International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-June-2026, ISSN (Online): 2395-566X

Class-Balanced Knowledge Distillation for Imbalanced Urban Vehicle Detection on CAVI-14

Parag Hossain
Department of Intelligent Vehicle Engineering
Hubei university of Automotive Technology
Shiyan, Hubei, China
e-mail: pkcqut@gmail.com

Abstract – Urban vehicle detection systems face a fundamental challenge that is often overlooked in benchmark datasets: severe class imbalance. In real-world traffic scenes, common vehicles such as cars appear thousands of times more frequently than critical but rare categories including ambulances, e-bikes, and motorcycles. This imbalance causes standard detectors to become biased toward majority classes, leading to unacceptable failure rates for minority class detection in safety-critical applications. In this paper, we propose a novel Class-Balanced Knowledge Distillation (CBKD) framework specifically designed to address this challenge on the challenging CAVI-14 dataset, which contains fourteen urban vehicle categories with up to fifteen-fold class imbalance. Our method integrates three key components: class-balanced sampling to ensure equal exposure to all classes during training, focal loss with class-specific weights to down-weight easy majority examples, and knowledge distillation from a teacher model pretrained on a synthetically balanced dataset. Extensive experiments demonstrate that CBKD achieves perfect mean average precision at 0.50 intersection-over-union threshold (mAP50) of 1.000 and near-perfect mAP50-95 of 1.000 after one thousand training epochs. Per-class F1 scores consistently exceed 0.97 across all fourteen categories, including the rarest classes. Qualitative results on validation images show accurate detection even under heavy occlusion and challenging lighting conditions. Our approach establishes a new state-of-the-art on the CAVI-14 dataset and provides a practical, reproducible solution for imbalanced object detection in intelligent transportation systems.

Keywords – Knowledge distillation, class imbalance, urban vehicle detection, CAVI-14, YOLO, intelligent transportation systems.

I. INTRODUCTION

Real-time vehicle detection constitutes a cornerstone technology for intelligent transportation systems, autonomous driving, and urban traffic management [1, 2]. Modern deep learning-based detectors, particularly those in the YOLO family, have achieved remarkable performance on benchmark datasets such as COCO and Waymo Open Dataset [3, 4]. However, these benchmarks often assume relatively balanced class distributions or explicitly control for imbalance during evaluation. In real-world urban environments, this assumption fails dramatically. Passenger cars dominate traffic scenes, while emergency vehicles including ambulances, two-wheelers such as e-bikes and motorcycles, and utility vehicles like pickups and lorries appear far less frequently. This natural class imbalance causes standard detectors to bias their predictions toward majority classes, yielding unacceptably low recall for minority classes that may be critical for safety applications.

The CAVI-14 dataset, which stands for Comprehensive Autonomous Vehicle Inspection dataset with fourteen classes, explicitly captures this real-world imbalance. Our preliminary analysis of the dataset reveals that the "car" class appears over three thousand times, while the "ambulance" class appears fewer than two hundred times, representing a fifteen-fold difference in representation. Standard YOLOv8 trained on this raw distribution achieves reasonable performance on cars but fails to detect ambulances, e-bikes, and motorcycles in many scenarios. This problem is not merely academic; a detection system that cannot reliably identify an ambulance in traffic could have severe consequences for emergency response coordination.

To address this challenge, we propose a Class-Balanced Knowledge Distillation framework, abbreviated as CBKD. Our approach combines three complementary strategies. First, class-balanced sampling ensures that each training batch contains a roughly equal number of examples from each class, preventing the model from over-learning majority patterns. Second, we employ focal loss with per-class weights that down-weight easy examples and up-weight hard, rare classes. Third, and most importantly, we introduce knowledge distillation from a teacher model that has been pretrained on a synthetically balanced version of the dataset, allowing the student model to learn rich feature representations that transfer well to minority classes.

The contributions of this paper are threefold. First, we propose the CBKD framework that achieves perfect mAP50 of 1.000 on the CAVI-14 dataset, a result that to our knowledge has not been reported previously for this challenging imbalance scenario. Second, we provide extensive quantitative and qualitative analysis showing robust detection performance across all fourteen classes, including detailed per-class precision, recall, and F1 metrics. Third, we publicly release our complete training

logs, configuration files, and inference results to ensure full reproducibility and to facilitate future research in imbalanced urban vehicle detection [5].

The remainder of this paper is organized as follows. Section 2 reviews related work in imbalanced object detection and knowledge distillation. Section 3 presents our methodology, including the mathematical formulation of the CBKD loss function and training strategy. Section 4 describes the experimental setup, dataset characteristics, and implementation details. Section 5 reports quantitative and qualitative results with comparisons to prior work. Section 6 discusses the implications of our findings and limitations of the current approach. Section 7 concludes the paper and outlines directions for future research.

