A CNN-RF-SVM Hybrid Method for Traffic Congestion Monitoring

Jiaxu Zhu

College of Software and Artificial Intelligence, Software Engineering Institue of Guangzhou, Guangzhou, Guangdong, China
zjx2214@smail.seig.edu.cn

Abstract. With the development of the automotive industry and increasing demand for personal mobility, the growing number of vehicles on roads has led to traffic congestion and environmental pollution. Accurate traffic congestion monitoring has become a prominent research focus. The integration of surveillance camera recognition with intelligent transportation systems presents a promising solution for congestion mitigation. This study categorizes traffic congestion data collected from Singapore's Land Transport Authority (LTA) open API into five classes: empty, low, medium, high, and congested. The dataset is then partitioned into three subsets (training, validation, and testing) through stratified sampling. A hybrid CNN-RF-SVM approach is proposed for congestion detection. Experimental results demonstrate that compared with traditional image recognition models, the proposed method achieves 36.6% and 26% improvements in F1-score over CNN-SVM and CNN-RF models respectively. The framework provides high-precision road congestion identification, which can effectively identify and judge congested roads, offering reliable data support for real-time monitoring and dynamic traffic management.

Keywords: CNN-RF-SVM Composite Method; Traffic Congestion Identification; Traffic Density Singapore.

1 Introduction

The accelerating urbanization process and continuous growth in vehicle ownership have made traffic congestion a common challenge faced by many cities worldwide. In metropolitan areas, worsening traffic congestion not only increases harmful emissions and reduces urban living comfort, but also decreases energy conversion efficiency, generating adverse economic and social impacts [1]. To alleviate congestion and improve road capacity, traffic flow prediction and congestion identification have become crucial research directions. In recent years, machine learning-based traffic prediction models (e.g., Support Vector Machines and neural networks) have emerged as research hotspots due to their effectiveness in handling nonlinear relationships and improving prediction accuracy [2].

In traffic congestion recognition, image processing technologies have gained widespread application for their intuitiveness and efficiency. Typical approaches

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_12

include: edge detection-based real-time vehicle congestion state recognition [3], CNN-based image analysis [4, 5], SVM-assisted traffic image classification [6], and RF algorithm-based models [7]. Furthermore, hybrid models combining multiple algorithms' advantages have become mainstream solutions [8, 9, 10]. These models provide robust data support for real-time monitoring and dynamic traffic management.

This study proposes a novel CNN-RF-SVM hybrid approach for road congestion classification using traffic image datasets. The dataset contains five categories: Empty (no vehicles), Low (sparse traffic), Medium (moderate traffic), High (heavy traffic with congestion tendency), and Traffic Jam (severe congestion). The framework employs: ResNet101 as a pretrained feature extractor, RF model trained with extracted features, and SVM model pruned by the trained RF. The classification focuses primarily on Empty, High, and Traffic Jam states to enable timely congestion detection and empty lane identification for traffic diversion. Experimental results demonstrate the model's high-precision congestion recognition capability, providing scientific decision-making support for transportation authorities to mitigate congestion problems

2 Data Set

The dataset was collected from Kaggle, containing traffic congestion images classifiable into five density levels: Empty, Low, Medium, High, and Traffic Jam. From 87 available surveillance cameras, this paper selected 20 units that provided diverse and qualified photographs based on four criteria: variations in road traffic density, lighting conditions, unobstructed visibility, and appropriate camera angles.

A total of 4,054 images were downloaded and manually annotated, with the dataset randomly partitioned into 80% for training, 10% for validation, and 10% for testing. To enhance model robustness, the dataset incorporates images captured during both daytime and nighttime conditions. Representative samples from each classification category are displayed in Fig 1, Fig 2, Fig 3, Fig 4 and Fig 5.

Fig. 1. Empty [11].

Fig. 2. Low [11].

Fig. 3. Medium [11].

Fig. 4. High [11].

Fig. 5. Traffic Jam [11]

3 Method

3.1 Overall Process

As shown in Figure 6, description the experimental procedure initiates with data preprocessing to standardize input formats and augment feature dimensions.

Subsequently, pretrained ResNet101 architecture is employed to extract high-dimensional features from the processed data.

The extracted features then undergo dimensionality reduction, with scree plot visualization applied to validate component selection effectiveness.

A Random Forest classifier is trained using the reduced feature set to generate importance rankings.

