ISSN: 2710-0588

Small Object Geological Carbonate Detection Algorithm Based on YOLOX

Junpeng Shi[1, *]

[1]Chengdu University of Technology, Chengdu, China

[*] Corresponding Author

Abstract

Detection of small object Carbonates poses a challenging task, primarily due to the minuscule nature of Carbonates making thcem difficult to distinguish from the background. Traditional methods often struggle when faced with these small Carbonates, as their scale is small and they exhibit minimal differences from the background, resulting in challenges in accurate detection and classification. To address this issue, this study proposes an Geological small object Carbonate detection algorithm based on spatial attention combined with self-attention mechanisms. This algorithm first utilizes spatial attention to assist the model in focusing on the regions of interest containing small object Carbonates, thereby reducing background interference and increasing attention towards small object Carbonates. Subsequently, the self-attention mechanism is employed to capture long-range dependencies across the entire image, aiding in understanding the relationship between Carbonate regions and the background, thus facilitating better differentiation between Carbonates and background. Finally, the proposed algorithm is evaluated on the public small object dataset TT-100k and the Geological Carbonate dataset NEU, respectively. Experimental results demonstrate that compared to the baseline model, the proposed algorithm achieves an improvement of 2.4% in small object average precision (APsmall) and 3.2% in overall average precision (AP0.5) at IoU=0.5 on the TT-100k dataset; and an improvement of 1.5% in APsmall and 1.8% in AP0.5 on the NEU dataset.

Keywords

Carbonate detection; small objects; attention mechanism.

1. Introduction

In modern society, Geological products ranging from aircraft wings to semiconductor chips are ubiquitous. Geological Carbonate detection, aimed at identifying various appearance flaws in Geological products, is one of the essential technologies for ensuring product quality and maintaining production stability. Traditional Carbonate detection methods, relying on manual inspection, are costly, inefficient, and struggle to cover large-scale quality inspection demands. In recent years, with the rapid advancement of artificial intelligence, Geological Carbonate detection algorithms based on deep learning techniques have made significant progress. These methods, by leveraging deep learning, can reduce the cost of traditional manual inspection, improve detection accuracy and efficiency, and thus play a vital role in smart manufacturing. Consequently, they have emerged as a burgeoning research focus in the field of computer vision. Widely applied in various production scenarios such as object detection in manufacturing workshops and Geological action recognition, Geological Carbonate detection faces numerous challenges compared to general object detection tasks. One of the primary challenges is the small scale of Carbonates. Due to their small size and complex morphology, small Carbonates are often challenging to accurately identify by the naked eye or traditional detection methods.

ISSN: 2710-0588

Therefore, Geological small object detection algorithms typically utilize advanced image processing techniques and deep learning models to capture and analyze high-resolution images of products, enabling precise detection and localization of minute Carbonates.

Based on the aforementioned analysis, this paper proposes the Small Object Area Perception (SOAP) module, which combines spatial attention and self-attention mechanisms. The spatial attention mechanism helps the model focus on the regions of interest containing object Carbonates, thereby reducing background interference and increasing attention towards small object Carbonates. Meanwhile, the self-attention mechanism captures long-range dependencies across the entire image, aiding in understanding the relationship between objects and background and facilitating better differentiation between Carbonates and background. The main contributions of this paper can be summarized as follows:

(1) The proposal of the SOAP module and its integration into the YOLOX baseline model, improving the detection accuracy of small object Carbonates in the object detection model.

(2) To validate the effectiveness of the proposed method, experiments are conducted on the publicly available TT-100k dataset and the NEU dataset. Experimental results demonstrate that on the TT-100k dataset, the proposed method improves the small object average precision (APsmall) by 2.4% and the overall average precision (AP0.5) by 3.2%. On the NEU dataset, the proposed method enhances APsmall by 1.5% and AP0.5 by 1.8%.

The overall structure of this paper is as follows: The second part provides a comprehensive review of current research on Geological small Carbonates. The third part elaborates on the details of the proposed method. The fourth part presents the experimental results of the proposed method on the public TT-100k dataset and NEU dataset. Finally, the fifth part provides a systematic summary of this paper and discusses future work. Organization of the Text.

