This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3330788

Consumer-Centric Insights into Resilient Small Object Detection: SCIoU Loss and Recursive Transformer Network

Le Wang, Yu Shi, Guojun Mao, Fayaz Ali Dharejo, Senior, IEEE Member, Sajid Javed and Moath Alathbah

***Abstract*—As an emerging consumer electronic product, the use of unmanned aerial vehicle(UAV) for a variety of tasks has received growing attention and favor in the enterprise or individual consumer electronics market in recent years. The deep neural network based object detectors are convenient to embed into the UAV product, however, the drone-captured images could bring the potential challenges of object occlusion, large scale difference and complex background to these methods because they are not desinged for the detection of small and tiny objects within the aerial images. To address the problem, we propose an improved YOLO paradigm called SR-YOLO with an Efficient Neck, Shape CIoU and Recursion Bottleneck Transformer for better object detection performance in consumer-level UAV products. Firstly, an efficient neck structure is presented to retain richer features through a small object detection layer and an up-sampling operator suitable for small object detection. Secondly, we design a new prediction box loss function called shape complete-IoU(SCIoU), which utilizes a width (height) limiting factor to alleviate the deficiency that the CIoU only focuses on aspect ratios by taking into account both the aspect ratio and the ratio of the two boxes' widths. Moreover, combined with recurrent neural network and multi-head self-attention mechanism at the cyclic manner, a recursive bottleneck transformer is constructed to relieve the impact of highly dense scene and occlusion problems exists in UAV images. We conduct the extensive experiments on two public datasets of VisDrone2019 and TinyPerson, where the results show that the proposed model surpasses the compared YOLO by 8.1% and 3.2% in mAP_{50} respectively. In addition, the analysis and case study also validate our SR-YOLO's superiority and effectiveness.**

***Index Terms*—unmanned aerial vehicle (UAV) image, small object detection, you only look once, Bottleneck Transformer.**

I. Introduction

Unmanned aerial vehicle(UAV) with object detection has permeated through various fields of consumer electronics [1]–[3], such as industrial inspection [4], environmental monitoring of aquaculture [5], agricultural crop monitoring [6], or civilian mountain search and rescue [7]. Although the UAV aerial images are of high resolution, they still suffer from dramatic changes in their scale, intricate background and occlusion.Currently, YOLO [8]–[11] series algorithms are gaining importance in object detection algorithms. YOLOv5 with high adaptability, easy deployment, and high accuracy is widely used in various target detection tasks and industrial production.However, the YOLO series algorithm, conceived for object detection in general natural scenes (typical natural scene datasets include MS COCO [12], PASCAL VOC [13], etc.). In order to keep YOLOv5 lightweight while better adapting it to UAVs for object detection. We improved YOLOv5 for the characteristics of the images in the actual UAV working scenarios. So that it can give consumers a better experience.

For the above reasons, we propose an improved efficient neck, a newly designed Shape CIoU and Recursion Bottleneck Transformer for small object detection in drone-captured images. Firstly, considering that small or tiny objects possess low pixels [14], [15] and there exists dramatic scale variation between different objects in aerial images, a new neck structure called Efficient Neck by integrating Content-Aware ReAssembly of Features [16](CARAFE) up-sampling operator and small object detection layer. Efficient Neck can better cope with the drastic changes in different object scales and obtain better detection results, such as detecting more tiny objects. Secondly, we design a novel prediction box loss function Shape Complete-IoU (SCIoU) to make the prediction box closer to the real box. Improvement of complete-IoU (CIoU) [17] aspect ratio penalty term failure by a newly designed shape penalty term. In addition, an improved Recursion Bottleneck Transformer (RBoT) module is proposed to obtain global information through recursively utilizing the self-attention mechanism to alleviate the issue of object occlusion in crowds or highly-density scenes.

By incorporating ther advanced components above, an innovative YOLO paradigm is derived to address the specific challenges associated with small even tiny object detection through enhancing the spatial resolution of the feature maps, refining the aspect ratio loss and obtaining global contextual signals. Specially, for the validation of our proposed method, we implement the two variants using our proposed SR-YOLO, namly SR-YOLOv5 and SR-YOLOv8(based on YOLOv5 and YOLOv8, respectively), and conduct the extensive experiments. We validate them on two benchmark datasets (VisDrone2019 and TinyPerson). The experimental results show that SR-YOLOv5 surpasses 8.1% on $mAP_{50}(\%)$, and the new model is more suitable for small object detection tasks than the

Le Wang†, Yu Shi† and Guojun Mao* are with Fujian Provincial Key Laboratory of Big Data Mining and Applications, College of Computer Science and Mathematics, FJUT 350118, China (email: lewang@fjut.edu.cn; 2211308023@smail.fjut.edu.cn; 19662092@fjut.edu.cn);
Fayaz Ali Dharejo and Sajid Javed are with Department of Electrical Engineering and Computer Science, Khalifa Univeristy, Abu Dhabi, UAE (email: fayaz.ali@ku.ac.ae) sajid.javed@ku.ac.ae);
Moath Alathbah * is with Department of Electrical Engineering, College of Engineering, King Saud University, Riyadh 11451, Saudi Arabia (email: malathbah@ksu.edu.sa)
*Corresponding Authors, †These authors contributed equally to this work.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3330788

original YOLOv5. Experiments have proved that SR-YOLO is able to recognize targets more accurately in a variety of complex backgrounds and poor lighting environments. Combined with the UAV Intelligent System, SR-YOLO can provide better intelligent identification services for electronics consumers. In summary, the main innovations of this paper are as follows:

A new neck network comprised of CARAFE, an additional feature fusion layer, and a detection head is realized, where CARAFE provides a large receptive field and will be flexible for up-sampling. The four-detection head structure can better alleviate the impact of object scale changes.

A new prediction box loss function is designed, which makes the prediction box closer to the real box while improving the width ratio between the prediction box and the real box. The new loss function enables the network to obtain higher-quality prediction boxes and maximize detection accuracy.

An improved recursion Transformer is proposed to enhance the recognition effect in occlusion and dense scenes. This new attention mechanism can help the network better obtain target context information in complex situations.

