ISSN: 2664-9640

UAV-YOLOv5: A Lightweight Object Detection Algorithm on Drone-captured Scenarios

Guoliang Li[1,2], Enguang Liu[3], Yanhong Wang[1,2,*], Yang Yang[1], and Haiying Liu[4]

[1] School of Control Science and Engineering, Shandong University, Jinan 250061, China

[2] Unimation Intelligent Technology Co., Ltd, Jinan 250101, China

[3] Shandong Transportation Institute, Jinan 250031, China

[4] School of Information and Automation, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China

*wangyanhong@sdu.edu.cn

Abstract

Aiming at the problems of common object detection algorithms on drone-captured scenarios, such as too large model, difficult deployment, low accuracy of small-scale object detection, this paper proposed a series of improved methods based on YOLOv5, which effectively improved the performance of the algorithm on drone-captured scenarios. A new dual-branch CSPNet (DB-CSPNet) structure was proposed, which effectively reduced the complexity and computation of the model. A new feature fusion path (FS-FPN) was proposed, which effectively improved the detection accuracy of the model. By integrating a attention mechanism (ACmix), the performance of the model is effectively improved. The experimental results shown that the proposed methods have a significant improvement effect on the accuracy of the object detection algorithm on drone-captured scenarios. The mAP@0.5 and mAP@0.5:0.95 of the algorithm which used the method 2 and 3 proposed in this paper can be improved by 2.5% and 1.6%. At the same time, the method 1 proposed in this paper can also achieve good lightweight effect, the model parameters and FLOPs can be reduced by 26.6% and 30.4%. The UAV-YOLOv5 implemented by all methods in this paper can also achieve a good balance between precision and lightweight. Compared with the default YOLOv5s, the mAP@0.5 and mAP@0.5:0.95 increased by 1.5% and 1.0%, and the parameters and FLOPs decreased by 3.7% and 7.0% respectively.

Keywords

Object Detection; Small Object; Drone; YOLO.

1. Introduction

With the development of deep learning technology, the constant proposal and improvement of various large public datasets[1, 2, 3], and the emergence of various high-performance GPUs and mobile edge devices, object detection algorithms using deep convolutional neural networks as the main structure have made considerable progress in many fields.

When UAV is used for aerial object detection, the detected objects are similar in category distribution, small in scale and large in number, which is different from ordinary life scenes. At the same time, the inference computing equipment of most UAVs are not enough to support the smooth operation of large neural networks. Therefore, it is of great research value to design a lightweight object detection algorithm suitable on drone-captured scenarios.

In the task of object detection, YOLO series algorithms play an extremely important role. YOLOv5[4] is the most mature version of YOLO series algorithms at present. Compared with other versions, YOLOv5 has strong industrial deployment capability, high interpretability, stable detection accuracy in different network scales and datasets. In this paper, we used YOLOv5 as the basic algorithm, and proposed a series of improved strategies to solve the problems of the object detector on drone-captured scenarios. We referred to the algorithm that integrated all the improved methods in this paper as UAV-YOLOV5.

On the basis of CSPNet[5], a new dual-branch CSPNet (DB-CSPNet) as the main network structure was designed to improved algorithm. By introducing new branches to divide the input information of the network, the detector achieved lightweight. Focusing on how to make the detector better detect small and medium-sized targets, the feature pyramid mechanism of YOLOv5 was adjusted, and a new feature fusion path FS (Focus-Small) - FPN was proposed, which greatly improved the detection accuracy of the algorithm. In order to further improve the performance of the model, and out of interest in the attention mechanism, a attention mechanism (ACmix[6]) was integrated in the network feature extraction and achieved good results.

The major innovations and contributions of this research are as follows:

- On the basis of CSPNet, a new type of dual-branch CSPNet (DB-CSPNet) structure was designed, which can greatly reduce the weight of the model and ensure the detection ability of the model.
- A new feature pyramid structure (FS-FPN) was proposed, which effectively improved the detection accuracy of the algorithm without increasing the number of feature pyramid layers.
- ACmix was introduced into the feature extraction network of YOLOv5, to help the model better detect objects on drone-captured scenarios and improve the model performance.

2. Related Work

2.1. Object Detector for UAV

Object detection on drone-captured scenarios is mainly focused on high-altitude ground detection. The number of detected objects is relatively large, the size is small, and the category distribution is similar. This requires that the detector has a good ability to detect small targets and dense small targets, and the requirements for the model's multi category prediction ability are relatively low. Considering that the inference computing unit of most UAVs is not enough to support the operation of large neural networks, this requires that the scale of the model should be relatively small, or the network size can be controlled by scaling settings. At the same time, the object detection on drone-captured scenarios belongs to real-time detection, so the real-time detection capability of the detector is highly required.

