2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Enhancing Multi-Category Smoke Detection Using Similarity Constrained Loss

Xingyuan Chen[1,+], Ding Xu[2,+], Yuzhe Huang[1], Qishen Chen[1] and Huahu Xu[1(✉)]

[1] Shanghai University, No. 99 Shangda Road, Shanghai, China
[2] Harbin Institute of Technology, No. 92, Xidazhi Street, Harbin City, China
[+] These authors contributed equally to this work.
hhxu25@shu.edu.cn

Abstract. Smoke detection plays a crucial role in ensuring public safety across various domains, including industrial settings, daily life, and disaster management. The effectiveness of smoke detection models heavily relies on the availability of comprehensive datasets and the optimization of loss functions. However, existing smoke detection research primarily focuses on fire-related scenarios, overlooking the significant differences in characteristics between smoke generated by different causes. To address this issue, we have developed a Multi-Category smoke detection dataset (MC-smoke dataset), which is organized based on the smoke's origin and main components. This dataset includes three categories of smoke and contains a total of 1,115 images. Furthermore, to alleviate the loss ambiguity issue present in existing object detection losses, we propose a novel Similarity-Constrained (SC) loss function. This function uses a similarity constraint coefficient in the bounding box to influence center regression and vertex regression losses, enabling more accurate smoke detection. Lastly, extensive experiments were conducted on both the MC-smoke dataset and the classic object detection dataset PASCAL VOC 2007, validating the substantial effectiveness enhancement achieved by the SC loss function. Additionally, comprehensive baseline and comparative experiments were conducted to affirm the suitability of the MC-smoke dataset for research about smoke detection training, testing, and validation.

Keywords: Smoke detection, Loss ambiguity, Bounding box loss, Multi-category smoke.

1 Introduction

Smoke is visible and perceptible, making it a crucial early warning indicator in emergency situations such as fires, traffic accidents, and hazardous chemical leaks [9,6]. Timely detection of smoke is critical to preventing dangers from escalating and causing greater losses. Sensor-based smoke alarm systems have been widely deployed in various settings. While smoke sensors that sample smoke particles exhibit high sensitivity in relatively confined spaces, they become increasingly problematic in open environments due to environmental fluctuations and a sharp decrease in detection capability as

space expands [3]. With the rapid development of computer technology, smoke detection techniques based on computer vision have received significant attention for their wide detection range and fast response speed. They effectively address the drawbacks of traditional smoke sensors and have gradually become a major focus of research.

Before deep learning, image-based smoke detection focused on manual feature extraction. Methods like dark channel prior [21] and bag-of-words model [28] improved segmentation and classification but lacked high detection precision. Deep learning has since advanced smoke detection by leveraging robust feature learning. For example, Gu et al. [10] proposed a deep dual-channel neural network to tackle the problem of changing the texture, color, and shape of smoke. Cao et al. [2] proposed an enhanced feature foreground network for pixel-level modeling of smoke features, to predict the bounding boxes of smoke plumes in videos. Hu et al. [14] introduced a joint weighting strategy for color and texture feature extraction, while Zhan et al. [33] fused visual and semantic features with a recursive feature pyramid. Li et al. [17] proposed a high-precision edge focused forest fire smoke detection network. Wang et al. [29] proposed a lightweight smoke detection network that incorporates refined edge cues, and a mutual context embedding module to improve smoke feature extraction and detection accuracy. Muhammad et al. [22] proposed a lightweight edge-intelligent assisted smoke detection method for detecting smoke in foggy environments.