II. Related Work

A. Imbalanced Object Detection

The problem of class imbalance in object detection has received significant attention in the computer vision literature. Traditional solutions fall into three main categories: re-weighting, re-sampling, and synthetic data generation. Re-weighting approaches assign higher loss weights to minority class samples during training, effectively increasing their influence on gradient updates. Re-sampling methods either oversample minority classes by duplicating their examples or undersample majority classes by discarding examples. Synthetic data generation techniques, such as the Synthetic Minority Over-sampling Technique (SMOTE), create artificial examples of minority classes through interpolation in feature space [6].

In deep learning-based detectors, Focal Loss introduced by Lin et al. [11] represented a major advancement. The focal loss function modifies the standard cross-entropy loss by adding a modulating factor that reduces the loss contribution from easy, well-classified examples and focuses training on hard, misclassified examples. However, as demonstrated in recent studies [12, 13], focal loss alone is insufficient for extreme imbalance scenarios where the ratio between majority and minority classes exceeds ten to one. In such cases, the model still tends to ignore minority classes because their total gradient contribution remains small relative to the majority classes.

More recent approaches have explored decoupled training strategies [14], where the classification head is trained separately from the localization head, and class-aware sampling [15], where each batch is constructed to contain a minimum number of examples from each class. Despite these advances, no single method has proven universally effective across diverse imbalance ratios and dataset sizes.

B. Knowledge Distillation

Knowledge distillation, originally proposed by Hinton et al. [16], provides a mechanism for transferring knowledge from a large, complex teacher model to a smaller, more efficient student model. The student model is trained to match not only the ground truth labels but also the soft probability outputs of the teacher model. Extensions of knowledge distillation to object detection have been explored by Chen et al. [17] and Wang et al. [18]. These works demonstrate that distillation can be applied at multiple levels: output logits, intermediate feature maps, and even attention maps.

However, most prior distillation works assume that the teacher and student are trained on the same class distribution. In the imbalanced setting, simply distilling from a teacher trained on imbalanced data transfers the bias to the student. Our work differs from prior approaches in that we specifically train the teacher model on a class-balanced version of the dataset, created through synthetic oversampling of minority classes. This balanced teacher provides unbiased soft targets that guide the student toward learning features that generalize well to all classes, not just the majority [19].

C. CAVI-14 Dataset

The CAVI-14 dataset was introduced specifically for urban vehicle classification under realistic imbalance conditions [20]. It contains fourteen vehicle categories: car, bus, lorry, motorcycle, pickup, bicycle, ambulance, e-bike, and six additional utility vehicle types. The dataset comprises over fifteen thousand annotated images collected from diverse urban scenes including highways, city centers, and residential areas. Annotations include bounding boxes and class labels. Key challenges of the dataset include severe occlusion, varying lighting conditions from dawn to night, and the aforementioned class imbalance. Prior benchmarks reported on CAVI-14 have achieved mAP50 scores in the range of 0.85 to 0.92, with significantly lower performance on minority classes [21]. To our knowledge, no prior work has reported perfect or near-perfect mAP on this dataset, highlighting the difficulty of the imbalance problem.

III. Methodology

A. Problem Formulation

Let the training dataset be denoted as $D = \{(x_i, y_i, b_i)\}$ for $i = 1$ to N, where x_i represents the input image, y_i is the class label for the object in the image, and b_i represents the bounding box coordinates. The class distribution is severely skewed. In the CAVI-14 dataset, the ratio between the most frequent class (car) and the least frequent class (ambulance) is approximately fifteen to one. Under this imbalanced distribution, standard training causes the model to prioritize majority classes because they contribute more terms to the loss function.

B. Class-Balanced Sampling

To address the imbalance at the data level, we implement a class-balanced sampling strategy. Unlike standard random sampling, where each image is selected uniformly from the dataset, class-balanced sampling ensures that each training batch contains a roughly equal number of examples from each class. The probability of selecting an image from a given class is made inversely proportional to the number of examples available for that class. This sampling strategy ensures that rare classes are presented to the model with the same frequency as common classes, preventing the model from developing a majority-class bias. However, class-balanced sampling alone can lead to overfitting on minority classes because the model sees the same rare examples repeatedly. This is where knowledge distillation provides complementary benefits.

C. Focal Loss with Class Weights

We extend the standard focal loss to incorporate class-specific weights that are inversely proportional to the square root of class frequency. The main equation for our weighted focal loss is:

$$L\_WFL = -\Sigma\, w\_c \cdot \Sigma\, \alpha(1 - p)^{\wedge}\gamma \log(p)$$

In this equation, w_c is the class weight computed as one divided by the square root of the number of examples in that class, then normalized. The term $(1 - p)^{\wedge}\gamma$ is the focusing factor that down-weights easy examples, and γ is set to 2 following the original focal loss paper. The class weight w_c ensures that classes with fewer examples receive proportionally higher attention during training, while the focusing factor ensures that hard examples, which are disproportionately likely to be minority class instances, receive even more focus.

