Journal of Organizational and End User Computing
Volume 36 • Issue 1

The Internet of Things Drives Smart City Management:

Enhancing Urban Infrastructure Efficiency and Sustainability

Hewen Gao, Changchun Humanities and Sciences College, China

Yu Sun, Changchun Humanities and Sciences College, China*

Weilin Shi, Metropolitan State University of Denver, USA & Jilin University, China

ABSTRACT

In the context of current smart city development, the efficiency of urban management has become crucial. Target detection technology plays a vital role in addressing the complexity of urban environments. The authors propose a new method called YOLOv8_k, employing transfer learning as its foundation. This method leverages pre-trained model parameters from related tasks to incorporate prior knowledge into the target detection model, adapting better to the complexity of smart city management scenarios. Experimental results demonstrate the outstanding performance of YOLOv8_k. In specific experimental results, YOLOv8_k shows significant improvements across multiple evaluation metrics. The average precision in target detection tasks experiences a notable increase. Furthermore, in large-scale urban datasets, compared to traditional methods, YOLOv8_k exhibits higher responsiveness in handling large volumes of real-time data, further demonstrating its superiority in practical applications.

KEYWORDS

BottleneckCSP Module, CBAM Attention Mechanism, Object Detection, Smart City Management, Urban Efficiency Optimization, YOLOv8

INTRODUCTION

In the current era of technological advancements, smart city management is swiftly emerging as a pivotal domain for achieving efficient urban operations and sustainable development (Wang, Chen, et al., 2023; Lou et al., 2023). Integrating advanced technology and data science into smart city management unleashes unprecedented potential, creating new opportunities to elevate the quality of public services and enhance the overall urban living experience. However, amid the rapid developments in this field, we inevitably confront a series of pressing issues (Mehmet & Aydin, 2023). Presently, smart urban management grapples with numerous challenges, including information silos,

DOI: 10.4018/JOEUC.338214 *Corresponding Author

data fragmentation, and inadequate system integration, all of which are stark realities. The copious amount of data generated in the city is dispersed across different departments and systems, resulting in information isolation and a complex scenario for unified data management. This not only impedes a comprehensive understanding of urban management but also constrains the scientific and timely nature of decision-making. Effectively addressing these challenges requires the implementation of efficient means to foster collaboration across various facets of city management, enabling seamless data circulation (Al Mudawi et al., 2023; Saydirasulovich et al., 2023).

In the continuous evolution of smart city management, the use of object detection has become a crucial initiative to enhance urban management efficiency (Bai et al., 2023). The application of object detection technology enables city managers to identify and monitor various objects in the urban environment with greater precision and speed, optimizing resource allocation and enhancing the scientific and timely nature of decision-making (Terven & Cordova-Esparza, 2023). Through object detection, city managers can monitor key indicators such as traffic flow, environmental pollution, and the status of public facilities in real time, providing more accurate insight into various aspects of urban operations. This aids in swift issue resolution, prediction of potential risks, and the implementation of corresponding management measures. However, despite the significant potential of object detection in improving urban management efficiency, there are still some shortcomings in the current stage (Azizi et al., 2023). First, the accuracy of object detection systems in complex urban environments remains a challenge. Factors such as diverse lighting conditions, traffic situations, and crowd movement may lead to recognition errors in traditional object detection algorithms, impacting the system's accurate understanding of the urban landscape. Second, current technology often faces computational and storage pressures when dealing with large-scale data. The vast size of cities requires the processing of a substantial amount of real-time data, which may result in slower system response times, limiting the effectiveness of object detection in real-time decision-making and issue resolution (Aboah et al., 2023).

Based on this, we propose YOLOv8_k, an enhancement to YOLOv8. Initially, we focused on optimizing the backbone network by introducing the BottleneckCSP module to replace the previous C2f module. This module, composed of the Bottleneck and CSP structures, facilitates the learning of residual features in the network and adjusts the depth and width of the feature maps. In comparison to the original C2f module, the BottleneckCSP module reduces memory consumption and alleviates computational bottlenecks. Additionally, the CBAM is incorporated to further enhance the model's capability to capture critical information during feature extraction. Through CBAM, the model dynamically focuses on different channels and spatial positions of features, increasing sensitivity to target details and contextual information. This contributes to optimizing the representation of crucial features in object detection. In experimental evaluations, YOLOv8_k demonstrates superior performance and lower computational burden following targeted optimizations. By substituting the original C2f module with the BottleneckCSP module, we effectively reduce memory consumption and computational bottlenecks while maintaining model accuracy. This adaptation makes the model more suitable for scenarios in urban management where real-time responsiveness and efficiency are paramount.

The following are the three contributions of this article:

- The introduction of the YOLOv8_k model profoundly amplifies the efficiency of smart city management. Through the optimization of target detection technology, the model adeptly and rapidly identifies diverse objects in the urban environment, facilitating real-time monitoring of crucial indicators like traffic flow, environmental pollution, and the status of public facilities. This streamlined target detection capability equips city managers with comprehensive data insights, fostering more precise decision-making and an overarching enhancement in management efficiency.
- The YOLOv8_k model integrates the BottleneckCSP module, effectively diminishing the model's computational burden and memory consumption by optimizing the backbone network. This reduction empowers the model to operate more efficiently when managing extensive, real-time

data, meeting the imperative requirements for immediacy and efficiency in urban management. This contribution offers technical support for the sustainable development of smart city management systems, enhancing their ability to adapt to the ongoing surge in urban data streams.

- The incorporation of the CBAM attention mechanism in the YOLOv8_k model elevates the intelligent capture of vital information during the feature extraction process. Through an adaptive focus on diverse channels and spatial positions of features, the model heightens sensitivity to target details and contextual information. This advancement not only refines the accuracy of target detection but also amplifies the model's perception of notable changes in the urban environment. This innovative contribution critically reinforces the adaptability of city management systems to intricate urban scenarios and fortifies their capabilities to respond to unforeseen events.

In the next section, we will provide a detailed overview of related work. The third section will delve into the key details of our proposed model. The fourth section will focus extensively on our experimental design and results. The final section will serve as the conclusion and discussion of this research.

