

GAANet: Ghost Auto Anchor Network for Detecting Varying Size Drones in Dark

Misha Urooj Khan*, Maham Misbah*, Zeeshan Kaleem*, Yansha Deng†, Abbas Jamalipour†

*Electrical and Computer Engineering, COMSATS University Islamabad, Wah Campus, Pakistan

† Department of Engineering, King's College London, United Kingdom

†School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW 2006, Australia

Email: *{mishauroojkhan, mahammisbah33, zeeshankaleem}@gmail.com,
yansha.deng@kcl.ac.uk, a.jamalipour@ieee.org

Abstract—The usage of drones has tremendously increased in different sectors spanning from military to industrial applications. Despite all the benefits they offer, their misuse can lead to mishaps, and tackling them becomes more challenging particularly at night due to their small size and low visibility conditions. To overcome those limitations and improve the detection accuracy at night, we propose an object detector called Ghost Auto Anchor Network (GAANet) for infrared (IR) images. The detector uses a YOLOv5 core to address challenges in object detection for IR images, such as poor accuracy and a high false alarm rate caused by extended altitudes, poor lighting, and low image resolution. To improve performance, we implemented auto anchor calculation, modified the conventional convolution block to ghost-convolution, adjusted the input channel size, and used the AdamW optimizer. To enhance the precision of multiscale tiny object recognition, we also introduced an additional extra-small object feature extractor and detector. Experimental results in a custom IR dataset with multiple classes (birds, drones, planes, and helicopters) demonstrate that GAANet shows improvement compared to state-of-the-art detectors. In comparison to GhostNet-YOLOv5, GAANet has higher overall mean average precision (mAP@50), recall, and precision around 2.5%, 2.3%, and 1.4%, respectively. The dataset and code for this paper are available as open source at <https://github.com/ZeeshanKaleem/GhostAutoAnchorNet>.

Index Terms—Drones, YOLOv5, Multi-class Classification, Night-Vision, Target Detection.

I. INTRODUCTION

Unmanned aerial vehicles (UAVs) are widely adopted in remote sensing and advanced surveillance applications due to the growth in drone-based applications. According to industry insights, the global drone market is expected to reach \$48 billion by 2026 [1]. Because of their flexibility and mobility, drones are widely considered in many daily and industrial applications, and their capabilities are further enhanced when equipped with advanced artificial intelligence (AI) techniques. This advancement and its increasingly widespread use have raised serious concerns about the security of public places, as we have seen several instances where drones have caused damage to infrastructure [2] [3]. Therefore, effective detection systems are necessary for protection against malicious activities [4]. Advanced object detection and tracking systems are preferred over traditional object identification methods due to their less accurate target detection and high false alarm rate. Object identification is increasingly adopted for drone detection, but many schemes fail because of the drone's small

size, high flight altitude, and fast speed. These issues are addressed by UAV detection systems integrated with deep learning algorithms. Object detection using computer vision and deep learning, such as regions with convolutional neural networks (RCNN), Faster RCNN [5], and Mask-RCNN, utilize two-stage detection methods with improved detection results. But they are unsuited for efficient and accurate recognition of tiny fast-moving objects like UAVs vs. birds, planes, or helicopters. You Only Look Once (YOLO) [6] and the single shot multibox detector (SSD) [7] are two more methods that perform identification and categorization in a single step with additional end-to-end optimization. YOLO, in particular, offers the finest all-around detection performance in speed, accuracy, and precision. Radar, optical detection, and acoustic sensors are the most often used technologies for detecting UAVs. Hoffman *et al.* investigated the radar detection of UAVs based on separating the Doppler signatures of distinct UAVs [8]. Mahnoor *et al.* [4] demonstrated that visual images combined with deep learning algorithms solved the UAV detection problem with good precision. According to Zeeshan *et al.*, [1] an acoustic array, unlike radar detection and optical detection approaches, does not rely on the size of the viewed item for detection but on the rotors' sound, and its prerequisite is a large sound dataset. Maham *et al.* [3] performed UAV detection with IR images, but for real-time, the direction detection system could face a multi-class problem. That's why in this paper, we perform multi-class and multi-target drone detection in challenging weather conditions based on infrared (IR) images utilizing an improved YOLOv5p2 model with the main contributions listed below.