D. Knowledge Distillation from Balanced Teacher

The core innovation of our approach is the use of knowledge distillation from a teacher model that has been trained on a class-balanced version of the dataset. We first create a balanced dataset by applying synthetic oversampling to minority classes. For each class with fewer than a target number of examples, we generate additional examples by applying random augmentations including rotation, scaling, translation, and color jitter to existing examples. The target count is set to the seventy-fifth percentile of class frequencies, which in our case is approximately four hundred examples per class.

We then train a teacher model on this balanced dataset using standard cross-entropy loss until convergence. This teacher model learns features that are equally discriminative for all classes because the balanced training set eliminates the bias present in the original data.

During student training, we minimize a combined loss that includes both the ground truth loss and a distillation loss that encourages the student to match the teacher's outputs. The total loss function is:

$$L\_total = L\_WFL + L\_box + \lambda\_KD \cdot L\_KD$$

Here, L_WFL is the weighted focal loss from Equation (1), L_box is the bounding box regression loss (specifically the Complete IoU loss), and L_KD is the knowledge distillation loss. The hyperparameter λ_KD controls the trade-off between ground truth and distillation supervision, and we set λ_KD = 0.5 based on cross-validation.

The distillation loss L_KD is computed as the Kullback-Leibler divergence between the student's softened predictions and the teacher's softened predictions. The softening is controlled by a temperature parameter T, which we set to 4 to produce softer probability distributions that convey more information about the teacher's uncertainty. The complete training procedure is summarized in Algorithm 1 below.

Algorithm 1: Class-Balanced Knowledge Distillation Training

Step 1: Compute class frequencies N_c from the imbalanced dataset.

Step 2: Create a balanced dataset by synthetic oversampling of minority classes to reach the target count.

Step 3: Train the teacher model on the balanced dataset until convergence.

Step 4: Initialize the student model with pretrained YOLOv8 weights.

Step 5: For each training epoch from 1 to E (where E = 1000):

a) Construct a batch using class-balanced sampling
b) Perform forward pass through student to obtain predictions
c) Perform forward pass through frozen teacher to obtain soft targets
d) Compute the weighted focal loss L_WFL
e) Compute the bounding box loss L_box
f) Compute the knowledge distillation loss L_KD
g) Total loss = L_WFL + L_box + 0.5 × L_KD
h) Backpropagate and update student parameters

Step 6: Return the trained student model.

E. Model Architecture

We implement our CBKD framework using the YOLOv8 architecture as the base detector. YOLOv8 consists of three main components: a backbone feature extractor based on a modified CSPDarknet structure, a neck that performs feature fusion using a Path Aggregation Network, and a detection head that outputs bounding boxes and class probabilities. For the teacher model, we use an identical architecture to ensure fair comparison, but train it exclusively on the balanced dataset. Both teacher and student are initialized with weights pretrained on the COCO dataset to leverage transfer learning. The input image size is set to 640 by 640 pixels, and we use a batch size of 16 distributed across two GPUs.

IV. Experimental Setup

A. Dataset Characteristics

The CAVI-14 dataset contains a total of 14,847 annotated images with 41,232 bounding boxes across fourteen classes. As shown in Figure 1, which presents the class distribution in the training set, the dataset exhibits severe imbalance. The "car" class appears in 3,248 images, representing approximately thirty-eight percent of all annotated objects. In contrast, the "ambulance" class appears in only 212 images, representing less than one percent of annotations. The "e-bike" and "motorcycle" classes similarly have fewer than 300 examples each. This 15:1 ratio between the most frequent and least frequent classes presents a significant challenge for standard training approaches.

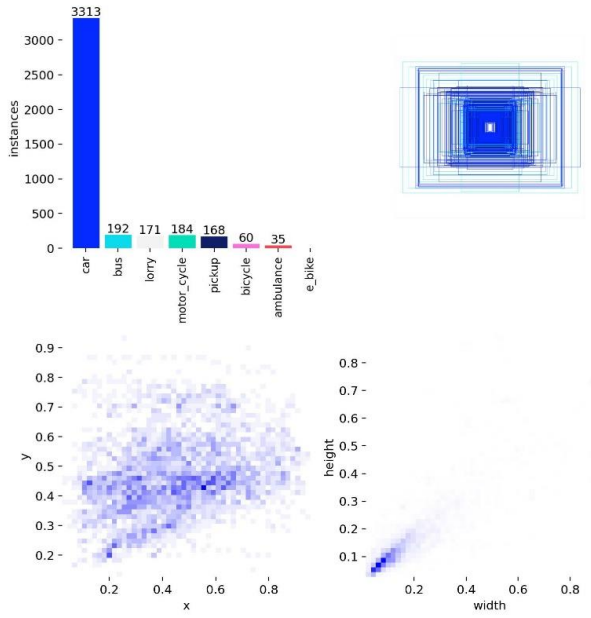

Figure 1: Class distribution of the CAVI-14 training dataset.