Considering the maturity, deployment capability, stability, generalization capability and low-end device friendliness of the algorithm, this paper used YOLOv5s as the basic algorithm and improved around it.

2.2. Mainstream Network Module

The structure of the object detector can be considered as the combination of the network module and the detector head according to the fixed normal form (such as feature extraction network add feature pyramid). Network module is an important part of the detector, and its performance greatly affects the final detection effect of the detector.

In recent years, excellent modules include CSPNet[5], RepVgg[7], etc. CSPNet achieves a good balance between detection accuracy and lightweight through clever gradient isolation strategy, and is widely used in YOLOv4[8], YOLOv5[4] and other algorithms. At the same time, some traditional modules, such as ResNet[9] and ResNeXt[10], can also play a very good lightweight effect after reconstruction using the idea of CSPNet.

ISSN: 2664-9640

Considering that the object detection on drone-captured scenarios needs both detection accuracy and real-time performance, and the model needs to be as light as possible, the module itself should have good interpretability and generalization ability. This paper attempts to make a series of improvements to CSPNet, and has achieved more excellent lightweight effect.

2.3. Feature Pyramid Network (FPN)

The original intention of FPN[11] is to enable the detector to detect objects of different scales. The source of this problem can be traced back to the design shortcomings of early object detection algorithms. The method of combining feature extraction network with FPN can not only enable the model to have multi-scale detection capability, but also improve the model's detection capability for specific scale targets to a certain extent. Thanks to the excellent performance of FPN, on this basis, various varieties of FPN emerge in endlessly, such as PAN[12], BiFPN[13],NAS-FPN[14], etc.

Based on the above, this paper designed a novel feature fusion path for the characteristics of the objects to be detected on drone-captured scenarios, which effectively improved the detection accuracy of the model for small and medium scale objects on the basis of ensuring the multi-scale detection capability of the model.

2.4. Attention Mechanism

The great achievements of attention mechanism in natural language processing also provide a new idea for the design of object detector.

This paper continues the idea of combining attention mechanism with convolution. Based on the network structure of YOLOv5, it tested and compared a variety of new visual attention modules, and finally selectd ACmix[6], which was combined with YOLOv5 to improve the object detection performance of the detector on drone-captured scenarios.

3. Approach

In order to solve the problems of common object detectors on drone-captured scenarios and improve the performance of the algorithm, this paper proposed a series of improved methods, and designed UAV-YOLOv5 according to these methods. In this section, we first gave the overall network structure of UAV-YOLOv5, and then introduced the methods proposed in this paper in turn, including DB-CSPNet, FS-FPN feature fusion path and ACmix structure which combined with YOLOv5.

3.1. The Overall Structure of UAV-YOLOv5

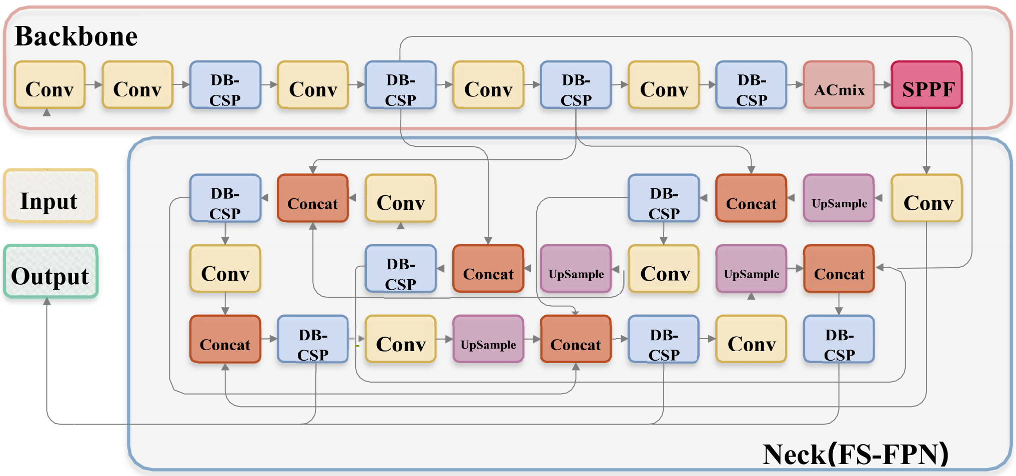
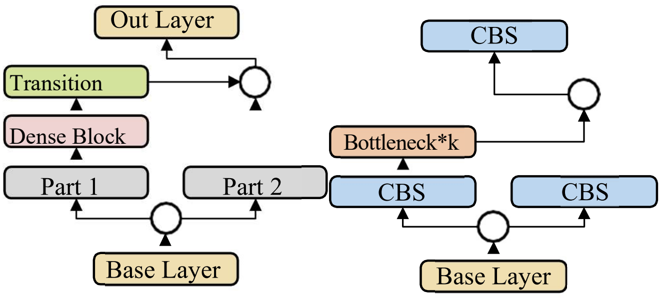