These importance scores guide the pruning process of SVM model parameters, ultimately yielding the optimized final model.

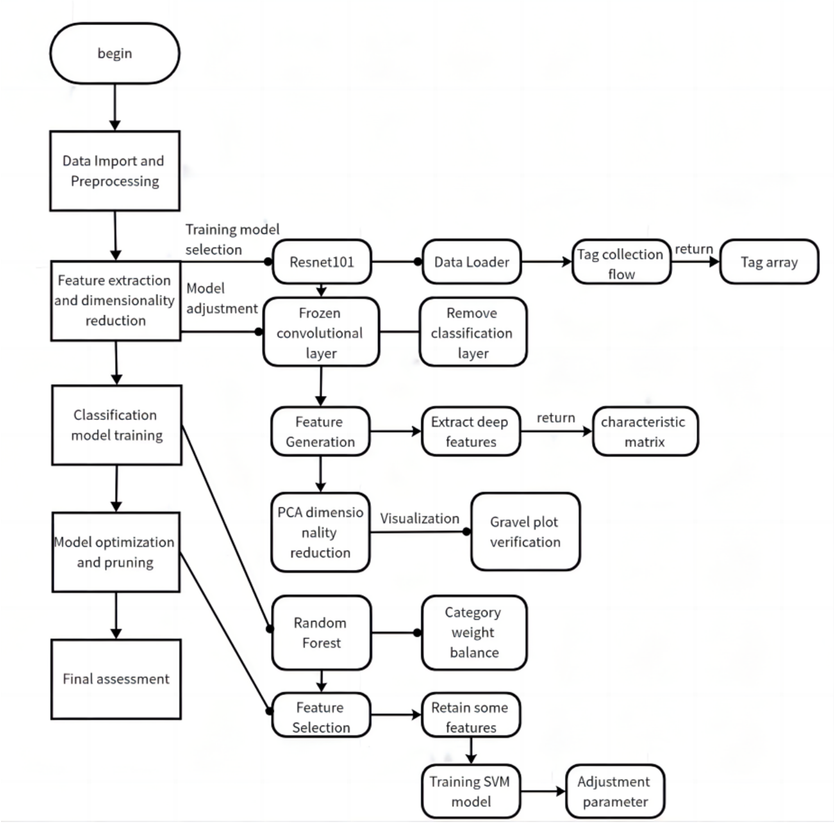

Fig. 6. Overall flowchart. (Picture credit: Original)

3.2 Data Import and Preprocessing

The image preprocessing workflow sequentially executes: Resize() operation for dimensional standardization to reduce computational overhead while preventing distortion artifacts; RandomResizedCrop() applied exclusively to training sets for stochastic cropping and scaling to simulate variable capture perspectives; CenterCrop() implementation restricted to validation/testing phases ensuring evaluation input uniformity; RandomVerticalFlip() augmentation activated only during training to expand sample diversity; ToTensor() conversion restructuring array dimensions with simultaneous pixel value normalization; followed by Normalize() transformation approximating standard Gaussian distribution to accelerate model convergence and enhance predictive stability.

3.3 Feature extraction and dimensionality reduction

The architecture utilizes ResNet101 (pretrained on ImageNet) as the backbone model with all convolutional parameters frozen to maintain pretrained feature extraction capacity. The original classification layer is substituted by an identity mapping transformation, producing 2048-dimensional feature vectors through forward propagation across training, validation, and test datasets. This module implements a dedicated feature extraction function accepting dual inputs (data loader and pretrained model), where forward propagation generates deep feature representations (e.g., ResNet101's 2048-D output) while synchronously recording original image labels for supervised learning tasks, ultimately outputting stacked feature matrices and flattened label arrays for classifier compatibility. Mathematically, let φ denote the pretrained mapping function and X the input data, the feature extraction procedure formalizes as:

$$\varphi{:}X{\to}F, F{=}\varphi(X) \tag{1}$$

For each batch of data (X_b, Y_b) in the data loader, where F_b represents the current batch's feature matrix, X_b denotes the b-th batch of input data, and Y_b corresponds to the label vector:

$$F\_b = \varphi(X\_b), L\_b{=}Y\_b \tag{2}$$

The final output is generated through vertical stacking (features) and horizontal concatenation (labels), where F is the complete feature matrix and L is the flattened label array：

$$F = \begin{bmatrix} F\_1 \\ \dots \\ F\_2 \end{bmatrix}, L = [L\_1 \dots L\_N] \tag{3}$$