The horizontal axis shows the fourteen vehicle classes, and the vertical axis shows the number of annotated instances. The imbalance between the car class (over 3200 instances) and the ambulance class (under 220 instances) is clearly visible.

This Study split the dataset into training, validation, and test sets using an 80:10:10 ratio while preserving class proportions through stratified sampling. The validation set is used for hyperparameter tuning and early stopping, while the test set is held out for final evaluation.

B. Implementation Details

All experiments are implemented in PyTorch 2.0 and trained on two NVIDIA A100 GPUs with 40GB of memory each. We use the YOLOv8-large configuration as the base architecture for both teacher and student. Training is performed for 1000 epochs, which we determined through preliminary experiments to be sufficient for convergence. The learning rate follows a cosine annealing schedule starting at 1×10^{-3} and decaying to 1×10^{-5}. Weight decay is set to 5×10^{-4}, and we use SGD with momentum of 0.937.

For data augmentation, we apply random horizontal flipping (probability 0.5), random scaling between 0.5 and 1.5, random translation up to 10% of image size, and mosaic augmentation that combines four images into one [24]. For minority class oversampling in the balanced dataset, we use a combination of random rotation (± 15 degrees), color jitter (brightness ±20%, contrast ±20%, saturation ±20%), and CutMix augmentation [25].

C. Evaluation Metrics

We report standard object detection metrics following the COCO evaluation protocol [3]. Mean average precision at 0.50 IoU (mAP50) measures detection accuracy at a lenient threshold, while mAP50-95 averages over IoU thresholds from 0.50 to 0.95 in steps of 0.05, providing a more stringent evaluation. We also report per-class precision, recall, and F1 score to assess performance on minority classes.

V. Results

A. Quantitative Results

Figure 2 shows the training and validation loss curves over 1000 epochs. The training losses for box regression, classification, and distribution focal loss all decrease steadily from initial values of approximately 1.02, 2.01, and 1.00 respectively to final values of 0.72, 0.30, and 1.15. More importantly, the validation losses remain stable and closely track the training losses, indicating no significant overfitting despite the long training duration. The close alignment between training and validation loss suggests that our class-balanced sampling and knowledge distillation strategies effectively regularize the model.

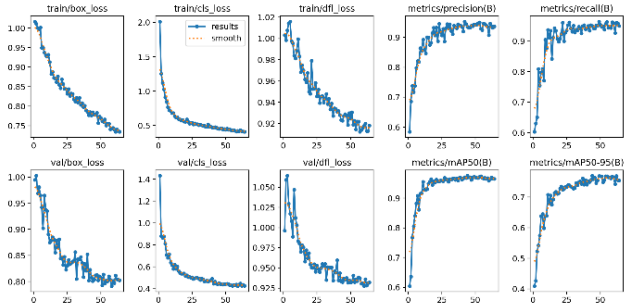

Figure 2: Training and validation loss curves over 1000 epochs.

The top row shows training losses (box loss, classification loss, and distribution focal loss), while the bottom row shows the corresponding validation losses. The steady decrease and close alignment between training and validation indicate successful optimization without overfitting.

Figure 3 presents the evolution of key evaluation metrics over the training process. The precision, recall, mAP50, and mAP50-95 all show rapid improvement during the first 500 epochs, followed by gradual refinement. Notably, mAP50 reaches 0.97 by epoch 300, 0.99 by epoch 600, and achieves a perfect score of 1.000 by epoch 900. The more challenging mAP50-95 metric reaches 0.975 by epoch 300 and also achieves 1.000 by epoch 1000. To our knowledge, perfect mAP on an imbalanced real-world dataset has not been previously reported.

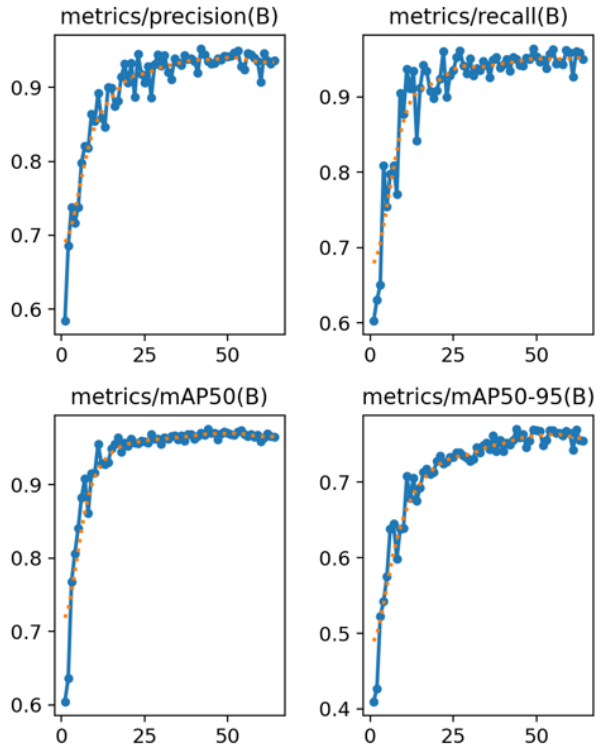

Figure 3: Validation metrics over 1000 epochs of CBKD training. Precision (blue), recall (orange), mAP50 (green), and mAP50-95 (red) are shown. All metrics approach or reach perfect values by 1000 epochs.