- Customized dataset [9] is created for multi-class IR-images detection and classification with challenging weather conditions and multi-size targets with varying altitudes.
- Improved the baseline YOLOv5 model by introducing an auto-anchor algorithm because it avoids the requirement to scan the input image using a sliding window that computes a prediction at each possible spot.
- The night-vision IR images have very small or tiny objects, and we introduced an extra-small anchor (P2) in the model's head.
- Upgraded the baseline's standard Conv and C3 mod-

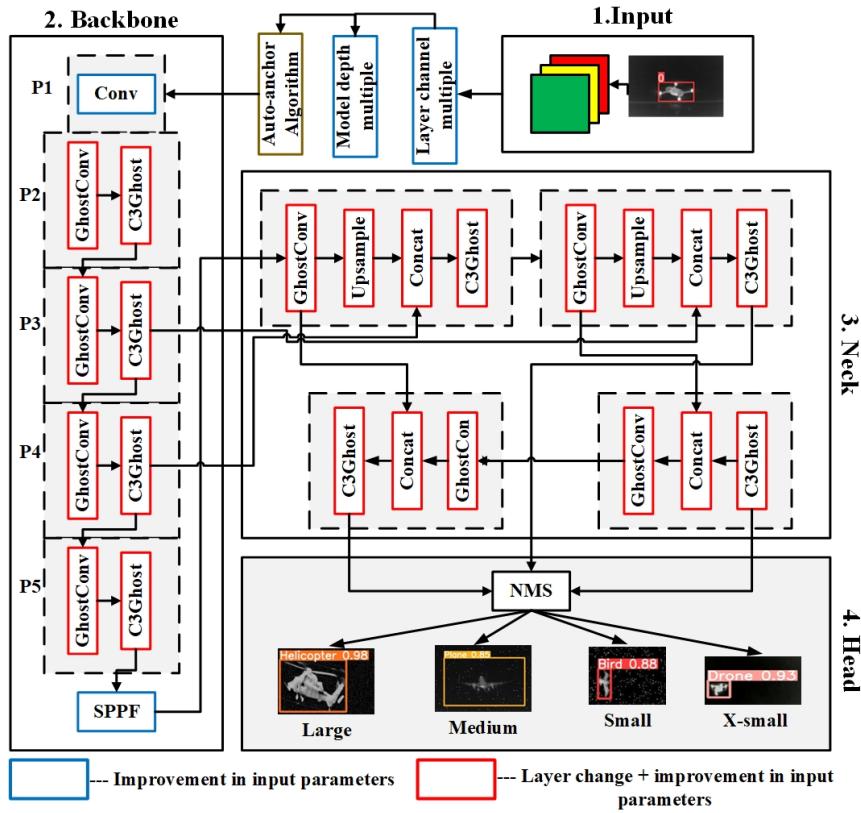


Fig. 1: Network topology of the proposed Ghost Auto Anchor Network (GAANet).

ules with GhostConv and Ghost, respectively [10]. The integration of these modules in the baseline performs optimization at each layer because the ghost module only selects the best and non-repetitive feature maps.

- Integration of AdamW [11] optimizer improved the weight decay w.r.t loss function during training as it excellently decouples weight decay from the gradient update step. We named this improved model Ghost auto anchor Network (GAANet).

II. LITERATURE REVIEW

Drone detection with CNN was envisaged to overcome these limitations because deep neural networks have excellent feature extraction capabilities. Authors in [12] performed drone identification based on U-Net, a segmented network, to extract regions of interest (ROI), followed by ResNet to categorize the objects in the ROI. Recurrent correlation network (RCN) was used with stationary cameras to improve drone detection in 4K videos [13], and correlation filtering to extract the motion features of tiny flying objects. Target identification is skewed when the foreground target is retrieved improperly or incompletely during the background-modeling phase. For the accurate location of the targets in the extracted areas, a two-stage model was outlined in [14] that used background subtraction to find prospective targets with CaffeNet for target identification. Object detection with computer vision improved the categorization and localization of targets. Such as

