Urban Work Zone Detection and Sizing: A Data-Centric Training and Topology-Based Inference Approach

Fan Zuo[1], Jingqin Gao[1,*], Kaan Ozbay[1], Zilin Bian[1], and Daniel Zhang[1]

***Abstract*— This study aims to address the challenges of automatically recognizing and sizing work zones in complex urban environments. We developed a deep-learning based work zone object detection model with a data-centric approach to iteratively enhance the model's performance by augmenting a custom training dataset collected from multiple sources, thereby overcoming the sparsity of annotated real-world work zone images. The training data is acquired from traffic cameras, mined from the web, and 3D-simulated work zone images. An innovative topology-based inference method is introduced, using XGBoost, for distinguishing true work zones from non-work operational zones with some work zone features. We also developed a reference-free work area size estimation method, which utilizes the standard heights of common construction equipment to provide a generalized real-pixel distance approximation. Our model's efficacy is demonstrated with an average mAP of 74.1% across all work zone classes, an accuracy of 98.4% for scene identification, and an accuracy of up to 89.52% for size estimation. Overall, our proposed approach significantly advances the capabilities of automated urban work zone detection and sizing, offering a cost-effective method to fill in the gap for the acquisition of work zone data in real-time by leveraging existing camera infrastructure.**

I. Introduction

The occurrence of work zone activities within highway networks and urban roadways can induce substantial traffic interruptions. These operational disruptions often result in roadway capacity reduction (i.e., by closing down one or more lanes) that might lead to severe congestion and roadway crashes. On the one hand, work zones have contributed to approximately ten percent of highway congestion in the U.S., resulting in an estimated annual loss of $700 million in fuel alone [1], while exacerbating the negative environmental effects of vehicle emissions and increasing safety risks [2]. For example, one study found that the crash rate increased by 24.4% under work zone conditions compared to non-work zone conditions [3]. On the other hand, while offline data sources such as work zone permit data may be available, the majority of U.S. cities have yet to establish a real-time approach to monitor actual activities during operational periods of work zones throughout the road network.

Given these concerns, real-time work zone detection becomes crucial. Knowing the location, duration, size of the work zones in real-time can provide vital insights into their impact on traffic flow and safety and help decision makers strategically allocate resources through the Transportation Management System (TMS).

*Corresponding author.

[1]Fan Zuo, Jingqin Gao, Kaan Ozbay, Zilin Bian, Daniel Zhang are with the C2SMART Center, Department of Civil and Urban Engineering, New York University (NYU), (e-mail: jingqin.gao@nyu.edu)

One viable approach to detecting work zones in real-time is using vision-based detection by means of existing traffic cameras. While these traffic cameras are extensively used for pedestrian and vehicle detection, their application to road work zone detection, especially in complex urban settings, remains limited. Urban work zones are uniquely challenging due to pedestrian and vehicular activities, complex surroundings, and lack of standardized setups.

The majority of existing studies using computer vision for work zones focus on off-street sites that might employ different types of equipment than those seen in roadside work zones, or they concentrate solely on detecting a single type of work zone component (e.g., traffic cones) [4]. Moreover, the mere recognition of work zone equipment does not necessarily confirm the presence of a work zone, as such equipment can sometimes serve other purposes, including regulating traffic (e.g., using barrels to separate traffic lanes). Consequently, there is a need for a new methodology that identifies work zone scenes in their entirety rather than simply detecting individual pieces of some work zone equipment.

Additionally, most of the existing Artificial Intelligence (AI) applications adopt model-centric approaches wherein data collection is perceived as a one-time event to improve the model architecture to enhance its performance [5]. However, given the inherent scarcity of open-source training samples for work zone detection, the current limitation in this domain is the lack of data rather than shortcomings in the model. This constrains the development of computer vision algorithms for this specific problem, which typically require substantial amounts of data.