Fig. 1 The overall structure of UAV-YOLOv5, including Backbone and feature pyramid(Neck)

The overall network structure is shown in Fig. 1. For the purpose of improving the lightweight of the detector, we designed DB-CSPNet, and replaced the default C3 (a specific implementation of CSPNet[5]) structure in YOLOv5 with this structure. For the purpose of improving the detection effect of the model on small and medium-sized objects, we designed FS-FPN to replace the PAN structure in the Neck part. Finally, in order to further improve the performance of the detector and explore the impact of attention mechanism on the traditional convolutional neural network, we introduced ACmix[6] at the end of Backbone, effectively improving the performance of the model.

3.2. DB-CSPNet

The core idea of CSPNet is to use two separate branches to calculate the gradient to avoid the gradient duplication problem, which not only improved the performance of the model, but also realized the lightweight of the model in a disguised way. The schematic diagram of CSPNet and the C3 module implemented by YOLOv5 according to this idea are shown in Fig. 2.



(a)A common CSPNet (b)C3 module

Fig. 2 Basic idea of CSPNet and the C3 module implemented by YOLOv5

The reason for the lightweight of CSPNet is that the horizontal branches indirectly reduce the number of input channels in the convolutional network, and the decrease in the number of dense layers parameters saves a lot of computing resources. The reason why CSPNet can still achieve high accuracy while ensuring the lightweight of the model is that the gradient isolation of the branches disguises the problem of vanishing gradients and exploding gradients, allowing the network to learn and utilize features more fully.

The basic idea of DB-CSPNet is to continue branching the branch of partial dense layers based on CSPNet, so DB-CSPNet can also be understood as the superposition of two CSPNet. By further branching horizontally, DB-CSPNet can achieve more significant lightweight.

Compared with the basic CSPNet, the DB-CSPNet was proposed in this paper has two main improvements:

- A new network branch was added, and the original network branch was adjusted to further develop the CSPNet gradient separation idea. At the same time, the horizontal channel separation also enabled the model to achieve lightweight.
- On the basis of the above, some minor adjustments had been made to the default branch, which improved the detection accuracy of the model without significantly increasing the amount of computation.

The structure of DB-CSPNet corresponding to the above two improvements is shown in Fig. 3. To facilitate differentiation, the modules corresponding to the above two improvements were named DB-CSPNet.1 and DB-CSPNet.2. This paper used the structure which was shown in Fig. 3(b) to design UAV-YOLOv5.

ISSN: 2664-9640

3.3. FS-FPN

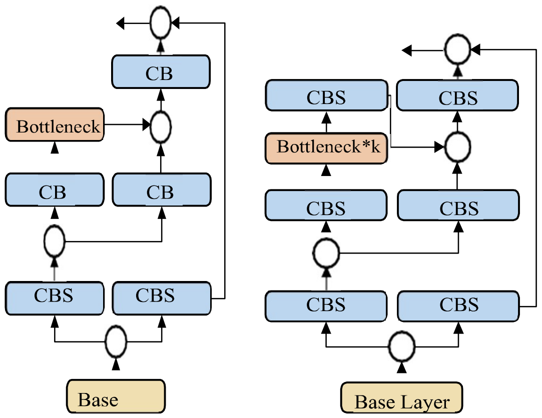
At present, the feature fusion network of the mainstream one-stage object detectors mostly use the PAN[12] structure, such as YOLOv5 selected in this paper, YOLOv7, the latest version of YOLO series, etc. Compared with the default FPN structure, PAN reintroduces a bottom-up feature fusion path, which enhances the feature representation capability of the feature pyramid by bringing the richer location information of the bottom feature map into the high-level feature map and combining it with the richer semantic features of the high-level feature map.