2. Related Works

2.1. Small Object Detection Algorithms in General Scenes

In general scenes, small object detection has been an important problem in the field of computer vision. Small objects refer to objects in images or videos that are relatively small in size and area. Compared to conventional object detection [13][14][15][16], small object detection is more challenging because they often have lower pixel density and complex background environments, making them easy to ignore or misclassify. Reference [17] proposed an H-DETR hybrid detector based on Transformer and enhanced it for dense small objects, resulting in an accurate and efficient algorithm model. [18] Used the Swin-Transformer [19] to upsample image features and changed the size of the sliding window to adapt to small object objects. The experimental results show that this method combines the bottleneck of Swin-Transformer with CenterNet, improving the detection accuracy of small objects. [20] Proposed an algorithm based on deep learning, HCF-Net. This method significantly improves the performance of infrared small object detection through multiple practical modules. [21] Enhanced the perception of small objects by adding a small object detection layer, improving the performance of locating and recognizing small objects. [22] Proposed a new detection model, IUDet, which transfers the features of the scale dimension to the spatial dimension to enhance information interaction. [23] Effectively improved the detection accuracy of small objects in remote sensing images based on context information and attention mechanism. [24] Proposed a model based on YOLOF, which can improve the accuracy of small object detection when the object scale changes. [25] Proposed a lightweight object detection algorithm for small object detection based on the YOLOv8 algorithm, which improves the accuracy of small object detection through two custom key strategies.

ISSN: 2710-0588

2.2. Carbonate Detection Algorithms in Geological Scenes

Unlike small object detection in general scenes, Carbonates in Geological scenes typically exhibit subtle or complex characteristics, making it difficult to distinguish between Carbonates and backgrounds. In contrast, small objects in general scenes usually have obvious appearance features and are relatively easy to detect and classify. Reference [26] proposed a multi-attention fusion mechanism, which alleviates the problem of unclear Carbonate features and enhances the ability of feature extraction. [27] Proposed a new automatic Carbonate detection algorithm. This algorithm uses a Gaussian mixture model to automatically divide the distribution of positive and negative samples based on similarity scores, effectively detecting Carbonates. [28] Utilized a multi-head self-attention (MHSA) as an independent attention block to enhance the feature extraction capability of the backbone network, improving the detection accuracy of Carbonates. [29] Designed a multi-scale segmentation head based on dynamic kernels, which can perceive Carbonate features in advance, fuse multi-scale Carbonate information, and enhance its ability to identify Carbonate information. [30] Used residual networks to divide VGG-16 into different residual modules, reducing the number of parameter learning iterations in deep residual networks to meet the real-time requirements of product appearance Carbonate detection. [31] Constructed a hybrid deep neural network model based on MobileNetv2 [32], YOLOv4 [33], and Openpose [34] to detect Geological small objects. [35] Introduced a depth information feature fusion module into the YOLOv7 [36] algorithm, thereby improving the model's detection accuracy of small object Carbonates. [37] Proposed the AENet model, which integrates input features at different levels to enhance the representation of Carbonate details.

3. Method

3.1. Overall network architecture

The overall architecture of the network is illustrated in Figure 1. Yolox-nano [38] model was chosen as the baseline model for this study. The selection of Yolox as the baseline model is primarily due to its excellent balance between speed and accuracy. Moreover, the Yolox model demonstrates outstanding versatility and scalability, making it applicable across various Geological scenarios and product categories. Its simple yet flexible network architecture enables easy integration into existing Geological production processes. While Yolox performs well in detecting objects of general scales, there are still limitations when detecting small object Carbonates. Geological Carbonates often exhibit subtle and complex characteristics, making it difficult to distinguish them from the background. Therefore, this study integrates spatial attention and self-attention mechanisms to design the Small Object Area Perception (SOAP) module, aiming to enhance the model's perception of small object Carbonates and prevent them from being treated as background noise by the model. This approach aims to improve the performance and reliability of the model in detecting small object Carbonates.

3.2. Small Object Area Perception module (SOAP)

Conventional object detection models often encounter challenges when dealing with small objects. Firstly, small objects possess limited information due to their small size, making them susceptible to being obscured by background noise, thus rendering them difficult to detect effectively. Secondly, their small size results in a relatively small number of pixels in the image, leading to indistinctive feature representations and making it difficult to distinguish them clearly from the background. Therefore, to address these challenges, this study combines spatial attention mechanisms and self-attention mechanisms to design the Small Object Area Perception (SOAP) module, aiming to enhance the detection performance of small object Carbonates. As illustrated in Figure 2 below.

