CHONGQING JIAOTONG UNIVERSITY

AITC

EMSA-Net : An Efficient Multi-Scale and Boundary Aware Network for Spillage Risk Vehicle Detection Under Low-Light Urban Conditions

Author : Ze Xu, Wenting Jia, Chunyang Yan, Ruikuang Wang, Yongfu He*

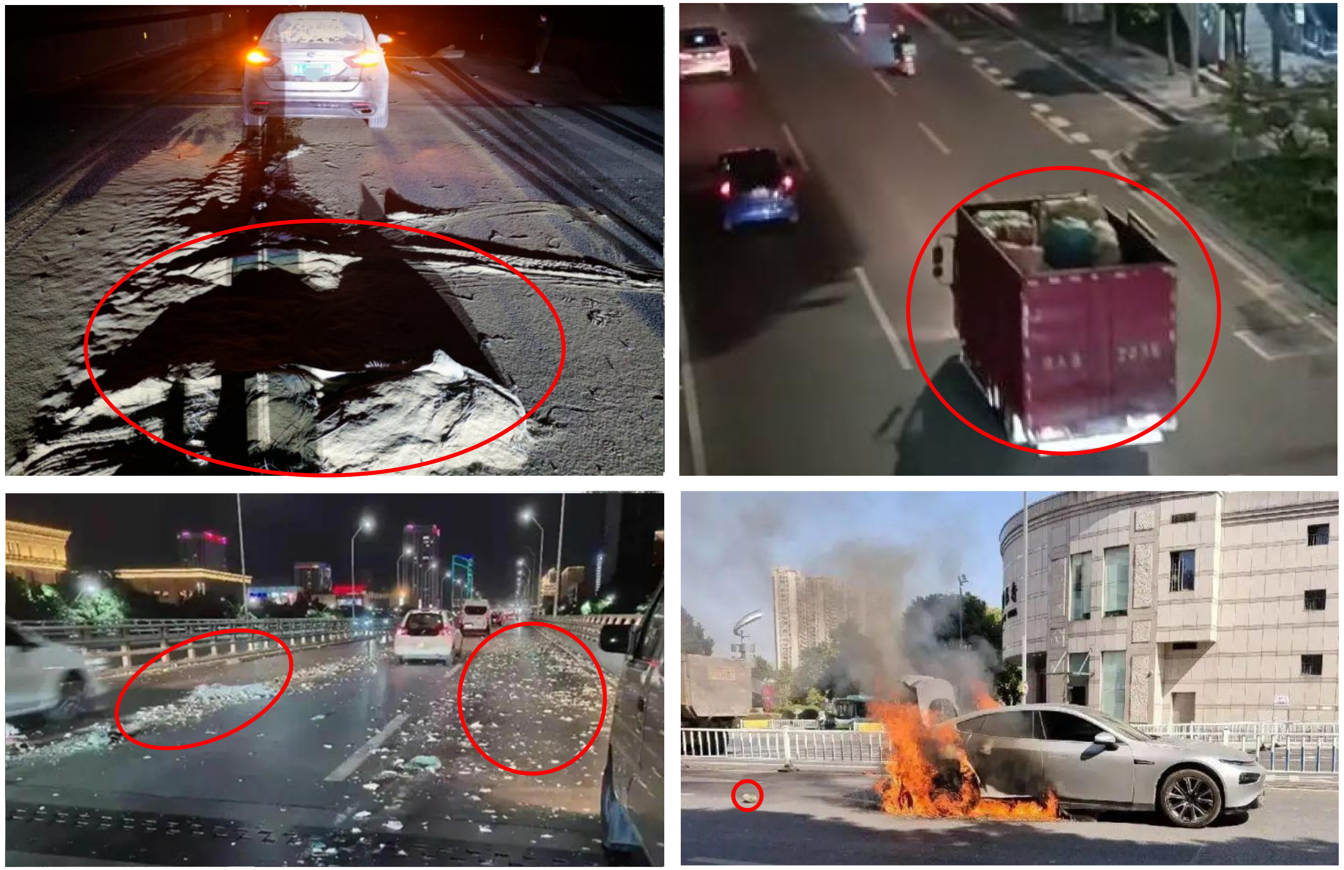
Chongqing Jiaotong University

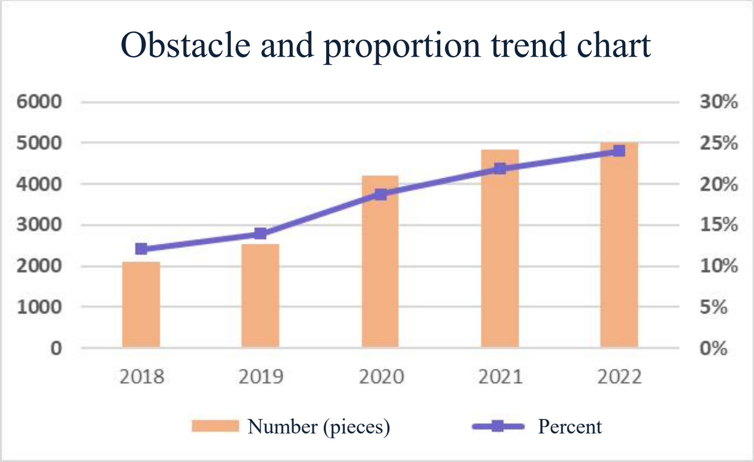
CONTENTS

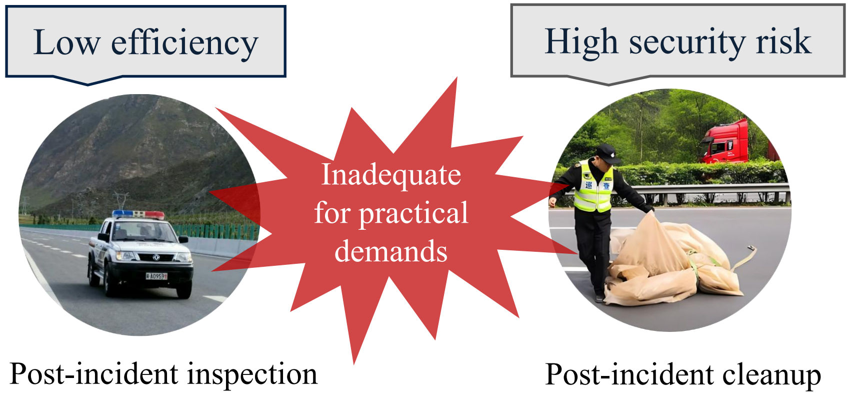
01.

Background

1.1 Research background



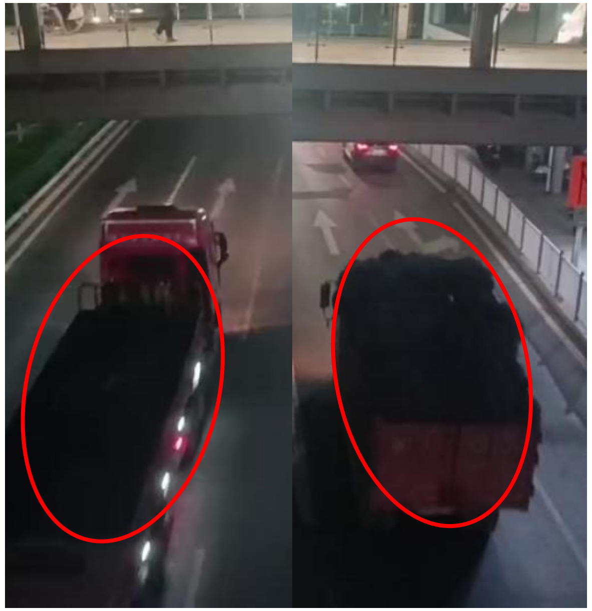




90% of road debris on the Shaoxing Expressway is caused by improper cargo loading, such as uncovered tarpaulins or unsecured goods. According to 2022 statistics, 78% of traffic accidents resulting from this debris occurred at night or in the early morning.

1.2 Research significance

Drastic variations in target scale, cargo form, and loading state



Severe boundary ambiguity caused by shadows under low-light conditions

Category	Model	Characteristics	Overview
Traditional Methods	Background Subtraction, GMM, etc.	Limited by the expressive power of handcrafted features, resulting in poor generalization.	Focuses on the passive identification of spillage events.
Deep Learning Methods	YOLOv3	with simple feature fusion and limited multi-scale capabilities.	Struggle to learn key features under adverse conditions, leading to reduced localization accuracy.
	YOLOv5	A powerful general-purpose detector, but lacks specialized modules to address boundary ambiguity.	

Our Goal is achieve proactive detection and intervention for Spillage Risk Vehicles(SRVs)

02.