II. Related Work

Object detection models: Generally, an image is taken as input and output with bounding boxes and labels on detected objects by object detection models, which thus can be mainly divided into two classes from the perspective of their identifying processes. The first type of models are two-stage detectors, in which the detection is separated into two phases, namely generating a large number of region proposal that may contain the object and append approximate location information, then classifying and regressing the RoIs, such as R-CNN [18], Fast R-CNN [19], Faster R-CNN [20], SPPNet [21], Mask R-CNN [22], FPN [23].

The second ones, called one-stage detectors, skip over the region proposal stage of the two-stage models and receive the complete image as input and regress the object border directly upon a dense sampling of locations, such as YOLOX [24], DETR [25], CornerNet [26], SSD [27], YOL5Ov6 [28], FCOS [29], EfficientNet [30], PP-YOLO [31].

Bounding box regression loss functions: In object detection, the positioning task needs to determine the position of the object in an image and then output its corresponding coordinate information. The positioning task relies on the bounding box regression module to locate the object. The bounding box regression refers to using a rectangular bounding box to predict the position of the object in an image, and then continuously adjust the position of the prediction box. The bounding box regression loss function has been developed rapidly with the introduction of IoU (intersection over union). GIoU [32] is proposed to solve the problem that IoU cannot optimize the non-intersection between the prediction box and the real box, aiming to make the prediction box and the real box intersected. Then DIoU and CIoU [33] further solved the degradation problem when the prediction and real boxes overlapped.

Consumer Electronic: The development of technology that combined drones and deep learning in the field of crop classification was investigated by Bouguettaya [34] et al, who pointed out the importance of the integration of UAV-based remote sensing technologies and deep learning algorithms, especially object detectors, which provide great convenience to the consumers in agriculture. Huang [35] proposed an algorithm based on YOLO-R to deal with the actual situation in road object detection for autonomous driving. Chen [36] et al. improved the YOLOv4 for pavement defect detection, aiding drivers to get more information about the road surface to reduce car damage and safety risk. In addition, Gung [37] et al. applied the defect detector in the enterprise production lines, where their model can effectively assist the workers in identifying the defective products.

YOLO series: YOLO series models are widely used in object detection because of its high accuracy and detection speed. The TPH-YOLO model proposed by Xingkui Zhu et al. [38] was in the light of the characteristics of aerial images. Based on YOLOv5, a Transformer layer was added in front of the detection head to form a Transformer Prediction Head (TPH) to replace the prediction head in the original model. Convolutional Block Attention Module (CBAM) was employed to find the area of attention in dense object scenes, but this structure requires huge computational costs and hardware resources. Raian Rahman et al. [39] trained four different models on the aerial view of urban traffic and then used non-maximum suppression (NMS) for the integration of the models. Although this method improves the accuracy, its reasoning time increases greatly. Munhyeong Kim et al. [40] proposed ECAP-YOLO, introduced an improved IECA module into the SPP module and network, and modified the number of detection layers to make the model more suitable for detecting small objects. Because the model deletes the detection layer of large objects, the detection effect of large objects is relatively poor. Tao Liu et al. [41] used YOLOv4 for sea surface object detection. They designed a new reverse depth separable convolution to replace the partial convolutions in the network, which improved the network a little in terms of accuracy, but if all the convolutions in the network are replaced, the effect is not ideal. Mohamad Haniff Junos et al. [42] utilized the improved YOLOv5 network to detect the oil palm image of the drone. This study added a tightly connected neural network, modified the activation function to switch, and added a new detection layer, resulting in the superior detection effect. By using the proposed object detection architecture SR-YOLO, it can demonstrate the promising results of recognizing and classifying both normal and small even tiny objects in aerial images.

III. Method

In order to enable the conventional object detector to address object occlusion, large scale difference and complex background simultaneously when recognizing and classifying the small even tiny objects in aerial images, we propose a novel SR-YOLO model through the three elaborated compo-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3330788

nents, namely an Efficient Neck, Shape CIoU and Recursion Bottleneck Transformer and implement its two variants (SR-YOLOv5 and SR-YOLOv8), which have been testified the detection performance in our experiments so that the consumer-level UAVs equipped with our method could bring better object detection experience to the consumers. The detail of the three proposed components is described as follows.

A. *Efficient Neck*

There are plenty of small (objects smaller than 32*32 pixels are defined in COCO [12] as small objects) and tiny (objects smaller than 20*20 pixels are defined as tiny objects in TinyPerson [43]) objects in aerial images. Meanwhile, objects from different classes could be of dramatic scale variance. However, the YOLO's neck network, for instance, the nearest-neighbor up-sampling used by YOLOv5's Neck prevents the effective propagation of high-level semantic information. YOLOv5's three-prediction head structure is only suitable for medium and large-sized object detection tasks. To fill these gaps, we propose a neck network structure composed of content-aware reassembly of features and small object path aggregation network neck to improve detection of tiny objects. This structure includes a small object detection layer and a CARAFE up-sampling.

On the one hand, the vanilla YOLOv5 uses 8, 16, and 32 times down-sampling, and the feature map obtained is hard to preserve the informative feature of small objects. Therefore, a small object detection layer is employed at four times down-sampling to predict a larger resolution image, and an additional detection head is added to alleviate the problem of drastic scale changes. The nearest neighbor up-sampling employed by YOLOv5 is simple and efficient, but it only uses the gray value closest to the sampling point. As a result, it produces discontinuous gray values after sampling, resulting in low image quality. Using this approach is difficult for deep networks to learn the features of smaller objects, which affects the model's detection accuracy. For this issue, we considered replacing the up-sampling component with one more suitable for aerial images.

Consider that CARAFE reassembles the feature information from the lower-resolution input feature map to generate a higher-resolution output feature map, allowing for more effective up-sampling for objects with small pixel. Through combining with CARAFE, small object detection layer could incorporate context information from surrounding regions to enhance the representation of small objects. Thus we employ this content-aware up-sampling operator and a small object detection layer to construct the Efficient Neck, as shown in Figure1.Compared with the original Neck using nearest neighbor up-sampling, the Efficient Neck part with CARAFE has a larger up-sampling kernel, which allows it to summarize contextual information over a large receptive field. CARAFE performs content-aware processing for specific instances, generating adaptive kernels on-the-fly. The up-sampling kernel and the feature map are semantically related such that CARAFE can be based on the input content. As a result, CARAFE better preserves feature information for small objects. Thus Efficient Neck can better utilize the surrounding and semantic information of the feature map. Hence, additional feature fusion layer is added and prediction head into Efficient Neck.As opposed to the three-prediction head in the vanilla model, a four-prediction head can better handle the drastic scale changes of different objects. The Efficient Neck has a significant impact on improving detection in scenarios that involve small objects whose scale changes drastically, resulting in missed and false detections. We verify this in the ablation experiments in Section 4.