In this paper, we introduce a deep learning based framework to effectively recognize urban work zone scenes and their sizes. The main contributions of this research are summarized as follows:

- We propose a data-centric training approach designed to iteratively improve the performance of work zone object detection by augmenting a customized training dataset fused from multiple data sources to overcome the sparsity of annotated real-world work zone images. These sources include 2,600 images with 15,000 work zone object labels from traffic cameras, web-mined images, and synthetic work zone images generated through a 3D simulator.
- We implemented a topology-based inference method using XGBoost to automatically identify work zone scenes. This innovative approach is designed to deal with the complexities of work zone scene detection

caused by the fact that recognizing individual or certain combination of work zone components alone (e.g., a traffic cone behind a car) may not necessarily represent a true work zone.
- We developed a reference-free work zone size estimation method, which utilizes the standard heights of common work zone equipment, to provide a generalized real-pixel distance rate method.

II. Related Works

A main challenge for work zone detection identified throughout the literature is the scarcity of publicly available, large-scale, domain-specific, annotated datasets of work zone imagery [4], [6], [7]. For example, Nath and Behzadan [6] used a CNN model that laid out a framework for detecting the most common types of off-street construction objects, namely, buildings, equipment, and workers. They recognized the lack of publicly available annotated work zone imagery dataset and introduced a systematic approach to visual data collection through crowd sourcing and web-mining and annotating the collected dataset for AI model training to overcome the limitation. They collected 3,500 images with 11,500 work zone elements and tested both YOLO-v2 and -v3, achieving a best-performing model with a 78.2% mAP.

Duan et al. [7] also stated that the lack of large-scale, open-source dataset for the construction industry limited the development of computer vision algorithms as they are often data-hungry. This study developed a new large-scale work zone image dataset, Site Object Detection dAtaset (SODA), with a total of 19,846 images and achieved a maximum mAP of 81.47%. The limitation of this dataset is it is mainly for off-street work zones and may not be suitable for detecting work zones that occur on the roadways.

Another study conducted by Katsamenis et al. [4] used Yolov5 for traffic cone detection using a training dataset of 500 traffic cones images. The data used in this paper was collected and manually annotated under the framework of the H2020 HERON project. The results showed that the proposed computer vision model could achieve a 91% accuracy in detecting traffic cones. However, work zones, especially urban work zones often composed by multiple types of construction objects and have no standard work zone set up, single object type detection may not be as effective as expected in such cases.

Given these research gaps, there is a need to construct a data-centric approach for automated urban work zone detection, which allows continuous improvement of training data in terms of work zone imagery, as well as an effective recognition method capable of identifying complex urban work zone scenes, rather than focusing solely on individual work zone components and /or construction equipment.

III. Methodology

The emerging field of Data-Centric AI is anticipated to introduce techniques for dataset optimization, thus enabling detection algorithms to be effectively trained even with relatively small datasets [5]. In this paper, we propose a data-centric framework for urban work zone detection and sizing which contains three modules: 1) a data-centric training that systematically augments training datasets with the goal of enhancing the accuracy of the work zone detection model, 2) a topology-based work zone scene inference that can identify work zones by understanding the positional relationships and connections among detected work zone objects (e.g., cones placed adjacent to a line of fences), and 3) a reference-free estimation for work zone size. We describe each part of the proposed detection methodology in the following sections. Fig. 1 shows the working flow of the proposed method.

Data-Centric Training for Work Zone Object Detection: We begin with collecting a customized dataset of 2,600 work zone images with about 15,000 labels from diverse sources including CCTVs, web-mined images, and a 3D simulator, offering a wide array of work zone scenarios (Fig. 1). This model is designed to identify several key work zone objects, such as traffic cones, construction workers and vehicles. Web-mined and 3D synthetic images primarily serve to fill the gaps in certain subcategories of the training data, which may be sparse in the CCTV images, yet their detection accuracy is vital for the model. For instance, web-mining is a good supplemental source for augmenting training images for construction vehicles. The quality of the data was incrementally improved by correcting label errors and pruning noisy labels based on data quality objectives.

Topology-Based Work Zone Scene Inference: Once work zone objects are identified, their topological arrangement is analyzed. We calculate a topology complexity score based on the positional relationships and connections among these objects. The score serves as an indicator of whether the scene represents an organized work zone or a random accumulation of work zone objects for non-work zone purpose. The score is then fed into a XGBoost classifier, which is trained on ground truth work zone scene data.