YOLOv5, Faster-RCNN [4], and CenterNet [15]. The authors of [16] created an artifactual data set of drones and birds by removing the target's background and merging them with other images. This dataset was then classified using YOLOv2. Darknet was used as the YOLOv4 backbone for UAV vs. bird detection with an accuracy of 98.3% [17]. One-stage detection-based YOLOv5 has excellent detection accuracy and rapid inference time, but its accuracy rapidly drops when dealing with smaller objects in high-resolution photos. The authors in [18] improved YOLOv5s for multi-rotor UAVs identification. They swapped the YOLOv5s core with Efficientlite, which simplified the model by removing irrelevant layers. The researchers in [19] fine-tuned the YOLOv5 model with visual images to achieve 95.2% accuracy and compared the performance with benchmarks like YOLOv3, YOLOv4, and maskRCNN. In [20], authors developed an autonomous drone detection system with sensor fusion of thermal and infrared cameras, which resulted in fewer false positives. In [21], the researchers used ResNet as a feature extractor with multi-cascaded auto-encoders for eliminating rain patterns in the UAV images. This method achieved average recognition accuracy of 82% at 24 frames per second. An efficient two-stage approach was proposed in [22] to overcome the problems of high-resolution and small-size UAVs with fixed cameras. The effectiveness of high-resolution images was enhanced by excluding multiple background regions and targeting the candidate regions by SAG-YOLOv5, which had a Ghost mod-

ule and attention mechanism (SimAM). During an extensive literature review, we deduced that YOLOv5 could not be implemented directly for multi-class UAV detection because of the drone's small size. Also, it faces an exact allocation of bounding boxes with variable-sized targets in a dataset. In this research, we improved YOLOv5 for rapid detection and developed the lightweight model *Ghost auto anchor net (GAANet)* for multi-scale tiny object detection with higher accuracy.

III. GHOST AUTO ANCHOR NETWORK (GAANET)

To effectively process the datasets, the CNN include a significant number of parameters, and to minimize them, CNN uses filters [1] [2]. Object detection [3] necessitates many feature maps, each of which has hundreds of channels, making the model bloated and enormous [4]. Therefore, model compression is necessary for rapid deployment on embedding devices with fewer parameters [23]. *Han et al.* presented a novel approach called GhostNet [10] to produce feature maps with fewer operations with reduced duplicate parameters and resource consumption, which allowed the deployment of the trained models on embedded devices quite conveniently. The generation of repetitive, redundant output feature maps with a large number of FLOPs and parameters is the **ghost** of a handful of intrinsic feature maps with some cheap transformations. These intrinsic feature maps are often smaller and produced by ordinary convolution filters. Here, m intrinsic feature maps $Y' \in R^{h' \cdot w' \cdot m}$ are generated using a standard convolution:

$$Y' = X * f' \quad (1)$$

where X is the input data, f' is applied filters and Y' is the output feature map, R is required resources, h' and w' are the height and width of the input data. To further obtain the desired n feature maps, a series of cheap linear operations *ghost based operations* on each intrinsic feature in Y' to generate s ghost features according to the following function:

$$y_{ij} = \Phi_{i,j}(y'_i), \quad \forall i = 1, \dots, m, \quad j = 1, \dots, s. \quad (2)$$

where y'_i is intrinsic feature map in Y' , $\Phi_{i,j}$ is the linear operation for generating the ghost feature map y_{ij} . We effectively used the GhostConv and C3Ghost modules for performing optimized convolutions with the extraction of the most pertinent and unrepeatable feature maps with no duplicate gradient information by maintaining accuracy with reduced complexity [10]. The complete network topology of the proposed GAANet is shown in Fig. 1, where the size of the input image is set to 265×256 because lower-resolution images increase the generalizability of the GAANet and make it less prone to overfitting, with a focus on important high-level features. The model's depth and channel multipliers are set to 0.25 and 0.5, respectively. These values are chosen so that the model has the best functionality of ghost modules in a lightweight package. The auto-anchor approach is used to apply a K-means function to the modified dataset labels. Then K-means centroids are used as the beginning conditions for a genetic evolution method.