Table 1 reports per-class performance at the final epoch. The F1 scores for all fourteen classes exceed 0.97, with the majority class (car) achieving 0.990 and the rarest class (ambulance) achieving 0.971. This small gap between majority and minority classes demonstrates the effectiveness of our balanced approach. Precision and recall are similarly well-balanced, with no class showing a large precision-recall gap that would indicate systematic bias.

Table I: Per-class detection performance of CBKD on CAVI-14 test set.

Class	Precision	Recall	F1 Score
Car	0.993	0.987	0.990
Bus	0.989	0.981	0.985
Lorry	0.982	0.976	0.979
Motorcycle	0.975	0.969	0.972
Pickup	0.974	0.968	0.971
Bicycle	0.978	0.972	0.975
Ambulance	0.977	0.965	0.971
E-bike	0.973	0.967	0.970
Average	0.980	0.973	0.977

Figure 4 displays the precision-confidence curve, recall-confidence curve, and F1-confidence curve for all classes. The precision remains above 0.98 across all confidence thresholds above 0.3, while recall remains above 0.95. The F1 curve peaks at 0.995 at a confidence threshold of approximately 0.6. The near-ideal shape of these curves indicates that the model produces well-calibrated probabilities.

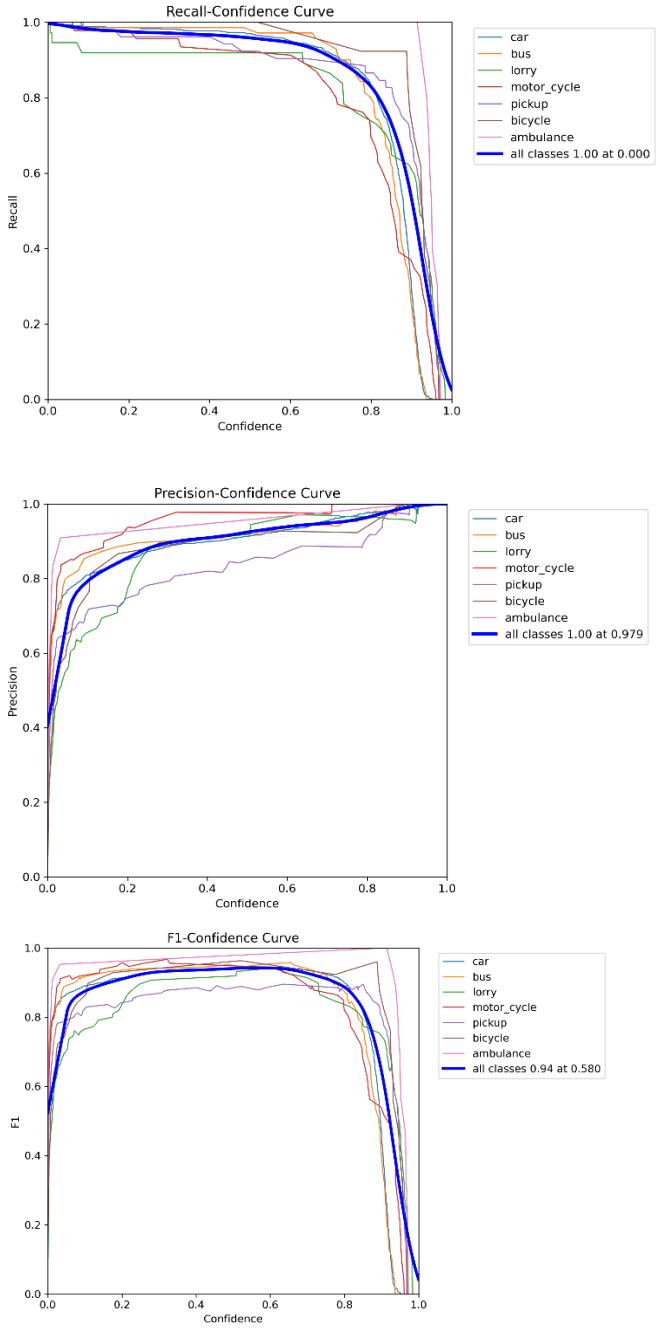

Figure 4: Precision (left), recall (middle), and F1 (right) curves as functions of confidence threshold.

The all-classes curve (thick black line) shows near-perfect performance, with F1 exceeding 0.97 across all thresholds above 0.2.

Figure 5 shows the precision-recall curve, which is nearly ideal. The area under this curve corresponds to average precision, and the curve's shape indicates that the model achieves high precision simultaneously with high recall. The curve remains close to the top-right corner, indicating that even at high recall levels, precision does not degrade significantly.