Reference-Free Work Zone Size Estimation: After the presence of a work zone is confirmed, an estimation algorithm is performed to approximate the size of the work zone. This is accomplished using a reference-free method, which utilizes the standard heights of common work zone equipment to establish a scale. With this scale, the distances and sizes of the work zone scene can be inferred, allowing for the estimation of the work zone's size.

A. Data-Centric Training Pipeline

The key of work zone detection and sizing lies in accurately recognizing work zone-related objects. For the purposes of this study, we focus on seven key objects, including traffic cones, barricades, barrels, chain fences, construction vehicles, signs, and workers. The YOLOv8 model that integrates cutting-edge backbone and neck architectures with the mosaic augmentation method is used as it enhances

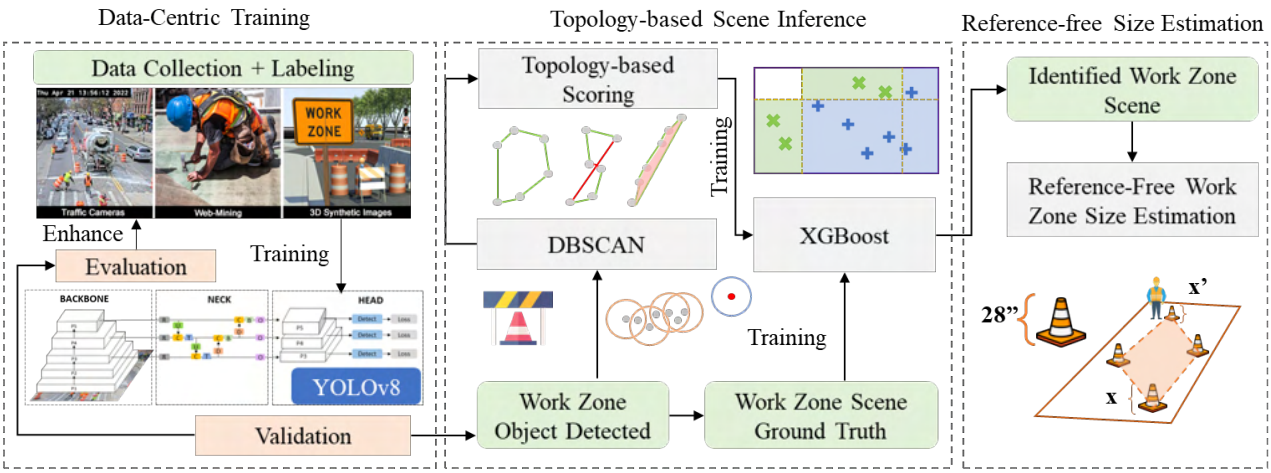

Fig. 1. Flowchart for the urban work zone detection and sizing.

both feature extraction and object detection, compared to previous YOLO versions [8]. Given the absence of a pre-trained YOLOv8 model tailored to our needs due to the lack of publicly available labeled roadway work zone data, we curated and manually labeled our own training set featuring work zone objects from the sources described in Fig. 1.

Data-centric model training recognizes that quality data is key to achieving better model performance, especially when dealing with real-world scenarios that are diverse and often unpredictable. Consider the objective of data-centric training in a simplified manner. Assume that we have a model, defined by its parameters θ, and a dataset $D = (x_i, y_i)$, where x_i is an instance (image in our case) and y_i is the corresponding label. In the typical model-centric training, we want to find optimal parameters θ^* that minimize a loss function L, averaged over all instances in the dataset:

$$\theta^* = argmin_\theta \frac{1}{N} \sum L(y_i, f(x_i; \theta)), \quad for\ all\ i\ in\ D \quad (1)$$

Where $f(x_i; \theta)$ is the output of the model given instance x_i and model parameters θ.

In contrast, for the data-centric training, we recognize that our dataset D itself might be sub-optimal, due to label errors or lack of image diversity. We hence introduce a notion of "dataset quality" $q(D)$, which outlines how good our data is. The goal then becomes to optimize not just the model parameters θ but also the dataset D itself:

$$(\theta^*, D^*) = argmin_{\theta,D} \frac{1}{N} \sum L(y_i, f(x_i; \theta)) - \lambda^* q(D), \quad for\ all\ i\ in\ D \quad (2)$$

Here, λ is a regularization parameter balancing model loss and data quality. Enhancing dataset D might include correcting labeling errors, ensuring data representation, and introducing edge cases for model generalization.