Methods

2.1 EMSA-Net

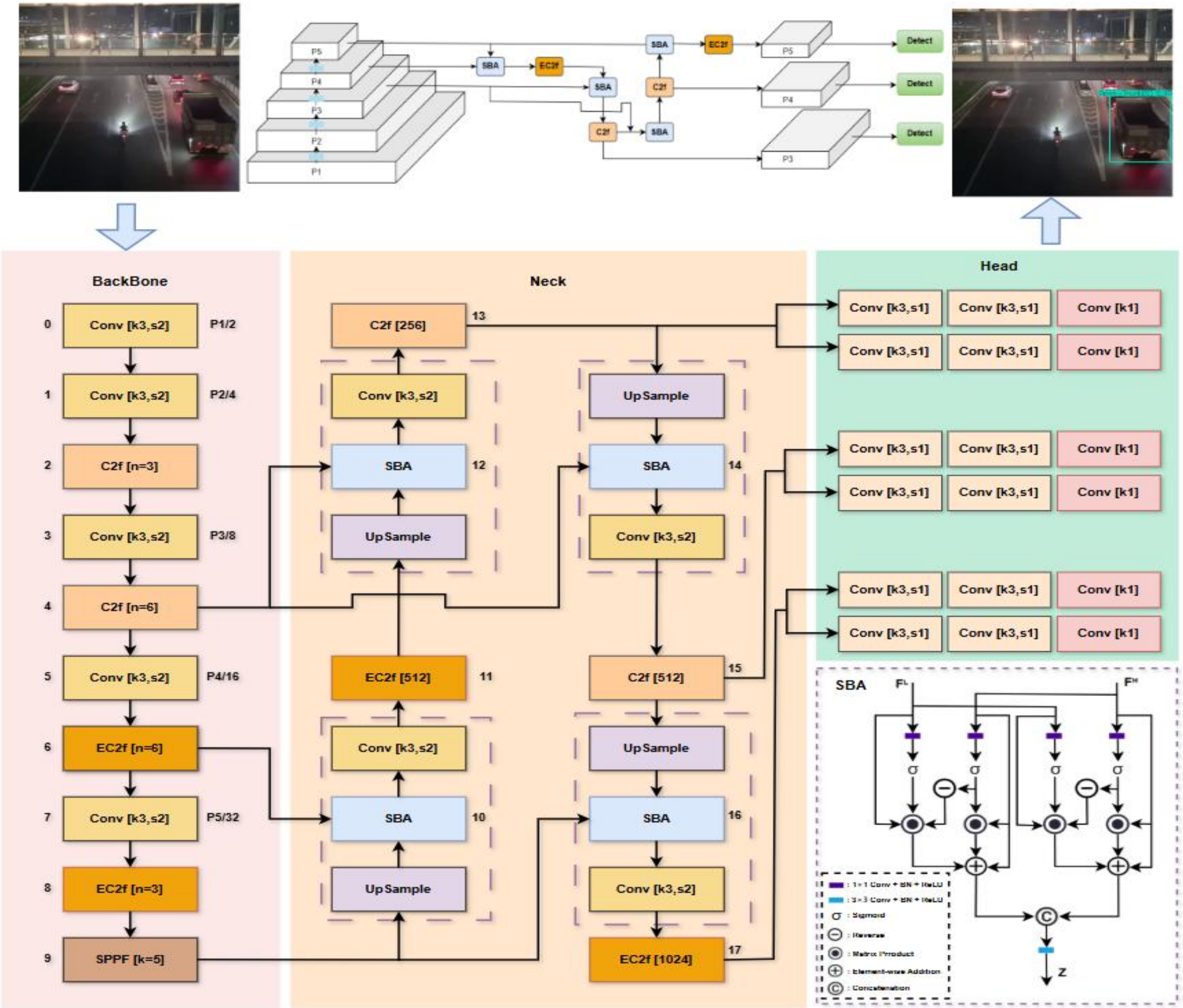
YOLOv8

✚

EMSConv

✚

SBA

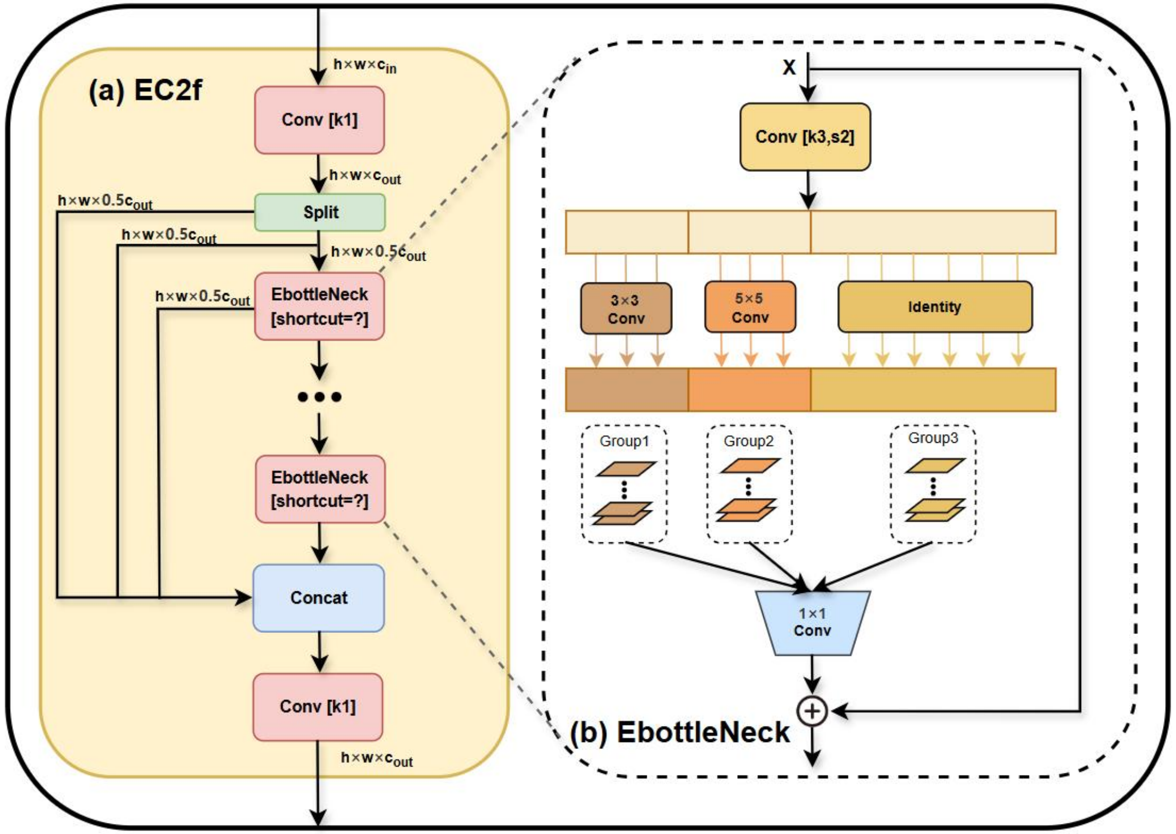


Overall Structure

1、EMSC: Replaces standard convolutions to efficiently extract rich multi-scale features, tackling scale variation.

2、SBA : Integrated into the FPN to perform adaptive feature recalibration, fusing semantics and boundaries to overcome low-light ambiguity.

2.2 EMSConv



EMSC in the backbone for efficient feature extraction

Key Process:

1、Channel Splitting: The input feature is split into an identity part X_{id} and a transformation part X_{trans}.