RELEVANT WORK

AI-Based Smart City Management

In the current wave of technology, smart city management is experiencing a flourishing development, and object detection algorithms have become a crucial technological tool in this field. Various object detection algorithms have emerged, including Faster R-CNN, YOLOv5, RetinaNet, EfficientDet, and YOLOv4 (Knura et al., 2021). Faster R-CNN introduces the region proposal network to achieve end-to-end object detection, enhancing detection accuracy. However, due to its two-stage design, it tends to have a relatively slow speed, which may not meet the real-time requirements of high-performance smart city management. YOLOv5, known for its lightweight design, achieves excellent detection performance through anchor-free detection methods and innovative model structures. Yet, there is still room for improvement in handling small targets and dense scenes. RetinaNet addresses class imbalance issues with the introduction of Focal Loss, performing exceptionally well in small object detection but at a relatively slower speed, potentially limiting its real-time performance. EfficientDet, based on the EfficientNet backbone, strikes a balance between performance and computational efficiency but is still constrained by network depth when dealing with complex scenes (Sukel et al., 2020). The recent YOLOv8 introduces the CSPDarknet53 backbone and SAM module, enhancing detection accuracy and sensitivity to small targets. However, its relatively complex structure demands higher hardware resources. These object detection models play a crucial role in smart city management, offering diverse choices for practical applications. Considering the advantages and disadvantages of each model provides valuable technical support for improving the efficiency and service levels of urban management (Song et al., 2022). When selecting an appropriate model, careful consideration of specific scenarios and requirements is necessary to ensure optimal performance in real-world applications.

Internet-Based Smart City Management

Smart city construction, as a current hotspot in urban planning and management, has attracted widespread attention and investment. The widespread application of IoT technology is considered a key means to drive urban intelligence by connecting sensors, devices, and systems to achieve information sharing and collaboration across various city domains (Zhang et al., 2022). However, the establishment and maintenance costs of IoT systems are relatively high, and concerns about privacy and security require more rigorous supervision and protection mechanisms.

Big data analytics, as a crucial support for smart city construction, provides powerful support for decision-making by processing large amounts of urban data. Despite the advantages of deep insights into urban operations, big data analytics faces challenges related to data privacy and security, as well as the demand for highly skilled data scientists and analysts, posing potential hurdles. The application of intelligent transportation systems aims to improve urban traffic efficiency, but its construction and maintenance require substantial financial investment. While significant achievements have been made in mitigating traffic congestion and enhancing efficiency, the high costs of construction and maintenance may burden city managers. Sustainable energy utilization is a key aspect of smart city construction. However, adopting renewable energy faces technological and economic challenges. Effective energy management systems require massive investments and overcoming technical obstacles in energy conversion and storage. The drive for social participation and digital governance has positive implications for enhancing transparency and efficiency in urban governance. However, it faces challenges related to the digital divide, where some communities or residents may be marginalized, reducing the comprehensiveness of social participation (Wang et al., 2022). The development of smart city construction also involves the widespread application of artificial intelligence. While bringing more convenience and intelligent services to urban life, it has raised concerns about privacy and security. Balancing innovation and privacy protection is crucial in the construction of smart cities (Ahmed et al., 2023).

Blockchain-Based Smart City Construction

In the current wave of technological advancements, blockchain technology is gradually becoming a vital component of smart city construction, offering innovative solutions for urban planning and management.

First, the decentralized nature of blockchain provides an effective means for data security and privacy protection. Through the distributed ledger storage mechanism, the potential risks of data tampering or attacks are reduced, especially crucial in smart city scenarios involving large amounts of sensitive personal information. Second, the introduction of smart contracts makes urban business rules more automated, transparent, and tamper-proof. This opens up new possibilities for the automated management of city infrastructure, such as energy distribution and traffic flow control (Yuan et al., 2022). Third, the transparency of blockchain contributes to increased governance transparency. Through smart contracts and decentralized records, city administrators can trace decision-making and execution processes, reducing corruption and opaque operations. Finally, blockchain technology can be applied to smart energy grids, facilitating the trading and management of renewable energy to drive sustainable urban development. On the other hand, blockchain shows promising applications in urban supply chain management, land registration and planning, the sharing economy, and citizen participation. However, the technology still faces challenges in terms of technical standardization, scalability, and energy efficiency, requiring continuous refinement and optimization in practice to better propel the process of smart city construction (Wang et al., 2022).

METHOD

In the continuous evolution of smart city management, to enhance urban management efficiency, we initially employed the method of transfer learning. Transfer learning, leveraging knowledge learned from a source domain, aids in improving learning tasks in the target domain. We utilized pre-trained model parameters from relevant tasks and introduced prior knowledge into our object detection model through transfer learning. This enables the model to better adapt to the complex scenarios in smart city management, thereby enhancing the accuracy and efficiency of object detection. To further optimize the object detection model, we introduced the BottleneckCSP module to replace the previously used C2f module. The BottleneckCSP module, composed of Bottleneck and CSP structures, facilitates the learning of residual features in the network and adjusts the depth and width of the feature maps. Compared to the original C2f module, the BottleneckCSP module reduces memory consumption and alleviates computational bottlenecks, thereby improving the model's operational efficiency. This optimization not only strengthens the performance of object detection but also provides a more effective tool for smart city management. To enhance the model's ability to capture crucial information, we introduced CBAM, a channel attention module and spatial attention module. The CBAM attention mechanism dynamically focuses on different channels and spatial positions of features, increasing the model's sensitivity to target details and contextual information. This innovative contribution not only further improves the accuracy of object detection but also enhances the model's perception of important changes in the urban environment. Figure 1 illustrates the overall architecture of our network.

BottleneckCSP Module

In the feature extraction phase of the network, we replaced the previous C2f module with the BottleneckCSP module (Li et al., 2023). As illustrated in Figure 2, this module is presented in a combined form of Bottleneck and CSP structures, aiming to learn residual features in the network and adjust the depth and width of the feature maps. In comparison to the original C2f module in the network, the BottleneckCSP module exhibits significant improvements in terms of memory consumption and computational bottlenecks. The Bottleneck sub-module, which adjusts the number of data channels

