ReDS-GLINet: Reparameterized Deep-Scale Global–Local Interaction Network for Tiny Object Detection

1st Minh Tai Pham Nguyen
Faculty of Advanced Program
Ho Chi Minh City Open University
Ho Chi Minh City, Vietnam
2151013086tai@ou.edu.vn

2nd Nguyen Phuc Cam Tu
Faculty of Fisheries
Nong Lam University
Ho Chi Minh City, Vietnam
npctu@hcmuaf.edu.vn

3rd Trong Nhan Le
Faculty of Computer Science and Engineering
Ho Chi Minh City University of Technology - VNU
Ho Chi Minh City, Vietnam
trongnhanle@hcmut.edu.vn

4th Minh Khue Phan Tran
Faculty of Information Technology
Ho Chi Minh City Open University
Ho Chi Minh City, Vietnam
khue.ptm@ou.edu.vn

***Abstract*—Tiny object detection has always been a crucial topic in computer vision. The aim for high detection ability on small targeting objects while remaining low in complexity is considered the ultimate challenge, as traditional techniques are struggling with computation cost and real-time capability. Therefore, this study is conducted, testing a proposed lightweight object detector, built from several innovations, on a public tiny shrimp dataset. Regarding the methods, firstly, the object detector adopts the unique feature extraction and aggregation block, the RepGCLI, which is the combination of reparameterization convolution and the Global Context and Local Interaction (GCLI) module. While the reparameterization convolution allows the model to have the flexibility to expand learning complex contexts and optimize inference speed during training and inference, the GCLI module assists the model in focusing on important characteristics by establishing the interaction of global and local features via 1D convolution to prevent information loss, subsequently reducing the noisy particles under complex environments. Lastly, the deep-scale detection framework is applied in the proposed detector, enabling the extraction and fusion of more abstract information to provide better feature visualizations of tiny targeting objects under sophisticated conditions. As a result, the proposed detector, named Reparameterized Deep-Scale Global–Local Interaction Network (ReDS-GLINet), has outperformed most of the current SOTA one-stage detectors in shrimp larvae detection while maintaining significant computational efficiency and parameter consumption. The mAP50 metric of this proposed detector reaches up to 90.4%, the required parameters are only 8.66 million, and FLOPS is 9.51G, which is half compared to the small-scale YOLOv8. The proposed detector indicates the potential of high-accuracy real-time application without sacrificing computation cost in tiny object detection.**

***Index Terms*—YOLO, Attention Module, Shrimp Larvae Detection.**

I. Introduction

Shrimp larvae are normally extremely small in size. Furthermore, they tend to appear in dense with translucent bodies, burdening the detection head of several advanced deep learning models in tiny object detection. Because traditional methods are labor-intensive, time-consuming, and prone to human error, leading to the failed insights of the actual shrimp population, recent development of deep learning has been proposed, mainly aimed at automatic counting of tiny shrimp approaches [1], [2]. However, existing models face many challenges, such as low contrast between larvae and water, various lighting conditions, and different background environments due to debris or other marine particles [3]. Furthermore, it is noticeable that the shrimp larvae counting process conventionally is combined with other IoT systems [4], [5]. Hence, the deep learning models are expected to not only have high performance in counting shrimp larvae, but they are also required to function ordinarily on resource-restrained devices.

In order to address the aforementioned problems, this study proposes the novel architecture of a one-stage detector, combining several innovations. Firstly, it is obvious that the differences between the targeting objects and background environments are crucial characteristics for recognizing the tiny targeting objects.Therefore, the unique feature extraction and aggregation blocks are proposed mainly utilizing the combination of the reparameterization convolution [6] and the global context and local interaction (GCLI) module [7]. While the reparameterization convolution can expand its convolution operations during training and merge them back to the original form during inference, allowing the adaptation of the training phase and inference phase for expanding learning ability or optimizing speed, the GCLI module ensures the interaction between global and local features, suppressing the noisy background environments and producing suitable attention weights that can lead the model to focus on potential targeting patterns of the tiny targeting objects under complex conditions. Secondly, to provide more information, the deeper scales of