2、Multi-Branch Transformation: X_{trans} is processed by parallel convolutions to capture diverse features.

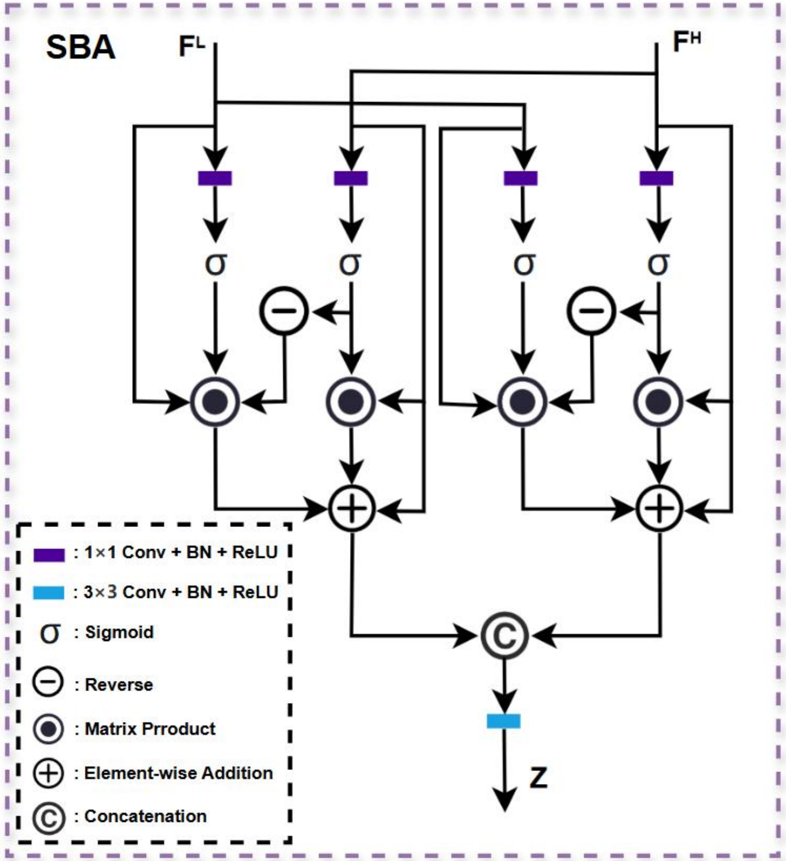
$$T_{trans,1} = Conv_{3\times3}(X_{trans,1}) \mid T_{trans,2} = Conv_{5\times5}(X_{trans,2})$$

3、Feature Aggregation: The identity and transformed features are concatenated and fused by a 1×1 Conv to enable cross-channel interaction.

$$T_{concat} = [X_{id}; T_{trans,1}; T_{trans,2}]$$

To address the drastic scale variation of Spillage Risk Vehicles (SRVs).