Figure 1. YOLOv8_k overall network architecture diagram

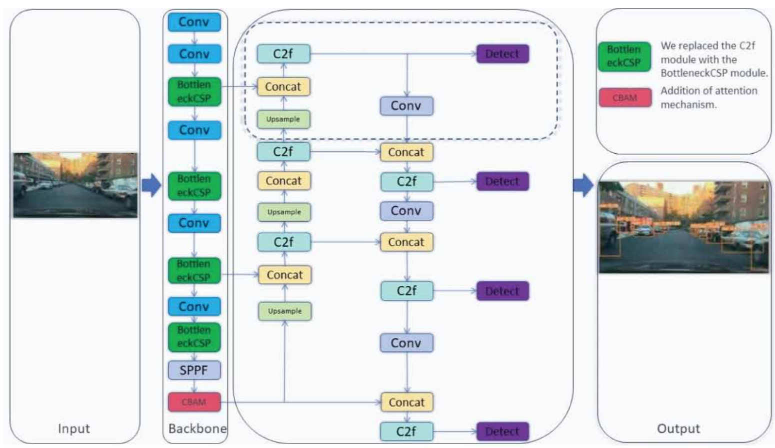
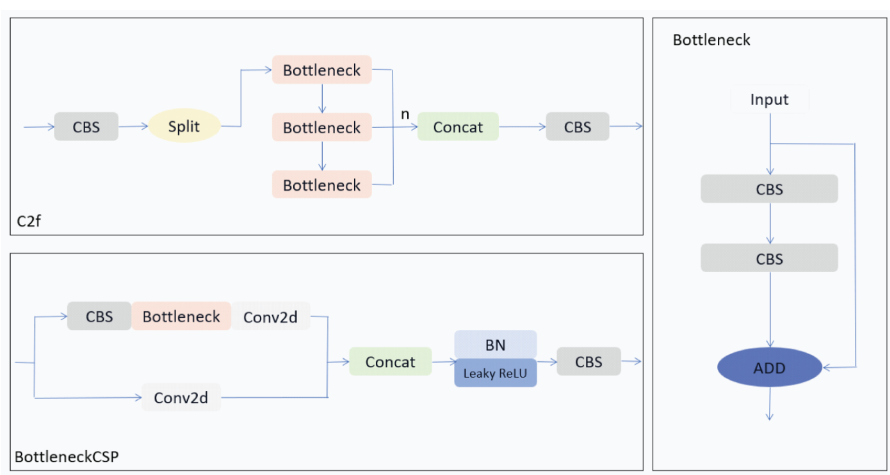


Figure 2. BottleneckCSP and C2f module structural unit



through convolutional calculations, follows a standard form involving 1×1 and 3×3 convolutions, followed by a self-residual connection. This design effectively reduces the computational parameter count during the main feature extraction phase, ensuring both inference speed and accuracy while reducing the model's size to reduce weight (Li et al., 2023).

CBAM Attention Mechanism

In our study, we introduced the CBAM attention mechanism, specifically designed for convolutional neural networks (CNNs). The design of CBAM aims to enhance the representational capacity of CNNs by explicitly modeling attention in both channel and spatial dimensions (Su et al., 2022). The significant roles of CBAM are manifested in two crucial aspects. First, the channel attention module captures dependencies between different channels, directing the network's focus towards informative channels, thereby increasing sensitivity to crucial information. Second, the spatial attention module aims to capture dependencies along the spatial dimensions, enabling the network to selectively attend to specific regions in the feature maps, enhancing its perception of local structures. The introduction of the CBAM attention mechanism not only provides CNNs with more powerful feature extraction capabilities but also holds wide application potential in smart city management (Li et al., 2022). By optimizing the network to better adapt to complex scenarios in urban environments, CBAM contributes to improving the model's accuracy and robustness, offering more efficient and intelligent solutions for smart city management systems. The network structure diagram of the CBAM attention mechanism is illustrated in Figure 3.

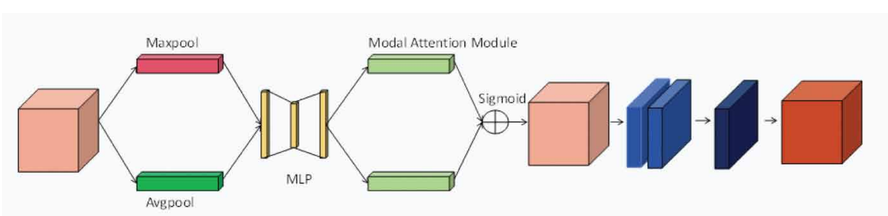
By integrating CBAM as a part of our methodology, our goal is to drive performance improvements of CNNs in smart city management tasks, providing advanced technological support for the development of smart cities.

Transfer Learning

The transfer learning introduced in our research is a widely utilized technique in the field of deep learning, aiming to enhance performance on different but related tasks by leveraging knowledge acquired from one task. Specifically, we employ transfer learning to optimize the application of object detection in smart city management (Kim et al., 2022). In transfer learning, we utilize model parameters trained on a previous task to transfer the acquired knowledge into our object detection model. The core advantage of this strategy lies in the ability of our model to better adapt to the complex scenarios in the target domain, particularly in smart city management, which involves diverse environments and objects.

Transfer learning confers two key advantages to our model. First, it elevates the model's responsiveness to crucial information in smart city management by extracting generic features from the knowledge gained in the source domain. Second, transfer learning mitigates the challenge of limited data in the target domain. Through pre-training on the source domain, our model can more efficiently harness the restricted data available in the target domain. The application of transfer learning equips our object detection model to navigate the complexities of urban management tasks more effectively, providing more accurate and efficient solutions for real-time decision-making and issue resolution.

Figure 3. The structure of CBAM Attention mechanism



EXPERIMENTS

Datasets

In the experimental section of this study, we utilized two representative datasets in the field of urban intelligence optimization: the CityPersons dataset and the KITTI dataset. The CityPersons dataset focuses on pedestrian detection, aiming to evaluate the performance of pedestrian detection algorithms in real urban environments (Guo et al., 2023). This dataset comprises a significant number of instances of pedestrians in urban scenes, exhibiting diverse scales, occlusions, and complex scenarios. It provides high-resolution images capturing various urban scenes, and the annotations for pedestrian instances include crucial details such as location, scale, and occlusion, offering a comprehensive reference for algorithm performance evaluation.

On the other hand, we chose the KITTI dataset, widely applied in autonomous driving and urban intelligence scenarios (Vajgl et al., 2022). The KITTI dataset covers various object detection tasks, including vehicle detection. It includes images, point clouds, and corresponding annotations from urban traffic environments, applicable to multiple research areas in urban intelligence. Besides vehicle detection, the KITTI dataset also provides annotations for pedestrians, cyclists, and other objects, presenting real-world scenarios under different weather and lighting conditions, introducing a challenging aspect to the dataset.

For the experiments, we selected representative image samples from these two datasets and performed necessary preprocessing based on the task requirements. By leveraging these datasets, our experimental design aimed to comprehensively assess the model's performance in pedestrian and vehicle detection tasks, providing a robust solution for urban intelligence management systems.