In this study, we refined the dataset based on model performance, adding data and correcting/pruning label errors specifically for subclasses falling short of desired accuracy. The data-centric training pipeline is illustrated in Fig. 2.

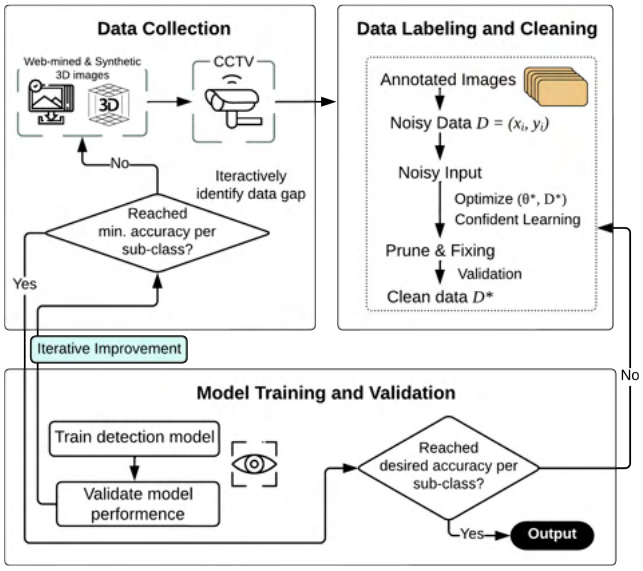

Fig. 2. Data-Centric training pipeline.

B. Topology-based Work Zone Inference using XGBoost

We introduced an innovative topology-based work zone inference using XGBoost to reliably identify work zone scenes under real-world conditions. Instead of relying solely on the presence of work zone objects, this methodology considers their arrangement and inter-connectedness within the work zone. A topology complexity score is derived from these detected objects, illustrating their layout complexity. Work zone objects are clustered using a density-based clustering algorithm, DBSCAN [9]. Any cluster with fewer than three items is considered as noise (i.e., non-work zone).

The topology complexity score views the detected work zone as a graph. From any random vertex, the algorithm searches for the nearest vertex with a degree less than two, then adds an edge between them. This process is repeated until every vertex in the graph is connected, and no vertex has a degree greater than two. The complexity is measured by

the features of the generated graph, including the number of edge cross points, the ratio of cycles to chordless cycles, and the length distribution of edges. These measurements then serve as input variables in XGBoost, a gradient-boosting algorithm known for its efficiency and performance. XGBoost is used as the classifier to identify whether a certain scene represents a work zone or not. The classifier training uses manually annotated ground truth data from traffic cameras. The detailed algorithm is provided in the Algorithm 1.

Algorithm 1 Topology Complexity Score

Require:

Initialize a list of qualified clusters C_q from the results of DBSCAN consists with C.
Initialize an empty Graph $G(V, E)$.
Initialize a list of available vertices V_a consists with all vertices v_a in G having a degree less than or equal to 1.
N is the total number of vertices in C, n is the total number of vertices in G.
Original Vertex: v_o; Destination Vertex: v_d; Number of cycles: N_c; Number of chordless cycles: N_{ch}; Number of crossing edges: N_{ce}

for each cluster C in C_q: **do**
  for Each vertex v in C **do**
    while $n < N$: **do**
      if $G = \varnothing$ **then**
        $v_o \leftarrow v$
      else if Degree of v_d ¡ 2 **then**
        $v_o \leftarrow v_d$
      else
        $v_o \leftarrow \forall v \in V_a$
      end if
      if V_a has more than two vertices **then**
        Find the nearest v_a and set it as v_d
        add an edge between v_o and v_d
        update G and V_a
      else if V_a has only two vertices **then**
        Set the last v_a other than v_o as v_d
        Add an edge between v_o and v_d
        Update G and V_a
      end if
    end while
    Calculate N_c, N_{ch}, N_{ce} and the length of edges in G, store these attributes.
  end for
  Final graph $G_f \leftarrow argmin_{N_c}(G)$
  Obtain topology scores from G_f
end for

C. Reference-free Work Zone Size Estimation

To overcome variance in camera positioning and perspective, we suggest a reference-free methodology for work zone size estimation, based on standard heights of common work zone equipment such as traffic cones. This method, initially proposed from our previous work [10] for pedestrian detection, eliminate the need for a physical scale reference in the scene which usually requires on-site human investigators.