ISSN: 2710-0588

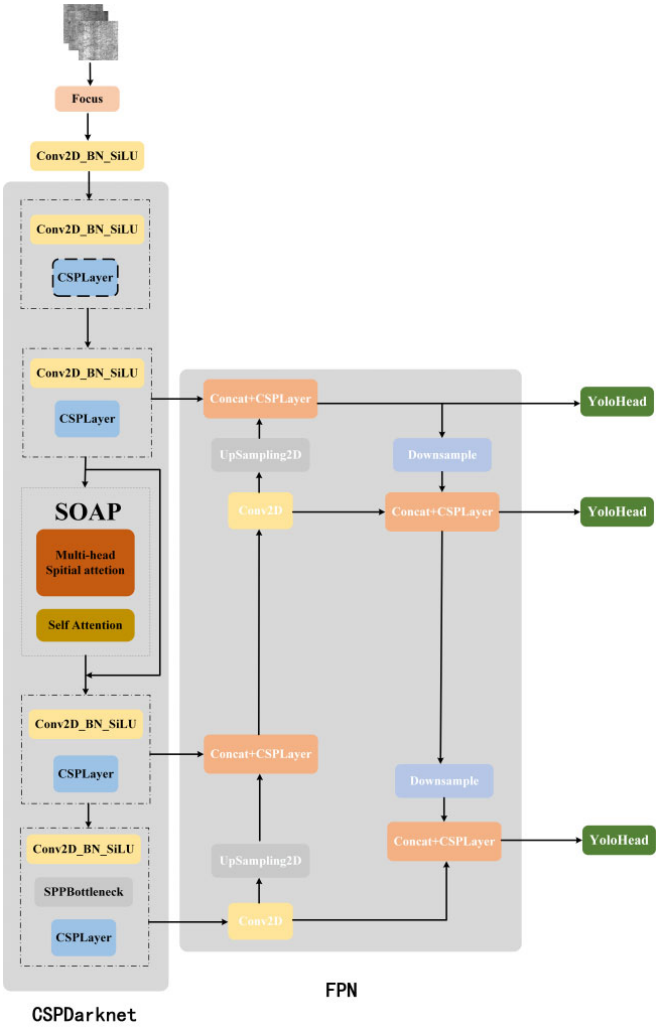

Figure 1. Overall Network Architecture

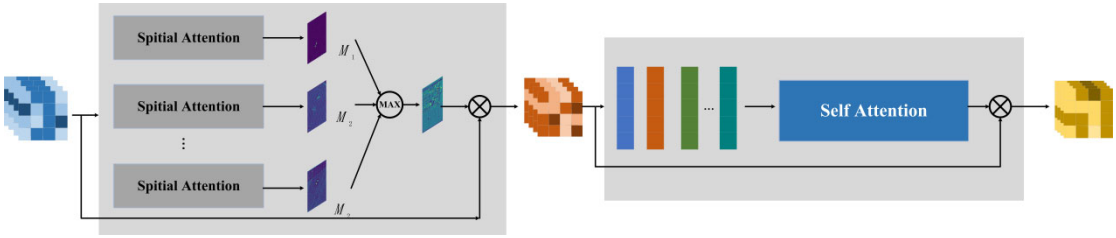

Figure 2. The network architecture of the SOAP module

First, utilizing spatial attention enables capturing spatial dependencies between different positions in the image, allowing the model to simultaneously attend to characteristics of different regions across the entire image space. The input feature map $F \in R^{H\times W\times C}$ is fed into distinct spatial attention modules, enabling it to focus on the small object region from various dimensions, thereby aiding in perceiving the features of the small object region. The computation formula is shown in Equation (3-1):

$$\begin{aligned} M_1 &= Spatial\_Attention(F) \\ M_2 &= Spatial\_Attention(F) \\ &\cdots \\ M_k &= Spatial\_Attention(F) \end{aligned} \tag{1}$$