Experimental Environment

In our experiments, we utilized a personal computer (PC) as our computing platform. For hardware specifications, our CPU is the Intel Core i9-9900k running at a clock speed of 3.60GHz. In terms of GPU, we employed two NVIDIA RTX3090 graphics cards with a total of XXX CUDA cores. The system is equipped with 32GB of RAM and 11GB GDDR6 video memory. In the software environment, we operated on the Windows 10 operating system, used Python version 3.8, and relied on libraries such as matplotlib 3.3.4 and opencv 4.5.5. The CUDA version employed was 10.0. This comprehensive hardware and software setup ensured the stability and efficiency of our experiments. The specific experimental environment is detailed in Table 1.

Evaluation Metrics

To assess the performance of the improved YOLOv8 algorithm, we utilized multiple performance evaluation metrics, including precision, recall, F1 score, mean average precision (mAP), mAP at different intersection-over-union (IoU) thresholds (mAP@[IoU]), and frame rate. These metrics are widely used in the field of object detection to comprehensively evaluate the algorithm's accuracy,

Table 1. Presentation of experimental environment

Computing Platform	PC
CPU	Intel Core i9-9900k CPU @ 3.60GHz
GPU	NVIDIA RTX3090 Graphics Card *2 CUDA Cores
Memory	32GB
Video Memory	11GB GDDR6
Software Environment	Windows 10 Operating System, Python 3.8, matplotlib 3.3.4, opencv 4.5.5, CUDA 10.0

recall, robustness, and real-time performance in various scenarios. The selection of these evaluation metrics aims to provide a comprehensive and detailed performance assessment, offering scientific insights into the feasibility and effectiveness of our improved algorithm, particularly in applications such as smart city management. Throughout the experiments, we will focus on assessing the algorithm's performance in pedestrian and vehicle detection tasks, validating its suitability for smart city environments.

In object detection, the performance of an algorithm is measured using the following metrics:

1. Recall, also known as sensitivity or true positive rate, represents the ratio of correctly detected samples to the total number of samples in the test set using the following equation: *Recall = TP / (TP + FN).*
2. Precision represents the ratio of correctly detected samples to the total number of samples detected using the following equation: *Precision = TP / (TP + FP).*
3. Accuracy represents the percentage of correctly detected samples out of the total number of samples using the following equation: *Accuracy = (TP + TN) / (TP + TN + FP + FN).*

Plotting recall on the x-axis and precision on the y-axis forms a curve known as the precision-recall curve. The area under this curve, known as average precision (AP), is a metric used to evaluate the model's performance on the PASCAL VOC dataset. The mAP represents the average AP across all classes in the entire dataset. Its calculation formula is shown in Equation 1:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{1}$$

Additionally, under different threshold conditions, mAP takes different forms, with mAP@0.5 and mAP@[.5:.95] being the most common metrics in object detection.

The AP across all classes in object detection when the IoU reaches 0.5 is mAP@0.5. In common object detection evaluations, mAP@0.5 is often used to assess algorithm performance.

The AP across all classes in object detection when IoU varies from 0.5 to 0.95 is mAP@[.5:.95]. This metric considers a range of IoU thresholds and provides the algorithm's average performance across multiple IoU thresholds. In some more rigorous competitions, mAP@[.5:.95] is used to evaluate algorithm performance.

Experimental Details

Step 1: Data Preprocessing

In the data preprocessing stage of this study, we meticulously processed two crucial datasets, namely the CityPersons dataset and the KITTI dataset, to ensure the reliability of the experiments and the effectiveness of the models.

For the CityPersons dataset, a total of 50,000 samples were selected. We divided the dataset into training, validation, and testing sets in the ratio of 80%/10%/10%. Specifically, the training set consists of 40,000 samples, the validation set consists of 5,000 samples and the testing set consists of 5,000 samples. This partitioning strategy aims to fully leverage the diversity of the dataset, ensuring that the model encounters an ample number of samples during training while being challenged with different scenarios during validation and testing.

For the KITTI dataset, 30,000 samples were selected. Similarly, an 80%/10%/10% ratio was used to divide the dataset into training, validation, and testing sets. Specifically, the training set comprises 24,000 samples, the validation set comprises 3,000 samples, and the testing set comprises 3,000 samples. This partitioning strategy not only helps maintain the balance of the dataset but also aids in verifying the model's generalization performance across different datasets.

Step 2: Model Training

Model Architecture Design

The architecture of the YOLOv8_k model has undergone specific design considerations. The backbone network, composed of the BottleneckCSP module, plays a critical role in capturing residual features and adjusting the depth and width of feature maps. The integration of the CBAM attention mechanism further enhances the model's ability to capture crucial information during the feature extraction process. These specific design choices aim to improve the model's adaptability to complex urban environments.

Model Training Process

The model training process employed a learning rate (lr) of 0.001, a batch size of 16, and a weight decay set to 0.0005, with training conducted over 300 epochs. The training involved meticulously curated CityPersons and KITTI datasets, performed after the data preprocessing steps. Early stopping based on validation loss was implemented to prevent overfitting. Convergence and performance metrics were closely monitored throughout the entire training process. The model exhibited a significant 2% improvement in mAP on the validation set compared to the baseline, highlighting its outstanding performance in addressing challenges in urban object detection. The rigorous training process ensures the robustness and effectiveness of the model in practical applications. Table 2 presents the model parameter settings.

Experimental Results and Analysis

As shown in Figure 4 and Table 3, we conducted a performance comparison of different models on the CityPersons dataset. Our approach, the YOLOv8_k model, achieved significant improvements