2.3 SBA



SBA in the neck for feature recalibration

Key Process:

1、**Semantics Guide Boundaries:** Deep features locate relevant boundaries in shallow features.

$$X_b' = X_b \odot \sigma(Conv_{1\times1}(X_s))$$

2、**Boundaries Sharpen Semantics:** Shallow features provide precise edges to refine coarse semantic maps.

$$X_s' = X_s \odot \sigma(Conv_{1\times1}(X_b))$$

3、**Fuse for Robust Output :** Concatenate and fuse the mutually enhanced features for a result rich in both detail and meaning.

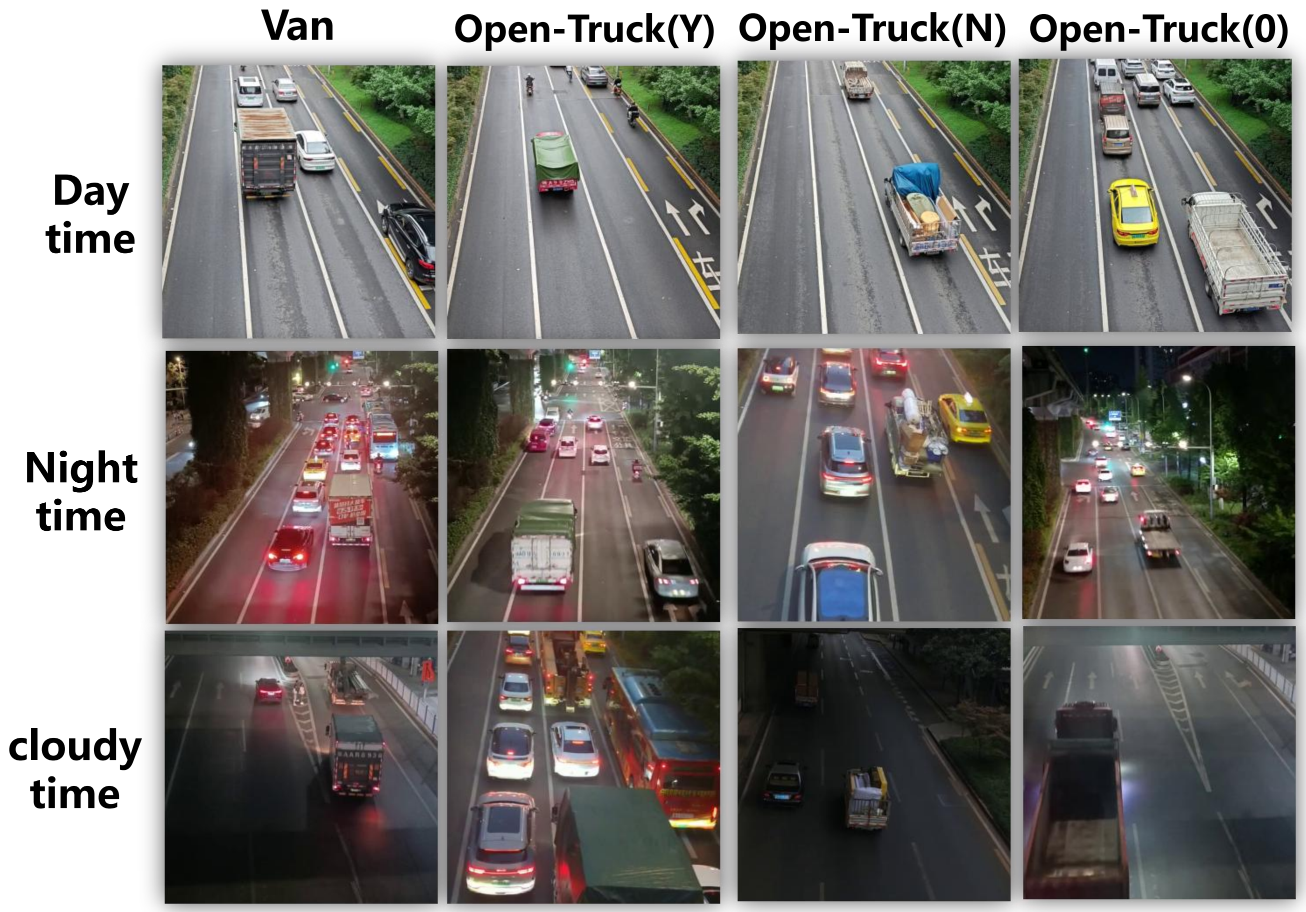
$$Z = ConvBlock([X_s', X_b'])$$

To address the severe boundary ambiguity under low-light conditions.