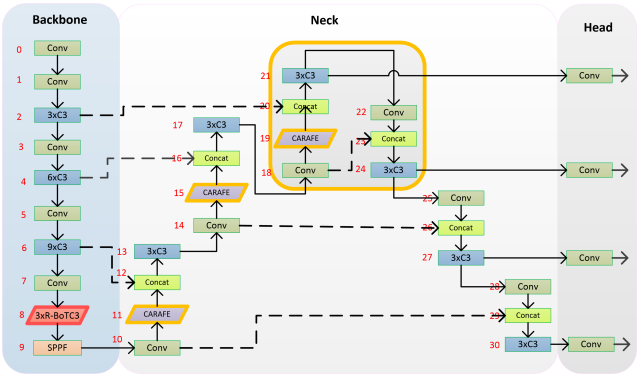


Fig. 1. Composition of the SR-YOLO structure.

B. *Shape CIoU*

The loss function for predicted box regression in object detection tasks generally consists of the classification loss and the bounding box regression loss. For bounding box regression loss, the most commonly used CIoU introduces two considerations based on the GIOU. The ratio between ρ and c (as shown in Figure 2) and the penalty/consistency term for the aspect ratio of the real box and the predicted box. ρ is the Euclidean distance between the centre b of the real box and the center b^{gt} of the predicted box. The dotted rectangle is the minimum bounding box containing two boxes, and c is minimum closure diagonal. The equation of CIoU is shown in equation1-3, where V is for calculating the aspect ratio between the predicted box and the ground truth box, and w^{gt} and h^{gt} represent the width and height of the real box, respectively. α is the balance factor of V.

$$I_{CIoU} = 1 - IoU + \frac{\rho^2(b, b_{gt})}{c^2} + V\alpha \tag{1}$$

$$V = \frac{4}{\pi^2} + (arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \tag{2}$$

$$\alpha = \frac{V}{(1 - IoU) + V} \tag{3}$$

However, as figure 3 shows, when the aspect ratio of the two sets of real boxes and the predicted box is equal, the aspect ratio will fail. The model's learning capability will weaken by this loss function. To address the above problem we design a SCIoU loss function as shown in equation 4. V_s is a newly

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3330788

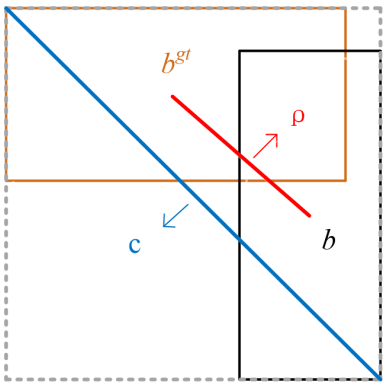


Fig. 2. diagonal diagram of minimum closure.

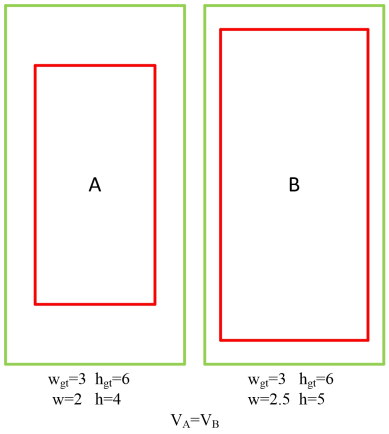


Fig. 3. Green boxes represent real boxes, and two red boxes represent predicted boxes of different sizes.

designed penalty with ratio-curbing factor (X) , as shown in equation 5. The ratio-curbing factor changes the shortcomings of CIoU that only focuses on the aspect ratio. With the addition of the ratio-curbing factor, SCIoU is able to focus on both the aspect ratio and the ratio of the widths of the two boxes. SCIoU allows the model to be optimized in terms of both the aspect ratio and the ratio of the two widths in order to avoid the degradation of the loss function as much as possible.

$$I_{SCIoU} = 1 - IoU + \frac{\rho^2(b, b_{gt})}{c^2} + V_s\alpha_s \tag{4}$$

$$V_s = \frac{4}{\pi^2} + (arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 X \tag{5}$$

Here the ratio-curbing factor is denoted as X, as shown in equation 6. When the value of X is lower, the predicted box is closer to the real box, the ratio of the width/height of the real box to the width/height of the predicted box tends to 1. Since CIoU has a combined effect on the aspect ratio, a similar effect can still be produced if the width w is replaced by h. Based on the above V_s and X, the new balance factor can be expressed in a new form, as shown in equation 7.

$$X_w = (1 - \frac{w}{w^{gt}})^2 \tag{6}$$

$$\alpha_s = \frac{V_s}{(1 - IoU) + V_s} \tag{7}$$

C. Recursion Bottleneck Transformer

In UAV images, it is prevalent that a number of objects are in close proximity to each other, thus from different perspectives it seems overlapped between objeacts and forms severe occlusions. On the other hand, image quality change caused by non-uniform environmental illumination is unavoidable during drone operations, which would increase the rate of missed and false detection. The convolution structure utilzed in YOLOv5 extracts limited local signals instead of contextual information. Global features can help the model to accurately localize in the presence of dense backgrounds and heavy occlusions in aerial images. Therefore global feature information is very important for small object detection. Transformer's dynamic attention mechanism, global modeling capability has a strong feature learning ability.Transformer can extract rich global feature information. DETR, the first object detection model based on transformer, greatly simplifies the object detection process in an end-to-end fashion. Srinivas [44] et al. replaced the convolution operator in the last residual block of ResNet with a multi-head self-attention (MHSA) module , allowing the network model to improve accuracy while reducing the number of parameters. Shen [45] et al. uses the recursive structure in recursive neural network to construct an iterative transformer block and applies it to the DEiT model. Their experimental results show that this simple recursive operation is capable of enhancing feature representation.