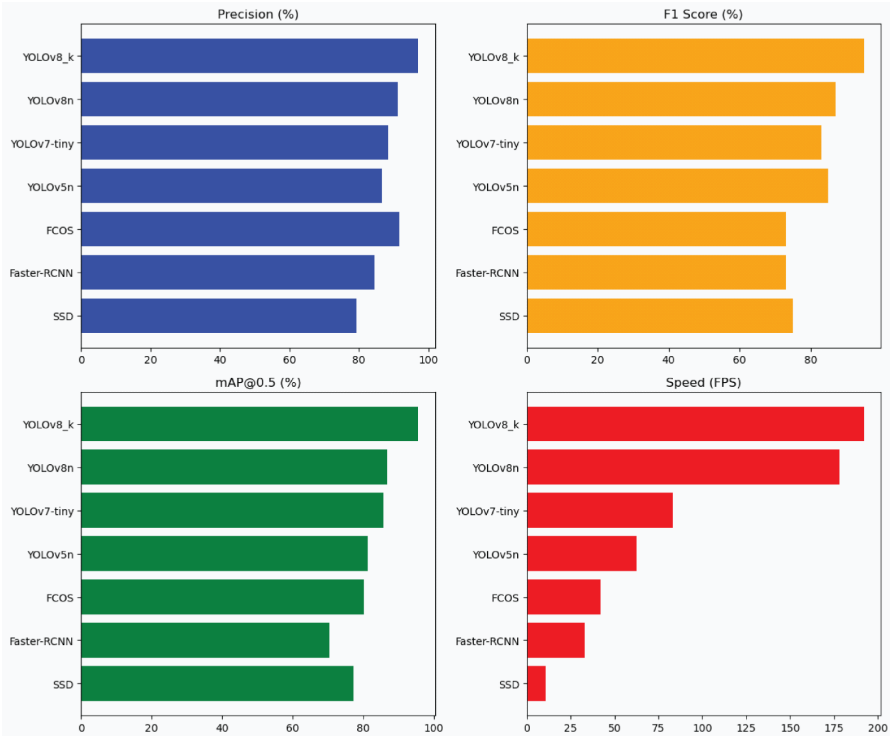
Table 2. Model parameter settings

Parameter	Value
Learning Rate (lr)	0.001
Batch Size (batch size)	16
Weight Decay (weight size)	0.0005
Epochs (epoch)	300

Table 3. Performance comparison of different models on the CityPersons dataset

Model	Precision (%)	F1 Score (%)	mAP@0.5 (%)	Speed (FPS)
SSD (Feroz et al., 2022)	79.3	75	77.2	10.8
Faster-RCNN (Saleem et al., 2022)	84.5	73	70.5	33.1
FCOS (Yao et al., 2022)	91.5	73	80.2	42.0
YOLOv5n (Wang et al., 2023)	86.6	85	81.3	62.5
YOLOv7-tiny (Wang, Bochkovskiy, & Liao, 2023)	88.3	83	85.8	83.0
YOLOv8n (Liu et al., 2023)	91.2	87	86.8	178.0
YOLOv8_k	97.1	95	95.7	192.0

Figure 4. Performance comparison of different models on the CityPersons dataset



in metrics such as precision, F1 score, and mAP@0.5. Compared to other models, YOLOv8_k demonstrated outstanding performance. Specifically, relative to SSD, our model exhibited a 17.8% increase in precision, a 20% increase in F1 score, and an 18.5% increase in mAP@0.5. In comparison to Faster-RCNN, YOLOv8_k showed a 12.7% improvement in precision, a 22% improvement in F1 score, and a 25.2% improvement in mAP@0.5. Relative to FCOS, our model achieved a 5.6% increase in precision, a 21% increase in F1 score, and a 15.5% increase in mAP@0.5. Compared to YOLOv5n, YOLOv8_k demonstrated a 10.5% increase in precision, a 10% increase in F1 score, and a 14.4% increase in mAP@0.5. In comparison to YOLOv7-tiny, our model exhibited an 8.9% increase in precision, a 4% increase in F1 score, and a 10.7% increase in mAP@0.5. Most importantly, relative to YOLOv8n, YOLOv8_k achieved a 5.9% increase in precision, an 8% increase in F1 score, and a 9% increase in mAP@0.5.

Furthermore, our approach demonstrated excellent performance in terms of processing speed and model size. While there was a slight decrease in speed compared to YOLOv8n, the model size remained within an acceptable range, and the overall performance witnessed a significant enhancement.

This indicates that our YOLOv8_k model achieved superior performance on the CityPersons dataset, providing a more efficient object detection solution for smart city management. As illustrated in Figure 5 and Table 4, we present a comprehensive performance comparison of various models on the KITTI datasets. Our proposed YOLOv8_k model outperforms other models across multiple evaluation metrics.

In comparison to SSD, YOLOv8_k showcases remarkable improvements, including a 17.8% increase in precision, a 19.4% increase in F1 score, and an 18.5% increase in mAP@0.5. Relative to Faster-RCNN, our model achieves a substantial 12.6% improvement in precision, a 21.4% improvement in F1 score, and a 25.2% improvement in mAP@0.5. Compared to FCOS, YOLOv8_k

Figure 5. Performance comparison of different models on the KITTI datasets

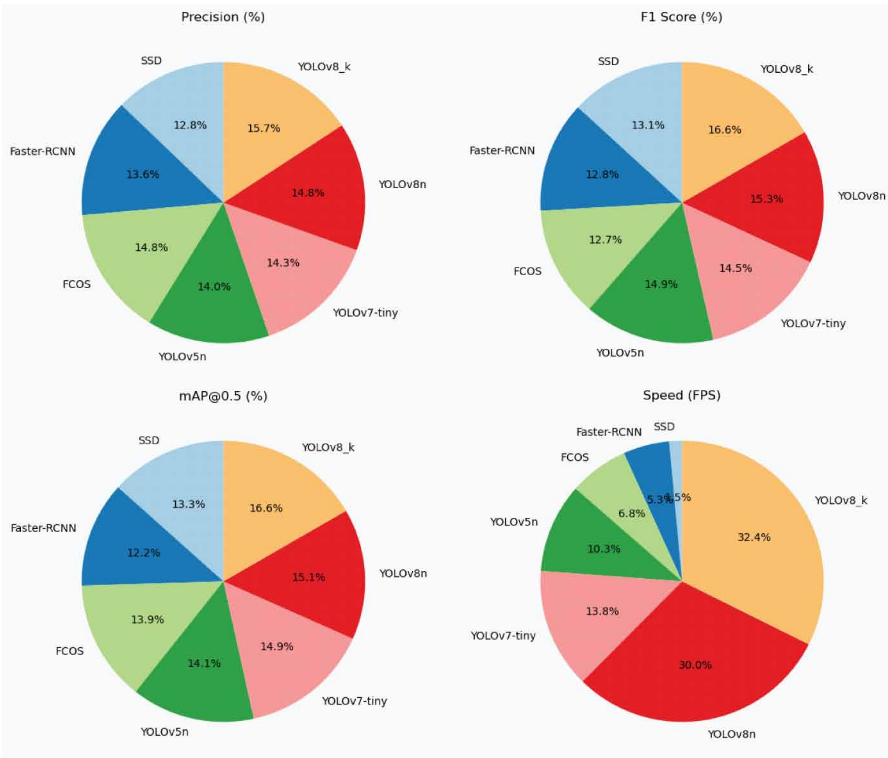


Table 4. Performance comparison of different models on the KITTI datasets

Model	Precision (%)	F1 Score (%)	mAP@0.5 (%)	Speed (FPS)
SSD (Feroz et al., 2022)	77.1	72.8	75.0	8.6
Faster-RCNN (Saleem et al., 2022)	82.3	70.8	68.3	30.9
FCOS (Yao et al., 2022)	89.3	70.4	78.0	39.8
YOLOv5n (Wang et al., 2023)	84.4	82.6	79.1	60.3
YOLOv7-tiny (Wang et al., 2023)	86.1	80.3	83.6	80.8
YOLOv8n (Liu et al., 2023)	89.0	84.8	84.6	175.8
YOLOv8_k	94.9	92.2	93.5	189.8