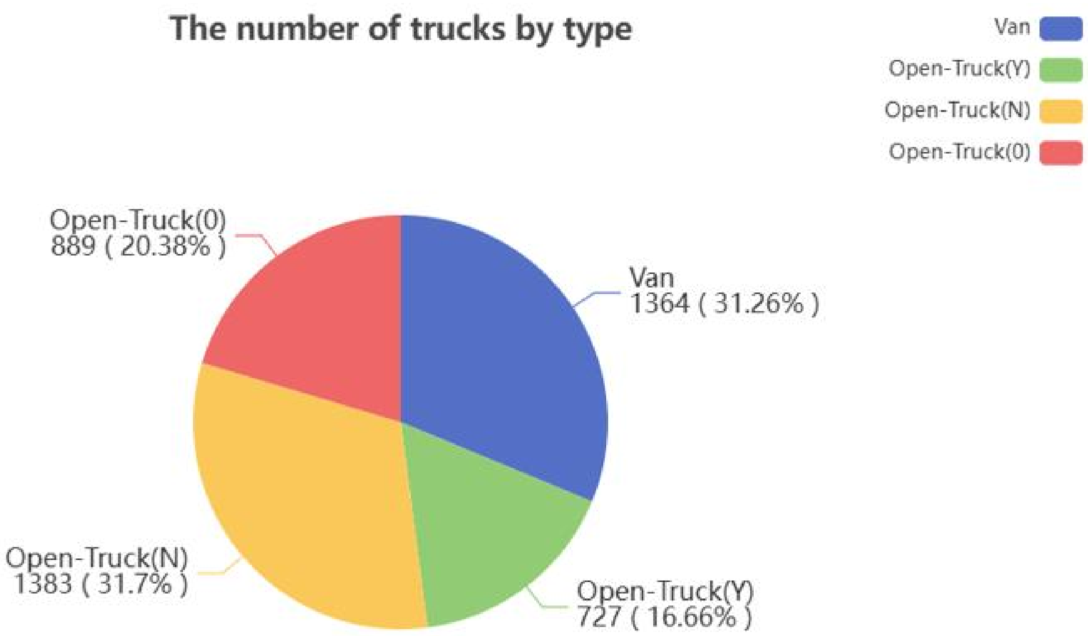
03.

EXPERIMENTS

3.1 UrbanDatasets



① Vehicle Type —— Van / Open-Truck
② Tarpaulin Cover —— Covered / Uncovered
③ Load Status —— Empty / Loaded



Uncovered trucks account for 52.08% of all trucks.

3.2 Comparison with State-of-the-Art

COMPARISON ON ALL TIME SCENE

Method	Recall(%)	F1-Score	mAP@0.5	mAP@0.5:0.95	FLOPs	FPS
YOLOv5	0.662	0.682	0.674	0.507	64.0G	136
YOLOv8	0.652	0.659	0.686	0.540	78.7G	191
YOLOv9	0.650	0.666	0.689	0.524	76.5G	175
YOLOv10	0.652	0.629	0.676	0.517	58.9G	217
YOLOv11	0.629	0.663	0.689	0.524	67.7G	189
YOLOv12	0.713	0.667	0.687	0.523	59.5G	161
RT-DETR	0.692	**0.703**	0.628	0.449	129.6G	97
EMSA-Net (Ours)	**0.715** ↑	0.684	**0.718** ↑	**0.555** ↑	85.6G	92

COMPARISON ON NIGHT TIME SCENE

Method	Recall(%)	F1-Score	mAP@0.5	mAP@0.5:0.95	FLOPs	FPS
YOLOv5	0.608	0.656	0.642	0.480	64.0G	134
YOLOv8	0.685	0.653	0.670	0.538	78.7G	212
YOLOv9	0.581	0.635	0.648	0.510	76.5G	193
YOLOv10	0.638	0.594	0.621	0.486	58.9G	245
YOLOv11	0.590	0.628	0.653	0.508	67.7G	212
YOLOv12	0.608	0.640	0.647	0.499	59.5G	163
RT-DETR	0.714	**0.671**	0.608	0.451	129.6G	92
EMSA-Net (Ours)	**0.724** ↑	0.666	**0.691** ↑	**0.549** ↑	85.6G	82