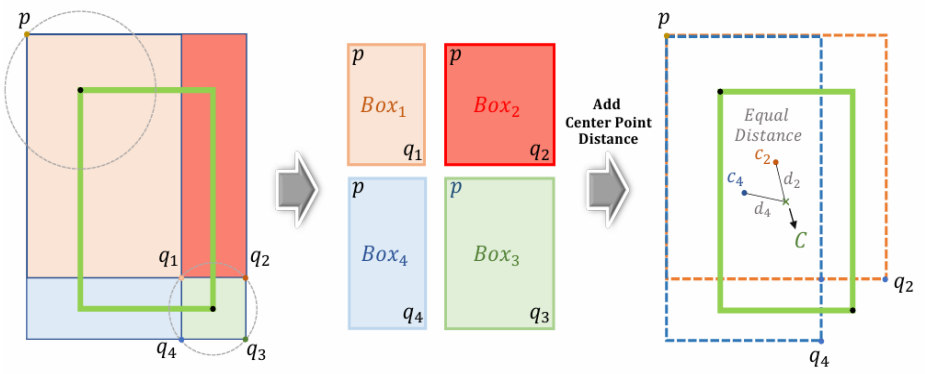


Fig. 1. Loss ambiguity based on positional regression bounding box loss. The bold rectangle represents the ground truth bounding box, with p and q_i, $i \in \{1,2,3,4\}$ denoting the two vertices of the predicted bounding box. c_i, $i \in \{2,4\}$are the centers of the predicted bounding boxes formed by p and q_i respectively, and C is the center of the ground truth bounding box. After introducing the center point distance loss [30], there is still a loss ambiguity problem, that is, c_2 and c_4 are respectively the same distance from C.

Deep learning-based smoke detection models typically determine a bounding box on a two-dimensional image plane that encloses all smoke pixels. However, to achieve a more precise bounding box, it is crucial to design an appropriate loss function for guiding model training. Unfortunately, both regression losses-based on the positions of bounding box vertices and losses based on the Intersection over Union (IoU) metric

suffer from a "loss ambiguity" issue. In other words, different predicted bounding boxes with varying positions and sizes may yield the same loss value. For example, as illustrated in Fig.1, the predicted rectangle defined by point p and the set $\{q_1, q_2, q_3, q_4\}$ may vary in position and size, yet its regression loss relative to the ground truth (depicted as the bold green rectangle) remains unchanged. In other words, the distances from the points to the bottom-right vertex of the ground truth are identical. Consequently, this loss ambiguity may mislead the training process and ultimately degrade detection performance.

In addition, high-quality datasets are essential components for training smoke detection models. Datasets in the field of smoke detection can be divided into real [1,8,18,27,32] and synthetic [13,15] smoke scenes based on their data sources. Most studies [14,17,26,29,33,35] and datasets [1,8,13,15,18,27,32] focus solely on smoke-induced by fires. However, dangerous situations such as collisions and laboratory gas leaks, in addition to combustion, can also produce smoke. Smoke generated by different causes exhibits significant differences in characteristics. Therefore, if a model is capable of accurately classifying smoke types in addition to simply detecting smoke, it would greatly facilitate the identification of the underlying causes of anomalous events and provide a scientific basis for the prompt initiation of targeted emergency response plans. For example, in the event of a chlorine gas leak, the characteristics of the detected smoke would differ markedly from those observed in fire incidents, offering valuable insights for qualitative analysis and the optimization of emergency responses.

In summary, the ambiguity in loss functions for bounding box position regression and the fact that existing datasets focus only on combustion scenes are two urgent problems in the field of smoke detection. To address these issues, we make two main contributions as follow:

- We construct a Multi-Category smoke detection dataset, called MC-smoke. Furthermore, we conduct comprehensive baseline and comparative experiments using the MC-smoke dataset to verify its potential as a new benchmark for research in multi-category smoke detection.
- We introduce a Similarity-Constrained (SC) loss function aimed at mitigating the issue of loss ambiguity by leveraging the similarity and center distance of bounding boxes to improve detection precision.

2 Related Work

2.1 Image-based Smoke Detection Dataset

A high-quality dataset plays a crucial role in deep learning-based smoke detection research, as it directly impacts the model's performance and generalization ability. Regrettably, the number of available datasets for smoke detection is limited, and they frequently focus on burning scenarios like wildfires and structural fires. For example, the Wildfire Observers and Smoke Recognition dataset (WOandSR) [27] and the FireCam dataset [8] exclusively collect smoke produced during combustion, as depicted in Fig.2(d) and Fig.2(c), respectively. Furthermore, the images of these two datasets are

Fig. 2. Image Examples from Different Smoke Detection Datasets. The images from various smoke detection datasets, all of which focus on fire-generated smoke. Furthermore, the single perspective limitation restricts the applicability of these datasets.

captured using telephoto equipment, resulting in reduced resolution of smoke in the images. While video smoke detection (VSD) [32] and USTC-SmokeRS [1] encompass images of diverse smoke categories, the former exhibits low resolution and lacks background information, as illustrated in Fig.2(a), whereas the latter is captured by satellite cameras and only applicable to certain scenarios, as depicted in Fig.2 (b). Consequently, the urgency to construct multi-category datasets for comprehensive smoke detection research has escalated. This enables researchers to assess and refine their models across a broad spectrum of smoke scenarios, thereby enhancing their efficacy in real-world settings.