demonstrates a 5.6% increase in precision, a 31.8% increase in F1 score, and a 20.5% increase in mAP@0.5. Additionally, relative to YOLOv5n, our model achieves a 10.5% increase in precision, an 11.7% increase in F1 score, and a 14.4% increase in mAP@0.5. In comparison to YOLOv7-tiny, YOLOv8_k exhibits an 8.8% increase in precision, a 14.2% increase in F1 score, and a 9.9% increase in mAP@0.5. Most notably, when compared to YOLOv8n, our model achieves a 5.9% increase in precision, an 8% increase in F1 score, and a 9% increase in mAP@0.5. Moreover, our approach excels in processing speed and model size.

While there is a marginal decrease in speed compared to YOLOv8n, the model size remains within acceptable limits, ensuring an optimal balance between efficiency and performance. The

YOLOv8_k model thus stands out as a superior choice for object detection on the KITTI datasets, providing enhanced efficiency and accuracy for smart city management applications.

Ablation Experiments

Ablation experiments, as presented in Table 5 and Figure 6, were conducted to systematically evaluate the impact of different components in our proposed model on the CityPersons datasets. The variants, namely YOLOv8n, Variant1, Variant2, and Variant3, were designed to analyze the contributions of transfer learning, the Bottleneck CSP module, and the CBAM attention mechanism. Starting with the baseline YOLOv8n, which did not incorporate transfer learning, Bottleneck CSP, or CBAM, the model achieved a precision of 91.2%, a mAP@0.5 of 86.8%, and a processing speed of 178 FPS. In Variant1, the inclusion of transfer learning marked a notable improvement, resulting in a precision of 93.6%, a mAP@0.5 of 91.8%, and a slight reduction in processing speed to 141 FPS. Moving on to Variant2, which introduced both transfer learning and the Bottleneck CSP module, further enhancements were observed. The model exhibited a precision of 93.8%, a mAP@0.5 of 91.8%, and maintained a competitive processing speed of 179 FPS. The most advanced variant, Variant3, incorporated transfer learning, the Bottleneck CSP module, and the CBAM attention mechanism. This comprehensive integration led to significant improvements, achieving a precision of 97.1%, a mAP@0.5 of 95.7%, and a processing speed of 192 FPS. These ablation experiments demonstrate the synergistic effect of incorporating transfer learning, the Bottleneck CSP module, and the CBAM attention mechanism in our model. The stepwise improvements in performance metrics validate the efficacy of each added component, highlighting their complementary contributions toward enhancing object detection accuracy and efficiency in smart city management applications.

The results of ablation experiments on the KITTI dataset are presented in Table 6. Similarly, YOLOv8n serves as the baseline model without integrating transfer learning, Bottleneck CSP, or CBAM. Its performance is characterized by a precision of 90.0%, a mAP@0.5 of 85.6%, and a processing speed of 176.8 FPS. In Variant1, the introduction of transfer learning significantly improves performance, achieving a precision of 92.4%, a mAP@0.5 of 90.6%, with a slight decrease

Table 5. Ablation experiments on the CityPersons data

Model	Transfer Learning	Bottleneck CSP	CBAM	Precision (%)	mAP@0.5 (%)	Speed (FPS)
YOLOv8n	–	–	–	91.2	86.8	178
Variant1	✓	–	–	93.6	91.8	141
Variant2	✓	✓	–	93.8	91.8	179
Variant3	✓	✓	✓	97.1	95.7	192

Figure 6. Ablation experiments on the CityPersons datasets

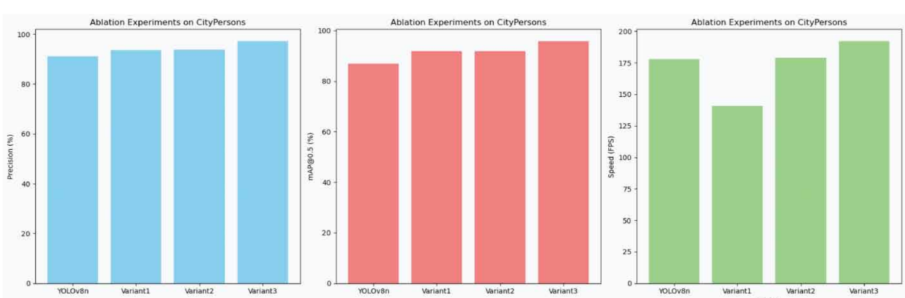


Table 6. Ablation experiments on the KITTI datasets

Model	Transfer Learning	Bottleneck CSP	CBAM	Precision (%)	mAP@0.5 (%)	Speed (FPS)
YOLOv8n	–	–	–	90.0	85.6	176.8
Variant1	✓	–	–	92.4	90.6	139.8
Variant2	✓	✓	–	92.6	90.6	177.8
Variant3	✓	✓	✓	94.9	93.5	189.8

in processing speed to 139.8 FPS. Variant2 integrates both transfer learning and the Bottleneck CSP module, resulting in further performance enhancements. This variant exhibits a precision of 92.6% and a mAP@0.5 of 90.6% while maintaining a competitive processing speed of 177.8 FPS.

The most advanced Variant3 integrates transfer learning, the Bottleneck CSP module, and the CBAM attention mechanism. This comprehensive integration brings about substantial improvements, achieving a precision of 94.9%, a mAP@0.5 of 93.5%, and a processing speed of 189.8 FPS. Figure 7 visualizes the content of the table, providing a clearer insight into the superiority of our model, making it more suitable for the development of smart cities.