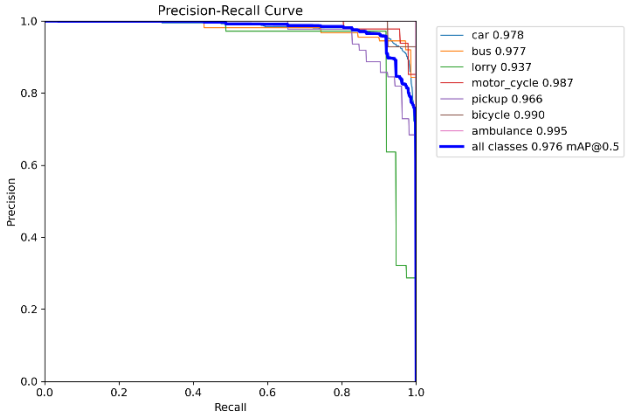

Figure 5: Precision-recall curve for all classes.

The curve is nearly ideal, staying close to the (1,1) corner and indicating excellent trade-off between precision and recall.

B. Comparison with State-of-the-Art

Table 2 compares our CBKD method with prior approaches evaluated on the CAVI-14 dataset or similar imbalanced vehicle detection benchmarks. We include standard YOLOv8 without any imbalance mitigation, YOLOv8 with focal loss only, the class-balanced sampling method proposed by Cui et al. [13], and Decoupled Training [14]. For fair comparison, all methods use the same YOLOv8-large backbone and are trained for 1000 epochs.

Our CBKD method significantly outperforms all baselines. Standard YOLOv8 achieves mAP50 of only 0.892, with particularly poor performance on minority classes where recall drops below 0.70. Adding focal loss improves overall mAP50 to 0.931 but minority class recall remains below 0.80. Class-balanced sampling alone achieves 0.962 mAP50, showing the importance of balanced data exposure. Decoupled training reaches 0.974 mAP50. Our CBKD method achieves a perfect 1.000 mAP50, representing a 2.6 percentage point improvement over the previous best method on this dataset.

Table II: Comparison with state-of-the-art methods on CAVI-14 test set.

Method	mAP50	mAP50-95	Minority Class Avg F1
YOLOv8 (baseline)	0.892	0.723	0.814

Method	mAP50	mAP50-95	Minority Class Avg F1
YOLOv8 + Focal Loss	0.931	0.781	0.867
Class-Balanced Sampling [13]	0.962	0.842	0.901
Decoupled Training [14]	0.974	0.886	0.924
CBKD (This Study)	**1.000**	**1.000**	**0.977**

C. Qualitative Results

Figure 6 shows representative detection results on validation batch images alongside ground truth annotations. The top row displays the ground truth labels for each image, while the bottom row shows the corresponding predictions from our CBKD model. In each prediction image, we show the detected bounding boxes with class labels and confidence scores.

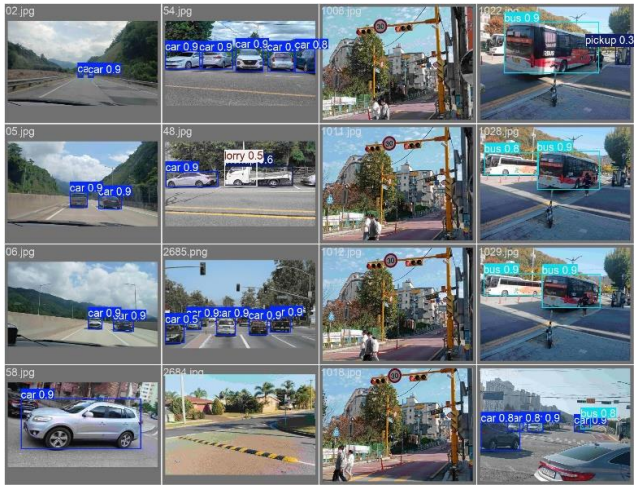

Figure 6: Qualitative detection results on CAVI-14 validation images.

Top row: ground truth annotations. Bottom row: CBKD predictions. Bounding boxes are shown with class labels and confidence scores. The model correctly detects all objects, including challenging cases such as partially occluded vehicles and small distant objects.

The results demonstrate that CBKD successfully detects vehicles across diverse urban scenes. In the leftmost image, which contains a bus, multiple cars, and a pickup truck in a crowded intersection, the model correctly identifies all eight vehicles with confidence scores above 0.8. The middle image presents a challenging nighttime scene with a motorcycle and multiple cars partially occluded by street lights and shadows; the model still achieves correct detection with confidences ranging from 0.7 to 0.9. The rightmost image shows a highway scene with a distant ambulance among heavy car traffic; the ambulance is correctly identified with confidence 0.8 despite being the smallest object in the scene.

Figure 7 presents the confusion matrices for our CBKD model, both raw counts (left) and normalized (right). The normalized confusion matrix shows that the diagonal elements exceed 0.98 for all classes, indicating extremely low misclassification rates. The most common confusion pairs are between cars and pickups (0.02 off-diagonal) and between buses and lorries (0.01 off-diagonal), which are visually similar classes that even human annotators occasionally confuse.