2.2 Bounding Box Regression Loss Function

Regression losses based on bounding box vertex positions and IoU-based loss functions are two common approaches in object detection. The former treats the four vertices of a bounding box as regression targets, computing coordinate differences between predicted and ground truth boxes using L1, L2, or smooth L1 losses [7,23]. However, this approach typically treats each vertex independently, thereby neglecting their inherent interrelationships and leading to loss ambiguity issues. Wen et al. [30] have introduced center point errors to enhance the discriminative power of deep features; although this modification partially alleviates the problem, loss ambiguity persists, as shown in Fig.1. [25,31,36] measure the matching degree by computing the intersection-over-union between predicted and ground truth boxes, thereby mitigating issues caused by scale and shape variations and providing more intuitive guidance for adjusting bounding box positions and sizes. Nevertheless, in special cases such as coinciding center points, these loss functions also encounter loss ambiguity. Moreover, even when other loss functions yield ambiguous results, our proposed SC loss effectively distinguishes between different bounding boxes.

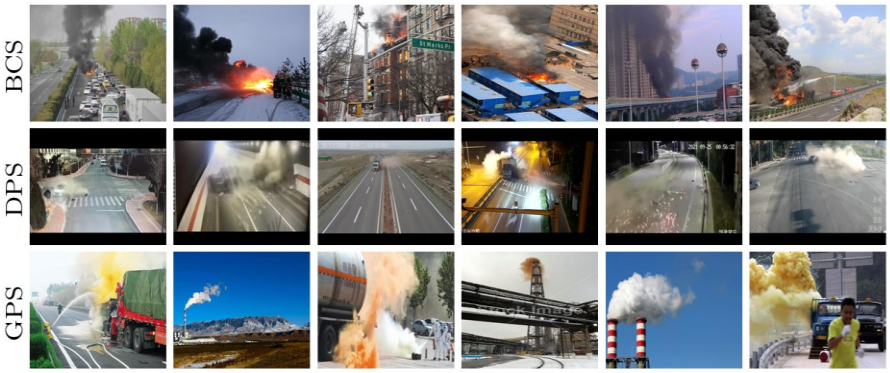


Fig. 3. Overview of the MC-Smoke dataset. "BCS": Burning Combustion Smoke, "DPS": Dust Particulate Smoke, "GPS": Gas Pollutant Smoke.

3 Proposed Dataset and Loss Function

3.1 Multi-category Smoke Dataset

Statement: The term "smoke" is commonly used to refer to particulate matter in the air. However, at the microscopic level, the particulate composition of smoke varies significantly depending on its source. These differences in the primary components lead to substantial variations in smoke attributes, including color, shape, and concentration. Therefore, this paper introduces a novel multi-category smoke detection dataset, as shown in Fig.3, aimed at encouraging researchers to explore the diversity of smoke in order to address different scenarios effectively.

Collection: The occurrence of anomalies is inherently random, which poses significant challenges in smoke data collection. To address this, we decided to gather smoke images from various sources, including the internet, search engines, and news articles, ensuring diversity in terms of categories, lighting conditions, and real-world scenarios. Although we collected as much data as possible, objective constraints such as time and the scarcity of original materials resulted in far fewer images of smoke not produced by combustion compared to those generated by combustion. Consequently, our dataset exhibits a long-tailed distribution.

Preprocessing: As with other classic datasets[1,4], the image resolution was uniformly modified to 640 × 480, and manual annotation methods were used to label the images. To ensure the quality of the collected images, we screened out duplicates and non-smoke images from the dataset, leaving a total of 1,115 pictures for annotation.

Properties: ①Viewpoints: The dataset includes smoke images captured from various viewpoints, providing a richer set of features for model training. ② Classification: The 1,115 smoke images are classified into three categories: "BCS" represents smoke produced by combustion (921 images), "DPS" refers to smoke from dust and particulate matter (90 images), and "GPS" denotes smoke from gaseous pollutants (104 images).

③ Label: The annotations were saved in an XML file following PASCAL VOC 2007 [4] format specification.