the architecture are introduced, consisting of $P \times 16$, $P \times 32$, and $P \times 64$ rather than $P \times 8$, $P \times 16$, and $P \times 32$. These deeper scales are capable of extracting more abstract information, which is important to handle the complex cases of detecting shrimp larvae, such as overlapping and dense appearance. As a result, the proposed model can effortlessly outperform most of the SOTA one-stage detectors in the large-scale YOLO family, reaching up to 90.4% mAP50 in the shrimp larvae dataset, while this model consumes nearly half the parameters and FLOPS compared to YOLOv8 [8].

The rest of the study is organized as follows: the related works are mentioned in section 2, while section 3 should provide the primary methodologies, section 4 should provide the experiment results, section 5 mentions the limitations and future works of the proposed model, and section 6 should provide a conclusion. Eventually, section 7 is for acknowledgment.

II. Related Works

Several noticeable works were proposed in the past in order to address numerous problems in complex environments of tiny object detection, especially in shrimp larvae detection. For instance, Duan et al. [9] introduced the improved YOLOv5 [10] by using the C2f block for feature extraction and Convolutional Block Attention Module (CBAM) [11]. Furthermore, the authors also proposed the region segmentation technique so that the receptive fields can be shrunk, highlighting the characteristics of targeting objects. Another work of Awalludin et al. [12] indicated the potential of image pre-processing techniques in tiny object detection by proposing the Canny Edge and Blob Processing for capturing the highlight features of shrimp larvae under complex environments. As traditional anchor-box approaches tend to be limited by the exceptional sizes of targeting objects that do not appear in the training set, Zhang et al. [13] introduced the unique anchor-free detector CAGNet in their works. This unique model employed the Coordinate Attention (CA) [14] in its backbone while having the spatial feature fusion module in the neck network. As the attention mechanisms assist the model to further focus on potential patterns of targeting objects under complex environments, the anchor-free heads allow the model to adaptively adjust the size of bounding boxes for the small or extremely small shrimp larvae. Acknowledging the appearance of various sizes of shrimp larvae, Dang et al. [1] employed the multiscale feature fusion network, allowing the deep learning model to learn the pattern of targeting shrimp larvae at different scales, providing higher detection ability in practical applications. Zhou et al. [15] stated the labor-intensive nature of both traditional approaches and annotation workloads for deep learning methods in counting shrimp larvae, so they introduced contrastive learning to reduce the dependence on large training datasets by extracting the information that contrasts the positive and negative of each pair. Although the proposed methodologies indicated promising results, there were still several problems remaining, including the low contrast in environments and the resource consumption of the proposed methods.

III. Methodologies

A. The Reparametizetion Glocal Context and Local Interaction Block (RepGCLI)

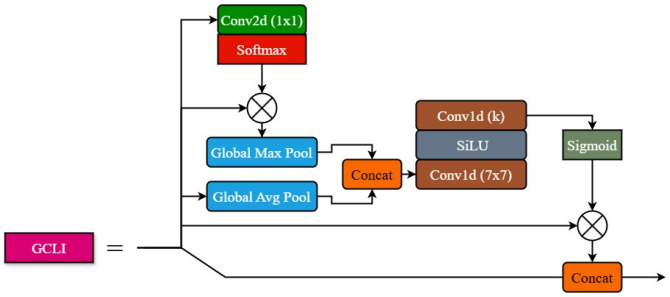

Fig. 1. Structure of the GCLI module.