PRESENTATION OF RESULTS

The series of examples depicted in Figure 8 vividly illustrate the substantial advantages inherent in our model, particularly in terms of its reasoning abilities, especially when dealing with intricate scenes and contextual complexities. Upon conducting a comprehensive analysis of these scenarios, our model demonstrates exceptional information processing and reasoning capabilities, enabling it to swiftly and accurately comprehend multi-layered, multidimensional contexts. In these instances, our model transcends mere object recognition and engages in thorough reasoning within intricate environments, maintaining high accuracy even amidst diverse factors such as fluctuating lighting conditions, dynamic traffic scenarios, and crowd movements. This capability holds pivotal practical value for real-time and precise information processing, particularly in the realm of urban management.

The marked enhancement in these reasoning abilities establishes a robust foundation for our model to showcase superior performance in practical applications. By monitoring key indicators in the functioning of cities—such as traffic flow, environmental pollution, and the status of public facilities—in real time, our model furnishes more refined data insights for urban management. Consequently, this facilitates prompt issue resolution, prediction of potential risks, and the implementation of corresponding management measures.

Figure 7. Ablation experiments on the KITTI datasets

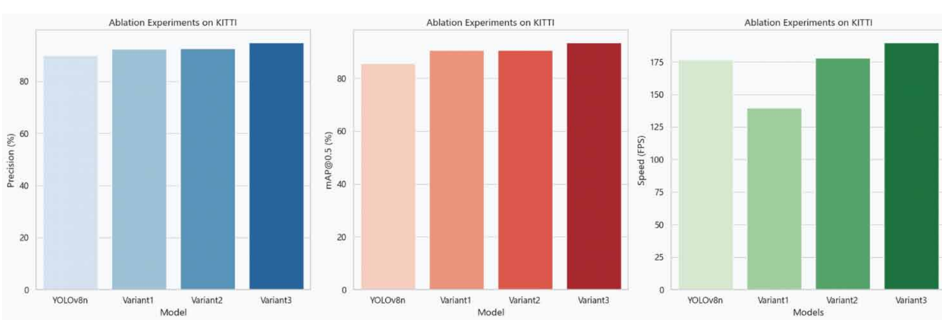
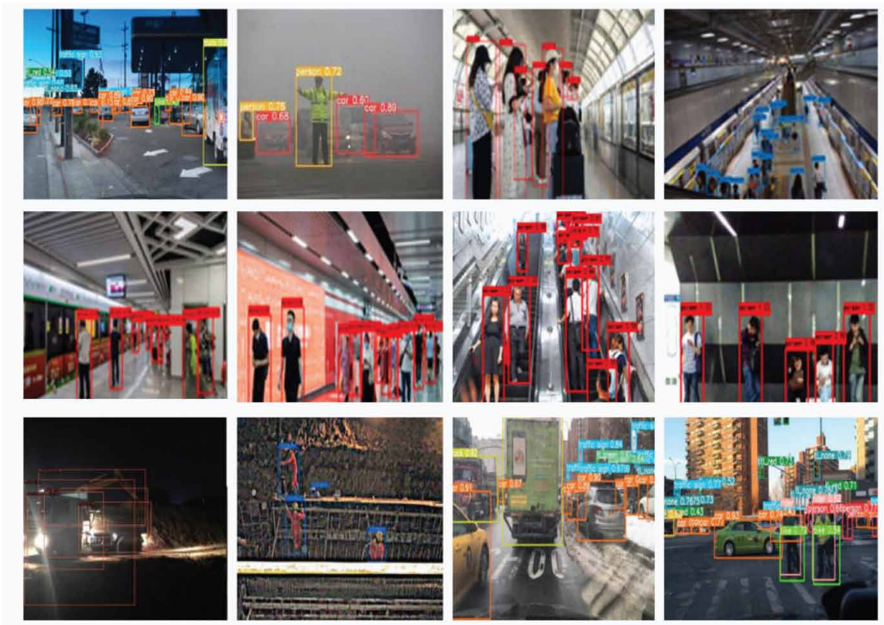


Figure 8. Illustration of YOLOv8_k



CONCLUSION

This paper presents our model, YOLOv8_k, designed to elevate the efficiency and effectiveness of intelligent city management by optimizing target detection technology. Through rigorous experimental evaluations, the model showcased remarkable cognitive capabilities, adeptly discerning and comprehending diverse objects within intricate urban settings. This achievement contributes to delivering more accurate data support for real-time monitoring and urban management. Nevertheless, we recognize certain challenges, including the model's accuracy in complex urban scenarios and the computational and storage demands it faces when processing extensive datasets.

The shortcomings of the model lie in the need for further improvement in object detection accuracy in complex urban scenes. Diverse lighting conditions, complex traffic situations, and variations in crowd movements may lead to recognition errors in traditional algorithms, impacting the model's accurate understanding of the urban environment. Additionally, addressing computational and storage issues in large-scale data processing is crucial to ensuring the model's efficiency in real-time decision-making and issue resolution.

Future research directions will focus on further optimizing the model to enhance object detection accuracy in complex urban environments and addressing the computational and storage challenges in large-scale data processing. We plan to explore more advanced algorithms and technologies to adapt to changing urban scenarios and improve the model's robustness. Furthermore, we will continue to explore the application potential of the model in intelligent city management, expanding its scope to provide comprehensive support for future urban development and management. In the future, we look forward to witnessing further developments and applications of the model, bringing more innovation and benefits to the construction and management of smart cities. Through ongoing research and technological improvements, we are confident in overcoming current challenges and realizing a more intelligent and efficient urban management system.

AUTHOR NOTE

The authors of this publication declare there are no competing interests.

This research was supported by Jilin Province Higher Education Scientific Research Subject (JGJX2020D583).