Inspired by the above research work, we propose a Recursive Bottleneck Transformer (RBoT), which has two variants of RBoT-a and RBoT-b, as shown in Figure 4.To be specific, RBoT-a is constructed with multiple MHSA layers in a manner of recursion based on the BoT**??**Alternatively, RBoT-b furnishes each subsequent MHSA layer with an additional residual connection operation, as shown in equation 8 and 9. Based on the pilot experimental results analysis, we only use one cycle (namely, two MHSA layers) when implementing this proposed module. Although RBoT-b is more complex than RBoT-a, the ablation experiment in Section 4.3 shows that it has not yet produced more parameters and the detection accuracy is slightly improved.

$$M_{(X)}a = MSHA(MSHA(f^{1\times1}(X))) + X \tag{8}$$

$$M_{(X)}b = MSHA(MSHA(f^{1\times1}(X)) + f^{1\times1}(X)) + X \tag{9}$$

$M_{(X)}a$ and $M_{(X)}b$ represent the outputs of RBoT-a and RBoT-b, respectively. MSHA means Multi-Head Self-Attention mechanism, and $f^{(1\times1)}$ is a convolution operation with a convolution kernel size of 1x1. The MHSA layer structure is shown in Figure 5, where four heads are used in RBoT, and only one is shown here for simplicity. The representation of the Self-Attention Layer is consistent with that in BoT.

It's worth noting that the RBoT module can be flexibly applied to the multiple versions of YOLO, thus actually we

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3330788

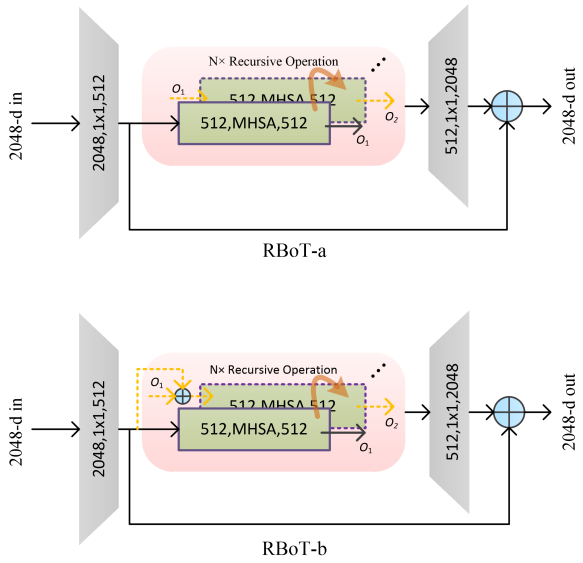


Fig. 4. The architecture of the Recursion Bottleneck Transformer.

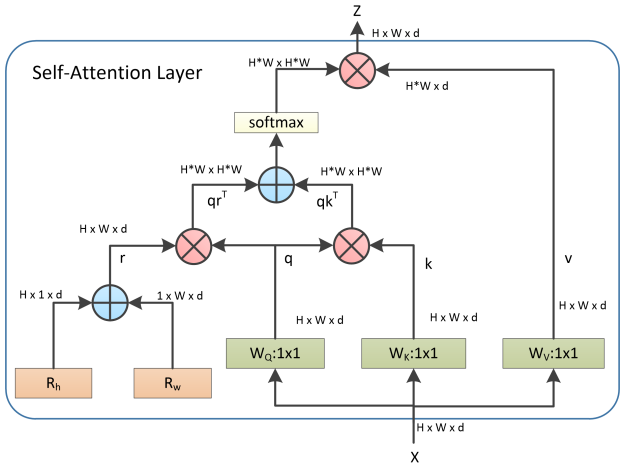


Fig. 5. Composition of the MHSA layer structure.

repalce the bottleneck module in C3 with the RBoT module in both YOLOv5 and YOLOv8 for practice, where a new RBoTC3 module is represented. The module obtains rich contextualized information and enhances the ability to capture differential local signals because it has strong global modeling and feature learning capabilities. RBoT makes the model more effective in complex background and partially occluded object detection. Compared to simultaneously replacing multiple C3 modules with RBoTC3, we only substitute the last C3 in the backbone network because the former way will significantly increase the computational cost, while the latter one can already retain the convolution translation invariance and local feature extraction ability of the model.

IV. Experiment and Analysis

To investigate the effectiveness of the model, we conduct the extensive experiments on the two classic and public UAV image datasets, VisDrone2019 [46] and Tiny Person [43].

A. Datasets and Experimental Setup

The VisDrone2019 dataset [46], released by Tianjin University, consists of images taken by various drones under changing weather, diverse scenes and different attitudes. These drone-captured images abound in small/tiny objects with complex environments and serious occlusion problems. The dataset covers ten predefined categories: pedestrians, people, cars, vans, buses, trucks, motorcycles, bicycles, and awning tricycles.

The TinyPerson dataset [43] is the first benchmark dataset for person detection with long-distance backgrounds released by the University of Chinese Academy of Sciences in 2019. It contains two types of person in the sea and on land (sea person and earth person). Low resolution (less than 20 pixels) and great variation of aspect ratio for the person are two main characteristics in this dataset, where the pose and angle of the person is variable, and some of the images have more than 200 tasks.

All the experiments in this paper were carried out on the A100 server with 80 G Video storage and Pytorch 1.12.1. In the training phase, the number of training rounds is set to 150, the batch size is 16, and the rest are the same as the default parameters of YOLOv5.

In order to accurately evaluate the performance of the algorithm, we select Average Precision (AP), Mean Average Precision (mAP) and Recall as the evaluation indicators of this experiment. AP can reflect the detection performance of a single object category. Mean Average Precision measures the comprehensive detection performance of all categories.

