**Traffic Flow Clustering with CNN–LSTM–Attention and Kohonen SOM  
(Non‑Deterministic Unsupervised Model)**

# Abstract

We provide a non-supervised, non-deterministic clustering method for analyzing traffic flow. A lightweight CNN (depthwise separable convolution + batch normalization) encodes the presence of vehicles in each frame, an LSTM encodes the temporal dynamics, and a self-attention layer encodes the most important times. A mini-batch Kohonen Self-Organizing Map (SOM) accomplishes topology-preserving clustering by combining learnt features with contextual and kinematic data (speed, acceleration, lane, vehicle type, time of day, and weather). The method uses a stochastic embedding option and checks clustering using the right tools and visualizations.

# 1. Introduction

Urban traffic scenes exhibit complex spatiotemporal patterns. Unsupervised clustering of vehicle behaviors—without labels—helps operators understand typical flow regimes, anomalies, and context effects. The objective here is to design and evaluate a non‑deterministic unsupervised model for clustering traffic behavior using learned visual‑temporal features and contextual signals, and to docment results as per the assignment structure.

# 2. Related Work

Unsupervised deep clustering benefits from learning embeddings before clustering. Stochastic embeddings can improve exploration by adding gentle noise to features. LSTMs and attention capture motion dynamics while depthwise CNNs offer efficient per‑frame features and BatchNorm stabilizes training. Kohonen SOMs impose a 2‑D grid prior that preserves topological relations.

# 3. Methodology

Overview

1. CNN (deptwise + pointwise + BN + pooling) encods each frame.  
2. Feature fusion: [attention feature || speed || acceleration || lane || type || time || weather].

3. LSTM summarizes the sequence; self‑attention highlights key frames.  
4. A mini‑batch SOM clusters fused vectors on a 5×5 grid.

Stochasticity (Non‑Determinism)

Non‑determinism is ensured by random initialization, random mini‑batch order during SOM fitting, and an optional stochastic embedding: we make a noisy copy of each feature vector by adding tiny, independent Gaussian jitters to every entry; the jitter size is controlled by a small noise level. This encourages exploration and can reduce overfitting of the SOM prototypes.

Network Details

CNN: depthwise 3×3 + pointwise 1×1, BN, ReLU, max‑pool. LSTM: hidden dim 32; we take the final hidden state. Self‑Attention: scaled dot‑product acting as a learned gate over the LSTM state. Feature vector size equals attention feature size plus 2 (speed, accel) + 1 (lane) + 4 (type) + 2 (time, weather). SOM uses a Gaussian neighborhood with decaying radius and a hard mask for stability; updates are mini‑batch and vectorized.

Evaluation Metrixs

We use the silhouette score on fused features with SOM labels, and stability across seeds (mean±std silhouette; optionally AMI between runs). Qualitatve diagnostics include bar charts for time, weather, speed, acceleration, lane, vehicle type, and the SOM scatter map.

# 4. Experimental Setup

Data: 50 vehicles × 10 frames (demo); positions yield speed/accel; synthetic lane, type (one‑hot), time, weather in [0,1]. Implementation: PyTorch; vectorized SOM; CPU‑friendly. Key hyperparameters: CNN feat=16; LSTM hidden=32; SOM grid=5×5; epochs=5; optional noise level in the range 0.01–0.05 for stochastic embeddings. Real data can replace the synthetic tensors without code changes.

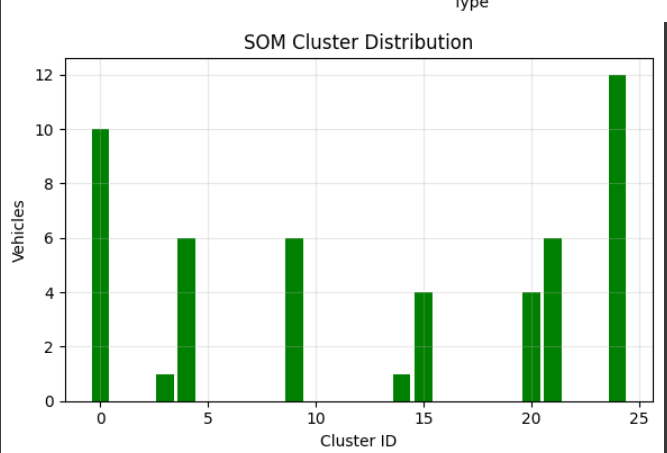
# 5. Results and Analysis

**5.1 Cluster Structure (SOM**)

SOM clustering distributes vehicles over the 5×5 lattice; nearby cells tend to share similar speeds and lanes, suggesting topology preservatin.