The reparameterization convolution was first introduced in the work of [6]. This unique convolution allows it to extend the learning capability by expanding another convolution with a kernel size of 1×1 during training, having more gradients flow to each convolution operation, while the extended convolution can then be merged back to the original form to save speed during inference. The merged convolution can still retain its essential weights but have its extended flow cut off, resulting in less gradient flow and increased inference speed. Apparently, as the requirement of low computation cost and parameter consumption, the reparameterization convolution is the crucial part of the architecture that can assist the model to adapt to low-platform devices. The GCLI module is, on the other hand, this unique attention module that was originally introduced to address the noisy labels in medical imaging [7]. However, considering the complex patterns in tiny object detection, the capability of the GCLI module can still have its vital role in the task. This unique module captures long-range dependencies and integrates 1D convolution to establish the relationship between global and local features without any dimensionality reduction, which may reduce the risk of information loss. The GCLI module further has the partial feature learning ability, which only allows a half number of features to be learnable, which can be beneficial to suppress the weighted redundancies caused by the attention mechanism. The structure of GCLI is illustrated in Figure 1. By combining the reparameterization convolution and GCLI module, the study proposed the RepGCLI block for feature extraction and aggregation. The Figure 2 below indicates the structure of RepGCLI.

This unique structure inherits the partial feature sharing in CSPNet [16] to maintain the feature diversity and reduce the burden of computation cost as the number of features is split in half. While the first number of features is used for reparameterization convolution, the rest of the features are passed through an identity branch. This unique mechanism is widely adopted in several advanced one-stage detectors [8], [17], [18] due to its high reliability and fewer parameter requirements. After the extraction phase, the output of reparameterization convolution and the identity branch are then aggregated and fed to the GCLI module to highlight the important features

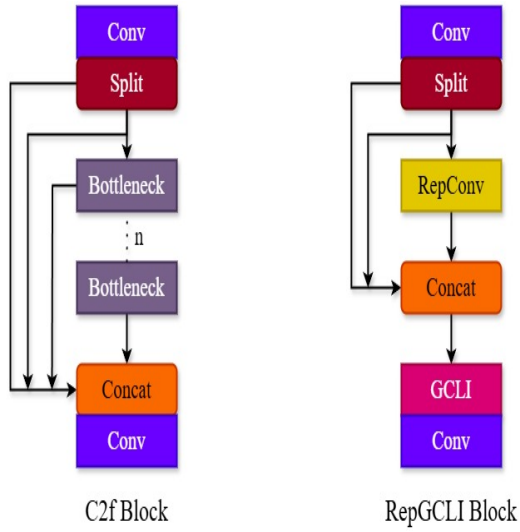

Fig. 2. Structure of the C2f block and the proposed RepGCLI block.

through suitable attention weights and suppress the irrelevant feature effect. At the end of this block, the weighted features are then passed through the final convolution for refining the fine-grained outputs before using them in the next stage.

B. The Deep-scale Detection Framework

The multi-scale detection framework is the primary component of many advanced one-stage object detectors that allow the deep learning models to detect and recognize the targeting objects regardless of different scales. While the popular scales for recent detectors are $P \times 8$, $P \times 16$, and $P \times 32$, the special sizes of shrimp larvae, which are extremely small, can reduce the capability and reliability of these scales. Hence, in order to detect the shrimp larvae effectively, it requires the unique deeper scales, providing more abstract information of complex patterns that can visualize more advantageous features of targeting objects. As a result, the $P \times 64$ scale is introduced, while the $P \times 8$ is suppressed for saving parameter and computation costs, this new modified version of multi-scale detection framework is called Deep-scale Detection Framework (DDF). The proposed DDF is proposed in Figure 3 below.

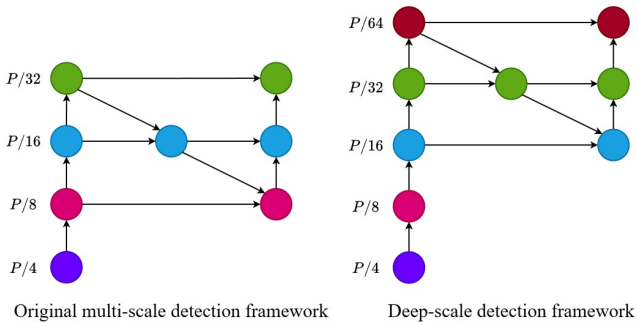

Fig. 3. Structure of the proposed DDF.