B. Experimental Result

In this section, we choosed YOLOv5s and YOLOv8nas our baseline model and show the comparison results between SR-YOLO and the other state-of-the-artmodels on the two datasets. Specifically,SR-YOLO is compared with RetinaNet [47], CenterNet [48], CornerNet [49], YOLOv4, YOLOv5s, YOLOv5m, YOLOx-s, the lastest YOLOv7-tiny [50] and the currently advanced YOLOv8 [51] on metrics mAP_{50}, $mAP_{50:95}$ and recall. From the Table I we can find that SR-YOLO obtains higher mAP_{50}, $mAP_{50:95}$ and Recall than all the other models without adding too many parameters. For example, compared to YOLOv5m, SR-YOLO has only 36.4% of its parameters, but $mAP_{50:95}$ are increased by 8.14%. SR-YOLO had 32% fewer parameters than YOLOv8s and was 2% higher at mAP50:95. It can be seen the superiority of SR-YOLO over the other algorithms on UAV aerial images since it can obtain better detection results with a lower number of parameters. Furthermore, SR-YOLO even surpass YOLOv8s while reducing its number of parameters by 32.1% In addition, to further verify the effectiveness of the proposed model, YOLOv3-tiny, YOLOv3, YOLOx-s, YOLOv5m were selected for comparison and validation on the TinyPerson dataset,on which the experimental results in Table II show that SR-YOLO outperforms YOLOv5s by 3.2 %, 1.01 %, and 3.5% in terms of

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3330788

TABLE I
RESULTS OF DIFFERENT MODELS ON VISDRONE.

Methods	parm	mAP_{50}	$mAP_{50:95}$	Recall
RetinaNet [47]	–	28.7	11.8	–
CenterNet [48]	32.67	22.7	12.4	–
Faster-RCNN [20]	–	33.2	17.0	–
CornerNet [49]	–	34.1	17.4	–
YOLOv4 [11]	64.36	31.2	16.8	–
YOLOXs [24]	8.94	33.9	18.7	–
YOLOv5s	7.04	33.5	18.5	34.2
YOLOv5m	20.89	38.4	22.1	38.0
YOLOv7tiny [50]	6.2	34.4	17.7	37.3
YOLOv8n [51]	3.012	33.50	19.27	33.62
YOLOv8s [51]	11.2	39.54	23.43	38.54
SR-YOLO	7.61	41.6	23.9	40.9
SR-YOLOv8n	2.84	37.99	22.54	37.18

TABLE II
RESULTS OF DIFFERENT MODELS ON TINYPERSON.

Methods	parm	mAP_{50}	$mAP_{50:95}$	Recall
CornerNet [49]	–	27.9	8.61	–
YOLOv3tiny [10]	8.47	10.60	2.66	–
YOLOv3 [10]	60.09	28.12	8.61	29.4
YOLOXs [24]	8.94	28.9	8.79	–
YOLOv5s	7.04	27.1	8.23	29.7
YOLOv5m	20.91	30.2	9.22	33
YOLOv7tiny [50]	6.2	24.2	7.1	27.8
SPD-CONV [52]	8.5	26.20	7.59	29.1
YOLOv8n [51]	3.012	23.50	7.70	25.90
SR-YOLO	7.61	30.3	9.24	32.8
SR-YOLOv8n	2.84	29.2	9.68	33.4

mAP_{50}, $mAP_{50:95}$ and recall, respectivelyIn summary, both of validations on the two benchmarks testify that our SR-YOLO outperforms the YOLOv5s model in all the metrics, proving the effectiveness of the proposed three components.SR-YOLOv8n similarly outperforms the YOLOv8n model in all metrics, demonstrating the generalizability of the proposed components.

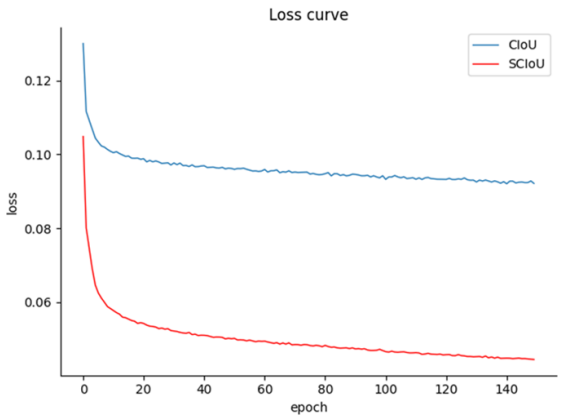


Fig. 6. box-Loss comparison chart.

C. Ablation Experiment

To comprehensively verify the effectiveness of Efficient Neck, SCIoU, and RBoT, we conducted a series of ablation studies and use YOLOv5s as the benchmark model (and denote SR-YOLOv5 as SR-YOLO for simplicity in the following description). As shown in Table III , "E.", "S.", "R." represent our Efficient Neck, Shape CIoU, and RBoT, respectively. We can observe that individually adding S. and E. to the baseline(YOLOv5s) implementation improves its performance by 0.6 and 4.8 in terms of mAP50.It is expected to witness the efficient neck component archives the most gain because of E's small object detection layer as well as up-sampling are designed to cue small object detection rates.

Based on this module, S. and R. are added separately, it shows that the $mAP_{50:95}$ of the model increased by 6.1% and 7.2 %. Finally, after combining all the innovative parts above, the SR-YOLO is 8.1%, 5.4%, and 6.7% higher than the benchmark method in terms of mAP_{50}, $mAP_{50:95}$ and recall, respectively. The comprehensive studied result demonstrates the effectiveness of the proposed model, which can significantly improve the detection accuracy compared to YOLOv5s with only 0.5728M increase in the number of parameters.

To observe the specific accuracy of SR-YOLO on each class of the objects, it shows the specific detection effect for them in Table IV. From the table it can be seen that it gets the best detection with mAP_{50} of 56.3% on the category of Car, followed by Bus with mAP_{50} of 42.8%, while objects with very small size such as Bicycle and Awning-tricycle have poorer detection, only gaining mAP_{50} of 15.9% and 13.5%. In the following experiment, we performed more detailed ablation experiments for each of the three components:

Efficient Neck: Firstly, we summed up the changes in parameters and accuracy after adding a small object detection layer (P2) and CARAFE on YOLOv5. After integrating the P2 layer, the parameter amount increases by 0.68 M, and the accuracy is greatly improved. As shown in Table V, YOLOv5 with Efficient Neck is 4.8 % higher than the benchmark method on mAP50.

SCIoU: To visually demonstrate the effect of SCIoU, we selected DIoU, CIoU and SCIoU to conduct comparative experiments on the model after adding Efficient Neck. From Table we can see that the SCIoU improves 0.5% and 0.3% in mAP50 compared with DIoU and CIoU, respectively.