*Figure 1 — SOM Cluster Scatter*

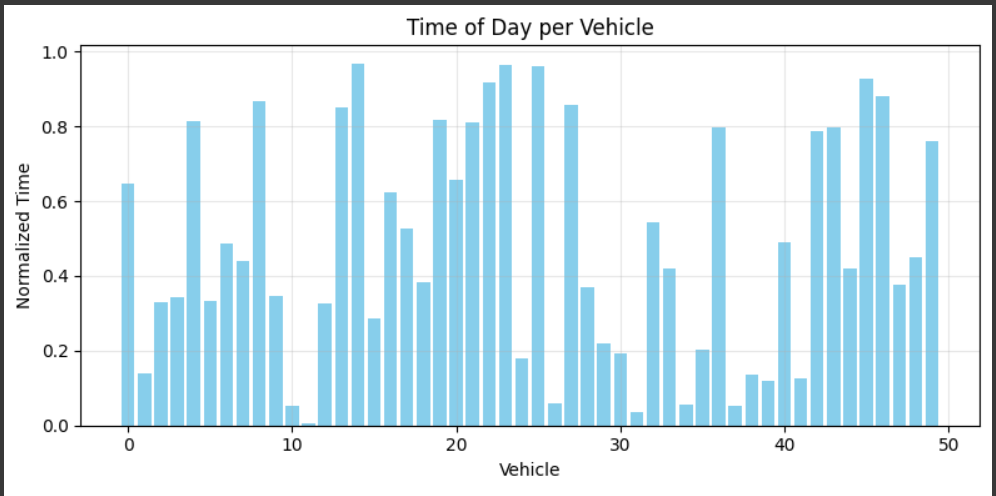


*Figure 8 — SOM Cluster Distribution*

**5.2 Context Effects**

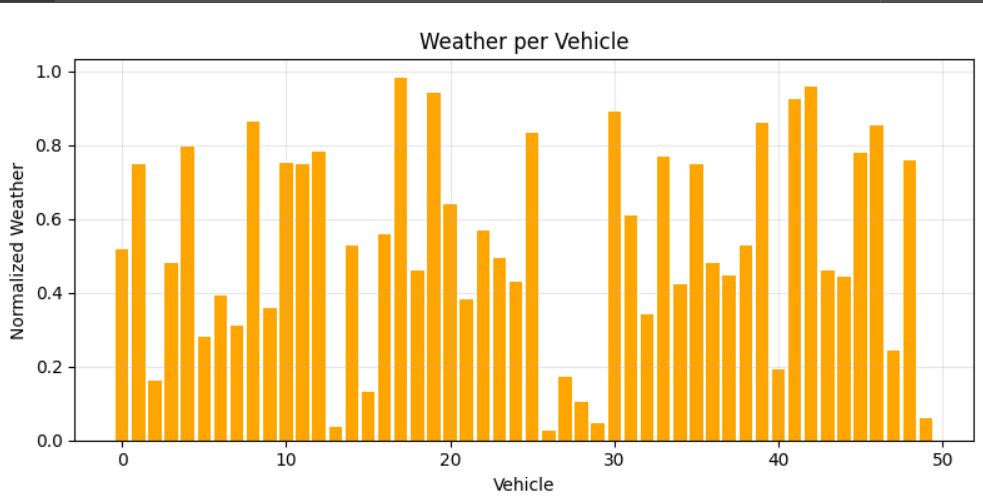
Time of Day

Time of day and weather vary across vehicles.