ISSN: 2710-0588

After obtaining spatial attention matrices of different dimensions, the maximum response portion of each attention matrix image is extracted to acquire the maximum response spatial attention matrix image, which is then applied to the input feature F. The calculation formula is depicted in Equation (2):

$$F' = MAX(M_1, M_2, \dots M_k) \odot F \tag{2}$$

Next, the obtained feature map F′ is passed through the self-attention module. Self-attention dynamically adjusts weights based on the importance of different parts of the input sequence. When perceiving the small object region, the self-attention mechanism can adaptively increase the attention to the small object region, thereby better perceiving the small object region. Initially, the feature map F′ is transformed through three independent linear transformations to generate vectors for query (Q), key (K), and value (V). The calculation formula is presented in Equation (3):

$$\begin{aligned} Q &= F' \cdot W_Q \\ K &= F' \cdot W_K \\ V &= F' \cdot W_V \end{aligned} \tag{3}$$

Where W_Q, W_K, W_V are learned weight matrices. Next, the dot product between query (Q) and key (K) is computed, followed by scaling each dot product to reduce the impact of large values on the softmax operation. The calculation formula is presented in Equation (3-4):

$$Attention(Q, K) = softmax(\frac{QK^T}{\sqrt{d'}}) \tag{4}$$

Where $\sqrt{d'}$ represents the scaling factor. Then, the self-attention weights are applied to the values (V) to obtain the weighted value feature map. The calculation formula is presented in Equation (3-5):

$$Attention\_Output = Attention(Q, K) \cdot V \tag{5}$$

Finally, the obtained attention feature map is applied to F′, resulting in the final output feature map F_o, as shown in Equation (3-6):

$$F_o = Attention\_Output \cdot F' \tag{6}$$

3.3. Loss function

In terms of the loss function, there are three types of loss functions in total, namely classification loss function, localization loss function, and confidence loss function. The classification loss function primarily measures the model's accuracy in classifying object categories. By computing the cross-entropy loss, this loss function penalizes the discrepancy between the model's prediction of object categories and the ground truth labels. The localization loss function primarily evaluates the model's accuracy in regressing object positions and sizes. Through the smooth L1 loss function, this loss function penalizes the difference between the model's prediction of object positions and sizes and the ground truth values. The confidence loss function mainly assesses the model's prediction accuracy of object presence. By employing Focal Loss or binary cross-entropy loss function, this loss function penalizes the discrepancy

ISSN: 2710-0588

between the model's prediction of object presence and the ground truth labels. The formulas for the three different loss functions are shown in Equations (3-7), (3-8), and (3-9):

$$Cls_{loss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} p_{i,c} log\ (\hat{p}_{i,c}) \tag{7}$$

Where $\hat{p}_{i,c}$ is the predicted probability by the model that the i-th anchor box belongs to class c, $p_{i,c}$ is the true label indicating whether the i-th anchor box belongs to class c. N represents the total number of anchor boxes, and C represents the number of classes for each anchor box.

$$Loc_{loss} = \frac{1}{N}\sum_{i}^{N}\sum_{j\in\{x,y,w,h\}} \begin{cases} 0.5\left(\hat{t}_{i,j} - t_{i,j}\right)^2, & if\left|\hat{t}_{i,j} - t_{i,j}\right| < 1 \\ \left|\hat{t}_{i,j} - t_{i,j}\right| - 0.5, & otherwise \end{cases} \tag{8}$$

Where $\hat{t}_{i,j}$ represents the predicted j-th parameter (either unknown or size) of the i-th anchor box, and $t_{i,j}$ represents the true i -th parameter of the i-th anchor box.

$$Conf_{loss} = -\frac{1}{N}\sum_{i=1}^{N} c_i\ log(\hat{c}_i) - (1 - c_i) log\ (1 - \hat{c}_i) \tag{9}$$

Where $\hat{c}_i$ represents the model's confidence prediction for the i -th anchor box, and c_i represents the confidence label of the i-th anchor box.

4. Experiment

We conducted ablation experiments and comparative experiments on both the publicly available TT-100k [39] small object dataset and the Northeastern University (NEU) [40] Carbonate dataset. Additionally, it visualized the feature maps after passing through the small object feature generator and the final detection results.

4.1. Dataset

TT-100k dataset: TT-100k is a public small object dataset for traffic sign detection. The image resolution is 2048×2048. After excluding categories with fewer than 100 samples, there are 45 categories remaining. This paper evaluates the detection performance of objects of different sizes in this dataset, including small-scale objects (area <32×32 pixels), medium-scale objects (32×32 <area <96×96), and large-scale objects (area> 96×96). The proportions of the three scales of objects in the dataset are (42, 50, 8)%. Due to the relatively high proportion of small objects, the TT-100k dataset is one of the best benchmarks for evaluating the performance of small object detection.