In addition, to demonstrate the role of the control ratio factor in SCIoU, we compared the status under the value changes of CIoU and SCIoU.Figure 6 shows that the proposed SCIoU model has a lower loss value and converges faster than the same structure with CIoU. The final result of SCIoU gains only half of the Loss value of CIoU (i.e., 0.044469)

RBoT: Finally, to compare the effect of the recursion Multi-Head Self-Attention module RBoT and the BoT module, we conducted our experiments under the four settings, namely adding no BoT(-BoT), original BoT (+BoT), RBoT-a and RBoT-b together with Efficient Neck and SCIoU. From Table VI it can be seen that mAP_{50} increased by 0.5%, and the parameter decreased by 0.11 M after adding BoT. Instead, our proposed RBoT has a lower number of parameters than the original BoT, reducing the model by 0.34 M. Meanwhile, the model equipped with both RBoT-a and RBoT-b is 1.6%

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3330788

TABLE III
THE EFFECTS OF THE YOLOV5 COMBINING DIFFERENT MODULES ON THE VISDRONE.

Methods	parm	mAP_{50}	$mAP_{50:95}$	Recall
YOLOv5	7.037	33.5	18.5	34.2
YOLOv5+SCIoU	7.037	34.1	18.9	34.7
YOLOv5+Efficient Neck	7.948	38.3	21.8	37.3
YOLOv5+Efficient Neck+SCIoU	7.948	39.6	22.7	38.6
YOLOv5+Efficient Neck+RBoT-b	7.61	40.7	23.1	39.8
YOLOv5+Efficient Neck+RBoT-b+SCIoU	7.61	41.6	23.9	40.9

TABLE IV
RECOGNITION EFFECT OF EACH CLASS ON SR-YOLO.

Class	Precision	Recall	mAP_{50}	$mAP_{50:95}$
all	51.8	40.9	41.6	23.9
Pedestrian	54.8	46.4	49.1	21.6
People	52.4	36.2	37.5	14.1
Bicycle	33.3	18.1	15.9	6.15
Car	69.7	81.2	81.9	56.3
Van	54.6	42.4	43.6	30.5
Truck	55.3	36.8	38.4	24.2
Tricycle	44.8	30.5	27.8	14.8
Awning-tricycle	24.8	16.7	13.5	8.26
Bus	71.9	55	61.7	42.8
Motor	56.5	45.3	46.2	19.8

TABLE V
PRECISION COMPARISON OF DIOU, CIOU AND SCIOU.

Methods	mAP_{50}	$mAP_{50:95}$	Recall
DIoU [33]	39.1	22.4	38.5
CIoU [33]	39.3	22.5	38.6
SCIoU	39.6	22.7	38.6

higher than with BoT on mAP_{50}, and 0.8%, 1.1% higher on $mAP_{50:95}$, respectively. Recall also improved significantly. Overall, the performance of RBoT-b is slightly better than RBoT-a.

D. Case study

Consider that it may encounter complicated situations when a UAV is working in real environment, we have selected several representative sets of UAV images from both VisDrone and TinyPerson for case study of intricate conditions. To analyze the above experimental results, this section shows the effect of the proposed model on the example images of VisDrone and TinyPerson datasets.

VisDrone: As shown in Figure 7, the left side is the YOLOv5 benchmark method, and the right side is the SR-YOLO algorithm, where the white box is the area with a more obvious contrast effect. Group (a) shows the most common images using low altitude electronic devices such as UAVs at transportation hubs. The targets in these scenarios are overlapped and occluded. Group (b) selected scenes with high contrast and complex backgrounds to test the model's ability to resist interference. Group (c) shows a scene of traffic monitoring using UAV equipment at night. This type of scene is darker and vehicles are blurred on the image due to high speed movement. (d) An image showing a situation where a vehicle is obscured by a tree while the UAV is checking street parking. In group (a), it is evident that the detection effect of YOLOv5 fails to detect many partially obscured cars and cyclists. In contrast, the improved model detects more obscured objects than the original model, and the boxes for detecting people on cars is accurate from the whole two-wheeler to people on cars. Group(b) shows that in the case of backlight, building shadow, and complex background, the original model can only two targets can be detected, while SR-YOLOv5 correctly detected four objects in such scenarios. In group (c), the complex background and the blurred vehicles at night with high speed are selected, in which the original model identifies the white objects on the roof as vehicles and the moving cars as trucks, while the model in this paper can correctly identify the category of the objects; in group (d), the original model does not detect the vehicles waiting for red lights in the distance and the small two-wheelers in the middle of the road, but SR-YOLO can see these two categories more accurately. From the investigation of the various cases that objects are hard to distinguish, we can see that our proposed model have been tested in different ambient lights, including backlight, low light and drop shadows of the building, indicating the robustness of our SR-YOLO under poor lighting conditions. Meanwhile, the model also shows its excellent performance in traffic intersections when there are lots of person and vehicles.These comparisons show that the improved model is more suitable for UAV aerial image detection.

TABLE VI
COMPARISON OF BOT, RBOT-A, RBOT-B PARAMETER QUANTITIES AND ACCURACY.

Methods	parm	mAP_{50}	$mAP_{50:95}$	Recall
-BoT	7.9479	39.5	22.6	38.6
+BoT	7.8366	40	22.8	39.1
+RBoT-a	7.61	41.6	23.6	41.1
+RBoT-b	7.61	41.6	23.9	40.9

TinyPerson: The comparison results of the YOLOv5 model and SR-YOLO detection are shown in Figure 8, with YOLOv5 on the left and the improved algorithm on the right. We selected (a) (b) (c) three sets of images using a drone Here we select the sub-figures (a), (b) and (c) to examine the distribution of people on a beach with tiny characters and a high amount of useless information in the background. In these three sets of images can be seen that the original YOLOv5 can barely detect tiny and dense people in the distance. In the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3330788



Fig. 7. Test effect comparison effect.

group (d) of fig 6 we have selected the noise images with aboundant scale-varying ships images where there is a lot of background noise (ships) for comparison. From the sub-figure, it can be seen that the proposed method can detect more person compared to the benchmark. Meanwhile, the original YOLOv5 recognizes a large number of sea person as earth person, instead, SR-YOLO accurately recognizes the type of people. In contrast, SR-YOLO can accurately detect more people and objects compared to the original model. (The experiments show that this model is better than the YOLOv5 model for tiny and dense object images.