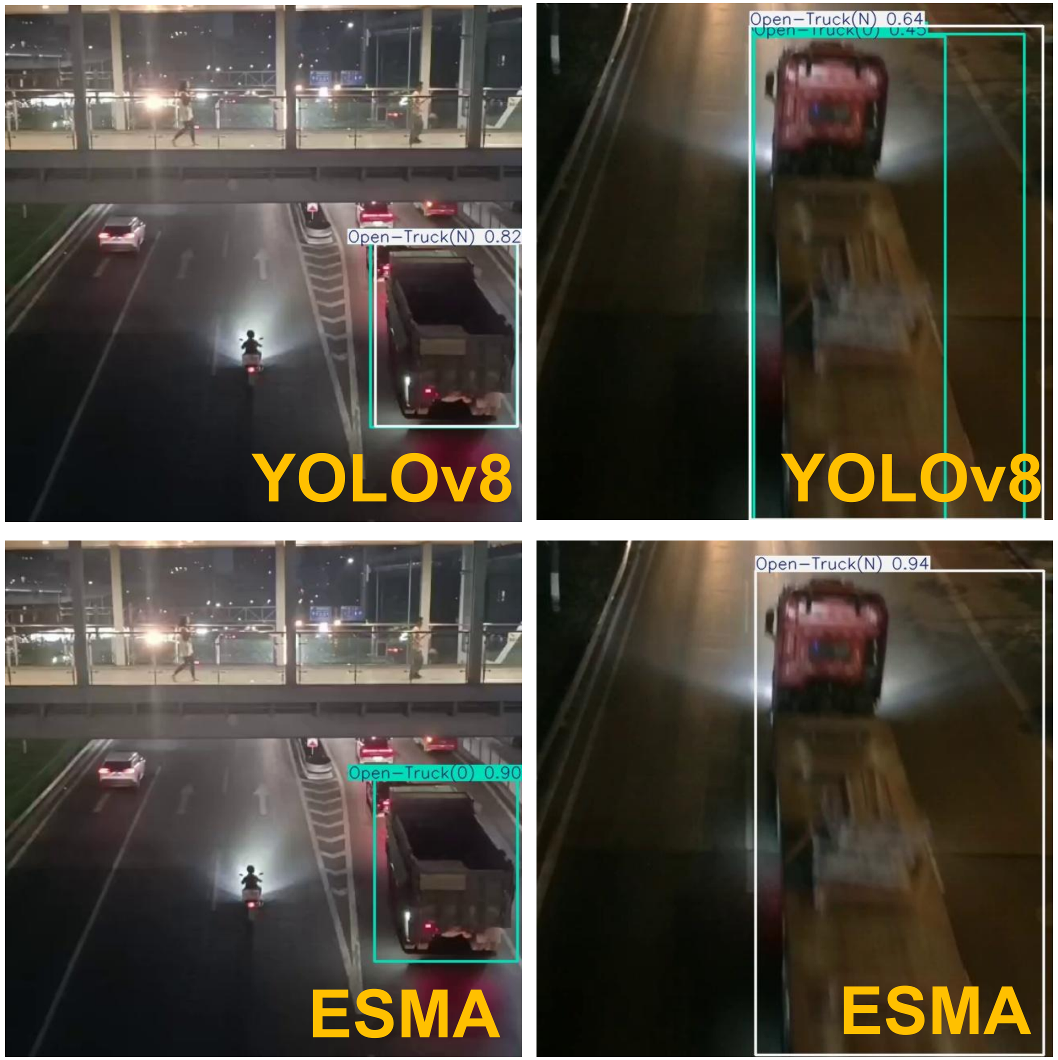
All-time scenarios

- Achieves a SOTA with 71.8% mAP@0.5 (+3.2% vs. YOLOv8).
- Demonstrates superior localization with a 1.5% gain in the stricter mAP@0.5:0.95, highlighting precision.

Night-time scenarios

- Advantage is most pronounced at night, reaching 69.1% mAP@0.5 (+2.1% vs. YOLOv8).
- Demonstrates superior localization with a 1.1% gain in the stricter mAP@0.5:0.95, highlighting precision.

3.3 Visualization Results



Spillage Risk Vehicles Inference Test Results

Scene Description:

A challenging nighttime scenario characterized by:

Low ambient light

Motion blur

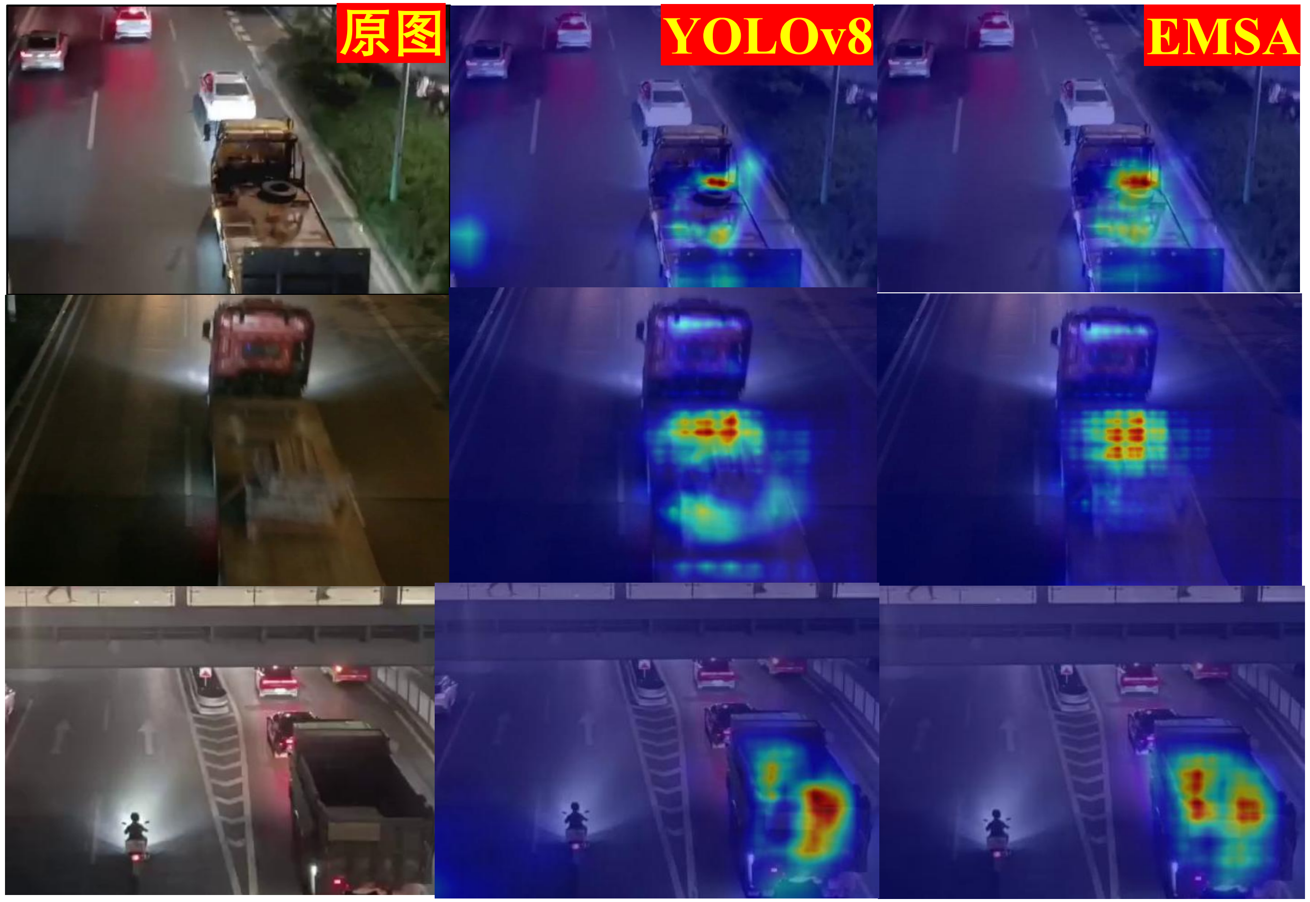
High-contrast glare from headlights

Visualization Analysis:

- **Baseline (YOLOv8):**
  Struggles with the complex conditions.
  Fails to detect the Open-Truck(Y) and misclassifies the Van with very low confidence (0.42).

- **Our Model (EMSA-Net):**
  Demonstrates exceptional robustness.
  Correctly identifies all targets with high confidence scores and tight, accurate bounding boxes.

3.4 Feature Maps Visualization Results



Inference Results of Spillage Risk Vehicle Feature Maps

Visualization Analysis：

- **Baseline (YOLOv8):**
The model's attention is diffuse and scattered.It wastes computational focus on irrelevant background areas, leading to uncertainty and errors.

- **Our Model (EMSA-Net):**
The model learns to focus its resources **directly on the most critical regions for risk assessment—the cargo area and the cargo itself.**

Effectiveness

Interpretability

Achieves more precise focus on spillage risk features.

04.

Conclusions

4.1 Conclusions

EMSA-Net

- A robust framework for Spillage Risk Vehicle detection in complex urban conditions.

Challenge 1 (Drastic Scale Variation):

- Varying distances and target types lead to extreme differences in object scale.

Challenge 2 (Severe Boundary Ambiguity):

- Low-light, glare, and shadows obscure target boundaries, hindering precise localization.

EMSC

Captures rich multi-scale features with high efficiency to handle scale variation.

SBA

Precisely localizes targets in low light by fusing boundary details with semantic information.

Overall Achievements

Achieved SOTA performance (71.8% mAP@0.5 and 55.5% mAP@0.5:0.95) while maintaining 92 FPS.

THANKS

EMSA-Net : An Efficient Multi-Scale and Boundary Aware Network for Spillage Risk Vehicle Detection Under Low-Light Urban Conditions

Author : Ze Xu, Wenting Jia, Chunyang Yan, Ruikuang Wang, Yongfu He*

Chongqing Jiaotong University