The NEU Carbonate dataset consists of six common types of surface Carbonates found in hot-rolled steel strips, including rolling oxide scales, patches, cracks, pitted surfaces, inclusions, and scratches. This dataset comprises a total of 1800 grayscale images, each with an original resolution of 200×200 pixels. Each type of Carbonate contains 300 samples. Similarly, this study evaluated the detection performance of various Carbonate sizes within this dataset.

ISSN: 2710-0588

4.2. Ablation experiment

Table 1. The ablation experiments in NEU dataset (%)

method	AP	$AP_{0.5}$	$AP_{0.75}$	AP_{small}	AR_{small}
Baseline	39.8	76.1	36.9	30.8	54.3
Baseline +SOAP	39.1	77.9	36.9	32.3	56.4

Table 2. The ablation experiments in TT-100k dataset (%)

method	AP	$AP_{0.5}$	$AP_{0.75}$	AP_{small}	AR_{small}
Baseline	42.6	71.9	44.1	24.8	32.5
Baseline +SOAP	44.3	75.1	47.0	27.2	35.6

As shown in Table 1 and Table 2, ablation experiments were conducted on the NEU Carbonate dataset and the TT-100K dataset to validate the effectiveness of the proposed method. The experimental results demonstrate that by incorporating the Small Object Awareness Perception (SOAP) module into the YOLOX-nano baseline model, the detection performance of the model on small objects can be significantly improved. Compared to the baseline model, on the NEU Carbonate dataset, the average precision (AP_{small}) and average recall (AR_{small}) for small objects were improved by 1.5% and 2.1%, respectively. At IoU of 0.5, the average precision was increased by 1.8%. On the TT-100K dataset, the AP_{small} and AR_{small} for small objects were increased by 2.4% and 3.1%, respectively. At IoU of 0.5, the average precision was improved by 3.2%. The results of the ablation experiments demonstrate that the proposed SOAP module effectively enhances the model's perception of small object Carbonate regions.

4.3. Comparative experiment

Table 3. The comparative experiment in NEU dataset (%)

Method	AP	$AP_{0.5}$	$AP_{0.75}$	AP_{small}	AR_{small}
NanoDet-m	26.6	62.7	17.1	28.1	54.6
NanoDet-m-1.5x	27.2	65.2	17.3	25.3	52.2
NanoDet-EfficientNet-Lite0	28.0	68.1	17.2	27.9	54.7
NanoDet-EfficientNet-Lite1	35.2	66.8	32.9	22.4	49.5
Yolov7-tiny	35.9	69.3	33.7	31.7	51.5
Yolov8-nano	**40.7**	75.6	**39.2**	30.1	39.3
our	39.1	**77.9**	36.9	**32.3**	**56.4**

The performance of our method was compared with other classic open-source lightweight object detection models on the NEU Carbonate dataset, and the comparative results are shown in Table 3. The proposed module significantly improves the performance of the detection model by enhancing the model's perception of small object regions. Among the NanoDet series, the models NanoDet-EfficientNet-Lite0 and NanoDet-m achieved the highest average precision (AP_{small}) for small objects and overall $AP_{0.5}$ at 0.5 IoU, reaching 68.1% and 28.1%, respectively. The model NanoDet-EfficientNet-Lite0 achieved the highest average recall (AR_{small}) for small objects at 54.7%. Compared to NanoDet-EfficientNet-Lite0, our method improved AP_{small} for small objects by 4.4% and $AP_{0.5}$ at 0.5 IoU by 9.8%. Compared to NanoDet-m, our method improved AR_{small} for small objects by 4.2%. Compared to Yolov7-tiny, our method improved

ISSN: 2710-0588

$AP_{0.5}$ at 0.5 IoU by 8.6%, AP_{small} for small objects by 0.6%, and AR_{small} for small objects by 4.9%. Compared to Yolov8-nano, our method improved $AP_{0.5}$ at 0.5 IoU by 2.3%, AP for small objects by 2.2%, and AR_{small} for small objects by 17.1%. The comparisons demonstrate the advantages of our method in detecting small object Carbonates. To further validate the generalization of our model, comparative experiments were also conducted on the TT-100k dataset, and the experimental results are shown in Table 4.