As illustrated in the mentioned figure, the backbone network is then extended to have a $P \times 64$ scale to expand the model extraction ability of different complex scenarios of shrimp larvae. This information is used for feature aggregation in the neck network and has its interaction with other shallow scales to provide advantageous feature visualization. At the end of the neck network, the final scales that are chosen for detection consist of $P \times 16$, $P \times 32$, and $P \times 64$. In order to prevent unnecessary computation cost and suppress extra parameters, the $P \times 8$, which is the shallowest scale, is suppressed. This deep-scale detection framework is proposed to provide more abstract extracted patterns that are useful for detecting tiny objects under different complex scenarios, where the shrimp larvae can overlap on each other or have low contrast.

C. The Proposed Reparameterized Deep-Scale Global–Local Interaction Network

Reparameterized Deep-Scale Global–Local Interaction Network (ReDS-GLINet) is the proposed one-stage detector for tiny object detection, which has its backbone and neck network integrated with the RepGCLI block for feature extraction and aggregation, while the deep-stage detection framework serves as the primary multi-scale object detection architecture of the model. The Figure 4 below indicates the overall ReDS-GLINet architecture.

The unique architecture is mainly inherited from the YOLOv8 [8] with a small scale in model parameters. All the C2f blocks, which are the original blocks in the YOLOv8 baseline model, are replaced with the RepGCLI block throughout the entire architecture. Unlike the traditional C2f block, which has its bottleneck serving as the main convolution operation, the RepGCLI blocks use the reparameterization convolution to extract features from the targeting objects, allowing the flexibility of the operation during training and inference in order to optimize the learning capacity and execution speed. Furthermore, the $P \times 64$ scale is adopted in the backbone network to enhance the feature extraction ability. During the neck network, $P \times 16$, $P \times 32$, and $P \times 64$ are the three main detection scales for feature aggregation to provide feature visualization for the final prediction heads. By having the RepGCLI with the proposed deep-scale detection framework, the ReDS-GLINet shall provide higher performance in tiny object detection while having the parameter consumption and computation cost reduced because of the bottleneck structure replacement. In addition, the involvement of the flexible convolution operation in reparameterization allows the model to then expand its learning capability and shrink for higher speed without affecting performance adaptively.

IV. Experiments

In order to provide the meaningful experiment results of different components in ReDS-GLINet, the ablation study is carried out to learn the effect of each part in the ReDS-GLINet compared to the original baseline YOLOv8. Furthermore, to evaluate the ability of the GCLI module, other SOTA attention modules, which include ECA [19], CBAM [11], SE [20], GCB