However, both bottom and high-level feature maps have different feature information, and their more important task is to ensure the multi-scale prediction capability of the model. In network design, the number of feature maps with different sizes will directly determine the number of network prediction heads, thus greatly affecting the detection effect of the model.

In practical applications, different scenes have different preferences for different scale, feature maps. For example, object detection on drone-captured scenarios has small size, large number and dense location distribution of objects to be detected. In this case, how to balance the output quantity, output size, location information and semantic information of the feature map in the feature pyramid has a very important guiding role in improving the detection effect of the model.

To improve the performance of the model, a simple and effective method is to add the corresponding feature map output in the feature pyramid, and then increase the number of detection heads. In the past, this method has been widely used in the detection of small objects and ultra-small objects. The problem with this method is that it will greatly increase the post-processing (NMS) time of the model and reduce the real-time detection capability of the model, although this has not been mentioned in most papers.

In this paper, we abandoned the idea of increasing the output of feature maps from the vertical, and changed to further stack and strengthen the feature maps in the horizontal. The well-designed fusion path was used to strengthen the feature information applicable to the drone-captured scenarios and enhanced the feature expression ability of the model, so as to improve the model performance on the premise of ensuring the real-time detection ability of the model.



(a)DB-CSPNet.1 (b) DB-CSPNet.2

Fig. 3 Two DB-CSPNet structures proposed in this paper

Although FS-FPN will increase the parameters and computation of the model to a certain extent, combined with the DB-CSPNet structure proposed in this paper, the final algorithm still has considerable lightweight effect.

3.4. Combined Attention Mechanism

In recent years, convolution and attention mechanism have made great achievements in the field of vision respectively, but they usually follow different design paradigms. Some recent studies have proved the feasibility of combining convolution with attention mechanism to improve the performance of the model. For example, ACmix[6] has achieved better results by ingeniously integrating convolution and attention mechanisms.

In this paper, we focused on ACmix and embedded it in the end of YOLOv5 feature extraction network with the idea of modularization (most attention mechanisms have poor sampling effect in shallow networks, and the common embedding position is the end of feature extraction network, so as to give full play to the complementary role of convolution and attention mechanisms), effectively improving the performance of the model.

4. Experiments

4.1. Implementation Details

All the experiments in this paper were completed on the ubuntu 20.04 operating system. The experiments used Python 3.8 programming language, PyTorch 1.10.0 deep learning framework, Cuda11.3 for computing acceleration, and all models were trained and verified on NVIDIA RTX 3090 GPU.

In comparison, we used mAP@0.5 and mAP@0.5:0.95 to evaluate the detection accuracy of the algorithm, and the parameters, FLOPs and inference time were to evaluate the lightweight effect of the algorithm.

4.2. Dataset

In order to fully verified the detection performance of UAV-YOLOv5 proposed in this paper on drone-captured scenarios, this paper used the VisDrone[15] dataset to train, and verify the model.

This dataset was collected by the AISKYEYE team in the Machine Learning and Data Mining Laboratory of Tianjin University, China. The benchmark dataset consists of 400 video clips formed by 265228 frames and 10209 still images, captured by various UAV cameras, covering a wide range of aspects, including location (shot from 14 different cities thousands of kilometers apart in China), environment (cities and countries), objects (pedestrians, vehicles, bicycles, etc.) and density (sparse and crowded scenes).

In this experiment, the split of the dataset is consistent with the official, the ratio of the number of pictures in the training set to that in the verification set is about 11:1.

4.3. Experimental Results and Comparison

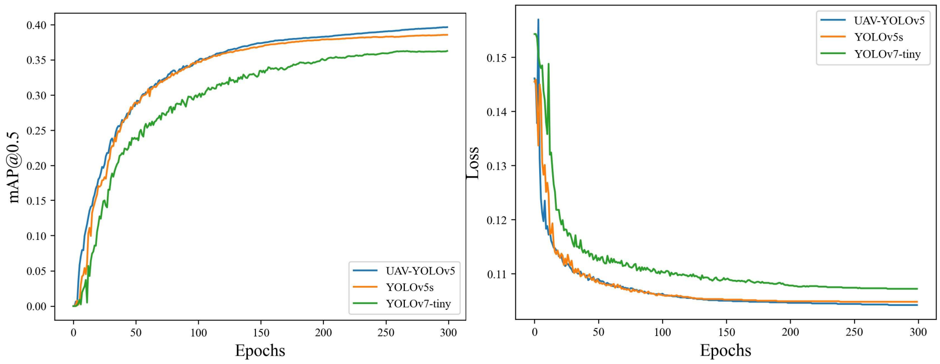
According to the experimental setup in 4.1, we used mAP@0.5, mAP@0.5:0.95, Param eters (M), FLOPs (G) and Time (ms) to evaluate the algorithm. In addition to comparing YOLOv5s with UAV-YOLOv5 proposed in this paper, we also compared it with other mainstream algorithms, including the YOLOv6n, YOLOv6l, YOLOv7-tiny, YOLOv7, and the classic YOLOv3, ResNet-CSP and ResNeXt-CSP. ResNet-CSP and ResNeXt-CSP are the new detector combined with classical algorithm ResNet, ResNeXt and CSPNet, which has excellent performance.