3.2 Similarity-Constrained Loss

Research in [36] suggests that an effective bounding box loss function should incorporate the center position and area of the bounding box. Besides, we contend that the shape of the bounding box can also serve as a crucial constraint. To address this issue, we propose to improve the smooth L1 loss by using the similarity of bounding boxes and name the improved loss function as Similarity-Constrained (SC) loss. The SC loss consists of the similarity constraint coefficient, bounding box vertex position regression, and bounding box center position regression, as shown in (1), where $\mathcal{L}_{\mathcal{V}}$ represents the distance between the vertices of the predicted bounding box and those of the ground truth, and $\mathcal{L}_{\mathcal{C}}$ represents the offset distance between the center point of the predicted bounding box and those of the ground truth to supplement the lack of information about smooth l1 loss. Finally, ζ represents the similarity constraint coefficient, that is, the position regression of the former two is constrained by the similarity between the predicted bounding box and the ground truth bounding box, so as to optimize the loss ambiguity problem.

$$\mathcal{L} = (\mathcal{L}_{\mathcal{V}} + \mathcal{L}_{\mathcal{C}}) \times \zeta \tag{1}$$

Position Regression: When computing the bounding box regression loss, we usually use the coordinates of the Left-Top (LT) vertex and the Right-Bottom (RB) vertex to represent any bounding box. Therefore, as shown in (2), the $\mathcal{L}_{\mathcal{V}}$ can be obtained by calculating the difference between the components of B_{gt} and B.

$$\mathcal{L}_{\mathcal{V}}(B_{gt}, B) = \sum_{i=1}^{4} \begin{cases} 0.5(a - a'), if\ |a - a'| < 1 \\ |a - a'| - 0.5, otherwise \end{cases} \tag{2}$$

where a and a' denote each corresponding component of B_{gt} and B. This formula avoids the problem of non-smoothness at the zero point and is more robust to outliers [7]. However, $\mathcal{L}_{\mathcal{V}}$ only includes the distance of the corresponding vertex components, ignoring the potential correlation information between vertex, and correlation information is also an important feature in deep learning, which helps to improve the performance of the model. Therefore, we introduced the measurement of the distance between the center of the bounding box.

The center of the bounding box can be calculated from the vertexes, which can reflect the position information of the bounding box globally, and can also contain the correlation information between the vertexes. Therefore, we introduce the center distance loss $\mathcal{L}_{\mathcal{C}}$ to enhance the associated information in SC loss. The loss function is the Euclidean distance between the ground truth box center and the predicted box center, as shown in (3).

$$\mathcal{L}_{\mathcal{C}}(B_{\mathrm{gt}}, B) = \sqrt{(x_c - x_c')^2 + (y_c - y_c')^2} \tag{3}$$

Similarity Constraint Coefficient: In order to optimize the problem of bounding box loss ambiguity, we introduced the similarity constraint coefficient ζ on the basis of position regression. Its core is to use the similarity between the predicted box and the ground truth box to constrain position regression.

The acquisition of the similarity constraint coefficient can be divided into three steps. Firstly, it is necessary to calculate the diagonal vectors of the predicted box ($\boldsymbol{p}$) and the ground truth box (g) respectively, as shown in (4)-(5). The diagonal vector not only contains the relationship information between vertices but also contains the characteristics of the bounding box. The two vector cosines are then calculated, as shown in (6). Notably, the included angle between the predicted box and the ground truth box diagonal vector ranges from 0° to 90°.

$$\boldsymbol{p} = P_{lt} - P_{rb} \tag{4}$$

$$g = P_{lt}^{gt} - P_{rb}^{gt} \tag{5}$$

$$\cos(p, g) = \frac{x_p x_g + y_p y_g}{\sqrt{x_p^2 + y_p^2} \cdot \sqrt{x_g^2 + y_g^2}} \tag{6}$$

Finally, convert the cosine value into the similarity constraint coefficient ζ, as shown in (7). The hyperparameter λ is set to 2, so the value range of the correlation coefficient is limited to 1 to 2. That is, the smaller similarity constraint coefficient indicates a closer match in shape and size between the predicted bounding box and the ground truth box. When two bounding boxes are completely similar, the model only needs to focus on the position regression of the vertices and center points of the bounding boxes.

$$\zeta = \lambda - \cos(p, g) \tag{7}$$

4 Experimental Results and Analysis

We conducted experiments following [36]. First, we performed baseline and comparative experiments on the MC-smoke dataset to evaluate its relevance to smoke detection and the impact of SC loss on model performance. Next, ablation experiments were conducted to assess the contributions of the three modules in the SC loss function. Finally, we evaluated the SC loss on the PASCAL VOC 2007 dataset using both classic and lightweight object detection models, alongside common losses, to further assess its effectiveness.