Table 4. The comparative experiment in TT-100k dataset (%)

Method	AP	$AP_{0.5}$	$AP_{0.75}$	AP_{small}	AR_{small}
NanoDet-m	37.6	52.0	43.7	17.3	41.8
NanoDet-m-1.5x	37.5	51.8	44.0	20.2	42.4
NanoDet-EfficientNet-Lite0	28.5	40.2	33.5	16.0	41.6
NanoDet-EfficientNet-Lite1	44.0	58.8	51.0	20.6	**45.9**
Yolov7-tiny	**47.7**	**76.8**	**52.0**	23.7	28.4
Yolov8-nano	44.3	71.3	48.0	22.6	27.4
our	44.3	75.1	47.0	**27.2**	35.6

From Table 4, it can be observed that in the NanoDet series, the highest average precision (AP_{small}) for small objects, average recall (AR_{small}) for small objects, and overall $AP_{0.5}$ at 0.5 IoU were achieved by the NanoDet-EfficientNet-Lite1 model, reaching 20.6%, 58.8%, and 45.9%, respectively. Compared to NanoDet-EfficientNet-Lite1, our method improved AP for small objects by 6.6%, $AP_{0.5}$ at 0.5 IoU by 16.3%, and decreased AR_{small} for small objects by 10.3%. Compared to YOLOv7, our method improved AP_{small} for small objects by 3.5% and AR_{small} for small objects by 7.2%, while decreasing $AP_{0.5}$ at 0.5 IoU by 1.7%. Compared to Yolov8-nano, our method improved AP_{small} for small objects by 4.6%, AR_{small} for small objects by 8.3%, and $AP_{0.5}$ at 0.5 IoU by 3.8%. Considering all performance metrics comprehensively, it can be concluded that our method is more suitable for practical applications.

4.4. Visualization results

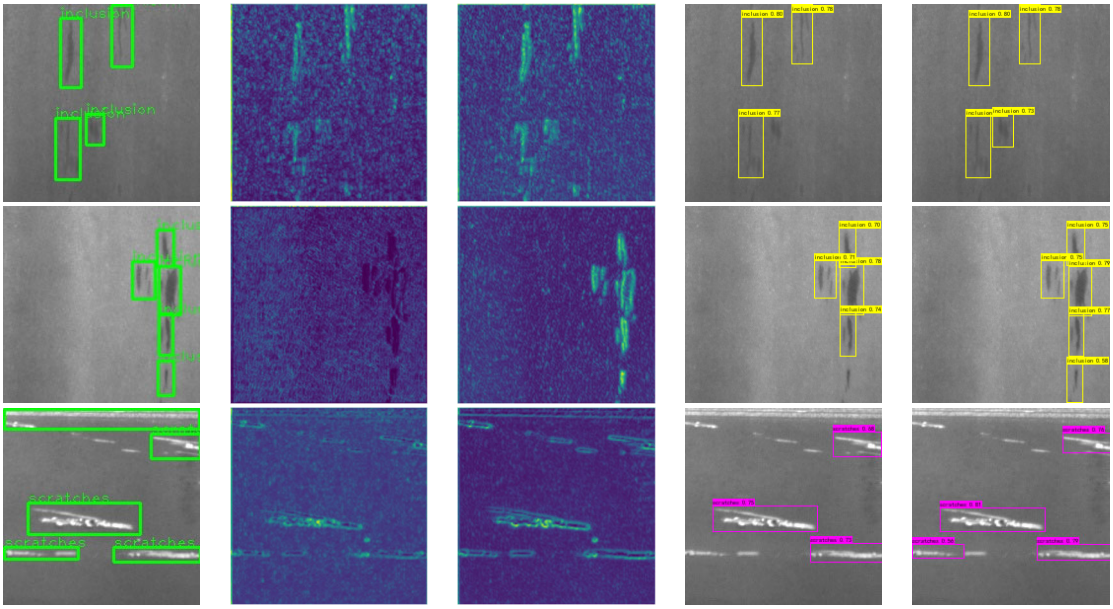

Figure 3. Visualization results

ISSN: 2710-0588

Figure 3 visualizes the results obtained after applying our algorithm to the original Carbonate images. The first column shows the original images, the second column displays the feature maps outputted by the baseline model, the third column presents the feature maps outputted by our algorithm, the fourth column shows the detection results of the baseline model, and the fifth column depicts the detection results of our algorithm. It can be observed that after processing the Carbonate images with our algorithm, the features in the small object regions become more pronounced. Additionally, the number of detected Carbonates increases. Therefore, the effectiveness of the SOAP module proposed in this paper can be inferred.

