

YOLO-Drone: Airborne real-time detection of dense small targets from high-altitude perspective

Li Zhu, Jiahui Xiong, Feng Xiong, Hanzheng Hu, Zhengnan Jiang
Youming Zhang

Abstract—Unmanned Aerial Vehicles (UAVs), specifically drones equipped with remote sensing object detection technology, have rapidly gained a broad spectrum of applications and emerged as one of the primary research focuses in the field of computer vision. Although UAV remote sensing systems have the ability to detect various objects, small-scale objects can be challenging to detect reliably due to factors such as object size, image degradation, and real-time limitations. To tackle these issues, a real-time object detection algorithm (YOLO-Drone) is proposed and applied to two new UAV platforms as well as a specific light source (silicon-based golden LED). YOLO-Drone presents several novelties: 1) including a new backbone Darknet59; 2) a new complex feature aggregation module MSPP-FPN that incorporated one spatial pyramid pooling and three atrous spatial pyramid pooling modules; 3) and the use of Generalized Intersection over Union (GIoU) as the loss function. To evaluate performance, two benchmark datasets, UAVDT and VisDrone, along with one homemade dataset acquired at night under silicon-based golden LEDs, are utilized. The experimental results show that, in both UAVDT and VisDrone, the proposed YOLO-Drone outperforms state-of-the-art (SOTA) object detection methods by improving the mAP of 10.13% and 8.59%, respectively. With regards to UAVDT, the YOLO-Drone exhibits both high real-time inference speed of 53 FPS and a maximum mAP of 34.04%. Notably, YOLO-Drone achieves high performance under the silicon-based golden LEDs, with a mAP of up to 87.71%, surpassing the performance of YOLO series under ordinary light sources. To conclude, the proposed YOLO-Drone is a highly effective solution for object detection in UAV applications, particularly for night detection tasks where silicon-based golden light LED technology exhibits significant superiority.

Index Terms—FPN, UAV, silicon-based golden LED light.

I. INTRODUCTION

Object detection is a fundamental and important problem in computer vision that involves automatically detecting visible objects of certain classes in images [1], [2]. With the advancement of deep learning in recent years, intelligent object detection systems have become widely used in a variety of applications, including autonomous driving, video surveillance, medical imaging, pedestrian safety, and so on [3], [4]. Unmanned aerial vehicles (UAVs), specifically drones equipped with remote sensing object detection technology, have recently found a wide range of applications including city surveillance, traffic security, person rescue, disaster management, etc. As a mobile platform, UAVs can acquire photography more cheaply and flexibly, making them a powerful supplement for satellite and airborne remote sensing [5], [6]. Real-world

scenarios, on the other hand, frequently involve small, blurry, low-resolution objects, complicated backgrounds, and varying lighting conditions for drone image. Furthermore, these objects are densely packed and located in arbitrary directions. Consequently, small object detection precision and accuracy in aerial images remain an unsolved challenge [6].

Current small object detection research focus on include network architectures based on multi-scale feature learning, context-based, and generative adversarial networks, as well as optimization methods such as data augmentation and training strategies [7]. Feature pyramid network (FPN) is crucial component of the small object detection network utilizing multi-scale feature learning [8], [9]. SSD (Single Shot MultiBox Detector) [10] and YOLO (You Only Look Once) [11]–[14] are both one-stage detectors with FPN hierarchies. These one-stage detectors create and classify object's occurrence areas simultaneously in one network. While the two-stage detectors, such as R-CNN (Regions with Convolutional Neural Network features) [15], SPP-Net [16], Fast R-CNN [17], and Faster R-CNN [18], firstly create area proposals of objects and then classify the objects only in those areas. The One-stage detector is fast because it uses a single feedforward neural network for region category prediction and fine-tuning. Although two-stage detectors have successfully implemented accurate models with mean averaging accuracy (mAP), their speed is insufficient for most real-time applications. For the purpose of enhancing the effectiveness of object detection, numerous researchers have put forth their enhancement strategies. Singh et al. used various training procedures for Scale Normalization of Image Pyramids (SNIP) to address the recognition and detection issues caused by excessive scale shifts [19], [20]. Kisanta et al. proposed a technique for copying small objects from one image to another and oversampling images with small objects [21]. Guan et al. integrated contextual data to improve object detection performance [22]. Lin et al. presented a multi-scale feature learning method for feature fusion to improve the effectiveness of object detection [8]. Bai et al. combined with GAN to detect objects [23]. However, object detection in complex scenarios from the UAV's perspective remains a challenge.

Object detection from a UAV perspective is the basis for a variety of sophisticated applications, including precise cargo delivery, target tracking, safety inspection, disaster warning, and so on [24]. Aerial images, which are often captured by drones, aircraft, or satellites, differ greatly from those of natural images. The issue of object detection in aerial images is challenging due to the following four considerations [25].

This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

Manuscript received April 19, 2021; revised August 16, 2021.

1. The size and shape of the objects vary depending on the angle that the aerial image is viewed. 2. With aerial images, sample imbalance is a prevalent problem. 3. The objects in aerial images are often obscured from each other, and the boundaries of objects are often truncated. 4. Small objects account for a higher proportion of all entities in the total aerial image than in natural photographs [26].

Although deep-learning-based approaches have made significant progress in object detection, miss-detection issues persist in UAV. Small objects lack the visual information required to differentiate them from the background or similar categories. Most existing algorithms have low positioning accuracy due to the small number of pixels in small objects. Furthermore, prior knowledge of small objects is extremely limited [27]. Taking these concerns into account, many novel frameworks have been proposed. Chen et al. proposed a data enhancement method based on super resolution GAN to detect small vehicles in UAV images [28]. Sommer et al. used Faster R-CNN deconvolution to upsample deep features and combined the upsampled features with shallow features to achieve a sufficiently high Feature map resolution [29], [30]. Jiang et al. proposed a single-shot detector for small object detection by fusing high-level semantic information into a shallow layer and introducing spatial attention [31]. Shivapriya et al. combined faster RCNN with the additive activation function (AAF) for better convergence and clear bounding variance in object detection [32]. Mandal et al. proposed the AVDNet, which uses ConvRes blocks at various scales to prevent features from disappearing for small objects in deeper layers [33]. Li et al. proposed a method for cropping the original image into small crops and then using a density map-guided algorithm to detect objects in each small crop [34].

It is worth noting that the light source is considered to be another important factor affecting the quality of aerial images, and in the object detection model, it directly affects the quality of the input images. Due to insufficient illumination, images in night object detection tasks have problems such as low brightness, low contrast, noise, and artifacts, and early researchers frequently focus on image enhancement techniques in low light. Kuang et al. proposed a novel bionic image enhancement method combined with weighted feature fusion technology to achieve night image enhancement [35]. Wang et al. used the hyperbolic tangent curve to adjust image brightness, and proposed block matching and optimized 3D non-sharpening filter in the YCbCr color space, which improved low-light image quality to some extent [36]. Qu et al. proposed an image conversion network based on CycleGAN for object detection under low light conditions [37]. Kim et al. combined an image quality improvement and object detection networks to enhance the image quality in low-light conditions [38]. Liu et al. proposed a compensation method for changes in illumination LogAbout, resulting in a significant improvement in detection accuracy [39]. However, many enhancement technologies consume more computing resources and low detection efficiency. Improving the light source itself is frequently overlooked when solving the problem of object detection in low light at night.

To improve the performance of object detection at night, it is necessary to use special light sources to enhance image

quality and to apply improved object detection models. The silicon-based golden LED developed by Jiang et al. is a novel type of special light source [40]–[44]. Its luminous properties outperform many other light sources like halogen lamps, fluorescent lamps, high intensity discharge lamps, and traditional LED. However, no studies of object detection using this light source have been conducted. However, there are few researches on object detection based on this light source in the current research.