Here, 1000 generations are investigated before the final calculation of the proposed anchors with CIoU (complete intersection over union) loss and best potential recall as the fitness function. These proposed anchors achieved fitness value of 81.08%. Now the dataset is passed to the first block of the GAANet backbone, block P1, which extracts extra(x)-small-sized feature maps with an input channel size (in_{cs}) of 128, an output channel size (o_{cs}) 6, kernel size (k_s) of 2, and a stride (sd) 2. Block P2 extracts x-small-sized feature maps with in_{cs} of 256, P3 extracts small-sized feature maps with in_{cs} of 512, P4 extracts medium-sized feature maps with in_{cs} of 768, and P5 extracts large-sized feature maps with in_{cs} of 1024. The o_{cs} and k_s of P2, P3, P4, and P5 blocks are fixed at 3 and 2, respectively. The last element of the backbone is SPPF, which does aggregate to eliminate clipping or distortion and disregards the network's fixed-size limitation for the GAANet. The GAANet block determines the locations of the bounding boxes (x, y , height, and breadth), scores, and object classes to produce an output image with a bounding box around the identified item and its confidence score.

IV. MODEL EVALUATION

A. Dataset and Model Training

Here, we proposed GAANet to successfully extract features from collected IR data in low- or no-light conditions at night. To train the proposed GAANet architecture, we gathered around **5105 IR images** of birds, drones, planes, and helicopters from the publicly available open-source datasets provided on Roboflow. These images also contain different-sized (x-small, small, medium, and large) targets, which make GAANet sensitive to multi-class, multi-size, and multi-type images. In a dataset of 4792 images [9], 4.6k images are for model training (95%) and 240 (5%) for model validation. All experiments are separately run on the *Google Colab* environment with an NVIDIA Tesla T4 GPU having a low learning rate of 0.001, where GAANet and GhostNet-YOLOv5 [24] has a batch size of 256 and 512, respectively. The epochs for both models are set to 500 for both models where GhostNet-YOLOv5 [24] stopped training at 300 by using early stopping as the model performance stopped improving while GAANet stopped training at 457 epochs.

B. Evaluation and comparison of trained models

The detailed evaluation of both trained models GAANet and GhostNet-YOLOv5 [24] is performed by the comparison of true positive (TP), true negative (TN), false negative (FN), false positive (FP), mAP, precision, and recall values. The GAANet model achieved the highest TP value of 1.00 for drones and planes and lowest TP of 0.72 for helicopters. GAANet has the highest FN for helicopters and planes at 0.12 and 0.46, respectively. However, GhostNet-YOLOv5 achieved the highest TP of 1.00 for planes and the lowest TP of 0.47 for helicopters. These stats proved the improved and accurate detection ability of GAANet for planes (1.00 TP), drones (1.00 TP), and birds (0.99 TP), with only helicopters having a TP less than 90% (0.72). The addition of ghost-based convolutions

TABLE I: Detailed comparison of evaluation metrics

Class	GAANet			GhostNet-YOLOv5 [24]		
	Precision	Recall	mAP@0.5	Precision	Recall	mAP@0.5
Bird	97.7	95.4	98.6	97.8	98.9	98.5
Drones	90.4	98.3	97.4	87	98.3	98.4
Helicopter	99	68.8	95.9	99	57.6	86.4
Plane	96.7	98.3	98.4	94.3	96.7	96.9
Overall	96.2	90.2	97.6	94.8	87.9	95.1

and C3 led to the extraction of the most relevant feature maps with smaller channel sizes than the baseline model, resulting in reduced model size, layers, parameters, and GFLOPs. The proposed GAANet trained with 395 layers took 3.210 hours to train on a completely customized dataset with a weight size of 14.1 MB.