The reference-free method divides the image into several hyper-planes, oriented perpendicularly to the horizontal plane and vanishing lines. Due to perspective effects, each hyper-plane exhibits a unique real-to-pixel distance ratio (RP-rate) as the number of pixels corresponding to a given real-world length varies between hyper-planes. Next, we propose that each specific type of work zone equipment in the image stands perpendicular to the horizontal plane and maintains a uniform actual height h_r. Fig. 3 provides an example illustrating the proposed area estimation method. Subsequently,

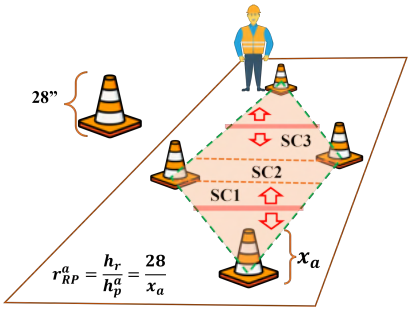

Fig. 3. Proposed work zone size estimation method.

the detected object (vertex) is ranked according to the RP rate, and one edge is then added to the adjunct vertices. Combining the final graph G_f generated from Algorithm 1 and removing the duplicated edges, a set of sub-cycles are generated. The calculation of the real distance is considered as the integration of the product between RP-rate and pixel distance Δp [10]. Similarly, the real area size of each chordless cycle can be formulated as the integration of the product between RP rate and the pixel area size Δs. Given that each cycle comprises at least three vertices, it's straightforward to divide a cycle into sub-cycles between two neighboring vertices using a supplementary vertical line. A horizontal line is added from each vertex inside the generated work zone area, and then the area is separated into sub-region (SC1 to SC3 in Fig. 3). Each sub-region is considered as many small rectangles, and the length of the rectangle is the real distance between the region's bounds, l, which can be calculated using the pixel distance multiplied by the RP rate. The rectangle width is calculated using the method introduced in [10] due to the perspective effect. Thus, each sub-region can be calculated as a double integral that is written as:

$$A_{SC} = \int_a^b \int_{r_a}^{r_b} lr \; dldr \tag{3}$$

where the a and b are the two vertices, l_a and l_b are the horizontal distance between the work zone bounds at vertex a and b; r_a and r_b are the RP rates of vertex a and b, which can be calculated as follow:

$$r^i_{RP} = \frac{h_r}{h^i_p} \tag{4}$$

where h_p^i is the pixel height of the detected equipment, and h_r is the real height of the equipment type. In this study, the height of a barrel, a cone and a barricade is assumed to be 37, 28, and 42 inches, respectively.

IV. Experiments and Results Analysis

To prove the effectiveness of our proposed framework in urban work zone detection, we evaluated the detection model performance using precision-recall (PR) curve and Mean Average Precision (mAP) over IoU 0.5 (mAP@0.5) based on different training datasets. Then, we assessed the performance of the work zone scene identification and size estimation based on confusion matrix, accuracy and F1 score.

A. Work zone object detection model performance

Our training data primarily derives from a subset of the 900+ fixed CCTV traffic cameras in New York City (nyctmc.org), providing a variety of urban work zone images under different lighting, weather, and traffic conditions, although with relatively low resolution (i.e., 240p). Free stock images sourced from the web and synthetic images from a 3D simulator are also incorporated as supplement data sources in addition to the CCTV images. The free stock images provide high-resolution depictions of specific classes, such as construction workers or vehicles, while the 3D simulator generates synthetic work zone images from various angles and work zone setups under controlled conditions. This approach bridges gaps in areas where real-world data may be scarce. For instance, if there are few examples of night-time work zones in the actual data, these scenarios can be simulated in 3D. This ensures the data-centric training dataset comprehensively covers all possible scenarios, a crucial aspect for training a robust model.