4.1 Implementation details And Evaluation Protocol

Dataset. Due to the imbalance in the MC-smoke dataset, we sampled 80% of the instances from each category to form the training set and reserved the remaining 20% for the test set. This approach ensures that the class distribution is maintained across both subsets, thereby reducing potential biases during model training and evaluation. In

Table 1. Baseline Experiment On MC-Smoke Dataset

Method	Params(M)	mAP^{50}	mAP^{75}	mAP
SSD [19]	26.3	77.26	53.69	46.85
Yolo v3 [5]	9.6	84.77	35.17	37.07
Mask RCNN [11]	44.5	91.54	48.50	49.76
Dynamic-Faster RCNN [34]	42.3	89.23	50.63	48.73
PAA [16]	37.5	89.96	59.93	51.90
Faster RCNN-Resnet50 [24]	44.1	89.44	51.15	47.37
Faster RCNN-SwinT [20]	48.3	95.18	64.05	56.96

addition to the proposed MC-smoke dataset, we also used the PASCAL VOC 2007 dataset [4] to evaluate the proposed SC loss. The PASCAL VOC 2007 dataset is widely used for object detection and includes 20 classes with a total of 9,963 images, comprising 5,011 training images and 4,952 test images.

Hyperparameters: To ensure a fair comparison between different methods, we used common hyperparameters in all experiments. Specifically, each experiment iterated for 200 epochs and used a batch size of 8 for each training. We selected AdamW as the optimizer and set the initial learning rate to 1e-4, and decay to 5e-4.

Evaluation Protocol: In object detection, accuracy is evaluated using the Intersection over Union (IoU) metric, which measures the overlap between predicted (A_{pre}) and ground truth (A_{gt}) boxes by dividing their intersection area by their union, as shown in (8).

$$\text{IoU} = \frac{A_{\text{gt}} \cap A_{\text{pre}}}{A_{\text{gt}} \cup A_{\text{pre}}} \tag{8}$$

A prediction is considered correct if IoU exceeds a set threshold, such as 0.5 for mAP^{50}. To evaluate overall performance, mean average precision (mAP) across IoU thresholds from 0.5 to 0.95, with higher mAP scores indicating better results. In experiments, mAP^{50}, mAP^{75}, and mAP are used as performance measures.

4.2 Experiments on MC-smoke

Experiments on MC-smoke: We evaluated six object detection algorithms with L1 loss on the MC-Smoke dataset: the single-stage models SSD [19] and Yolo V3 [3], and the models Faster RCNN [5], MaskRCNN [11], Dynamic-Faster RCNN [34], and PAA [16], which enhances performance by optimizing anchor assignment rules. Notably, SSD uses VGG as its backbone, YoloV3 uses DarkNet, and the others use ResNet50 [12]. To assess the impact of different backbone networks, we replaced the Faster RCNN's backbone with Swin Transformer [20]. The final results are presented in Table 1. Among the four models based on ResNet50, performance differences were minimal. The Swin Transformer-based Faster RCNN achieved the best performance, with mAP^{50}, mAP^{75}, and mAP scores of 95.18, 64.05, and 56.96, respectively, significantly

Table 2. Comparison of Faster RCNN-SwinT using different loss functions on MC-smoke Dataset. (*Baseline value for relative improvement.)

Losses	mAP^{50}	mAP^{75}	mAP
IoU Loss [31]	94.86	49.35	51.49
CIoU Loss [36]	94.94	57.49	53.32
L1 Loss [11]	95.18	64.05	56.936*
Smooth L1 Loss [24]	95.62	65.08*	55.33
SC Loss (Ours)	**96.39***	**74.09**	**61.72**
Relative improv (%)	**0.81%**	**13.84%**	**8.35%**

higher than the other models. We attribute this to Swin Transformer's powerful feature extraction capabilities, which model both global and local features, thereby enhancing overall performance.