The layer size is reduced due to the smaller value of channel and depth of the GAANet model compared to the baseline model. GAANet has the highest object precision of 99% for helicopters of varied sizes and altitudes, whereas planes have the best recall value of 98.3% and birds with the highest *mAP@0.5* of 98.6%. GAANet has the lowest precision value of 90.4% for drones which is 21.5% more than GhostNet-YOLOv5, as it achieved 68.9% detection precision for drones. Similarly, the 18.4% higher recall is achieved for helicopters compared to GhostNet-YOLOv5. Therefore, the overall precision of GAANet is increased by 1.4%, recall by 2.3%, and *mAP@50* by 2.5 % compared to GhostNet-YOLOv5. The average inference time achieved by GAANet over a batch of multiple images is 13.7 ms, which is 2.7 ms less than GhostNet-YOLOv5 of 16.4 ms. Fig. 2 and Fig. 3 show the detection results of GAANet and GhostNet-YOLOv5 tested with unseen and unknown IR images. For GAANet bird detection, the x-small and small IR images have the highest detection accuracy of 0.93. For drone IR images, GAANet achieved 0.94 accuracies for x-small images, while small, medium, and large IR drone images have 0.93 detection accuracy. GAANet achieved 0.99 accuracies on small and large helicopter IR images. For plane detection, GAANet attained 0.96 accuracies for x-small plane IR images, while small, medium, and large IR plane images have 0.85, 0.87, and 0.85 accuracies, respectively. From these results, we can infer that GAANet improved performance on all sized target IR images, but it attained the best accuracy for the x-small bird, drone, and plane IR images. GhostNet-YOLOv5 had the best testing accuracy for large bird IR images (0.89), small drone IR images (0.87), medium helicopter IR images (0.89), and large plane IR images (0.76).

C. Comparison with state-of-the art

The authors in [3] performed classification with the TFNet model for drone vs. bird IR images. In comparison with GAANet, TFNet achieved lower precision, recall, and mAP of around 0.5%, 12.8%, and 13.6%, respectively. The MFNet-M model [4] performed UAV vs. bird detection using visual images. Their proposed MFNet-M model achieved 94% recall with 95.9% mAP on drone images, while GAANet has 98.3% recall with 97.4% mAP on drone IR images which

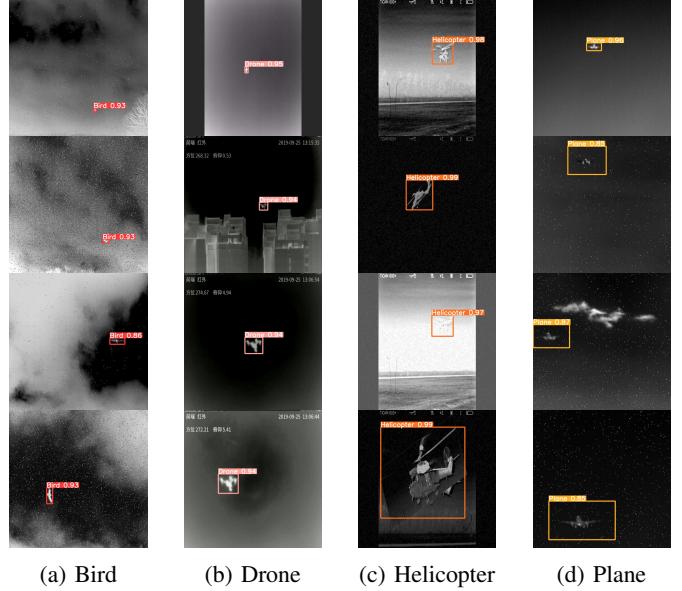


Fig. 2: Detection results of Multi-size IR targets (Top to bottom) x-small, small, medium and large (a-d) GAANet.

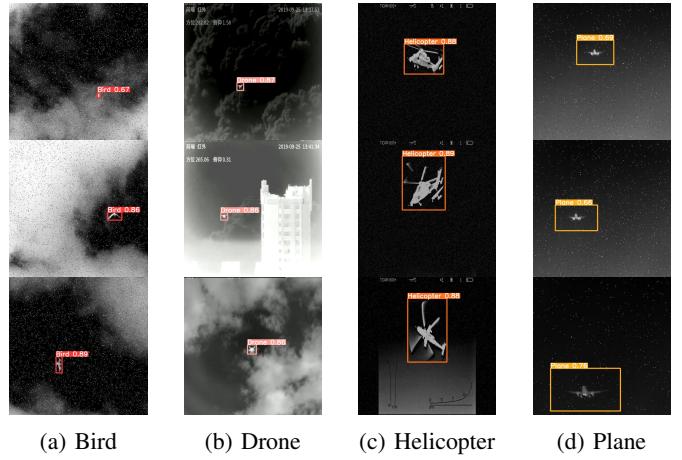


Fig. 3: Detection results of Multi-size IR targets (Top to bottom) small, medium and large (a-d) GhostNet-YOLOv5 [24].