Among the 2,600 images collected and annotated, 890 CCTV, 850 stock, and 280 synthetic 3D images were used as the training data, and 580 CCTV images were used as the test/validation data. We used the YOLOv8 as our model, trained on four customized training models for 300 epochs, and evaluated on the test set. The four training models are: 1) *baseline model* uses original CCTV data without data-centric processing, 2) *DC-CCTV* uses CCTV data with data-centric processing, 3) *DC-CCTV+Stock* uses CCTV and free stock data with data-centric processing, and 4) *DC-CCTV+Stock+3D* uses CCTV, free stock and synthetic 3D data with data-centric processing.

Fig.4 displays the PR curve resulting from different training model combinations. The performance, compared to the baseline, saw substantial improvement after implementing data-centric training in DC-CCTV, which included data cleaning and re-labeling. Furthermore, when additional data sources were added to the training data pool, overall model performance increased due to data enhancement processing (Fig.4 (c) and (d)), especially for subclass barricade, construction vehicle, work zone sign, and chain fence. Table I presents the mAP@0.5 scores for each work zone object class. Interestingly, a marginal decline is observed in performance for certain specific types upon data augmentation. A possible explanation is that the inclusion of stock and synthetic 3D images broadened the model's generalization thus reduced the risk of overfitting.

TABLE I
Training Performance Comparison

Detection Type	Base	Data-Centric Training		
	CCTV	***CCTV***	***CCTV+Stock***	***CCTV+Stock+3D***
barricade	0.572	0.644	0.680	0.725
worker	0.516	0.610	0.620	0.579
const. veh	0.272	0.675	0.700	0.737
cone	0.692	0.755	0.773	0.764
barrel	0.830	0.871	0.882	0.872
sign	0.424	0.750	0.727	0.799
chain fence	0.291	0.655	0.680	0.713
ALL	**0.514**	**0.708**	**0.723**	**0.741**

All values are mAP@0.5

B. Work zone scene identification model performance

For XGBoost training, we selected 684 CCTV images, 399 containing unique work zones and 285 containing work zone objects that do not constitute work zones (e.g., cones for lane control). Features like work zone shape, equipment count, number of crossing edges, the ratio of the cycle over the chordless cycle, and the number of edge outliers were manually labeled for model training. We tested the trained model on 853 CCTV images, with the confusion matrix presented in Table II. Upon analyzing the confusion matrix, it is noted that the accuracy of the model is 98.4%, whereas the F1 score stands at 0.713. The discrepancy between these metrics can largely be attributed to the skewed distribution of the dataset, where the majority of data points are not associated with a work zone. Despite the inherent bias indicated in the F1 score, the results showcase practical applicability and offer satisfactory performance in real-world contexts.

TABLE II
Work Zone Detection Confusion Metrics

	True WZ	**Trun Non-WZ**
Predicted WZ	15	5
Predicted Non-WZ	7	826

WZ: Work Zone

C. Work zone size estimation model performance

Figure 5 presents illustrative examples of work zone inference and area estimation. The first two images accurately identify work zones and their sizes, while the last two images avoid false positives despite the presence of work zone equipment. This demonstrates our model's accuracy even in complex environments. We evaluated our method using ten real-world images, manually estimating work zone sizes via satellite images. The method achieved an accuracy within a range of 67.71% to 89.52%. The processing duration for an individual image spans from 3 to 15 milliseconds, considering input images with dimensions of 640×640 pixels.

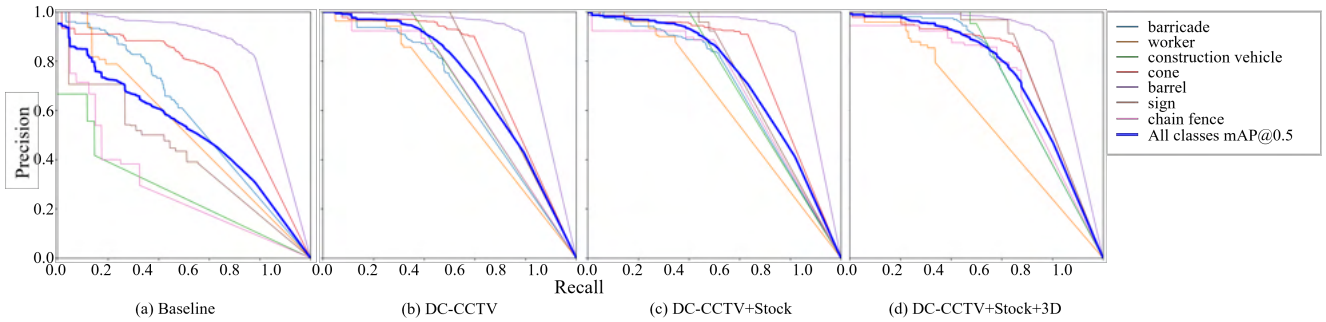

Fig. 4. PR curve of model validation from baseline model and data-centric (DC) trained models using different training data sources combinations.