5. Conclusion

In order to address the challenge of detecting small objects in Geological Carbonate detection, this study introduces the SOAP module by combining spatial attention and self-attention mechanisms to enhance the model's perception of small object regions. Subsequently, experiments were conducted on the publicly available small object dataset TT-100K and the Geological Carbonate dataset NEU to validate the effectiveness of the proposed approach through ablation and comparative studies. In the future, we plan to extend this research method to Carbonate detection in other Geological domains to verify the versatility of the proposed module. Moreover, Geological Carbonates encompass not only the issue of small objects but also challenges related to unclear foreground and background, inter-class similarity, and intra-class variance. Therefore, addressing all these challenges in the Carbonate detection process will be the focus of our future research endeavors.

References

[1] Hou R, Chen G, Han Y, et al. Multi-modal feature fusion for 3D object detection in the production workshop[J]. Applied Soft Computing, 2022, 115: 108245.

[2] Zhang P. Rapid detection method of workshop staff based on deep learning[C]//2021 2nd International Conference on Computer Science and Management Technology (ICCSMT). IEEE, 2021: 243-248.

[3] Łysakowski M, Żywanowski K, Banaszczyk A, et al. Using AR and YOLOv8-Based Object Detection to Support Real-World Visual Search in Geological Workshop: Lessons Learned from a Pilot Study[C]//2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). IEEE, 2023: 154-158.

[4] Ramachandran S, Sistu G, Kumar V R, et al. Woodscape Fisheye Object Detection for Autonomous Driving--CVPR 2022 OmniCV Workshop Challenge[J]. arXiv preprint arXiv:2206.12912, 2022.

[5] Liang Z, Hu Y, Yang R. Research on Geological Human Action Recognition based on improved slowFast[C]//2023 6th International Conference on Intelligent Autonomous Systems (ICoIAS). IEEE, 2023: 94-99.

[6] Wang Z, Yan J. Multi-sensor fusion based Geological action recognition method under the environment of intelligent manufacturing[J]. Journal of Manufacturing Systems, 2024, 74: 575-586.

[7] Wu L, Ma X G. Leveraging Foundation Model Automatic Data Augmentation Strategies and Skeletal Points for Hands Action Recognition in Geological Assembly Lines[J]. arXiv preprint arXiv:2403.09056, 2024.

[8] Sturm F, Sathiyababu R, Hergenroether E, et al. Semi-Supervised Learning Approach for Fine Grained Human Hand Action Recognition in Geological Assembly[J]. 2023.

[9] Juanjuan Z, Xiaohan H, Zebang Q, et al. Small Object Detection Algorithm Combining Coordinate Attention Mechanism and P2-BiFPN Structure[C]//International Conference on Computer Engineering and Networks. Singapore: Springer Nature Singapore, 2023: 268-277.

[10] Qingyong Y, Chenchen H, Likun C, et al. CLAHR: Cascaded Label Assignment Head for High-Resolution Small Object Detection[J]. IEEE Access, 2024.

ISSN: 2710-0588

[11] Chen J, Wang X. Dense small object detection algorithm based on improved YOLO5 in UAV aerial images[J]. Comput. Eng. Appl.,(3), 2023: 100-108.

[12] Sun S, Mo B, Xu J, et al. Multi-YOLOv8: An Infrared Moving Small Object Detection Model Based on YOLOv8 for Air Vehicle[J]. Neurocomputing, 2024: 127685.

[13] Sun Z, Hua Z, Li H, et al. Flying Bird Object Detection Algorithm in Surveillance Video[J]. arXiv preprint arXiv:2401.03749, 2024.

[14] Zhou H, Lin Y, Yang L, et al. Benchmarking deep models on salient object detection[J]. Pattern Recognition, 2024, 145: 109951.