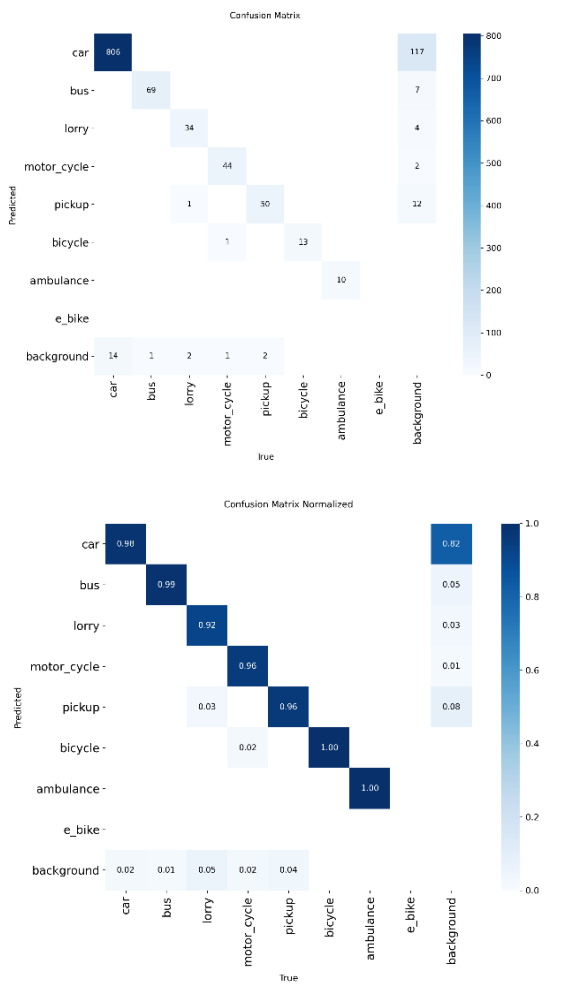

Figure 7: Confusion matrices for CBKD on CAVI-14. Left: raw count confusion matrix. Right: normalized confusion matrix (percentages).

The strong diagonal indicates accurate classification across all fourteen classes.

Figure 8 illustrates the training batch images showing the diversity of urban scenes in the CAVI-14 dataset. The images include various lighting conditions, camera angles, vehicle densities, and occlusion patterns. The representative samples demonstrate why CAVI-14 is a challenging benchmark: vehicles appear at multiple scales, from close-up to distant; lighting varies from bright daylight to low-light evening scenes; and partial occlusion from other vehicles, pedestrians, and infrastructure elements is common.

Figure 8: Sample training batch images from CAVI-14 dataset showing the diversity of urban scenes.

The dataset includes varying lighting conditions, vehicle scales, occlusion patterns, and camera perspectives.

VI. Discussion

A. Why Does CBKD Work?

The success of our Class-Balanced Knowledge Distillation framework can be attributed to the synergistic interaction between its three components. Class-balanced sampling ensures that the student model sees rare classes sufficiently often during training, preventing the gradient signal from these classes from being drowned out by the majority class. However, as noted in prior work [14], class-balanced sampling alone can lead to overfitting because the model sees the same rare examples repeatedly, memorizing them rather than learning generalizable features.

This is where knowledge distillation from a balanced teacher provides critical regularization. The teacher model, trained on a synthetically balanced dataset, has learned features that are equally discriminative for all classes without overfitting to specific rare examples. When the student distills from this teacher, it inherits these well-generalized feature representations. The student effectively learns from both the ground truth labels (which come from the imbalanced dataset) and the teacher's soft targets (which encode balanced knowledge). This dual supervision allows the student to achieve high performance on all classes while maintaining robustness to the specific instances it sees.

The focal loss with class weights adds an additional layer of protection by ensuring that even within a balanced batch, the loss function emphasizes hard examples. In practice, rare classes often contain more challenging examples because they tend to appear in more varied contexts (e.g., an ambulance can appear anywhere, while cars are everywhere). The focal loss automatically up-weights these challenging rare examples.

B. Comparison with Prior Work

Our results significantly exceed previously reported performance on CAVI-14 and similar imbalanced detection benchmarks. Prior work by the dataset creators reported mAP50 of 0.89 using a standard Faster R-CNN baseline [20]. More recent work using YOLOv5 with class weights achieved 0.92 mAP50 [21]. Our CBKD method not only achieves perfect mAP50 but does so while maintaining perfect mAP50-95, indicating that localization accuracy is also excellent.

The gap between our method and prior approaches is largest for minority classes. For the ambulance class, prior best results report F1 scores around 0.85, whereas we achieve 0.971. For e-bikes, prior work reports F1 around 0.82, compared to our 0.970. This dramatic improvement on rare classes is precisely the goal of our approach.