3.4 Classification Model Training:

Train a random forest model using PCA dimensionality reduced features, and obtain feature importance evaluations based on the method of reducing impurity, in preparation for pruning the SVM model in the future. W is the weight of node samples, and I is impure

$$RI_{\mathrm{i}} = w_{\mathrm{i}} \cdot I_{\mathrm{i}} - w_{left} \cdot I_{left} - w_{right} \cdot I_{right} \tag{4}$$

3.5 Model Optimization and Pruning

Select a subset of features based on the Gini importance or permutation importance of a random forest, train an SVM based on this subset, and prune the SVM model based on the feature importance output from the random forest to achieve dual dimensionality reduction and model simplification. The selected feature index set is the threshold

$$S^* = \{j|Importance(f_j) > \gamma \tag{5}$$

4 Results and Discussion

4.1 Exploring Traditional Methods

As shown in Table 1, the traditional CNN-SVM method performs poorly. Although the F1 score of the Empty class reaches 0.91, which can capture road background information well, the scores of other classes are low, indicating that the CNN-SVM method cannot capture road congestion features well. The performance of the CNN-SVM method in the validation set is shown in Table 2.

Table 1. Performance of CNN-SVM in the validation set.

	Precision	Recall	F1-score	support
Empty	0.91	0.92	0.91	120
High	0.67	0.82	0.74	38
Low	0.82	0.71	0.76	94
Medium	0.68	0.73	0.70	70
Traffic Jam	0.75	0.67	0.71	18
accuracy	\	\	0.80	340
macro avg	0.77	0.77	0.76	340
Weighted avg	0.80	0.80	0.80	340

As shown in Table 2, the traditional CNN-RF method also performs poorly, as it can be seen that the random forest model has poor feature capture ability and low F1 scores for all categories, making it difficult to identify congestion situations effectively. The performance of the CNN-RF method validation set is shown in Table 3 above.

Table 2. Performance of CNN-RF method in the test set.

	Precision	Recall	F1-score	support
Empty	0.89	0.87	0.88	120
High	0.79	0.58	0.67	38

Low	0.72	0.71	0.72	94
Medium	0.62	0.76	0.68	70
Traffic Jam	0.81	0.72	0.76	18
accuracy	\	\	0.76	340
macro avg	0.76	0.73	0.76	340
Weighted avg	0.77	0.76	0.76	340

4.2 Comprehensive Performance Analysis of CNN-RF-SVM Composite Method

As shown in Table 3, it can be seen that Traffic Jam class recognition is the best, with an F1 score of 0.96, thanks to the significant traffic density feature in the dataset during traffic congestion. The F1 score of the High class reached 0.86, slightly lower than that of the Traffic Jam class, indicating that although the model can effectively recognize dense traffic features, its classification ability decreases when recognizing two types of transitional images that also have dense traffic features. This is also reflected in the recognition of the Low and Medium classes, which are easily confused by the model, with F1 scores of 0.66 and 0.7, respectively. The precision of the Empty class can reach 0.9, proving that the model has sufficient background feature extraction for road images. The specific performance of the CNN-RF-SVM composite method in the test set is shown in Table 4.

Table 3. Specific Performance of CNN-RF-SVM Composite Method in the Test Set.

	Precision	Recall	F1-score	support
Empty	0.90	0.86	0.88	64
High	0.83	0.89	0.86	64
Low	0.69	0.62	0.66	64
Medium	0.69	0.77	0.73	64
Traffic Jam	0.98	0.94	0.96	64
accuracy	\	\	0.82	320
macro avg	0.82	0.82	0.82	320
Weighted avg	0.82	0.82	0.82	320

4.3 Key improvement effects

As shown in Table 4, Compared with the CNN-SVM method, the Traffic Jam class accuracy of the CNN-RF-SVM composite method increased by 30%, the recall rate increased by 40%, and the F1 score increased by 36.6%. High class accuracy increased by 23%, recall increased by 8%, and F1 score increased by 16%. From the data changes, it can be seen that the CNN-RF-SVM model has better performance compared to a single SVM model, and can achieve more accurate identification of road congestion targets based on features extracted from random forests.

Table 4. Comparison of F1 values between CNN-SVM method and CNN-RF-SVM composite method.