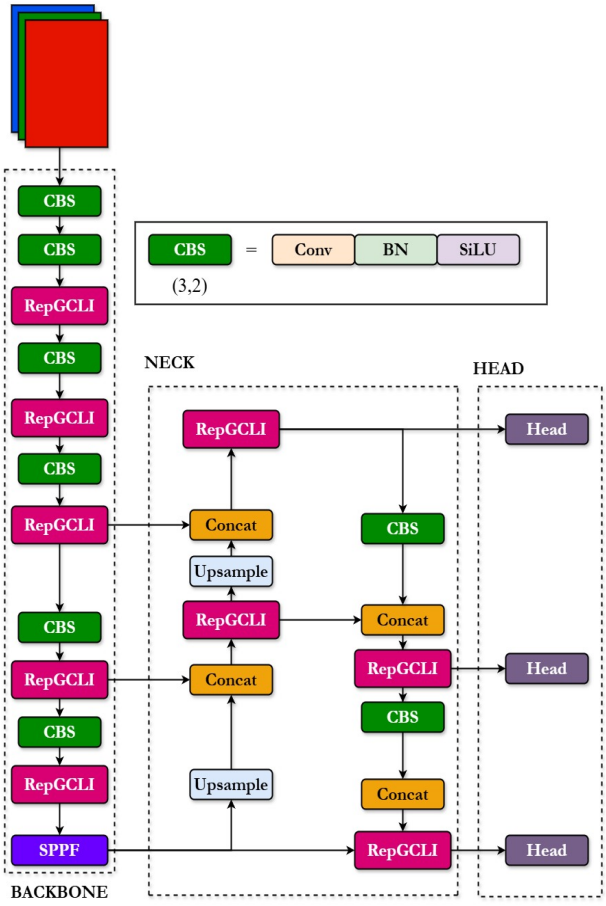

Fig. 4. Structure of the proposed ReDS-GLINet.

[21], simAM [22], and CA [14], are added and suggested to replace the position of the GCLI module in the RepGCLI. The proposed ReDS-GLINet are then evaluated with other one-stage detectors, including YOLOv5 [10], YOLOv6 [23], YOLOv7 [17], YOLOv8 [8], YOLOv9 [18], YOLOv10 [24], YOLOv11 [25], and YOLOV12 [26]. The models are evaluated in performance using precision, recall, mAP50, mAP50-95, and the number of parameters and FLOPS in resource consumption and computation cost.

A. Experimental dataset

The dataset used in the experiment is public access on the Roboflow platform [27], [28] with about 860 images in total. The shrimp larvae in the dataset are small to very small and appear dense with different lighting conditions and a complex background caused by irrelevant particles that are similar to the shrimp larvae. The author of the dataset conducts the top-down position of the camera angle at different heights with 1280 pixels resolution. Furthermore, the backgrounds of the containing boxes for shrimp larvae are various in color, making the detection become challenging. This study further re-splits the training, validation, and test sets as follows: 232 images for training, 205 images for validation, and 423 images for testing. By providing a higher number of data for testing, the final result of the model can be more general for various cases in practical detection. Several images of the dataset are illustrated in Figure 5.

Fig. 5. Several samples of the shrimp larvae dataset.

B. Environment and Hyperparameters Setups

All the models are trained and evaluated on a single machine using an Intel Core i7 12700K with 32GB RAM and a GPU GeForce RTX 3090 24GB VRAM, PyTorch version 2.4.0 with CUDA version 11.8, and Ubuntu 23.0 OS. All the experimental models are trained from scratch without using any pretrained model with the hyperparameter setups indicated in Table I.

TABLE I
THE SETTING OF HYPERPARAMETERS FOR TRAINING AND TESTING PHASE.

Hyperparameter	Value
Learning rate	0.01
Input size	1280 × 1280
Batch size	2
Epoch	200
Weight decay	0.0005
Momemtum	0.937

C. Ablation Experimental Results

This ablation study is carried out to gain insight about the effect of each component in ReDS-GLINet. Table II illustrates the performance of the model with each component attached.

As indicated in the table, the YOLOv8s is the small-scale baseline of YOLOv8 originally achieves a precision of 80.2% and a recall of 83.76%. Although it can achieve the mAP50 up to 86.04%, the computation cost and number of parameters are remarkably high, reaching up to 28.60 GFLOPS and 11.13 params. By introducing the DDF, the performance of the model is significantly enhanced, improving the precision to 86.66%, while the mAP50 is achieved up to 90.3%. In spite of the noticeable ability in detection capability of the model, the params are increased remarkably from 11.3 to 15.40. However, as the deeper scale requires fewer computation costs because of the smaller size of feature maps, the GFLOPS is decreased incredibly from 28.60 to 16.99. These results indicate that the deeper scales are exceptionally crucial to provide sustainable information for tiny object detection while being beneficial in resource consumption. Additionally, in the final configuration, the RepGCLI is introduced to replace all the C2f blocks in the baseline small-scale YOLOv8. This configuration successfully maintains the overall performance of the model, which the mAP50 is slightly increased to 90.4%, while it witnesses a small decline in mAP50-95. However,

TABLE II
THE PERFORMANCE OF DIFFERENT COMPONENTS IN REDS-GLINET ON THE EXPERIMENTAL TEST SET.