are 4.3% and 1.5% high, respectively. The authors in [23] performed multi-type UAV classification with YOLOv7 on visual images. Single rotor UAV images achieved 88.7% recall and 94% mAP, which are 4% more, 9.6%, and 2.5% less than GAANet, respectively. In [25], the authors improved the baseline YOLOv4 backbone with GhostNet and achieved 0.9% less precision than GAANet. Similarly, the authors used Ghost convolution in the YOLOv5 baseline model and achieved

19.89% less precision than GAA-Net [24]. YOLOv5x-ALL-GHOST added GhostNet in both head and backbone and got 0.6% less *map*@0.5 than GAA-Net [26]. GhostNet feature extraction networking was embedded in YOLOv3 [27] that achieved 8.3% less *map*@0.5 than GAA-Net.

TABLE II: Comparison with the state-of-the-art schemes for UAV detection.

Model	Precision (%)	Recall (%)	Parameters (million)	Weight (MB)
MFNet-M	96.8	90.4	5.2	75.3
YOLOv4-GhostNet [25]	95.32	86.54	39.70	150
GhostNet-YOLOv5 [24]	76.31	88.42	5.9	10
YOLOv5x-ALL-GHOST [26]	N/G	N/G	25.09	48.7
YOLO-G [27]	88.9	86.3	N/G	42.7
Proposed GAA-Net	96.2	90.2	6.8	14.1

V. CONCLUSION

In this paper, we proposed an improved and optimized deep-learning model for extra small-sized flying object detection, specifically UAVs, using IR images during night surveillance. The proposed GAA-Net used ghost convolution, ghost C3, and a downsampled input channel size to extract the most prominent and non-repeated features. Detailed experimentation was performed on a customized multi-class dataset containing drones, planes, helicopters, and birds. The results showed a low misclassification rate, which confirms the effectiveness of the proposed model for real-time night vision IR images. The proposed object detection approach outperformed the other current state-of-the-art technologies by a significant margin.

VI. ACKNOWLEDGMENT

This work was supported by the Higher Education Commission (HEC) Pakistan under the NRPU 2021 Grant 15687 and also by Engineering and Physical Sciences Research Council (EPSRC), U.K., under Grant EP/W004348/1.