The algorithm comparison results are shown in TABLE 1,, and part of the train process is shown in Fig.4.

According to the comparison of the experimental results in TABLE 1, it is not difficult to see that compared with the default YOLOv5s, the mAP@0.5 and mAP@0.5:0.95 of UAV- YOLOv5 was increased by 1.5 and 1.0 respectively. Considering the large number of small and medium-sized objects in the dataset, the increase of this value is enough to prove that UAV-YOLOv5 can achieve a better detection effect on drone-captured scenarios.

Table 1. Comparison of experimental results

Methods	mAP@ 0.5	mAP@ 0.5:0.95	Parameters(M)	FLOPs(G)	Time(ms)
YOLOv5s	38.1	20.4	7.04	15.8	14.1
YOLOv6n	36.4	19.1	4.63	11.3	9.76
YOLOv7-Tiny	35.8	18.6	6.03	13.1	10.7
UAV-YOLOv5	39.6	21.4	6.78	14.7	13.5
ResNeXt-CSP	43.8	24.3	29.0	51.9	17.1
ResNet-CSP	45.3	25.2	33.9	63.9	20.9
YOLOv3	47.6	26.8	61.54	154.7	47.9
YOLOv5x	47.9	27.2	86.23	203.9	22.8
YOLOv6l	47.3	27.1	59.54	150.5	21.3
YOLOv7	48.6	27.5	36.53	103.3	19.6
YOLOv7x	49.6	28.5	70.84	188.2	19.7
UAV-YOLOv5l	47.7	26.7	32.39	71.3	19.2



(a) mAP@0.5 (b) Loss Function

Fig. 4 Some experimental results, including the comparison of the mAP@0.5 of the UAV-YOLOv5, YOLOv5s,YOLOv7-tiny and the convergence comparison of loss function of these algorithms

It can be seen from the experimental data in TABLE 1 that the values of Parame ters (M) and FLOPs (G) of UAV-YOLOv5 were decreased by 3.7% and 7.0% respectively. And compared to the default algorithm, UAV-YOLOv5 had higher inference speed. The experimental results show that UAV-YOLOv5 can not only effectively improves the object detection accuracy on drone-captured scenarios, but also gives consideration to the lightweight effect.

To sum up, on drone-captured scenarios, compared with the default algorithm, UAV- YOLOv5 had improved both the detection accuracy and the lightweight effect. And the latest lightweight version of YOLOv7 also had a certain gap compared with the UAV- YOLOv5 proposed in this paper.

Although the UAV-YOLOv5 proposed in this paper was mainly for lightweight networks, the UAV-YOLOv5l obtained after amplification had no significant difference in accuracy compared with the latest YOLOv7 and YOLOv7x, and the UAV-YOLOv5l had significantly reduced the computation and parameters (compared with YOLOv7, YOLOv7x, the FLOPs of the UAV-YOLOv5l was decreased by 62.1% and 30.9%), and on this basis, the detection accuracy was the same as YOLOv3, it's enough to show that the method proposed in this paper not only has a significant comprehensive improvement effect on lightweight networks, but also can achieve good results on large networks.

The improvement effects of the methods proposed in this paper will be further ex plained in the ablation experimental analysis.

ISSN: 2664-9640

4.4. Ablation Studie

UAV-YOLOv5 mainly includes three improvements, DB-CSPNet, FS-FPN (named B in this section) and the combination of ACmix (named C in this section). DB-CSPNet also contains two improvements (named A.1 and A.2 in this section). This section will give the ablation experimental data of all the above methods. TABLE 2 is a comparison of comprehensive data including all evaluation indicators.