V. Conclusion

In this paper, we have presented a novel small object detection model SR-YOLO, which consists of an efficient neck, a newly designed prediction box loss function and a recursive bottleneck transformer, exploring the ability to tackle the object occlusion, light variation and intricate background in UAV images. It suggests a new YOLO paradigm for utilizing in electronic products like drones because it can be flexibly integrated into different versions of YOLO. Moreover, the proposed method is efficient due to its number of parameters and accuracy. Our study shows that designing and training lightweight modules of object detectors for application of electronics is an interesting direction which deserves broader focused researches. This work only make an attempt in exploration, and could be further improved in more aspects: how to combine the attention mechanism and feature fusion to capture better detailed and semantic features, and how to further improve the real-time by efficient fusion strategy as well as keep the accuracy for a lighter deployment in consumer electronics.



Fig. 8. Comparison of TinyPerson Dataset Detection Effects.

VI. Acknowledgment

The authors would like to acknowledge the support provided by Researchers Supporting Project number (2023J01954) the Natural Science Foundation of Fujian Province, China. The support provided by Researchers Supporting Project number (RSPD2023R868), King Saud University, Riyadh, Saudi Arabia. Also Khalifa University of Science and Technology under Faculty Start-Up grants FSU-2022-003 Award No. 8474000401.

References

[1] Muhammad Zawish, Lizy Abraham, Kapal Dev, and Steven Davy. Towards resource-aware dnn partitioning for edge devices with heterogeneous resources. In GLOBECOM 2022-2022 IEEE Global Communications Conference, pages 5649–5655. IEEE, 2022.

[2] Muhammad Zawish, Nouman Ashraf, Rafay Iqbal Ansari, and Steven Davy. Energy-aware ai-driven framework for edge-computing-based iot applications. IEEE Internet of Things Journal, 10(6):5013–5023, 2022.

[3] Muhammad Zawish, Steven Davy, and Lizy Abraham. Complexity-driven cnn compression for resource-constrained edge ai. arXiv preprint arXiv:2208.12816, 2022.

[4] Kangcheng Liu and Ben M Chen. Industrial uav-based unsupervised domain adaptive crack recognitions: From database towards real-site infrastructural inspections. IEEE Transactions on Industrial Electronics, 70(9):9410–9420, 2022.

[5] Yang Zhiling. Review of the application of machine vision in aquaculture uavs. In 2023 7th International Conference on Machine Vision and Information Technology (CMVIT), pages 68–72. IEEE, 2023.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3330788

[6] Lucas Prado Osco, Mauro dos Santos de Arruda, Diogo Nunes Gonçalves, Alexandre Dias, Juliana Batistoti, Mauricio de Souza, Felipe David Georges Gomes, Ana Paula Marques Ramos, Lúcio André de Castro Jorge, Veraldo Liesenberg, et al. A cnn approach to simultaneously count plants and detect plantation-rows from uav imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 174:1–17, 2021.

[7] Dunja Božić-Štulić, Željko Marušić, and Sven Gotovac. Deep learning approach in aerial imagery for supporting land search and rescue missions. International Journal of Computer Vision, 127(9):1256–1278, 2019.

[8] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 779–788, 2016.

[9] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7263–7271, 2017.

[10] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.

[11] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.

[12] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, pages 740–755. Springer, 2014.

[13] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. International journal of computer vision, 88:303–338, 2010.

[14] Aftab Ahmed, Jiandong Guo, Fayaz Ali, Farha Deeba, and Awais Ahmed. Lbph based improved face recognition at low resolution. In 2018 international conference on Artificial Intelligence and big data (ICAIBD), pages 144–147. IEEE, 2018.

[15] Farah Deeba, Hira Memon, Fayaz Ali Dharejo, Aftab Ahmed, and Abddul Ghaffar. Lbph-based enhanced real-time face recognition. International Journal of Advanced Computer Science and Applications, 10(5), 2019.

[16] Jiaqi Wang, Kai Chen, Rui Xu, Ziwei Liu, Chen Change Loy, and Dahua Lin. Carafe: Content-aware reassembly of features. In Proceedings of the IEEE/CVF international conference on computer vision, pages 3007–3016, 2019.

[17] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, and Dongwei Ren. Distance-iou loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI conference on artificial intelligence, volume 34, pages 12993–13000, 2020.

[18] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580–587, 2014.

[19] Ross Girshick. Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 1440–1448, 2015.

[20] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28, 2015.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9):1904–1916, 2015.

[22] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 2961–2969, 2017.

[23] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2117–2125, 2017.

[24] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430, 2021.

[25] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16, pages 213–229. Springer, 2020.

[26] Hei Law and Jia Deng. Cornernet: Detecting objects as paired keypoints. In Proceedings of the European conference on computer vision (ECCV), pages 734–750, 2018.

[27] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, pages 21–37. Springer, 2016.

[28] Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, et al. Yolov6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976, 2022.

[29] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In Proceedings of the IEEE/CVF international conference on computer vision, pages 9627–9636, 2019.

[30] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning, pages 6105–6114. PMLR, 2019.

[31] Xiang Long, Kaipeng Deng, Guanzhong Wang, Yang Zhang, Qingqing Dang, Yuan Gao, Hui Shen, Jianguo Ren, Shumin Han, Errui Ding, et al. Pp-yolo: An effective and efficient implementation of object detector. arXiv preprint arXiv:2007.12099, 2020.

[32] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. 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, pages 658–666, 2019.

[33] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, and Dongwei Ren. Distance-iou loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI conference on artificial intelligence, volume 34, pages 12993–13000, 2020.

[34] Abdelmalek Bouguettaya, Hafed Zarzour, Ahmed Kechida, and Amine Mohammed Taberkit. Deep learning techniques to classify agricultural crops through uav imagery: A review. Neural Computing and Applications, 34(12):9511–9536, 2022.

[35] Yu-Fang Huang, Tsung-Jung Liu, and Kuan-Hsien Liu. Improved small object detection for road driving based on yolo-r. In 2022 IEEE International Conference on Consumer Electronics - Taiwan, pages 279–280, 2022.

[36] Jia-Jiun Gung, Chia-Yu Lin, Pin-Fan Lin, and Wei-Kuang Chung. An incremental meta defect detection system for printed circuit boards. In 2022 IEEE International Conference on Consumer Electronics-Taiwan, pages 307–308. IEEE, 2022.