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)	Params (Million)	FLOPS (G)
YOLOv8s	80.28	83.76	86.04	58.14	11.13	28.60
YOLOv8s + DDF	86.66	84.15	90.3	**58.98**	15.40	16.99
YOLOv8s + DDF + RepGCLI	**86.80**	**84.15**	**90.4**	58.81	**8.66**	**9.51**

its efficiency is significantly improved, as the number of parameters drops from 15.40 to 8.66 million, and the GFLOPS is nearly halved from 16.99 to 9.51, making it much more lightweight and computationally efficient. This configuration is the final form of ReDS-GLINet. In summary, while the DDF improves performance at the cost of computational efficiency, the introduction of RepGCLI astonishingly reduced the complexity without negatively affecting the overall performance, making it well-suited for real-world applications, particularly in resource-constrained environments.

D. Model Experimental Results

This section is carried out to evaluate the performance of the proposed detector with other SOTA one-stage detectors, mainly focusing on the YOLO family to test the real-time capability. Table III illustrates the overall performance of each detector on the experimental shrimp larvae test set.

The table indicates that the proposed ReDS-GLINet demonstrates the highest performance among other one-stage detectors, with the model boosting the precision to 86.80% and recall to 84.15%, while mAP50 is 90.4% and mAP50-95 is 58.81%. The results illustrate the significant detection capability compared to other current detectors. In comparison, the YOLOv7 indicates the competitive mAP50 as it reaches up to 85.91%, while the mAP50-95 is achieved at 58.56%. However, this model has the highest computation cost with significant params. Similarly, the YOLOv8, while this model reaches slightly higher than the YOLOv7 in mAP50 metrics, still suffers from a noticeable computational burden. Although the latest models, YOLOv11 and YOLOv12, have remarkably low parameters and computation costs, their performances are much lower than others, especially YOLOv12, whose mAP50 is the lowest among the category. As YOLOv11 and YOLOv12 are mainly powered with the self-attention mechanism, these results indicate that such a mechanism is not suitable in the case of tiny-object datasets. On the other hand, the earlier version, such as YOLOv5, which is the most lightweight detector as it only consumes 2.50 million parameters and 7.2 GFLOPS, still achieves the remarkable performance of mAP50 and mAP50-95. The YOLOv6 is about double in computation cost or params, but it offers slightly lower performance compared to YOLOv5. Overall, the ReDS-GLINet is the most suitable detector under various complex scenarios of the experimental dataset. This unique detector offers a notable performance in detection while maintaining computational efficiency. Its unique feature extraction block and deep-scale detection framework allow it to outperform other one-stage detectors in both high accuracy and efficiency. Several detection samples of the ReDS-GLINet are presented in Figure 6 below.

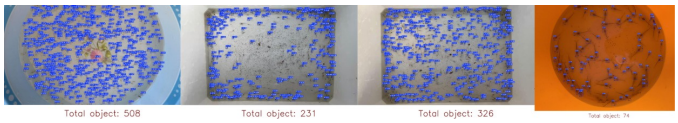

Fig. 6. Several detection samples of ReDS-GLINet in the test set.

V. LIMITATIONS AND FUTURE WORKS

In spite of the significant performance of the ReDS-GLINet, the extra parameters and high computation cost are still considerable problems compared to other current one-stage detectors. The YOLOv5 small scale uses only 2.5 million parameters, compared to the ReDS-GLINet's 8.66 million parameters. The proposed model also failed to find the targeting objects in a number of very complicated situations, such as when there were a lot of them overlapping or when particles that weren't related to the targeting objects showed up. Figure 7 indicates several failed detection cases for the proposed model. Therefore, there is still room for further experimentation by proposing several data augmentation techniques or image processing to filter the irrelevances. In addition, it is noticeable to conduct work with a more advanced deep learning model in order to improve the overall detection performance to further reduce unnecessary computation costs and resource consumption.