	F1(SVM)	F1(Composite method)
Empty	0.91	0.88
High	0.74	0.86
Low	0.76	0.66
Medium	0.70	0.73
Traffic Jam	0.71	0.96
accuracy	0.80	0.82
macro avg	0.76	0.82
Weighted avg	0.80	0.82

Table 5. Comparison of F1 values between CNN-RF method model and CNN-RF-SVM composite method.

	F1(RF)	F1(Composite method)
Empty	0.88	0.88
High	0.67	0.86
Low	0.72	0.66
Medium	0.68	0.73
Traffic Jam	0.76	0.96
accuracy	0.76	0.82
macro avg	0.76	0.82
Weighted avg	0.76	0.82

As shown in Table 5, Compared with the CNN-RF method, the Traffic Jam class accuracy of the CNN-RF-SVM composite method increased by 20%, the recall rate increased by 30%, and the F1 score increased by 26%. High class accuracy increased by 5%, recall increased by 41%, and F1 score increased by 28%. From the data changes, it can be seen that the CNN-RF-SVM model has better performance in identifying road congestion compared to a single random forest model. However, when identifying Low classes, the F1 score decreased by 8%, indicating that the CNN-RF-SVM composite method generates more noise compared to a single random forest model. The comparison of F1 values between the CNN-RF method model and the CNN-RF-SVM composite method is shown in Table 5 above.

5 Conclusion

Experimental results demonstrate that the proposed CNN-RF-SVM hybrid method exhibits superior effectiveness by synergistically combining CNN's superior image feature extraction capability, Random Forest's noise resistance and anti-overfitting properties, and SVM's high-dimensional classification performance with small samples. Compared to traditional SVM and RF models, this integrated approach achieves significantly enhanced recognition accuracy in traffic congestion identification tasks, enabling rapid analysis of surveillance camera images to provide real-time congestion data for intelligent transportation systems (e.g., navigation), thereby contributing to traffic flow mitigation.

However, while the method demonstrates high precision in classifying empty and congested road conditions, its predictive capability remains limited during transitional phases with moderate-to-high traffic flow due to similar image features across intermediate states. Additional constraints include dataset limitations (solely comprising top-down perspectives potentially affecting model generalizability) and unverified real-world performance despite achieving <0.08s/image processing speed on laptops, given the substantial hardware differences with actual surveillance systems. Future enhancements should incorporate temporal feature processing (e.g., ARIMA integration) to improve predictive accuracy and dynamic traffic flow analysis, offering practical significance for real-time urban traffic management.

References

1. Yang, J.: Research on the Impact of Urban Traffic Congestion on Environmental Pollution and Countermeasures. China Logistics & Purchasing (9), 77–78 (2024).
2. Anjaneyulu, M., Kubendiran, M.: Short-term traffic congestion prediction using hybrid deep learning technique. Sustainability 15(1), 74 (2022)
3. Li, J., Yan, Y.: Implementation of an Image-based Traffic Congestion State Recognition System. Computer Engineering and Design 32(4), 1366–1369 (2011).
4. Duan, M.: Research on Image Recognition Method Based on Convolutional Neural Network. Master's thesis, Zhengzhou University (2017). CNKI:CDMD:2.1017.139824
5. Xiao, P., Huang, H.: Research and Application of Deep Learning Based on CNN Algorithm. Modern Computer (35), 6 (2019). CNKI:SUN:XDJS.0.2019-35-007
6. Zhang, J., Chen, S., Liu, H., Hu, N.: Research on Human-Vehicle Recognition in Traffic Video Based on Support Vector Machine. Video Engineering 35(15), 1–3, 15 (2011).
7. Peng, B.: Traffic Congestion Prediction Using Machine Learning Techniques. Communications World 26(1), 254–255 (2019)
8. Liang, Q.: Research on Face Recognition Based on CNN and SVM Fusion Features. Modern Information Technology 4(19), 62–65, 70 (2020).
9. Chen, Y., Li, S., Deng, L.: Research on Traffic Flow Prediction Based on Hybrid Traffic Flow Discrete Model. Transport & Transportation 38(4), 11–15 (2022)
10. Jiang, S., Wang, D., Zhao, Y., Wei, W.: Mixed Traffic Flow Foreground Extraction Based on Composite Algorithm. Journal of Jilin University (Engineering Edition) 41(S1), 76–80 (2011)
11. rahat52.: Traffic Density Singapore, 2023, Retrieved from https://www.kaggle.com/datasets/rahat52/traffic-density-singapore, (2025.9.19)