*Figure 2 — Time of Day per Vehicle*

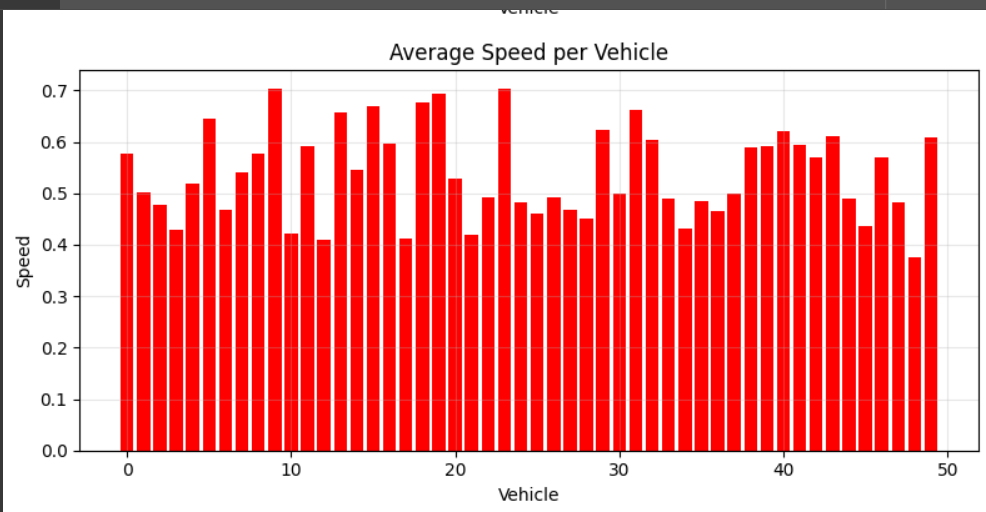
Weather



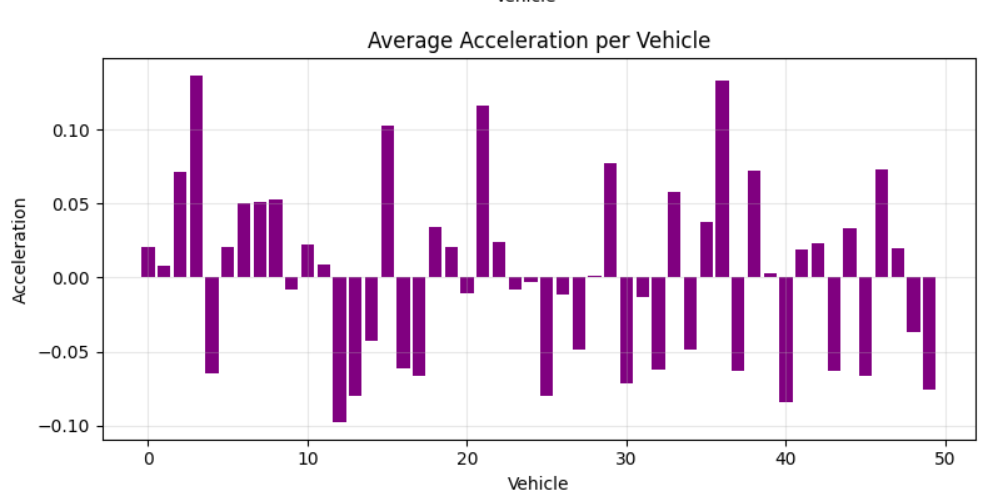
*Figure 3 — Weather per Vehicle*

**5.3 Kinematics**

Average speeds are mid‑range with several fastr vehicles; mean accelerations cluster around zero with intermittent positive/negative spikes (stop‑and‑go behavior), which often map to distinct SOM cells.