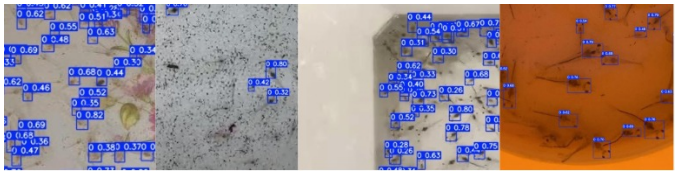

Fig. 7. Several failures cases of ReDS-GLINet.

VI. CONCLUSION

In conclusion, the ReDS-GLINet has outperformed most of the current SOTA one-stage detectors in tiny object detection, with its highest mAP50 reaching up to 90.4%. In addition, the proposed detector can achieve significant performance while maintaining remarkable computational efficiency. The ReDS-GLINet only requires 9.51 GFLOPS and 8.66 params in order to function normally, making it a suitable option for detecting tiny objects such as shrimp larvae on low-platform devices. Furthermore, this study also illustrates the potential of using deeper scales in deep learning architectures to construct

TABLE III
THE PERFORMANCE OF DIFFERENT ONE-STAGE DETECTOR ON THE EXPERIMENTAL TEST SET.

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)	Parameters (Million)	FLOPS (G)
YOLOv5	78.86	83.04	84.60	56.23	**2.50**	**7.2**
YOLOv6	78.85	81.07	83.40	54.66	4.23	11.9
YOLOv7	79.79	84.09	85.91	58.56	11.41	34.8
YOLOv8	79.82	84.46	86.38	59.28	11.13	28.6
YOLOv9	79.57	84.74	86.34	59.30	7.28	27.4
YOLOv10	78.29	81.93	83.88	56.81	8.06	24.8
YOLOv11	78.45	82.27	84.07	56.55	9.41	21.3
YOLOv12	77.38	79.91	81.85	53.99	9.23	21.21
ReDS-GLINet (our)	**86.80**	**84.15**	**90.4**	**58.81**	8.66	9.51

robust feature visualizations for the final detection heads for recognizing tiny targeting objects, while the GCLI module shows its ability to suppress noise of complex patterns in under sophisticated environments making it an ideal choice for a combination of other convolution operations in small-scale deep learning models.

VII. ACKNOWLEDGMENT

The work is acknowledged from Ho Chi Minh City University of Technology (HCMUT), VNU-HCM, Ho Chi Minh Open University and Nong Lam University. We would like to further express our sincere gratitude to the author who published the shrimp larvae dataset online as open access on open platform.

REFERENCES

[1] T.-H. Dang, N.-H. Dang, and V.-T. Tran, "Shrimp larvae detection and counting in aquaculture using multiscale feature fusion networks," *Computers and Electronics in Agriculture*, vol. 233, p. 109850, 2025.

[2] W.-C. Hu, L.-B. Chen, M.-H. Hsieh, and Y.-K. Ting, "A deep-learning-based fast counting methodology using density estimation for counting shrimp larvae," *IEEE Sensors Journal*, vol. 23, no. 1, pp. 527–535, 2022.

[3] B. Gong, L. Jing, and Y. Chen, "Tsd: Random feature query design for transformer-based shrimp detector," *Computers and Electronics in Agriculture*, vol. 221, p. 108949, 2024.

[4] Y.-K. Hsieh, J.-W. Hsieh, W.-C. Hu, and Y.-C. Tseng, "Aiot-based shrimp larvae counting system using scaled multilayer feature fusion network," *IEEE Internet of Things Journal*, 2024.