Table 2. Results of ablation experimen

Methods	mAP@ 0.5	mAP@ 0.5:0.95	Parameters(M)	FLOPs(G)	Time(ms)
YOLOv5s	38.1	20.4	7.04	15.8	14.1
YOLOv5s+A.1	36.5	19.4	5.13	10.9	13.2
YOLOv5s+A.2	37.1	19.8	5.17	11.0	12.8
YOLOv5s+A.2+B	38.9	20.8	5.68	13.0	13.1
UAV-YOLOV5	39.6	21.4	6.78	14.7	13.5

The ablation experiment proved that the DB-CSPNet.1 with new branches can achieve the best lightweight effect, the parameters and calculation amount were greatly reduced (FLOPs and Parameters were reduced by 31% and 27.1% respectively). At the same time, 2 the reduction of detection accuracy was also in an acceptable range. In some scenarios 2 where detection accuracy is not emphasized, the improved lightweight effect can play an excellent role.

After some adjustments to the structure of DB-CSPNet.1, we get DB-CSPNet.2, which has better detection accuracy than the former. The mAP@0.5 and mAP@0.5:0.95 of DB- CSPNet.2 were increased by 0.6 and 0.4 respectively.

To sum up, the introduction of DB-CSPNet has reduced the overall calculation and parameters of the model, not only achieving better lightweight effect, but also providing a good buffer space for subsequent improvement. Even if the subsequent improvement will increase some calculation and parameters, the final UAV-YOLOv5 can still achieve good lightweight effect.

After the embed of ACmix, the number of parameters and computation of the model have been improved to a certain extent, but considering the improvement in detection performance (mAP@0.5 was increased by 0.7), the overall effect is still in line with expectations. By introducing attention mechanism in the high-sampling layer, combined with the advantages of convolutional neural network and attention mechanism, the two complementary, can effectively improve the performance of the model.

5. Conclusion

We analyzed and referred to the excellent ideas of CSPNet, and designed a double branch CSPNet structure (DB-CSPNet), which effectively realized the large scale lightweight of the model. For the purpose of improving the detection effect of the detector on small and medium targets on drone-captured scenarios. We proposed a new feature pyramid structure (FS-FPN). The ablation experiment proves that FS-FPN can significantly improve the detection accuracy of the algorithm, and also provided a new idea for our subsequent research, that is, to strengthen the feature fusion network to improve the performance of the model. We have tried some of the latest attention mechanisms and embedded ACmix into YOLOv5 with the idea of modularization. Although the lightweight effect of the model had been offset to a certain extent, the combination of convolution and attention mechanism is undoubtedly an effective way to improve the performance of the algorithm. Although this requires a lot of experiments to verify, especially the detection of drone-captured scenarios which is different from the general living environment.

The UAV-YOLOV5 proposed in this paper aims to give consideration to both the lightweight and the detection accuracy of the algorithm. Although it has achieved good results, the trade-off in the lightweight aspect also affects the detection accuracy improvement of the model to a considerable extent. In general, UAV-YOLOV5 still has a lot of room for improvement. In the subsequent research, we will focus on a larger scale network, take improving the detection accuracy as the main research direction, and look for a more significant detection accuracy improvement strategy. We hope that this study can provide some useful references for subsequent researchers.

Acknowledgments

This research was funded by QLUTGJHZ2018019.

References

[1] Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[2] Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

[3] Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European conference on computer vision (ECCV), 2014, pp. 740–755.

[4] Glenn, J. 6.2 - YOLOv5 Classification Models, Apple M1, Reproducibility, ClearML and Deci.ai integrations, 2022. https: //github.com/ultralytics/yolov5/releases/tag/v6.2.

[5] Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR) Workshops, June 2020.

[6] Pan, X.; Ge, C.; Lu, R.; Song, S.; Chen, G.; Huang, Z.; Huang, G. On the Integration of Self-Attention and Convolution. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 815–825.

[7] Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. RepVGG: Making VGG-Style ConvNets Great Again. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2021, pp. 13733–13742.

[8] Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.109342020.

[9] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

[10] Xie, S.; Girshick, R.; Dollar, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[11] Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[12] Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[13] Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.

[14] Ghiasi, G.; Lin, T.Y.; Le, Q.V. NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[15] Zhu, P.; Wen, L.; Du, D.; Bian, X.; Fan, H.; Hu, Q.; Ling, H. Detection and Tracking Meet Drones Challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence 2021, pp. 1–1. https://doi.org/10.1109/TPAMI.2021.3119563.