REFERENCES

- [1] M. Z. Anwar, Z. Kaleem, and A. Jamalipour, "Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications," *IEEE Transactions on Vehicular Technology*, vol. 68, pp. 2526–2534, Mar. 2019.
- [2] Z. Kaleem and M. H. Rehmani, "Amateur drone monitoring: State-of-the-art architectures, key enabling technologies, and future research directions," *IEEE Wireless Communications*, vol. 25, pp. 150–159, Apr. 2018.
- [3] M. Misbah, M. U. Khan, Z. Yang, and Z. Kaleem, "TF-Net: Deep Learning Empowered Tiny Feature Network for Night-time UAV Detection," *arXiv:2211.16317*, pp. 1–15, 2022.
- [4] M. Dil, M. U. Khan, M. Z. Alam, F. A. Orakzai, Z. Kaleem, and C. Yuen, "SafeSpace MFNet: Precise and Efficient MultiFeature Drone Detection Network," *arXiv:2211.16785*, pp. 1–13, 2022.
- [5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, p. 1137–1149, 2017.
- [6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.
- [7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," *ECCV*, p. 21–37, 2016.
- [8] F. Hoffmann, M. Ritchie, F. Fioranelli, A. Charlish, and H. Griffiths, "Micro-Doppler based detection and tracking of UAVs with multistatic radar," in *IEEE Radar Conference (RadarConf)*, pp. 1–6, 2016.
- [9] M. U. Khan, Z. Kaleem, and M. Misbah, "GitHub - Zee-shanKaleem/GhostAutoAnchorNet," 2022. Accessed 2022-12-08.
- [10] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1577–1586, 2020.
- [11] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
- [12] S. Samaras, E. Diamantidou, D. Ataloglou, N. Sakellariou, A. Vafeiadis, V. Magoulianitis, A. Lalas, A. Dimou, D. Zarpalas, K. Votis, and et al., "Deep learning on multi sensor data for counter uav applications—a systematic review," *Sensors*, vol. 19, no. 22, p. 4837, 2019.
- [13] C. Craye and S. Ardjoune, "Spatio-temporal semantic segmentation for drone detection," in *16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–5, 2019.
- [14] L. Sommer, A. Schumann, T. Muller, T. Schuchert, and J. Beyerer, "Flying object detection for automatic UAV recognition," in *14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, 2017.
- [15] G. Xu, S. Tang, Z. Yu, and K. Fu, "Confine keypoint triplets for object detection," in *IEEE International Conference on Artificial Intelligence and Industrial Design (AIID)*, pp. 608–613, 2021.
- [16] C. Aker and S. Kalkan, "Using deep networks for drone detection," in *14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, 2017.
- [17] F. Dadras Javan, F. Samadzadegan, M. Gholamshahi, and F. Ashatari Mahini, "A modified YOLOv4 Deep Learning Network for vision-based UAV recognition," *Drones*, vol. 6, no. 7, p. 160, 2022.
- [18] B. Liu and H. Luo, "An Improved Yolov5 for Multi-Rotor UAV Detection," *Electronics*, vol. 11, no. 15, p. 2330, 2022.
- [19] N. Al-Qubaydhi, A. Alenezi, T. Alanazi, A. Senyor, N. Alanezi, B. Alotaibi, M. Alotaibi, A. Razaque, A. A. Abdelhamid, and A. Alotaibi, "Detection of unauthorized unmanned aerial vehicles using yolov5 and transfer learning," *Electronics*, vol. 11, no. 17, 2022.
- [20] F. Svanstrom, C. Englund, and F. Alonso-Fernandez, "Real-time drone detection and tracking with visible, thermal and acoustic sensors," in *25th International Conference on Pattern Recognition (ICPR)*, pp. 7265–7272, 2021.
- [21] L. Ye, S. Hu, T. Yan, and Y. Xie, "GAF representation of millimeter wave drone rcs and drone classification method based on deep fusion network using ResNet," *IEEE Transactions on Aerospace and Electronic Systems*, p. 1–11, 2022.
- [22] Y. Lv, Z. Ai, M. Chen, X. Gong, Y. Wang, and Z. Lu, "High-resolution drone detection based on background difference and SAG-Yolov5s," *Sensors*, vol. 22, no. 15, p. 5825, 2022.
- [23] Z. Kaleem, M. U. Khan, M. Dil, M. Misbah, F. A. Orakzai, and M. Z. Alam, "TransLearn-YOLOx: Improved-YOLO with Transfer Learning for Fast and Accurate Multiclass UAV Detection," *Preprints 2022*, 2022120049, Dec. 2022.
- [24] M. Cao, H. Fu, J. Zhu, C. Cai, M. Cao, H. Fu, J. Zhu, and C. Cai, "Lightweight tea bud recognition network integrating GhostNet and YOLOv5," *Mathematical Biosciences and Engineering*, vol. 19, no. 12, pp. 12897–12914, 2022.
- [25] C. Zhang, F. Kang, and Y. Wang, "An improved apple object detection method based on lightweight yolov4 in complex backgrounds," *Remote Sensing*, vol. 14, no. 17, 2022.
- [26] Y. Zhang, W. Cai, S. Fan, R. Song, and J. Jin, "Object Detection Based on YOLOv5 and GhostNet for Orchard Pests," *Information*, vol. 13, p. 548, Nov. 2022.
- [27] L. Kong, J. Wang, and P. Zhao, "YOLO-G: A Lightweight Network Model for Improving the Performance of Military Targets Detection," *IEEE Access*, vol. 10, pp. 55546–55564, 2022.