In this study, a new network for object detection from the perspective of UAVs, the YOLO-Drone is proposed for small and dense objects. And a night silicon-based golden (UAV-LED-G) dataset is made for night detection. The performance of the proposed YOLO-Drone was verified on two benchmark datasets, UAVDT and VisDrone2019-DET, as well as a night dataset UAV-LED-G under the special light source silicon-based golden LED.

The following are the main contributions of this study.

1) The proposed YOLO-Drone has high object detection performance and fast inference speed on UAV datasets. The mAP@.5 of YOLO-Drone in UAVDT, VisDrone2021-DET, and UAV-LED-G is 34.04%, 34.30%, and 87.71%, respectively, and the inference speed reaches 53 FPS, which is significantly better than ClusDet, DMNet, YOLOv3, and YOLOv5 series.

2) Two new basic feature networks are proposed in this study, the backbone network Darknet59 and the neck network MSPPFPN. The branch structure of the four-scale detector is built based on Darknet59, which improves the detection ability of small object by extracting low-dimensional high-frequency information. The MSPP-FPN structure combines one spatial pyramid pooling (SPP) and three atrous spatial pyramid pooling (ASPP) to improve feature aggregation in the object detection neck stage.

3) The silicon-based golden LED is used for the first time to detect objects. The first large dataset for object detection collecting under this light source is established. This specific light source has been shown to perform better than others in challenging tasks requiring object detection at night.

4) This study developed one multi-rotor autonomous cruise UAV and one hybrid-winged autonomous cruise UAV independently using YOLO-Drone and edge computing platforms. They've achieved real-time detection of small objects from the perspective of drones, which can be applied in multi-task and multi-scene scenarios such as security, transportation, intelligent city, and so on.

II. METHODS

A. Object Detection Algorithm

In this study, a new end-to-end object detection algorithm, named YOLO-Drone, is proposed for small and dense object from the perspective of UVAs. The architecture of YOLO-Drone is shown in Figure1. Darknet59 is referred to as the backbone of YOLO-Drone. In the proposed Darknet59, four-scale detectors are constructed and the depth of network is deepened. A novel feature extractor MSSP-FPN with one SPP and three ASPPs on four detector branches respectively is

proposed as the neck of YOLO-Drone in order to improve the feature aggregation performance. In YOLO-Drone, the GIoU is introduced in the location loss for the purpose of increasing the regression accuracy of the object frame. The structures of backbone, neck and head in the YOLO-Drone are shown in Figure2.

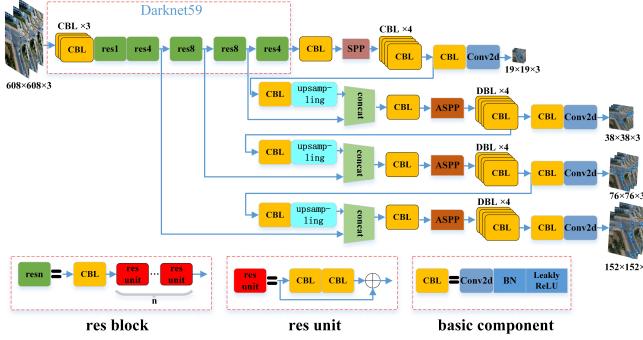


Fig. 1: The overall network structure of YOLO-Drone.

1) Backbone of YOLO-Drone: Darknet59: The first YOLO-Drone technique is to replace the Darknet53 backbone in YOLOv3 with the Darknet59 backbone. YOLO-Drone have four-scale detectors that extract 32x, 16x, 8x, and 4x down-sampling feature maps. A 4x down-sampling detector is introduced to improve the performance. From a high-altitude UAV perspective, there are a number of small objects to be detected. Theoretically, the feature maps of small-scale detectors with high-dimensional semantic information are generally used to detect large-scale objects. Compared with the three-scale detectors without shallow large-scale feature maps, the feature maps of large-scale detectors with low dimensional high-frequency information retain more detailed information of small-pixel objects, which is required for representations of small objects. As shown in Figure 2, the 76x76 feature maps derived from the third detector are double upsampling, and then spliced with the output of the third stage of the basic feature extraction network. The fourth super large-scale detector based on the 152x152 feature map is established mainly for the detection of small objects.

Type	Filters	Size	Output
Convolutional	32	3 × 3	608 × 608
Convolutional	32	3 × 3	608 × 608
Convolutional	32	3 × 3	608 × 608
Convolutional	64	3 × 3 / 2	304 × 304
Convolutional	32	1 × 1	
Convolutional	64	3 × 3	304 × 304
Convolutional	128	3 × 3 / 2	152 × 152
Convolutional	64	1 × 1	
Convolutional	128	3 × 3	152 × 152
Convolutional	256	3 × 3 / 2	76 × 76
Convolutional	128	1 × 1	
Convolutional	256	3 × 3	76 × 76
Convolutional	512	3 × 3 / 2	38 × 38
Convolutional	256	1 × 1	
Convolutional	512	3 × 3	38 × 38
Convolutional	1024	3 × 3 / 2	19 × 19
Convolutional	512	1 × 1	
Convolutional	1024	3 × 3	19 × 19
Convolutional			19 × 19

(a) Backbone: Darknet59 (b) Neck: MSPP-FPN (c) Head: YOLO

Fig. 2: The structure of YOLO-Drone's backbone, neck and head

The residual (Res) unit [45] is used as the calculation unit of Darknet59. As shown in Figure 1, Darknet59 has 6 scales of feature map information. The Res unit consists of a 1×1 convolution, a 3×3 convolution, and a shortcut connection. 1×1 convolution is used to decrease the computational complexity by reducing the feature maps of the previous output to 1/2 of the original one on the channel dimension. The shortcut connection fixes the problem of gradient disappearance. In YOLO-Drone, three cascading 3×3 convolutions are introduced in the pre-first down-sampling phase, 1 Res unit in the 2x down-sampling phase, 4 Res units in the 4x down-sampling phase, 8 Res units in the 8x down-sampling phase, 8 Res units in the 16x down-sampling phase and 4 Res units in the 32x down-sampling phase. Different from the three-scale YOLOv3, YOLO-Drone has deeper convolution layers and larger receptive fields, which boosts the effectiveness of extracting shallow features of small objects.

2) Neck of YOLO-Drone: MSPP-FPN: YOLO-Drone combines one SPP and three ASPPs on the four detector branches to improve the feature aggregation ability. Figure 3 depicts the neck of the YOLO-Drone, known as MSPP-FPN, which consists of FPN with SPP and ASPPs. MSPP-FPN provides multi-scale receptive fields for each detector branches and aggregates feature information from the global to the details. The SPP [16] module combines features of various scales, aggregates responses from multiple maximum pooling windows of various sizes, and improves the invariance of feature scales, making it resistant to object deformation. The ASPP module improves the feature region of the object of interest by integrating response maps of various receptive fields, combining location maps with different response levels.

