Models, Data Sets and Challenges,” arXiv preprint arXiv:2401.10643, 2024.

2. S. V. Huynh, N. H. Nguyen, N. T. Nguyen, V. T. Q. Nguyen, C. Huynh, and C. Nguyen, “A Strong Baseline for Vehicle Re-Identification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2021, pp. 1–10.

3. R. Kumar, A. Subramanian, and V. Gandhi, “Strength in Diversity: Multi-Branch Representation Learning for Vehicle Re-Identification,” in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), 2019, pp. 1–10.

4. Y. Li, J. Zhang, and X. Wu, “Vehicle Re-Identification in Aerial Images and Videos: Dataset and Approach,” in Proc. IEEE Int. Conf. Image Process. (ICIP), 2020, pp. 1–5.

5. Y. Wang, R. Li, and Y. Shao, “Vehicle Re-Identification Method Based on Efficient Self-Attention CNN-Transformer and Multi-Task Learning Optimization,” Sensors, vol. 25, no. 10, pp. 1–16, 2025.

6. L. Lu, Z. Fu, D. Chu, W. Wang, and B. Xu, “CLIP-SENet: CLIP-based Semantic Enhancement Network for Vehicle Re-Identification,” arXiv preprint, arXiv:2502.04567, 2025.

7. R. Athilakshmi, E. V. S. S. Prasanna, S. Yasir, “Multi-Camera Trajectory Based Vehicle Re-Identification for Robust Traffic Monitoring,” in Proc. 2025 ICAECT: Advances in Electrical, Computing, Communication and Sustainable Technologies, 2025.