[5] W.-C. Hu, L.-B. Chen, P.-J. Yu, M.-Y. Wu, K.-H. Chen, and X.-R. Huang, "A deep learning-based underwater image enhancement scheme for turbid underwater aquaculture environments," in *2024 6th International Conference on Computer Communication and the Internet (ICCCI)*. IEEE, 2024, pp. 107–112.

[6] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "Repvgg: Making vgg-style convnets great again," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13 733–13 742.

[7] M. T. P. Nguyen, M. K. P. Tran, T. Nakano, T. H. Tran, and Q. D. N. Nguyen, "Partial attention in global context and local interaction for addressing noisy labels and weighted redundancies on medical images," *Sensors (Basel, Switzerland)*, vol. 25, no. 1, p. 163, 2024.

[8] Ultralytics, "Ultralytics yolov8. 2023," *URL https://github.com/ultralytics/ultralytics*, 2023.

[9] H. Duan, J. Wang, Y. Zhang, X. Wu, T. Peng, X. Liu, and D. Deng, "Shrimp larvae counting based on improved yolov5 model with regional segmentation," *Sensors*, vol. 24, no. 19, p. 6328, 2024.

[10] Ultralytics, "YOLOv5: A state-of-the-art real-time object detection system," https://docs.ultralytics.com, 2021.

[11] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[12] E. A. Awalludin, M. M. Yaziz, N. A. Rahman, W. N. J. H. W. Yussof, M. S. Hitam, and T. T. Arsad, "Combination of canny edge detection and blob processing techniques for shrimp larvae counting," in *2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, 2019, pp. 308–313.

[13] G. Zhang, Z. Shen, D. Li, P. Zhong, and Y. Chen, "Cagnet: an improved anchor-free method for shrimp larvae detection in intensive aquaculture," *Aquaculture International*, vol. 32, no. 5, pp. 6153–6175, 2024.

[14] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13 713–13 722.

[15] H. Zhou, S. H. Kim, S. C. Kim, C. W. Kim, S. W. Kang, and H. Kim, "Instance segmentation of shrimp based on contrastive learning," *Applied Sciences*, vol. 13, no. 12, p. 6979, 2023.

[16] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "Cspnet: A new backbone that can enhance learning capability of cnn," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 390–391.

[17] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 7464–7475.

[18] C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, "Yolov9: Learning what you want to learn using programmable gradient information," in *European conference on computer vision*. Springer, 2024, pp. 1–21.

[19] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 534–11 542.

[20] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[21] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "Gcnet: Non-local networks meet squeeze-excitation networks and beyond," in *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 2019, pp. 0–0.

[22] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, "Simam: A simple, parameter-free attention module for convolutional neural networks," in *International conference on machine learning*. PMLR, 2021, pp. 11 863–11 874.

[23] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie *et al.*, "Yolov6: A single-stage object detection framework for industrial applications," *arXiv preprint arXiv:2209.02976*, 2022.

[24] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "Yolov10: Real-time end-to-end object detection," *arXiv preprint arXiv:2405.14458*, 2024.

[25] Ultralytics, "Ultralytics yolov11. 2024," *URL https://github.com/ultralytics/ultralytics*, 2024.

[26] Y. Tian, Q. Ye, and D. Doermann, "Yolov12: Attention-centric real-time object detectors," *arXiv preprint arXiv:2502.12524*, 2025.

[27] P. N. P. Hi, "shrimp5 dataset," https://universe.roboflow.com/ph-nguyn-phan-hi/shrimp5 , jan 2024, visited on 2025-03-01. [Online]. Available: https://universe.roboflow.com/ph-nguyn-phan-hi/shrimp5

[28] ——, "shrimpv4 dataset," https://universe.roboflow.com/ph-nguyn-phan-hi/shrimpv4 , feb 2024, visited on 2025-03-24. [Online]. Available: https://universe.roboflow.com/ph-nguyn-phan-hi/shrimpv4