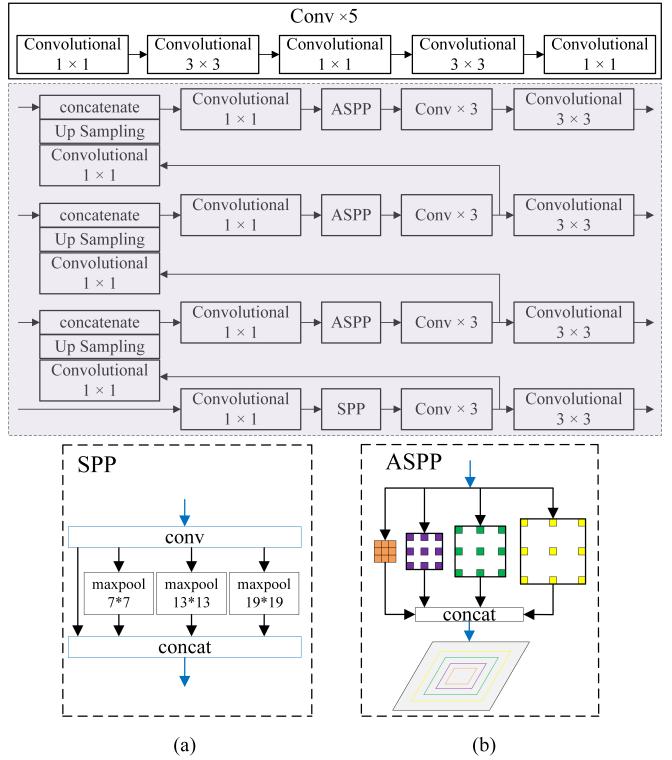


Fig. 3: The structure of MSPP-FPN.

$$\begin{aligned}
Loss_{YOLO-Drone} &= Loss_{loc} + Loss_{conf} + Loss_{class} \\
&= \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [2 - \frac{A^o}{(\hat{x}_2 - \hat{x}_1) * (\hat{y}_2 - \hat{y}_1) + (x_2 - x_1) * (y_2 - y_1) - A^o} \\
&\quad - \frac{(\hat{x}_2 - \hat{x}_1) * (\hat{y}_2 - \hat{y}_1) + (x_2 - x_1) * (y_2 - y_1) - A^o}{(\max(\hat{x}_2, x_2) - \min(\hat{x}_1, x_1)) * (\max(\hat{y}_2, y_2) - \min(\hat{y}_1, y_1))}] \\
&\quad + \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] \\
&\quad - \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] \\
&\quad + \sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c \in classes} [\hat{P}_i^j \log(P_i^j) + (1 - \hat{P}_i^j) \log(1 - P_i^j)]
\end{aligned} \tag{1}$$

The proposed SPP module in MSPP-FPN is a part of the 32x downsampling detector branch as shown in Figure 3 (a). The SPP is inserted into the first 1×1 convolution after the Darknet59 output to reduce dimensionality. The SPP consists of three max-pooling windows with different size of 5×5 , 9×9 and 13×13 . The stride in the max-pooling layer is 1, so the feature maps after pooling can be contacted directly with the original ones. The SPP module in MSPP-FPN can improve the ability to represent those significant features in high-dimensional feature maps.

The proposed ASPP modules in MSPP-FPN are part of the 16x, 8x and 4x down-sampling detector branches. The ASSP is located after the shallow feature maps obtained from the fusion of top-down features, as shown in Figure 3 (b). The ASPP is more concerned with details and specific information. To increase the receptive field, it employs multiple parallel atrous convolutions with varying dilated rates rather than some of the standard convolutions in YOLOv3. Figure 3 (b) depicts the relationship between the dilated convolution receptive fields at different scales in the proposed ASPP. In theory, those positioning maps produced by dilated convolutions with varying dilated rates are complementary [46]. Because the low-response areas are pulled up by the adjacent low-response areas, the proposed ASSP can improve detection accuracy by enhancing the object feature regions.

3) *Loss function*: The YOLO-Drone's loss function consists of position loss $Loss_{loc}$, confidence loss $Loss_{conf}$, and classification loss $Loss_{class}$. The GIoU value is used to calculate the distance between the predicted and true bounding boxes. Then, there is a strong positive relationship between minimizing position loss and increasing IoU value. Equation 1 depicts the total loss function of the YOLO-Drone.

where the prediction box parameters are denoted as $B^p = (\hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2)$. and the real frame parameters are denoted

as $B^g = (x_1, y_1, x_2, y_2)$, where (x_1, y_1) is the upper left corner coordinate, and (x_2, y_2) is the lower right one. A^o is the overlap area of B^p and B^g ; λ_{coord} and λ_{noobj} are weight coefficients, S^2 is the number of grid that input images are divided into, B is the number of predicted bounding boxes. I_{ij}^{noobj} and I_{ij}^{obj} represent whether there is an object in the bounding box (if there is an object then $I_{ij}^{noobj}=0$, otherwise $I_{ij}^{noobj}=1$). Cross entropy functions are used in the confidence and classification losses. When no object is present in the cell, the model does not penalize the classification error.

YOLOv3's position loss, on the other hand, employs mean square error (MSE) as the objective function, as illustrated in Equation 2. The sum of the mean square errors of the center coordinate (x, y) and width coordinate (\sqrt{w}, \sqrt{h}) between the prediction box and the corresponding real one is given by Equation 2. $Loss_{YOLOv3-loc}$ disregards the intersection over union (IoU) between the prediction and real boxes, resulting in no strict positive correlation between loss minimization and IoU maximization. Furthermore, despite the fact that (w, h) is square rooted, the regression window is still sensitive to scales.

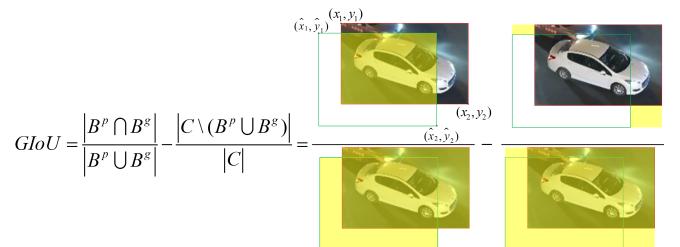


Fig. 4: The Schematic diagram of GIoU.

Given problems of $Loss_{YOLOv3-loc}$, YOLO-Drone employs GIoU as an evaluation index, which improves the network's ability to regress the object location. The GIoU

$$\begin{aligned}
Loss_{YOLOv3-loc} = & \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \left[\left(x_i - \hat{x}_i^j \right)^2 + \left(y_i - \hat{y}_i^j \right)^2 \right] \\
& + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \left[\left(\sqrt{w_i^j} - \sqrt{\hat{w}_i^j} \right)^2 + \left(\sqrt{h_i^j} - \sqrt{\hat{h}_i^j} \right)^2 \right]
\end{aligned} \tag{2}$$

calculation process is displayed in Figure 4. The minimum box that contains both the true and predicted boxes is found first. Then, calculate the ratio of the minimum box that does not include the two boxes in comparison to the entire minimum box. GIoU equals IoU minus the above ratio. In contrast to IoU, when the predicted and real boxes do not intersect, GIoU can still reflect the distance between the two boxes, whereas IoU loss disappears.

B. Design of Embedded Autonomous UAV Patrol System

In this paper, two autonomous cruise UAV systems are built: one is a multi-rotor UAV for detection at low and medium altitudes (<70m), and the other is a hybrid-winged UAV for detection at high altitude (>70m). Figure 5 illustrates the design of the autonomous cruise UAV systems. Once the drone system has a certain altitude, the visible/infrared integrative camera starts taking pictures of the target area. The system then transmits the recorded video using the YOLO-Drone algorithm to the intelligent terminal platform. The real-time object detection results are sent back to the ground workstation by the image transmission module and displayed there in real-time.

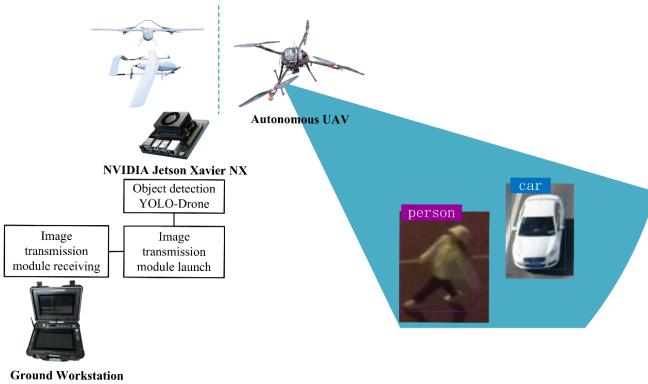


Fig. 5: Architecture of autonomous UAV patrol system proposed by this paper.