Fig. 5. Output of work zone detection, inference, and area estimation. The top two images highlight detected work zones in orange, with estimated sizes at the bottom right. The bottom two images show correct detection of non-work zones that contain work zone equipment.

Nevertheless, the area estimation methodology comes with certain limitations that demands future work. First, it functions optimally when the identified boundary equipment forms a closed enclosure. Further improvement is needed for work zones that utilize natural boundaries such as road curbs or walls in addition to typical enclosures like barrels or cones. Also, the reference-free estimation approach assumes uniform heights for each object type, which could introduce bias if actual sizes deviate.

V. Conclusion

In this study, a multi-facet framework for work zone detection was introduced, integrating data-centric AI training, topological analysis, gradient boosting classification, and reference-free size estimation. Through iterative data improvement, label correction, and an innovative topology-based inference method, the approach enables precise work zone scene identification. The individual accuracy and robustness of each module were validated through several experiments.

In summary, the proposed holistic approach, empirically validated in NYC, enables real-time work zone identification and size estimation in complex urban environments. It promises more informed work zone management, improving safety and mobility, and offers a cost-effective mechanism for remote data generation. Future plans include integration with traffic and incident detection and work zone permit databases to enhance Transportation Management Systems (TMS). However, limitations in detecting certain work zones, oversensitivity to specific elements, and the need for topology-based inference tuning must be addressed in future research to fully realize the algorithm's potential.

Acknowledgment

This work was funded by the C2SMART Center, a Tier 1 Center awarded by U.S. DOT under the University Transportation Centers Program. The authors acknowledge the help from Liu Yang, Angela Zhang and Lukelo Luoga for annotating the training dataset. The contents of this paper reflect the views of the authors and the U.S. Government assumes no liability for the contents or use thereof.

References

[1] P. K. Edara, C. Sun, A. Robertson *et al.*, "Effectiveness of work zone intelligent transportation systems," Iowa State University. Institute for Transportation, Tech. Rep., 2013.

[2] P. Edara, R. Rahmani, H. Brown, C. Sun *et al.*, "Traffic impact assessment of moving work zone operations," Smart Work Zone Deployment Initiative, Tech. Rep., 2017.

[3] O. Ozturk, K. Ozbay, and H. Yang, "Estimating the impact of work zones on highway safety," Tech. Rep., 2014.

[4] I. Katsamenis, E. E. Karolou, A. Davradou, E. Protopapadakis, A. Doulamis, N. Doulamis, and D. Kalogeras, "Tracon: A novel dataset for real-time traffic cones detection using deep learning," in *Novel & Intelligent Digital Systems*. Springer, 2022, pp. 382–391.

[5] M. Motamedi, N. Sakharnykh, and T. Kaldewey, "A data-centric approach for training deep neural networks with less data," *arXiv preprint arXiv:2110.03613*, 2021.

[6] N. D. Nath and A. H. Behzadan, "Deep convolutional networks for construction object detection under different visual conditions," *Frontiers in Built Environment*, vol. 6, p. 97, 2020.

[7] R. Duan, H. Deng, M. Tian, Y. Deng, and J. Lin, "Soda: Site object detection dataset for deep learning in construction," *arXiv preprint arXiv:2202.09554*, 2022.

[8] G. Jocher, A. Chaurasia, and J. Qiu, "Yolo by ultralytics," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[9] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[10] F. Zuo, J. Gao, A. Kurkcu, H. Yang, K. Ozbay, and Q. Ma, "Reference-free video-to-real distance approximation-based urban social distancing analytics amid covid-19 pandemic," *Journal of Transport and Health*, vol. 21, p. 101032, 2021.