[37] Wen-Hui Chen, Hsiu-Jui Hsu, and Yu-Chen Lin. Implementation of a real-time uneven pavement detection system on fpga platforms. In 2022 IEEE International Conference on Consumer Electronics - Taiwan, pages 587–588, 2022.

[38] Xingkui Zhu, Shuchang Lyu, Xu Wang, and Qi Zhao. Tph-yolov5: Improved yolov5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF international conference on computer vision, pages 2778–2788, 2021.

[39] Raian Rahman, Zadid Bin Azad, and Md Bakhtiar Hasan. Densely-populated traffic detection using yolov5 and non-maximum suppression ensembling. In Proceedings of the International Conference on Big Data, IoT, and Machine Learning: BIM 2021, pages 567–578. Springer, 2022.

[40] Munhyeong Kim, Jongmin Jeong, and Sungho Kim. Ecap-yolo: Efficient channel attention pyramid yolo for small object detection in aerial image. Remote Sensing, 13(23):4851, 2021.

[41] Tao Liu, Bo Pang, Lei Zhang, Wei Yang, and Xiaoqiang Sun. Sea surface object detection algorithm based on yolo v4 fused with reverse depthwise separable convolution (rdsc) for usv. Journal of Marine Science and Engineering, 9(7):753, 2021.

[42] Mohamad Haniff Junos, Anis Salwa Mohd Khairuddin, Subbiah Thannirmalai, and Mahidzal Dahari. Automatic detection of oil palm fruits from uav images using an improved yolo model. The Visual Computer, pages 1–15, 2021.

[43] Xuehui Yu, Yuqi Gong, Nan Jiang, Qixiang Ye, and Zhenjun Han. Scale match for tiny person detection. In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pages 1257–1265, 2020.

[44] Aravind Srinivas, Tsung-Yi Lin, Niki Parmar, Jonathon Shlens, Pieter Abbeel, and Ashish Vaswani. Bottleneck transformers for visual recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 16519–16529, 2021.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2023.3330788

[45] Zhiqiang Shen, Zechun Liu, and Eric Xing. Sliced recursive transformer. In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIV, pages 727–744. Springer, 2022.

[46] Dawei Du, Pengfei Zhu, Longyin Wen, Xiao Bian, Haibin Lin, Qinghua Hu, Tao Peng, Jiayu Zheng, Xinyao Wang, Yue Zhang, et al. Visdrone-det2019: The vision meets drone object detection in image challenge results. In Proceedings of the IEEE/CVF international conference on computer vision workshops, pages 0–0, 2019.

[47] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision, pages 2980–2988, 2017.

[48] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. Objects as points. arXiv preprint arXiv:1904.07850, 2019.

[49] Hei Law and Jia Deng. Cornernet: Detecting objects as paired keypoints. In Proceedings of the European conference on computer vision (ECCV), pages 734–750, 2018.

[50] Xuehui Yu, Yuqi Gong, Nan Jiang, Qixiang Ye, and Zhenjun Han. Scale match for tiny person detection. In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pages 1257–1265, 2020.

[51] Ultralytics, "YOLOv8," 2022, Accessed 23 February 2023. [Online].Available: https://github.com/ultralytics/ultraly.

[52] Raja Sunkara and Tie Luo. No more strided convolutions or pooling: A new cnn building block for low-resolution images and small objects. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pages 443–459. Springer, 2022.



Le Wang received received his Ph.D degree from the Institute of Computer Network Information Center Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China in 2021. He joined College of Computer Science and Mathematics of FJUT as an instructor and researcher.



Yu Shi received his bachelor degree from Wuxi Taihu University in 2021. He is currently pursuing a Master Degree in the College of Computer Science and Mathematics, Fujian University of Technology, FuZhou, China. His current research interests include computer vision and machine learning.



Guo-jun Mao is currently the dean at Fujian University of Technology, a member of the standing committee of Chinese Association for Artificial Intelligence, a member of National Science and Technology Awards committee, etc.: previously he was a professor at Central University of Finance and Economics and a senior visiting scholar at University of Vermont. He received his Ph.D. degree from the Beijing University Of Technology in 2003. He works in the field of artificial intelligence, data mining, and big data and distributed computing.



Fayaz Ali Dharejo is Senior IEEE Member and currently Postdoc fellow at Khalifa University. He received his PhD Degree in Computer Applied Engineering from the Institute of Computer Network Information Center Chinese Academy of Sciences, University of Chinese Academy of Sciences Beijing, China. He received B.E. degree in Electronic Engineering from QUEST, Pakistan and received M.E. degrees from University of Electronic Science and Technology of China, Chengdu, China. He has published more than 30 articles in reputed ISI impact factor journals including; IEEE TFS, ACM TIST, International Journal of Intelligent Systems, IEEE TCBB, IEEEE GRSL and etc. He is a member of professional bodies such as PEC, ACM, IEEE, IEEE Societies, such as IEEE Geoscience and Remote Sensing Society, IEEE Computer Society and IEEE Signal Processing Society. His research interests are image enhancement, lightweight models, and computer vision.



Sajid Javed is a faculty member at Khalifa University (KU), UAE. Prior to that, he was a research fellow at KU from 2019 to 2021 and at the University of Warwick, U.K, from 2017-2018. He received his B.Sc. degree in computer science from the University of Hertfordshire, U.K, in 2010. He completed his combined Master's and a Ph.D. degree in computer science from Kyungpook National University, Republic of Korea, in 2017. His research interests include visual object tracking in the wild, multi-object tracking, background-foreground modeling from video sequences, moving object detection from complex scenes, and cancer image analytics including tissue phenotyping, nucleus detection, and nucleus classification problems. His research themes involve developing deep neural networks, subspace learning models, graph neural networks.



MOATH ALATHBAH received his Ph.D. degree from Cardiff University, U.K. He is currently an Assistant Professor with King Saud University, Saudi Arabia. His research interests include the development of photoelectronic, integrated electronic active and passive discrete devices, the design, fabrication, and characterization of MMIC, RF and THz components, smart antennas, microstrip antennas, microwave filters, meta-materials, 5G antennas, MIMO antennas miniaturized multiband antennas/wideband and microwave/millimeter components using micro and nano technology.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/