The hybrid-winged autonomous cruise drone system workflow diagram is shown in Figure 6, similar to the multi-rotor autonomous cruise drone system. Geographic Information System (GIS) is used to plan the route, and the hybrid-winged UAV will follow it as it travels. The visible/infrared integrative camera records the video streaming data, which is then sent to the edge computing platform via an image grabber. In our autonomous cruise drone systems, the proposed YOLO-Drone runs on the NVIDIA Jetson Xavier NX, which has 384-core

Volta GPU, 6-core ARM v8.2 64-bit CPU, and 8GB 128-bit LPDDR4x memory (<https://www.nvidia.cn/autonomous-machines/embedded-systems/jetson-xavier-nx/>).

The detection results are returned to the ground station via the image transmission module and displayed in real time.

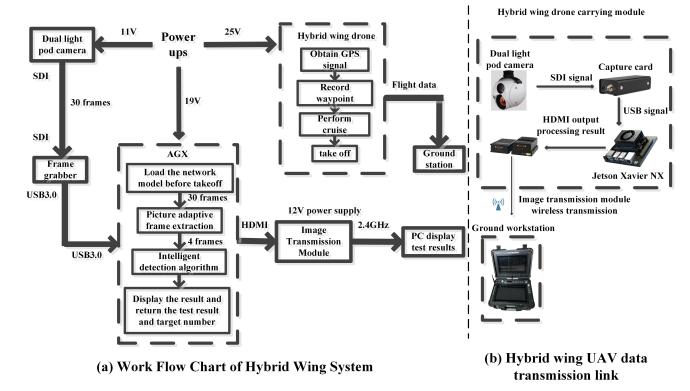


Fig. 6: Workflow chart and transmission link of the hybrid-winged autonomous cruise UAV.

The electric long-endurance hybrid-winged autonomous cruise UAV designed in this paper combines fixed-wing and quadcopter modes. It has the characteristics of long flight time, long range, fast speed and vertical take-off and landing (VTOL). The hybrid-winged UAV's ability to operate in challenging environments like mountains, hills, jungles, and densely populated areas is greatly improved by the VTOL mode. Compared with other vertical VTOL aircrafts, hybrid-winged UAVs have the advantages of good technical feasibility and low cost. In order to fully couple the two flight modes, the rotor shaft of the hybrid-winged UAV is mounted under the wing, and the quadcopter is symmetrically distributed on both sides of the fuselage. Figure 7(a) shows the hybrid-winged UAV developed in this paper.

The multi-rotor UAV designed in this paper consists of aracks, batteries, motors, electronic speed controllers (ESC), flight controllers, propellers, remote controls, a camera, etc. A signal is transmitted from the remote control to the remote control receiver, which then sends the signal to the flight controller. The PWM waveform (pulse width modulation waveform), which has been algorithmically processed, is sent from the flight controller to the ESC. The ESC completes the flight actions by adjusting the supply current to various motors to control the rotation speed of the propeller. The UAV is powered by batteries. Figure 7(b) shows the multi-rotor UAV developed in this paper.



Fig. 7: Two autonomous cruise UAV systems developed in this study. (a) the hybrid-winged autonomous cruise UAV (b) the multi-rotor UAV.

Fig. 8 shows the ground station software of the autonomous cruise UAV system. The main interface of the ground station software, the display area for fixed-wing modal member parameters, the interface of route planning, the generated mapping route, and the generated cruise route are depicted in Figure 8(a-e).

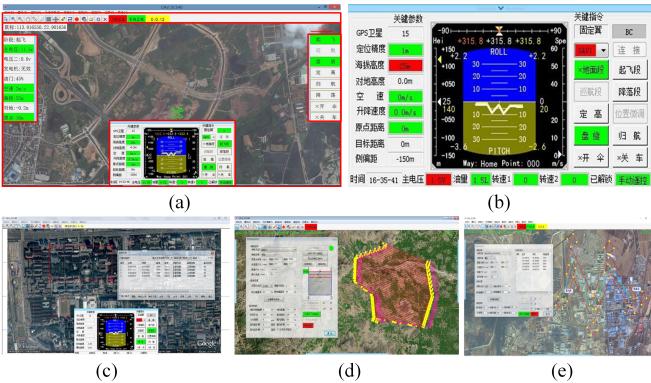


Fig. 8: The ground station software of the autonomous cruise UAV.

III. EXPERIMENT AND ANALYSIS

A. Dataset

In this paper, three datasets were used to evaluate the performance of the proposed object detection algorithm, including one private dataset of pedestrians created at night under silicon-based golden light LED (UAV-LED-G), and two benchmark datasets from a drone perspective, namely UAVDT [6] and VisDrone2021-DET [47]. In this study, three datasets were used to evaluate the performance of the proposed YOLO-Drone: two benchmark datasets from the UAV perspective, namely UAVDT [47] and VisDrone2021-DET [47], and one private UVA dataset collected at night under silicon-based golden light LED(UAV-LED-G).

UAVDT is a UAV-based vehicle detection and tracking dataset that was presented at ICCV2018. It consists of 80,000 images with a resolution of 1080×540 pixels, and over 2,700 annotated vehicles. These UAV images are captured under complex scenarios, including various weather conditions, flying altitude, camera views, vehicle categories and occlusions.

In our experiments, 40409 UAVDT images were divided into a training set of 23829 images and a test set of 16580.

VisDrone2021-DET is a large-scale aerial image dataset for object detection, single-object tracking, and multiple-object tracking. It contains 10,209 static images with resolution of 2,000 × 1,500 pixels. All images are captured by multiple UAV platforms in various lighting and weather conditions. VisDrone2021-DET is labeled with ten different categories: pedestrian, person, car, van, bus, truck, motor, bicycle, awning-tricycle, and tricycle.

The UAV-LED-G dataset was acquired by the multi-rotor autonomous cruise UAV platform developed in this study. The drone flies at 6 different altitudes, ranging from 15 to 65 meters in increments of 10 meters. The collection location is the sports field of Nanchang University's Qianhu Campus, and the collection time is at night under silicon-based golden LEDs [40]–[44]. The image is preprocessed by framing, dedistortion, etc. For annotation, 10 domain experts were invited to label our preprocessed dataset using LabelImg tool for 3 months. The dataset goes through multiple rounds of double-checking. The UAV-LED-G consists of 8416 images with a resolution of 1,920×1,680 that have been divided into training set(6312 images) and test set(2104 images) in a 3:1 ratio. The small objects to be detected in UAV-LED-G are pedestrians, and their pixel range [4*15,102*126] accounts for only 0.00289%–0.61979% of the total images, which makes the detection of these pedestrians challenging. Figure 9 shows some examples of frames from UAV-LED-G.

The silicon-based golden LED is a kind of zero-blue light, no phosphor healthy light source composed of silicon substrate gallium nitride-based yellow LED and silicon plate aluminum gallium indium red LED. The silicon-based golden light LED is more energy-efficient [48] and has the following advantages over fluorescent LEDs. (1) It does not contain blue-violet light with a wavelength shorter than 500 nm, avoiding light pollution caused by blue-rich hazards, protecting the driver's eyes, and thus improving road safety.

(2) Because it only contains long-wavelength red and yellow light components, it has better penetration ability in adverse weather conditions like rain, fog, dust, haze, and so on. This enables the objects to maintain high clarity and brightness even in difficult weather conditions.

(3) It adopts pure LED synthesis to prevent phosphor light decay, guarantee the stability of light quality, and increase the lifespan of the light source [49]. The LM80 claims that a luminous flux rate of more than 97% is maintained by its 6,000-hour light source.

(4) It is energy efficient. In comparison to ordinary LEDs, which have a luminous efficiency of 60-90 lm/W, the silicon-based golden light LEDs have a luminous efficiency of 150-180 lm/W.