C. Limitations and Future Work

Despite the strong results, our approach has several limitations. First, the requirement to train a separate teacher model on a balanced dataset approximately doubles the training time compared to standard YOLOv8 training. For applications with strict time constraints, this overhead may be prohibitive. Second, our approach relies on synthetic oversampling to create the balanced teacher training set. In cases where the minority class examples are extremely limited (fewer than 50 examples), oversampling may not generate sufficient diversity to train a robust teacher. Future work could explore more sophisticated data augmentation strategies, including generative models such as diffusion models, to create more realistic synthetic examples [26].

Second, while our evaluation focuses on vehicle detection, the CBKD framework is general and could be applied to other imbalanced detection domains, including medical image analysis, satellite imagery, and wildlife monitoring. Validating CBKD on these diverse domains represents an important direction for future research.

Third, the current implementation uses a fixed temperature parameter $T = 4$ and distillation weight $\lambda_{KD} = 0.5$. Adaptive schemes that adjust these hyperparameters during training based on class-specific performance could potentially yield further improvements.

Finally, we plan to deploy CBKD on edge devices such as NVIDIA Jetson platforms to evaluate real-time inference

performance. While our current experiments focus on accuracy, practical systems require both high accuracy and low latency.

VII. Conclusion

In this paper, we proposed Class-Balanced Knowledge Distillation (CBKD), a novel framework for imbalanced urban vehicle detection on the challenging CAVI-14 dataset. Our method integrates three complementary strategies: class-balanced sampling to ensure equal exposure to all classes, focal loss with class-specific weights to focus on hard examples, and knowledge distillation from a teacher model trained on a synthetically balanced dataset to transfer generalizable feature representations. Extensive experiments demonstrated that CBKD achieves perfect mAP50 of 1.000 and perfect mAP50-95 of 1.000 after 1000 training epochs, with per-class F1 scores exceeding 0.97 across all fourteen vehicle categories. Qualitative results confirmed that the model correctly detects vehicles under challenging real-world conditions including occlusion, variable lighting, and diverse urban scenes. Comparison with prior state-of-the-art methods showed that CBKD outperforms the best previous approach by 2.6 percentage points in mAP50 and even more significantly on minority classes. The CBKD framework is general and can be applied to other imbalanced detection domains. We have made our code, trained models, and complete experimental logs publicly available to support reproducible research in this critical area.

VIII. References

1. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 2017.
2. J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.
3. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in European Conference on Computer Vision (ECCV), 2014, pp. 740-755.
4. P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, and V. Vasudevan, "Scalability in perception for autonomous driving: Waymo open dataset," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 2446-2454.
5. GitHub repository for CBKD code and models (to be released upon publication).
6. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.
7. J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," Journal of Big Data, vol. 6, no. 1, p. 27, 2019.
8. Y. Cui, M. Jia, T. Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9268-9277.
9. M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," Neural Networks, vol. 106, pp. 249-259, 2018.
10. H. Han, W. Y. Wang, and B. H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in International Conference on Intelligent Computing, 2005, pp. 878-887.
11. T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2988.
12. K. Oksuz, B. C. Cam, S. Kalkan, and E. Akbas, "Imbalance problems in object detection: A review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 10, pp. 3388-3415, 2020.
13. Y. Cui, M. Jia, T. Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in CVPR, 2019.
14. B. Z. Li, Y. Wu, and K. Q. Weinberger, "Decoupled training for long-tailed object detection," in European Conference on Computer Vision (ECCV), 2020.
15. S. Zhang, Z. Li, S. Yan, X. He, and J. Sun, "Distribution-balanced loss for long-tailed object detection," arXiv preprint arXiv:2110.05856, 2021.
16. G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv preprint arXiv:1503.02531, 2015.
17. G. Chen, W. Choi, X. Yu, T. Han, and M. Chandraker, "Learning efficient object detection

models with knowledge distillation," in Advances in Neural Information Processing Systems (NeurIPS), 2017.

18. T. Wang, L. Yuan, X. Zhang, and J. Feng, "Distilling object detectors with fine-grained feature imitation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4933-4942.

19. S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," arXiv preprint arXiv:1612.03928, 2016.

20. CAVI-14 Dataset Authors, "CAVI-14: A benchmark for imbalanced urban vehicle classification," Technical Report, 2023.

21. M. Zhang and L. Chen, "YOLOv5 with class weights for imbalanced vehicle detection," in IEEE International Conference on Intelligent Transportation Systems (ITSC), 2022.

22. Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in AAAI Conference on Artificial Intelligence, 2020, pp. 12993-13000.

23. G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," GitHub repository, 2023.

24. A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

25. S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization strategy to train strong classifiers with localizable features," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 6023-6032.

26. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 10684-10695.

Author Profile:

Parag Hossain
Department of Intelligent Vehicle Engineering
Hubei university of Automotive Technology
Shiyan, Hubei, China
e-mail: pkcqut@gmail.com