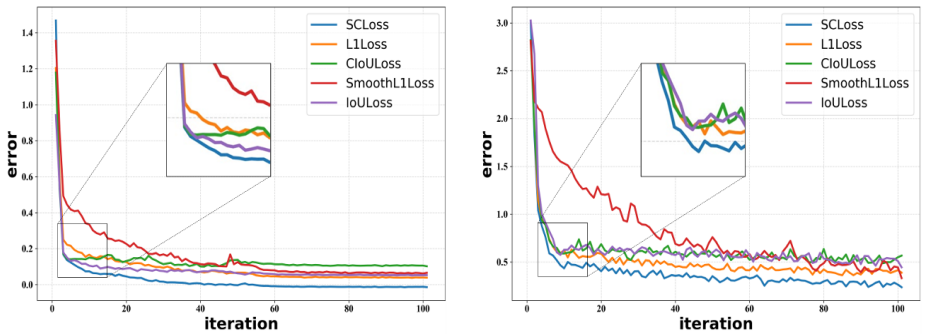
Similarity-Constrained Loss on MC-smoke: In this section, we integrate four popular object detection loss functions, which include L1, Smooth L1 [24], IoU [31], and CIoU [36] losses, into the Swin-Transformer-based Faster RCNN and compare with the proposed similarity-constraint (SC) loss to further verify the effectiveness of the SC loss.

As shown in Table 2, the model trained with the proposed SC loss achieved the highest scores in all three evaluation metrics. Specifically, the mAP^{50}, mAP^{75}, and mAP of the model based on the SC loss function were 96.39, 74.09, and 61.72, respectively. When the IoU threshold was set to 0.5, the model based on the SC loss function outperformed the other models by only 0.77, indicating that other loss functions can achieve satisfactory results when the accuracy requirement is low. However, when the IoU threshold was set to 0.75, the model based on the similarity loss function showed a significant improvement, with a 9.01 higher score than the second-best model.

Table 3. Ablation Experiment on MC-smoke Dataset.

No.	$\mathcal{L}_v$	$\mathcal{L}_c$	ζ	mAP^{50}	mAP^{75}	mAP
1	√	×	×	95.62	65.08	55.33
2	√	×	√	93.66	60.86	53.62
3	√	√	×	96.15	70.06	57.57
4	×	√	×	91.62	49.98	48.90
5	×	√	√	94.57	71.44	57.77
6	√	√	√	**96.39**	**74.09**	**61.72**

Ablation Experiment on MC-smoke: To assess the influence of the three components in the SC loss function-center regression loss $\mathcal{L}_c$, vertex regression loss $\mathcal{L}_v$, and similarity constraint coefficient ζ-on the overall performance of the loss function, we conducted two ablation experiments. In the initial experiment, the model was trained solely using the vertex regression loss function. Subsequently, we combined the center regression loss and similarity constraint coefficient with the vertex regression loss for further training, and compared the experimental outcomes, detailed in Table 3 (No.1-No.3). Combining all three modules resulted in the model achieving its optimal performance. Additionally, combining the center regression loss with the vertex regression loss led



(a) DPnet (b) Faster RCNN

Fig. 4. Variation of Training Loss with Iterations. Two distinct models [24,37] were employed for training with five different loss functions.

to a marginal improvement in model performance, with the mAP^{75} evaluation index increasing by 4.98. Conversely, combining the vertex regression loss with the similarity constraint coefficient alone resulted in a slight decrease in the scores of all three evaluation indicators, with reductions of 1.96, 4.22, and 1.71 compared to the baseline experiment, respectively.

To further assess the impact of the similarity constraint coefficient, we conducted a second ablation experiment, with the results presented in Table 3 (No.4-No.5). In this experiment, we maintained the center regression loss unchanged and introduced the similarity constraint coefficient and vertex regression loss for further experimentation. The results indicated that while the model's performance trained solely on the center regression loss was subpar, combining it with the similarity constraint coefficient led to significant improvement. The scores of the three evaluation indicators increased by 2.95, 21.46, and 8.87, respectively. Nonetheless, the performance did not surpass that of the model when all three modules were combined simultaneously.

Table 4. Loss Function Comparison on PASCAL VOC 2007.