B. K-means cluster anchor

YOLO-Drone employs the k-means method for clustering in order to choose the optimized prior boxes necessary for the network to learn features while preventing errors brought on by the scale size of anchors. In this study, the Euclidean

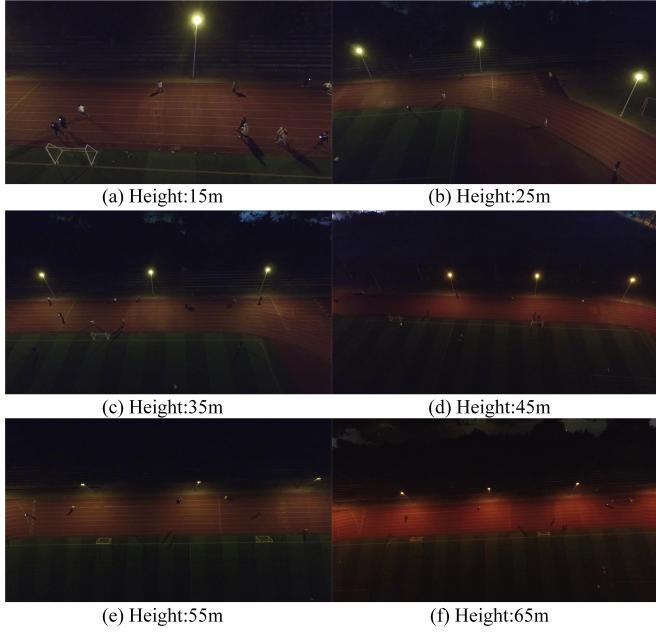


Fig. 9: Pedestrian dataset example under silicon-based golden light LED light source at night.

distance in the k-means was replaced with the Intersection over Union (IOU) score of the candidate bounding box and the actual bounding box. The IOU score can reduce the potential large error caused by a large anchor, allowing for an accurate assessment of the clustering effect. The distance function calculation formula is shown in Equation 6:

$$d = 1 - IoU \quad (3)$$

12 classes of the four scales were ultimately chosen for clustering, taking into account the four-scale detector, recall balance and model complexity. The average distribution results of these 12 classes in UAVDT, VisDrone2021-DET, and UAV-LED-G are displayed in TABLE I.

TABLE I: K-means clustering results on three types of datasets

Dataset	Scale	Anchor1	Anchor2	Anchor3
UAVDT	P5	5*11	7*18	11*23
	P4	10*12	17*17	10*35
	P3	23*23	16*35	35*38
	P2	21*60	48*87	131*157
VisDrone2021-DET	P5	5*10	12*6	12*14
	P4	10*30	16*32	23*22
	P3	33*23	30*61	62*45
	P2	59*119	116*90	131*157
UAV-LED-G	P5	4*8	5*11	6*7
	P4	4*14	5*17	6*14
	P3	7*9	7*19	8*12
	P2	8*16	8*25	10*28

C. Experimental results and analysis

1) Comparisons with the SOTA on UAVDT dataset: In the experiments, the proposed YOLO-Drone was compared with 7 state-of-the-art algorithms(R-FCN,SSD,RON,Faster

RCNN,Faster RCNN+FPN,ClusDet,DMNet) and 6 YOLO-based algorithms (YOLOv3, YOLOv4, YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x) for object detection on the UAVDT. Three evaluation indicators, mAP@0.5 , mAP@0.75 and mAP@.5:.95, were calculated, as shown in Table II. Among them, the mAP@0.5 and mAP@0.75 are noted as AP@0.5 and AP@0.75 respectively, and the mAP@.5:.95 is the average value at IoU from 0.5 to 0.95 with step size 0.05. Compared with the aforementioned 7 object detection algorithms, the three evaluation indicators obtained by YOLO-Drone are all the highest. Most of the performances of YOLO-Drone are superior to those of the YOLO-based algorithms, and some are comparable to those of YOLOv4. The mAP of YOLO-Drone is as high as 34.0%, which is greater than all 13 comparison algorithms mentioned above, and 10.6%, 7.5%, 9.4%, 10.1%, 1.4% and 2.9% higher than Faster RCNN+FPN, ClusDet, DMNet, YOLOv3, YOLOv4, and YOLOv5x, respectively.

2) Ablation Experiments on UAVDT dataset: 9 ablation experiments were completed on UAVDT to investigate the effects of the main modules in YOLO-Drone, and the results are shown in Table 3. The following are the experimental conditions.

Case (a) uses a four-scale detector (D4) based on YOLOv3.

Case (b) introduces the proposed feature extraction network Darknet59 to YOLOv3.

Case (c) introduces a SPP module (SPPx1) in the 32x downsampling detection branch based on Case (a).

Case (d) introduces a SPP module (SPPx1) in the 32x downsampling detection branch based on Case (b).

Case (e) introduces 3 ASPP modules (ASPPx3) in the 16x, 8x and 4x downsampling detection branches respectively based on Case (d). It is also equivalent to introduce the MSPP-FPN based on Case (b).

Case (f) replace the location loss from MSE to GIoU on the basis of Case (a).

Case (g) uses a SPP module (SPPx1) in the 32x downsampling detection branch based on Case (f).

Case (h) the entire YOLO-Drone. It is also the same as using GIoU on a Case (e) basis.

Table III demonstrates the additive gain effect of the main modules D4, Darknet59, MSPP-FPN, and GIoU in YOLO-Drone. The fact that Case (c) on mAP is 3.76% greater than YOLOv3 demonstrates the benefits of Darknet59's four-scale detector module. Case (f) is 1.50% higher than Case (c) on mAP. Case (i) again improved mAP by 4.87% compared to Case (f). In total, YOLO-Drone increased mAP by 10.31% over YOLOv3, with improvements in the car, truck, and bus of 14.12%, 2.16%, and 14.10%, respectively.

Inference speed: YOLO-Drone achieves 53 FPS on the NVIDIA RTX 2080ti GPU, meeting real-time requirements (>30 fps) and outperforming DMNet [26].

3) Results of multiple scenarios on the UAVDT dataset:

The performance of YOLO-Drone was evaluated on 9 complex scenarios in UAVDT, including daytime, night, foggy, forward-looking, side, bird's-eye, high-altitude, medium-altitude, low-altitude views. As can be seen from the Table IV, the performance of YOLO-Drone is significantly superior to YOLOv3

TABLE II: The comparison of the performance on UAVDT dataset

Methods	backbone	#Image	mAP@.5:.95	AP50	AP75
R-FCN [50]	ResNet50	15096	7.0	17.5	3.9
SSD [10]	N/A	15096	9.3	21.4	6.7
RON [51]	N/A	15096	5.0	15.9	1.7
FRCNN [18]	VGG	15096	5.8	17.4	2.5
FRCNN [18]+FPN [28]	ResNet50	15096	11.0	23.4	8.4
ClusDet [52]	ResNet50	25427	13.7	26.5	12.5
DMNet [26]	ResNet50	32764	14.7	24.6	16.3
YOLOv3	Darknet53	16580	9.4	23.9	4.3
YOLOv4	CSPDarknet53	16580	18.9	32.6	20.4
YOLOv5s	N/A	16580	17.9	31.2	18.4
YOLOv5m	N/A	16580	16.8	29.5	17.3
YOLOv5l	N/A	16580	17.7	30.1	18.6
YOLOv5x	N/A	16580	17.8	31.1	18.6
YOLO-Drone	Darknet59	16580	18.1	34.0	17.6

TABLE III: Ablation experiment on UAVDT dataset

case	D4	Darknet59	SPPx1	ASPPx3	GioU	all	car	truck	bus	FPS
YOLOv3						23.91	50.31	2.96	18.46	130
(a)	✓					24.29	54.73	3.49	14.65	114
(b)	✓	✓				27.67	56.31	2.47	24.22	84
(c)	✓		✓			26.49	62.18	2.45	14.84	-
(d)	✓	✓	✓			28.53	57.28	3.99	24.32	62
(e)	✓	✓	✓	✓		29.17	59.76	2.30	25.44	80
(f)	✓				✓	29.27	67.72	4.43	15.66	-
(g)	✓		✓		✓	30.21	68.52	4.69	17.41	-
(h)	✓	✓	✓	✓	✓	34.04	64.43	5.12	32.56	53