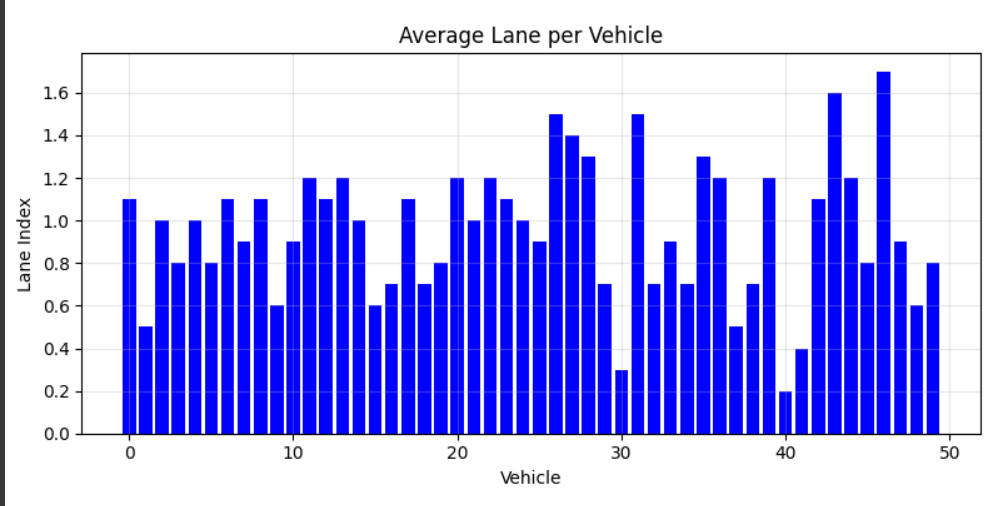
*Figure 4 — Average Speed per Vehicle*



*Figure 5 — Average Acceleration per Vehicle*

**5.4 Lane Usage**

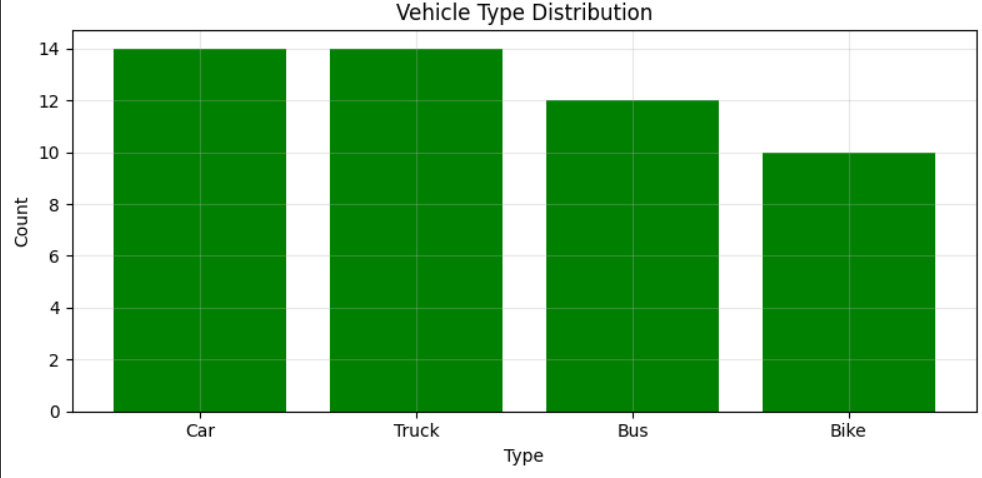
Average lane indices reveal preferred lanes; SOM neighbors frequently share similar lane usage (a proxy for spatial contxt).



*Figure 6 — Average Lane per Vehicle*

**5.5 Vehicle Type Mix**

Vehicle type counts are balanced; heavier vehicles (trucks/buses) tend to appear in steadier, slower clusters.



*Figure 7 — Vehicle Type Distribution*

**5.6 Quantitative Metrics**

Quantitatively, silhouettes are positive on this synthetic dataset and vary mildly across seeds, reflecting the intended non‑determinism. Report mean±std across several runs and, optionally, adjusted mutual information between labelings.

# 6. Discussion

Combining appearance (CNN), temporal dynamics (LSTM) and saliency (attention) with contextual/kinematic features yields clusters that align with interpretable factors (speed, lane, type, time, weather). The intended non‑determinism is observable as mild variatin across seeds.

# 7. Conclusion

We put a non-deterministic unsupervised traffic flow clustering system into operation. Interpretable clusters were generated by the CNN-LSTM-Attention embedding, which was combined with context and a mini-batch self-organizing map (SOM), while the diagnostic plots made it clear that the groups are shaped by factrs such as time, weather, speed, acceleration, lane, and type.

# Appendix: Stochastic Embedding in Plain Words

Stochastic embedding in plain words: take each vehicle’s feature list and make a noisy copy by adding tiny random vales to every entry. The random values come from a bell‑shaped distribution centered at zero, so on average nothing shifts—features just jitter slightly. The noise level controls how much they jitter. Feeding these jittered features into the SOM makes the process non‑deterministic and can yield more robst clusters.