Losses.	DPNet (2.5M)			Faster RCNN (44M)		
	mAP^{50}	mAP^{75}	mAP	mAP^{50}	mAP^{75}	mAP
IoU Loss [31]	83.06	47.62	48.73	74.13	46.22	44.28
CIoU Loss [36]	85.11	51.74	51.16	74.43	46.76	44.60
L1 Loss [11]	81.56	49.94	47.95	74.15	45.96	43.60
Smooth L1 Loss [24]	85.84	51.10	50.21	74.21	45.75	43.61
SC Loss (Ours)	**87.24**	**53.33**	**52.02**	**74.94**	**46.80**	**43.33**
Relative improv (%)	**1.63%**	**3.07%**	**1.68%**	**0.6%**	**0.08%**	**1.08%**

4.3 Experiment on PASCAL VOC 2007

As shown in Table 4, we evaluate the effectiveness of the proposed SC Loss by integrating it into both a lightweight model (DPNet [37]) and a classical detection model (Faster RCNN [24]) on the PASCAL VOC 2007 dataset. The results demonstrate that SC Loss consistently outperforms conventional loss functions when used with DPNet, achieving notable improvements across all metrics. Specifically, SC Loss improves the mAP^{50} by 1.63%, the mAP^{75} by 3.07%, and the overall mAP by 1.68% compared to the second-best loss function, showcasing its ability to better constrain model training and mitigate the effects of regression ambiguity. In contrast, when applied to the Faster RCNN model, SC Loss, although it did not achieve the same magnitude of improvement, still outperformed the other four loss functions in terms of performance. The error convergence curves in Fig.4 demonstrate that the proposed SC Loss achieves the most stable and consistent error reduction across both DPNet and Faster RCNN.

The results on the PASCAL VOC 2007 dataset and MC-Smoke dataset demonstrate that the proposed similarity-constrained loss function outperforms traditional loss functions by optimizing the loss ambiguity problem, achieving lower final error values and greater stability, and validating its effectiveness in improving object detection performance.

5 Conclusion

In this paper, we introduce the SC loss, which refines bounding box regression by incorporating both similarity and center distance, thereby addressing loss ambiguity and enhancing model performance. Additionally, we present a multi-category smoke dataset by smoke characteristics, enabling researchers to explore the complexities of multi-category smoke detection. In the future, we aim to further improve the performance of the lightweight model, facilitating its deployment on monitoring terminals for faster and more accurate smoke detection.

References

1. Ba, R., Chen, C., Yuan, J., Song, W., Lo, S.: Smokenet: Satellite smoke scene detection using convolutional neural network with spatial and channel-wise attention. Remote Sens. 11(14), 1702 (2019)
2. Cao, Y., Tang, Q., Wu, X., Lu, X.: Effnet: Enhanced feature foreground network for video smoke source prediction and detection. IEEE Trans Circuits Syst Video Technol 32(4), 1820–1833 (2021)
3. Chaturvedi, S., Khanna, P., Ojha, A.: A survey on vision-based outdoor smoke detection techniques for environmental safety. ISPRS Journal of Photogrammetry and Remote Sensing 185, 158–187 (2022)
4. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. Int J Comput Vis 88, 303–308 (2009)
5. Farhadi, A., Redmon, J.: Yolov3: An incremental improvement. In: IEEE Conf. Comput. Vis. Pattern Recognit. vol. 1804, pp. 1–6. Springer Berlin/Heidelberg, Germany (2018)