TABLE IV: Divide the scene detection results on the UAVDT dataset

method	object	daylight	night	fog	low-alt	medium	high	front	side	bird
YOLOv3	car	58.64	60.17	38.02	59.70	56.85	31.72	49.50	60.77	40.34
	truck	5.63	4.19	9.55	26.16	2.30	2.15	1.92	14.46	4.72
	bus	7.50	47.35	5.25	67.86	22.93	4.41	21.22	13.08	23.28
	all	23.92	37.24	17.61	51.24	27.36	12.76	24.21	29.44	22.78
YOLO-Drone	car	71.00	73.35	52.27	72.26	71.01	50.80	66.57	73.19	53.76
	truck	5.48	4.63	12.22	24.28	2.97	10.35	2.70	8.24	11.94
	bus	7.72	79.30	2.57	52.76	38.73	3.44	38.26	17.92	11.05
	all	28.07	52.43	22.35	49.77	37.57	20.32	35.84	33.11	25.58

in most scenarios. YOLO-Drone's detection results on car outperformed YOLOv3 in all 9 scenarios. The mAP obtained by YOLO-Drone in the first 8 of the aforementioned scenarios increases by 4.15%, 15.19%, 5.9%, 5.8%, 3.67%, 11.63%, 6.4%, and 10.21%, respectively, in comparison to YOLOv3. The mAP of the YOLO-Drone decreased by 1.47% at low altitude views. It is worth noting that there are fewer images at low altitude viewing angles (about 30%), resulting in class imbalances. Meanwhile, in low-altitude images, the size of the truck and bus exceeds 100×50, and these large sizes may mismatch the anchor frame calculated by YOLO-Drone. Because of the aforementioned two factors, the performance of YOLO-Drone slightly degraded at low altitude view. In fact, because medium altitude is the most common flying altitude for UAVs, YOLO-Drone, which focuses on dense and small objects in medium altitudes, is an appropriate and optimized object detection algorithm for UAV.

Figure 10 depicts YOLO-Drone and YOLOv3 object detection in a UAVDT composite scenario with three types of intersection factors: weather, viewing angle, and altitude. It demonstrates how much better YOLO-Drone's object de-

tection ability is than YOLOv3 under the three challenging conditions of high altitude, night, and fog. YOLO-Drone, in comparison to YOLOv3, has the following key advantages.

1. YOLO-Drone is better able to detect smaller objects at greater distances in daylight, medium-altitude, and front-view scenes (Fig10 a1 and a2).

2. YOLO-Drone detects more objects in the foggy, high-altitude, bird's-eye view (Fig10 b1 and b2).

3. YOLO-Drone can detect densely packed tiny objects more accurately in high-altitude views.

4. It is clear that the irradiation of ordinary light sources at night causes significant light pollution, resulting in the false detection of YOLOv3. In contrast, YOLO-Drone is significantly less affected by light pollution (Fig10 d1 and d2).

In summary, YOLO-Drone has superior performance and robustness in complex scenes with a variety of lights, weathers, altitudes, and viewing angles.

4) Comparison of results on the VisDrone2021-DET dataset: The mAP@.5 values of the YOLO-Drone, YOLOv3, and YOLOv5 series are compared on the VisDrone2021-DET, as shown in Table V. YOLO-Drone has the highest mAP

TABLE V: Algorithms are trained on VisDrone2021-DET-train and tested on VisDrone2021-DET-test-dev

Method	all	pedestrian	people	bicycle	car	van	truck	tricycle	awn-tricy	bus	motor
YOLOv3	25.71	16.43	8.38	6.99	64.37	33.82	29.49	12.46	13.22	51.89	20.06
YOLO-Drone	35.45	30.85	18.17	11.78	76.15	42.19	41.27	21.41	19.27	59.87	33.51
YOLOv5s	28.90	26.30	17.00	8.84	70.50	32.60	29.40	13.60	14.30	52.40	24.20
YOLOv5m	29.70	28.20	18.20	7.45	71.30	33.70	30.50	13.90	14.80	52.80	25.80
YOLOv5l	31.20	29.60	18.80	9.36	72.90	35.90	31.90	15.40	15.70	54.80	27.30
YOLOv5x	33.10	30.30	20.20	11.20	73.90	37.60	36.80	18.70	16.40	56.60	29.70

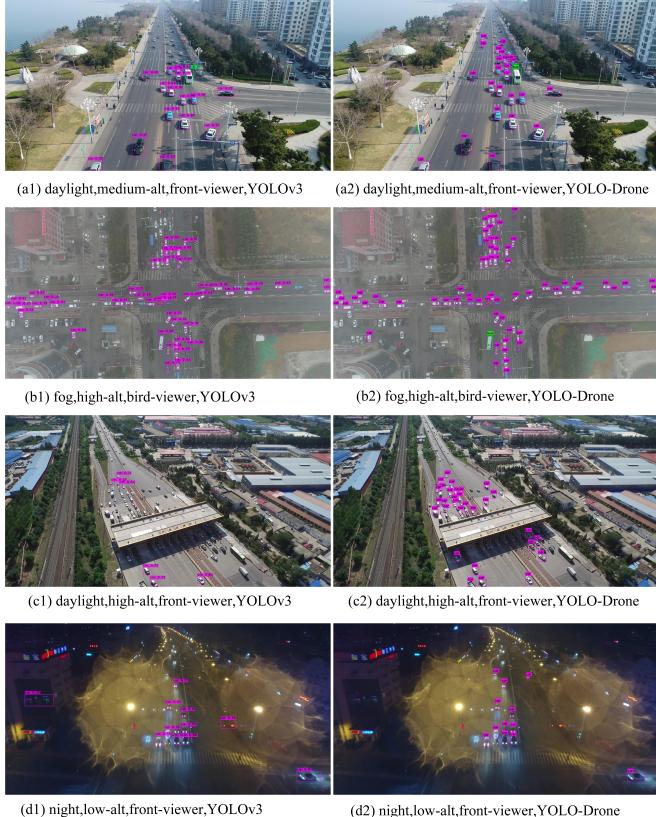


Fig. 10: Comparison of actual measurement results of YOLO-Drone (right) and YOLOv3 (left) on the UAVDT dataset.

on the all 10 categories of objects, which is 9.74%, 6.55%, 5.75%, 4.25%, and 2.35% higher than the mAP of YOLOv3, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, respectively. YOLO-Drone's AP for 9 object categories outperform the YOLOv5 series, and its detection of the tenth category 'person' is comparable to the best two algorithms, YOLOv5m and YOLOv5l.

Furthermore, as shown in Table IV, the proposed YOLO-Drone is compared to the 8 state-of-the-art object detection algorithms on VisDrone2021-DET. According to the experimental results, the mAP of YOLO-Drone is 35.45%, which is 3.05%, 5.15%, 26.7%, 3.54%, 30.67%, 14.08%, 6.69%, and 1.33% higher than the aforementioned eight algorithms, Clus-Det, DREN, Faster RCNN, Cascade RCNN, SSD, RetinaNet, RefineDet, and CornerNet, respectively. As can be seen, in the object detection task from the perspective of UAVs, YOLO-Drone outperforms not only the YOLO series but also other significant one-stage and two-stage algorithms.

5) *The application of YOLO-Drone and silicon-based golden light LED at night:* Due to the intensity of night illumination, strong light pollution, light reflection and diffuse reflection, and the lack of obvious characteristics of small objects at high altitude, detecting night objects from the perspective of a UAV is a difficult task. Night scene images from UAVDT and VisDrone2021-DET were screened for this study. The three nightly datasets are UAV-LED-G (pedestrian, 9580 images), VisDrone2021-N (pedestrian, 600 images), and UAVDT-N (car, 5023 images). According to Table VII, the average number of boxes for each image in the three datasets is 9.37, 8.30, and 8.29, the box size ranges are [4×15, 102×126], [3×4, 77×250], and [7×8, 401×225], and the average box size is 35×36, 35×36, and 56×56, respectively.