[15] Liu Z, Wang B, Li Y, et al. UnitModule: A lightweight joint image enhancement module for underwater object detection[J]. Pattern Recognition, 2024, 151: 110435.

[16] Yuan J, Xiao L, Wattanachote K, et al. FGNet: Fixation guidance network for salient object detection[J]. Neural Computing and Applications, 2024, 36(2): 569-584.

[17] Feng C, Wang C, Zhang D, et al. Enhancing Dense Small Object Detection in UAV Images Based on Hybrid Transformer[J]. Computers, Materials & Continua, 2024, 78(3).

[18] Huo D, Kastner M A, Liu T, et al. Small object detection for birds with Swin transformer[C]//2023 18th International Conference on Machine Vision and Applications (MVA). IEEE, 2023: 1-5.

[19] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.

[20] Xu S, Zheng S C, Xu W, et al. HCF-Net: Hierarchical Context Fusion Network for Infrared Small Object Detection[J]. arXiv preprint arXiv:2403.10778, 2024.

[21] Liang J, Ai J. Improving Accuracy of Small Object Detection in Screen based on YOLOv7[C]//2023 10th International Conference on Dependable Systems and Their Applications (DSA). IEEE, 2023: 851-859.

[22] Chai E, Chen L, Hao X, et al. Mitigate the scale imbalance via multi-scale information interaction in small object detection[J]. Neural Computing and Applications, 2024, 36(4): 1699-1712.

[23] Zhou H P, Zhang J, Sun K L, et al. Small Object Detection in Remote Sensing Based on Contextual Information and Attention[J]. Journal of Computers, 2024, 35(1): 145-157.

[24] Jing R, Zhang W, Liu Y, et al. An effective method for small object detection in low-resolution images[J]. Engineering Applications of Artificial Intelligence, 2024, 127: 107206.

[25] Ma S, Lu H, Liu J, et al. LAYN: Lightweight Multi-Scale Attention YOLOv8 Network for Small Object Detection[J]. IEEE Access, 2024.

[26] Yue X, Zhong G, Chu B. A multi-attention fusion mechanism for collaborative Geological surface Carbonate detection[C]//Fourteenth International Conference on Graphics and Image Processing (ICGIP 2022). SPIE, 2023, 12705: 209-218.

[27] Wang C, Wei X, Jiang X. An automated Carbonate detection method for optimizing Geological quality inspection[J]. Engineering Applications of Artificial Intelligence, 2024, 127: 107387.

[28] Li G, Shao R, Wan H, et al. A model for surface Carbonate detection of Geological products based on attention augmentation[J]. Computational Intelligence and Neuroscience, 2022, 2022.

[29] Li X, Li L, Wang P, et al. You only train twice: A lighter and faster method for Geological weld Carbonate detection based on dynamic kernel network[J]. Measurement, 2024, 231: 114642.

[30] Hu B, Zhou D, Wu Q, et al. An Improved Carbonate Detection Algorithm for Geological Products via Lightweight Convolutional Neural Network[C]//Proceedings of the 6th International Conference on Advances in Image Processing. 2022: 13-19.

[31] Zhou X, Xu X, Liang W, et al. Intelligent small object detection for digital twin in smart manufacturing with Geological cyber-physical systems[J]. IEEE Transactions on Geological Informatics, 2021, 18(2): 1377-1386.

[32] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.

ISSN: 2710-0588

[33] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[34] Cao Z, Simon T, Wei S E, et al. Realtime multi-person 2d pose estimation using part affinity fields[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7291-7299.

[35] Yi F, Zhang H, Yang J, et al. YOLOv7-SiamFF: Geological Carbonate detection algorithm based on improved YOLOv7[J]. Computers and Electrical Engineering, 2024, 114: 109090.

[36] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7464-7475.

[37] Wan Y, Yi L, Jiang B, et al. AENet: attention enhancement network for Geological Carbonate detection in complex and sensitive scenarios[J]. The Journal of Supercomputing, 2024: 1-24.

[38] Ge Z, Liu S, Wang F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

[39] Zhu Z, Liang D, Zhang S, et al. Traffic-sign detection and classification in the wild[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2110-2118.

[40] Bao Y, Song K, Liu J, et al. Triplet-graph reasoning network for few-shot metal generic surface Carbonate segmentation[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-11.