6. Gaur, A., Singh, A., Kumar, A., Kumar, A., Kapoor, K.: Video flame and smoke based fire detection algorithms: A literature review. Fire technology 56, 1943–1980 (2020)
7. Girshick, R.: Fast r-cnn. In: Proc IEEE Int Conf Comput Vis. pp. 1440–1448 (2015)
8. Govil, K., Welch, M.L., Ball, J.T., Pennypacker, C.R.: Preliminary results from a wildfire detection system using deep learning on remote camera images. Remote Sens. 12(1), 166 (2020)
9. Gragnaniello, D., Greco, A., Sansone, C., Vento, B.: Fire and smoke detection from videos: A literature review under a novel taxonomy. Expert Systems with Applications p. 124783 (2024)
10. Gu, K., Xia, Z., Qiao, J., Lin, W.: Deep dual-channel neural network for image based smoke detection. IEEE Trans Multimedia 22(2), 311–323 (2019)
11. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conf. Comput. Vis. Pattern Recognit. pp. 770–778 (2016)
13. He, L., Gong, X., Zhang, S., Wang, L., Li, F.: Efficient attention based deep fusion cnn for smoke detection in fog environment. Neurocomputing 434, 224–238 (2021)
14. Hu, Y., Zhan, J., Zhou, G., Chen, A., Cai, W., Guo, K., Hu, Y., Li, L.: Fast forest fire smoke detection using mvmnet. KNOWL-BASED SYST 241, 108219 (2022)
15. Khan, S., Muhammad, K., Mumtaz, S., Baik, S.W., de Albuquerque, V.H.C.: Energy-efficient deep cnn for smoke detection in foggy iot environment. IEEE Internet of Things Journal 6(6), 9237–9245 (2019)
16. Kim, K., Lee, H.S.: Probabilistic anchor assignment with iou prediction for object detection. In: Proc. Eur. Conf. Comput. Vis. pp. 355–371. Springer (2020)
17. Li, R., Hu, Y., Li, L., Guan, R., Yang, R., Zhan, J., Cai, W., Wang, Y., Xu, H., Li, L.: Smwe-gfpnnet: A high-precision and robust method for forest fire smoke detection. Knowledge-Based Systems 289, 111528 (2024)
18. Lin, J., Fu, C., Huang, Q., Zhu, Y.: Contextual interaction enhancement network for smoke detection. In: 2024 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE (2024)
19. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: Eur. Conf. Comput. Vis. pp. 21–37. Springer (2016)
20. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proc IEEE Int Conf Comput Vis. pp. 10012–10022 (2021)
21. Long, C., Zhao, J., Han, S., Xiong, L., Yuan, Z., Huang, J., Gao, W.: Trans mission: a new feature for computer vision based smoke detection. In: Artificial Intelligence and Computational Intelligence: International Conference, AICI 2010, Sanya, China, October 23-24, 2010, Proceedings, Part I 2. pp. 389–396. Springer (2010)
22. Muhammad, K., Khan, S., Palade, V., Mehmood, I., De Albuquerque, V.H.C.: Edge intelligence-assisted smoke detection in foggy surveillance environments. IEEE Trans Industr Inform 16(2), 1067–1075 (2019)
23. Qi, C.R., Liu, W., Wu, C., Su, H., Guibas, L.J.: Frustum pointnets for 3d object detection from rgb-d data. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 918–927 (2018)
24. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. 28 (2015)

25. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: A metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 658–666 (2019)
26. Sathishkumar, V.E., Cho, J., Subramanian, M., Naren, O.S.: Forest fire and smoke detection using deep learning-based learning without forgetting. Fire ecology 19(1), 9 (2023)
27. Toni Jakovcevic, D.K.: Wildfire observers and smoke recognition (Mar 2023), http://wildfire.fesb.hr/
28. Vidal-Calleja, T.A., Agammenoni, G.: Integrated probabilistic generative model for detecting smoke on visual images. In: 2012 IEEE Int Conf Dev Learn Epigenetic Robot. pp. 2183–2188. IEEE (2012)
29. Wang, J., Zhang, X., Zhang, C.: A lightweight smoke detection network incorporated with the edge cue. Expert Systems with Applications 241, 122583 (2024)
30. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: Computer vision–ECCV 2016: 14th European conference, amsterdam, the netherlands, October 11–14, 2016, proceedings, part VII 14. pp. 499–515. Springer (2016)
31. Yu, J., Jiang, Y., Wang, Z., Cao, Z., Huang, T.: Unitbox: An advanced object detection network. In: ACM SIGMM Rec. pp. 516–520 (2016)
32. Yuan, F.: Video smoke detection (2023)
33. Zhan, J., Hu, Y., Zhou, G., Wang, Y., Cai, W., Li, L.: A high-precision forest fire smoke detection approach based on argnet. COMPUT ELECTRON AGR 196, 106874 (2022)
34. Zhang, H., Chang, H., Ma, B., Wang, N., Chen, X.: Dynamic r-cnn: Towards high quality object detection via dynamic training. In: Eur. Conf. Comput. Vis. pp. 260–275. Springer (2020)
35. Zhang, L., Lu, C., Xu, H., Chen, A., Li, L., Zhou, G.: Mmfnet: Forest fire smoke detection using multiscale convergence coordinated pyramid network with mixed attention and fast-robust nms. IEEE Internet of Things Journal 10(20), 1816818180 (2023)
36. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D.: Distance-iou loss: Faster and better learning for bounding box regression. In: Proc Conf AAAI Artif Intell. vol. 34, pp. 12993–13000 (2020)
37. Zhou, Q., Shi, H., Xiang, W., Kang, B., Wu, X., Latecki, L.J.: Dpnet: Dual-path network for real-time object detection with lightweight attention. arXiv: CoRR (2022)