The proposed YOLO-Drone is applied to nighttime object detection. As shown in Table VII, the detection APs of YOLO-Drone in UAV-LED-G, UAVDT-N and VisDrone2021-N are 87.71%, 47.82% and 73.35%, respectively, indicating that its performance is superior. YOLO-Drone's APs outperform YOLOv3 by 21.67%, 13.18% and 0.4% on VisDrone2021-N, UAVDT-N, and the new light source dataset UAV-LED-G, respectively. Figure 11 shows the results of YOLO-Drone's nighttime object detection under various lighting conditions. In conclusion, it can be seen from Table VII and Figure 11 that YOLO-Drone has obvious advantages in object detection at night scenes, as well as high accuracy and robustness, particularly for small object detection.

In order to investigate the differences between the silicon-based golden LED and the ordinary light sources in the night detection task, object detection methods under various light sources were compared. The detection APs of YOLO-Drone, YOLOv3 and YOLOv4 on UAV-LED-G are greater than 87%, according to Table VII. Thus, the silicon-based golden light LED has a significant advantage in night detection tasks, allowing most algorithms to achieve high accuracy. The APs of YOLO-Drone and YOLOv3 on UAV-LED-G are 87.71% and 87.41%, respectively, outperforming those of VisDrone2021-N (39.89% and 61.26% higher, respectively) and UAVDT-N (14.36% and 27.24% higher, respectively) in the common light source datasets. This holds true when there are lots of small objects and bounding boxes. The left and middle columns of Figure 11 show the results of YOLO-Drone at six different flight altitudes under the silicon-based golden LED. The right column of Figure 11 depicts the YOLO-Drone of VisDrone2021-N under ordinary light sources.

The following conclusions can be drawn from Figure 11 and Table VII .1. In the night object detection task, the silicon-based golden LED is effective for most mainstream object

TABLE VI: The comparison of the performance on VisDrone2021-DET dataset

Method	#Image	mAP
YOLO-Drone	10209	35.45
ClusDet [52]	10209	32.4
DREN [53]	10209	30.3
Faster RCNN [18]	10209	8.75
Cascade RCNN [54]	10209	31.91
SSD [10]	10209	4.78
RetinaNet [55]	10209	21.37
RefineNet [56]	10209	28.76
CornerNet [57]	10209	34.12

TABLE VII: The performance of the algorithm on the night dataset

method	YOLO-Drone	YOLOv3	YOLOv4	YOLO-Drone	YOLOv3	YOLO-Drone	YOLOv3
category	pedestrian	pedestrian	pedestrian	pedestrian	pedestrian	car	car
Data set	UAV-LED-Y	UAV-LED-Y	UAV-LED-Y	Visdrone2021-N	Visdrone2021-N	UAVDT-N	UAVDT-N
Box min	4*15	4*15	4*15	3*4	3*4	7*8	7*8
Box max	102*126	102*126	102*126	77*250	77*250	401*225	401*225
Box mid	35*26	35*26	35*26	18*30	18*30	25*38	25*38
Box avg	35*36	35*36	35*36	35*36	35*36	56*56	56*56
box/picture	9.37	9.37	9.37	8.30	8.30	8.29	8.29
AP	87.71	87.41	87.37	47.82	26.15	73.35	60.17

detection neural networks (i.e. YOLO series), achieving significantly higher detection accuracy than ordinary light sources. 2. The silicon-based golden LED has a greater ability to penetrate in harsh environments, preserving or even enhancing the fine features of objects at night. 3. The silicon-based golden LED reduces the rate of missed detection of nighttime objects by avoiding light pollution caused by blue-rich hazards. 4. Under the silicon-based golden LED, the detection of multi-scale objects has improved, especially those at the medium and small scales. This is consistent with the underlying theory of multi-scale object detection algorithms based on pyramidal structures. Therefore, the silicon-based golden LED has promising applications in object detection tasks.

IV. CONCLUSION

In view of the difficulty in detecting small objects in the actual complex high-altitude scenes, most of the existing algorithms are difficult to detect effectively. This paper proposes a small pixel object detection algorithm YOLO-Drone based on the YOLOv3 architecture. For different datasets, the algorithm uses K-means clustering on the training set to find the best prior; in order to improve the network's ability to detect dense small objects, a small object detector is added to the original network. The four-fold down-sampling feature is used to fuse high-level feature information, and the background information and semantic information of small objects are fully integrated, which improves the model's ability to detect small object features; in order to better adapt to small object detectors, Darknet59 is proposed. It is obtained by increasing the number of convolutions in the shallow stage on the basis of Darknet53; the SPP module is embedded in the first detector, and multiple maximum pooling windows of different sizes in parallel are used to improve the robustness of the network to the detection of deformed objects. In the second, third, and fourth detectors, the ASPP module is embedded, and the adjacent high-response areas increase the low-response area by parallel convolution of multiple holes with different hole rates; using GIoU-based position loss, there is a strong correlation between minimizing position loss and maximizing IoU, so as to achieve a more accurate regression of the object frame. The results on the UAVDT and VisDrone2021-DET datasets show that its performance exceeds most existing algorithms. The results on the night-time detection dataset show that under the same light source, the YOLO-Drone algorithm has high accuracy. Under different light sources, the silicon-based golden light LED light source has a significant advantage in night-time detection tasks.

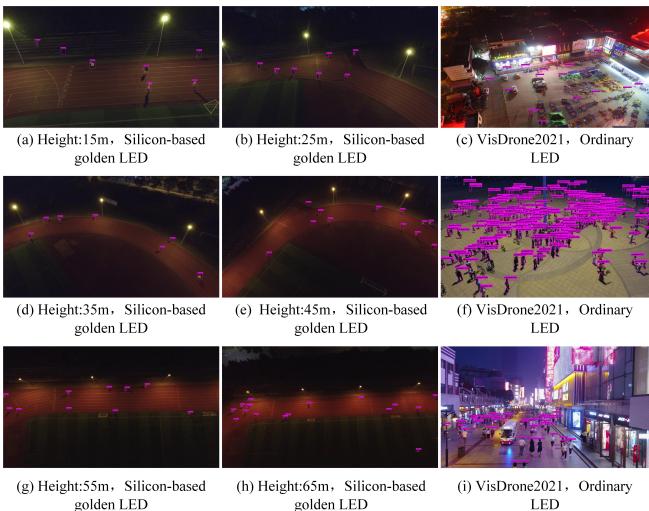


Fig. 11: Comparison of measured results under different light sources.

ACKNOWLEDGMENTS

This should be a simple paragraph before the References to thank those individuals and institutions who have supported your work on this article.