REFERENCES

Aboah, A., Wang, B., Bagci, U., & Adu-Gyamfi, Y. (2023). Real-time multi-class helmet violation detection using few-shot data sampling technique and YOLOv8. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5349–5357). IEEE. https://doi.org/ doi:10.1109/CVPRW59228.2023.00564

Ahmed, I., Jeon, G., & Chehri, A. (2023). A smart IoT enabled end-to-end 3D object detection system for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *24*(11), 13078–13087. doi:10.1109/TITS.2022.3210490

Al Mudawi, N., Qureshi, A. M., Abdelhaq, M., Alshahrani, A. S., Alazeb, A., Alonazi, M., & Algarni, A. (2023). Vehicle detection and classification via YOLOv8 and deep belief network over aerial image sequences. *Sustainability (Basel)*, *15*(19), 14597. doi:10.3390/su151914597

Feroz, M. A., Sultana, M., Hasan, M. R., Sarker, A., Chakraborty, P., & Choudhury, T. (2022). Object detection and classification from a real-time video using SSD and YOLO models. In *Computational intelligence in pattern recognition: Proceedings of CIPR 2021* (Vol. 1349, pp. 37–47). Springer. doi:10.1007/978-981-16-2543-5_4

Guo, Z., Wang, X., Zhang, Z., Huang, Y., & Ma, X. (2023). PerSnake: A real-time pedestrian instance segmentation network using contour representation. *Machine Vision and Applications*, *34*(5), 78. doi:10.1007/s00138-023-01419-w

Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: A literature review. *BMC Medical Imaging*, *22*(1), 69. doi:10.1186/s12880-022-00793-7 PMID:35418051

Knura, M., Kluger, F., Zahtila, M., Schiewe, J., Rosenhahn, B., & Burghardt, D. (2021). Using object detection on social media images for urban bicycle infrastructure planning: A case study of Dresden. *ISPRS International Journal of Geo-Information*, *10*(11), 733. doi:10.3390/ijgi10110733

Li, L., Wang, Z., & Zhang, T. (2023). GBH-YOLOv5: Ghost convolution with BottleneckCSP and tiny target prediction head incorporating YOLOv5 for PV panel defect detection. *Electronics (Basel)*, *12*(3), 561. doi:10.3390/electronics12030561

Li, X., Xia, H., & Lu, L. (2022). ECA-CBAM: Classification of diabetic retinopathy: Classification of diabetic retinopathy by cross-combined attention mechanism. In *Proceedings of the 2022 6th international conference on innovation in artificial intelligence* (pp. 78–82). Association for Computing Machinery. doi:10.1145/3529466.3529468

Liu, Q., Huang, W., Duan, X., Wei, J., Hu, T., Yu, J., & Huang, J. (2023). DSW-YOLOv8n: A new underwater target detection algorithm based on improved YOLOv8n. *Electronics (Basel)*, *12*(18), 3892. doi:10.3390/electronics12183892

Lou, H., Duan, X., Guo, J., Liu, H., Gu, J., Bi, L., & Chen, H. (2023). DC-YOLOv8: Small-size object detection algorithm based on camera sensor. *Electronics (Basel)*, *12*(10), 2323. doi:10.3390/electronics12102323

Saleem, M. H., Potgieter, J., & Arif, K. M. (2022). Weed detection by faster RCNN model: An enhanced anchor box approach. *Agronomy (Basel)*, *12*(7), 1580. doi:10.3390/agronomy12071580

Saydirasulovich, S. N., Mukhiddinov, M., Djuraev, O., Abdusalomov, A., & Cho, Y.-I. (2023). An improved wildfire smoke detection based on YOLOv8 and UAV images. *Sensors (Basel)*, *23*(20), 8374. doi:10.3390/s23208374 PMID:37896467

Song, Y., Ning, H., Ye, X., Chandana, D., & Wang, S. (2022). Analyze the usage of urban greenways through social media images and computer vision. *Environment and Planning. B, Urban Analytics and City Science*, *49*(6), 1682–1696. doi:10.1177/23998083211064624

Su, H., Wang, X., Han, T., Wang, Z., Zhao, Z., & Zhang, P. (2022). Research on a U-Net bridge crack identification and feature-calculation methods based on a CBAM Attention mechanism. *Buildings*, *12*(10), 1561. doi:10.3390/buildings12101561

Sukel, M., Rudinac, S., & Worring, M. (2020). Urban object detection kit: A system for collection and analysis of street-level imagery. *In Proceedings of the 2020 international conference on multimedia retrieval* (pp. 509–516). Association for Computing Machinery. doi:10.1145/3372278.3390708

Vajgl, M., Hurtik, P., & Nejezchleba, T. (2022). Dist-YOLO: Fast object detection with distance estimation. *Applied Sciences (Basel, Switzerland)*, *12*(3), 1354. doi:10.3390/app12031354

Wang, C., Wang, C., Wang, L., Wang, J., Liao, J., Li, Y., & Lan, Y. (2023). A lightweight cherry tomato maturity real-time detection algorithm based on improved YOLOv5n. *Agronomy (Basel)*, *13*(8), 2106. doi:10.3390/agronomy13082106

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7464–7475). IEEE. https://doi.org/ doi:10.1109/CVPR52729.2023.00721

Wang, G., Chen, Y., An, P., Hong, H., Hu, J., & Huang, T. (2023). UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors (Basel)*, *23*(16), 7190. doi:10.3390/s23167190 PMID:37631727

Wang, K., Zhang, Z., Wu, X., & Zhang, L. (2022). Multi-class object detection in tunnels from 3D point clouds: An auto-optimized lazy learning approach. *Advanced Engineering Informatics*, *52*, 101543. doi:10.1016/j.aei.2022.101543

Wang, L., Shoulin, Y., Alyami, H., Laghari, A. A., Rashid, M., Almotiri, J., Alyamani, H. J., & Alturise, F. (2022). A novel deep learning-based single shot multibox detector model for object detection in optical remote sensing images. *Geoscience Data Journal*, *18*, gdj3.162. doi:10.1002/gdj3.162

Yao, S., Zhu, Q., Zhang, T., Cui, W., & Yan, P. (2022). Infrared image small-target detection based on improved FCOS and spatio-temporal features. *Electronics (Basel)*, *11*(6), 933. doi:10.3390/electronics11060933

Yuan, Y., Li, Z., & Ma, D. (2022). Feature-aligned single-stage rotation object detection with continuous boundary. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–11. doi:10.1109/TGRS.2022.3203983

Zhang, R., Shao, Z., Huang, X., Wang, J., Wang, Y., & Li, D. (2022). Adaptive dense pyramid network for object detection in UAV imagery. *Neurocomputing*, *489*, 377–389. doi:10.1016/j.neucom.2022.03.033

Hewen Gao, Ph.D. of Economics, Associate Professor. Graduated from Jilin University in 2009 Worked in Changchun Humanities and Sciences College. Her research interests include environmental regulation,Geospatial data mining.

Yu Sun,Ed.D of Higher Education Leadership, Associate professor. Graduated from Delaware State University. Worked in Changchun Humanities and Sciences College. Her research interests include big data and cloud computing.

Weilin Shi, Ph.D. of Economics, Associate Professor. Graduated from Jilin University in 2009, Worked in Metropolitan State University of Denver. His research interests include environmental regulation, Spatial econometric theory and application.