REFERENCES

- [1] G. Y. Zou Z, Shi Z, “Object detection in 20 years: A survey,” *arXiv preprint arXiv:1905.05055*, 2019.
- [2] H. D. Wu X, Li W, “Deep learning for unmanned aerial vehicle-based object detection and tracking: a survey,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 1, pp. 91–124, 2021.
- [3] B. P. Richards B A, Lillicrap T P, “A deep learning framework for neuroscience,” *Nature neuroscience*, vol. 22, no. 11, pp. 1761–1770, 2019.
- [4] Z. A. Lliu M, Wang X, “Uav-yolo: Small object detection on unmanned aerial vehicle perspective,” *Sensors*, vol. 20, no. 8, p. 2238, 2020.
- [5] S. A. Mittal P, Singh R, “Deep learning-based object detection in low-altitude uav datasets: A survey,” vol. 104. IEEE, 2020, p. 104046.
- [6] Y. H. Du D, Qi Y, “The unmanned aerial vehicle benchmark: Object detection and tracking,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 370–386.
- [7] Z. F. Tong K, Wu Y, “Recent advances in small object detection based on deep learning: A review,” *Image and Vision Computing*, vol. 97, p. 103910, 2020.
- [8] G. R. Lin T Y, Dollár P, “Feature pyramid networks for object detection,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, 2017.
- [9] L. L. Deng C, Wang M, “Extended feature pyramid network for small object detection,” 2021.
- [10] E. D. Liu W, Anguelov D, “Ssd: Single shot multibox detector,” in *Journal of Physics: Conference Series*. IOP Publishing, 2016, pp. 21–37.
- [11] G. R. Redmon J, Divvala S, “You only look once: Unified, real-time object detection,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.
- [12] F. A. Redmon J, *YOLO9000: better, faster, stronger*. University of Derby (United Kingdom), 2017.
- [13] ——, “Yolov3: An incremental improvement,” *arXiv preprint arXiv:1804.02767*, 2018.
- [14] L. H. Y. M. Bochkovskiy A, Wang C Y, “Yolov4: Optimal speed and accuracy of object detection,” *arXiv preprint arXiv:2004.10934*, 2020.
- [15] D. T. Girshick R, Donahue J, “Rich feature hierarchies for accurate object detection and semantic segmentation[,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.
- [16] R. S. He K, Zhang X, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [17] G. R, “Fast r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.
- [18] G. R. Ren S, He K, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *Advances in neural information processing systems*, vol. 28, 2015.
- [19] L. D. B. Singh, “An analysis of scale invariance in object detection snip,” in *IEEE Conference on Computer Vision and Pattern Recognition*. Springer, 2018, pp. 3578–3587.
- [20] L. D. B. Singh, M. Najibi, “Sniper: efficient multi-scale training, neural information processing systems,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9333–9343.
- [21] M. J. Kisant M, Wojna Z, “Augmentation for small object detection,” in *arXiv preprint arXiv:1902.07296*, 2019.
- [22] Z. J. Guan L, Wu Y, “Scan: Semantic context aware network for accurate small object detection,” in *international Journal of Computational Intelligence Systems*. Springer Berlin/Heidelberg, Germany, 2018, pp. 951–961.
- [23] D. M. Bai Y, Zhang Y, “Sod-mtgan: Small object detection via multi-task generative adversarial network,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 206–221.
- [24] A. O. Motlagh N H, Taleb T, “Low-altitude unmanned aerial vehicles-based internet of things services: Comprehensive survey and future perspectives,” in *IEEE Internet of Things Journal*, vol. 3, no. 6, 2016, pp. 899–922.
- [25] L. Y. Huang H, Deng J, “A fully convolutional network for weed mapping of unmanned aerial vehicle (uav) imagery,” *PloS one*, vol. 13, no. 4, p. e0196302, 2018.
- [26] Z. S. Li C, Yang T, “Density map guided object detection in aerial images,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 190–191, 2020.
- [27] K. A. Bouguettaya A, Zarzour H, “Vehicle detection from uav imagery with deep learning: A review,” in *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [28] N. Y. Chen Y, Li J, “mall object detection networks based on classification-oriented super-resolution gan for uav aerial imagery,” in *2019 Chinese Control And Decision Conference (CCDC)*. IEEE, 2019, pp. 4610–4615.
- [29] S. T. Sommer L, Schumann A, “Multi feature deconvolutional faster r-cnn for precise vehicle detection in aerial imagery,” *2018 IEEE winter conference on applications of computer vision (WACV)*. IEEE, pp. 635–642, 2018.
- [30] S. A. Sommer L, Schmidt N, “Search area reduction fast-rcnn for fast vehicle detection in large aerial imagery,” *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 3054–3058, 2018.
- [31] S. S. Jiang D, Sun B, “A feature fusion and spatial attention-based single shot detector for small object detection,” *Electronics*, 2020, 9(9): 1536, 2020.
- [32] S. A. Shivappriya S N, Priyadarsini M, “Cascade object detection and remote sensing object detection method based on trainable activation function,” *Remote Sensing*, 2021, 13(2): 200, 2021.
- [33] M. P. Mandal M, Shah M, “A small-sized vehicle detection network for aerial visual data,” *IEEE Geoscience and Remote Sensing Letters*, 2019, 17(3): 494–498, 2019.
- [34] Z. S. Li C, Yang T, “Density map guided object detection in aerial images,” 2020.
- [35] L. Y. J. Kuang H, Zhang X, “Nighttime vehicle detection based on bio-inspired image enhancement and weighted score-level feature fusion,” *IEEE Transactions on Intelligent Transportation Systems*, 2016, 18(4): 927–936, 2016.
- [36] C. G. Wang W, Peng Y, “Low-illumination image enhancement for night-time uav pedestrian detection,” *IEEE Transactions on Industrial Informatics*, 2020, 17(8): 5208–5217, 2020.
- [37] X. R. Qu Y, Ou Y, “A small-sized vehicle detection network for aerial visual data,” 2019.
- [38] K. S. H. Kim I S, Jeong Y, “Low illumination enhancement for object detection in self-driving,” 2019.
- [39] M. J. Liu H, Gao W, “A novel method to compensate variety of illumination in face detection,” 2002.
- [40] X. L. Jiang F, Zhang J, “Efficient ingan-based yellow-light-emitting diodes,” *Photonics Research*, 2019, 7(2): 144–148, 2019.
- [41] W. X. L. Zhang T R, Wang F, “Aging mechanism of gan-based yellow leds with v-pits,” *Chinese Physics B*, 2019, 28(6):67305–067305, 2019.
- [42] M. C. L. Zeng J M, Wang X L, “Effect of barrier temperature on photoelectric properties of gan-based yellow leds,” *Chinese Physics Letters*, 2020, 37(3):038502, 2020.
- [43] W. X. Yang X, Zhang J, “Enhance the efficiency of green-yellow led by optimizing the growth condition of preparation layer,” *Superlattices and Microstructures*, 2020:106459, 2020.
- [44] Z. J. Tao X, Liu J, “Performance enhancement of yellow ingan-based multiple-quantum-well light-emitting diodes grown on si substrates by optimizing the ingan/gan superlattice interlayer,” *Optical Materials Express*, 2018, 8(5): 1221–1230, 2018.
- [45] V. V. Szegedy C, Ioffe S, “Inception-v4, inception-resnet and the impact of residual connections on learning,” 2017.
- [46] S. H. Wei Y, Xiao H, “Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation,” 2018.
- [47] D. D. Zhu P, Wen L, “Vision meets drones: Past, present and future,” *Vision meets drones: Past, present and future*.
- [48] Z. W. Zhang Y X, Zhi-Cheng L U, “Study of the fabrication and spectral analysis of silicon-based nanocolumn gan-led,” *Spectroscopy and Spectral Analysis*, 2019, 2019.
- [49] H. M. Davis J L, “A decade of gains in led packaging technologies as captured in lm-80 data,” 2020.
- [50] H. K. Dai J, Li Y, “R-fcn: Object detection via region-based fully convolutional networks,” *Advances in Neural Information Processing Systems*, 2016, 2016.
- [51] Y. A. Kong T, Sun F, “Ron: Reverse connection with objectness prior networks for object detection,” 2017.
- [52] C. P. Yang F, Fan H, “Clustered object detection in aerial images,” 2019.
- [53] C. X. Zhang J, Huang J, “How to fully exploit the abilities of aerial image detectors,” 2019.

- [54] V. N. Cai Z, “Delving into high quality object detection,” 2018.
- [55] G. R. Lin T Y, Goyal P, “Focal loss for dense object detection,” *Proceedings of the IEEE international conference on computer vision*. 2017: 2980-2988, 2017.
- [56] B. X. Zhang S, Wen L, “Single-shot refinement neural network for object detection,” *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018: 4203-4212, 2018.
- [57] D. J. Law H, “Detecting objects as paired keypoints,” *Proceedings of the European conference on computer vision (ECCV)*. 2018: 734-750, 2018.