UDC 004.93 (045)
DOI:10.18372/1990-5548.69.16425

[1]V. M. Sineglazov,
[2]V. V. Kalmykov

IMAGE PROCESSING FROM UNMANNED AERIAL VEHICLE USING MODIFIED YOLO DETECTOR

[1,2]Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine
E-mails: [1]svm@nau.edu.ua ORCID 0000-0002-3297-9060, [2]kvvad@ukr.net

Abstract—*Identifying objects from drone images is a state-of-the-art task for artificial neural networks. Since drones are always moving at different altitudes, the scale of the object varies greatly, making it difficult to optimize the networks. Moreover, flying at high speeds and low altitudes leads to blurred images of densely populated objects during movement, which is a problem when solving the problem of recognizing and classifying small sized objects. This paper addresses the above problem solutions and solves them by applying an additional prediction model to identify objects of different scales. We also modify the loss function to penalize larger objects more and vice versa to encourage recognition of smaller objects. To achieve improvements, we use advanced techniques such as multiscale testing, image blurring, object rotation, and data distortion. Experiments with a large data set show that our model has good performance in drone images. Compared to the baseline model (YOLOv5), our model shows significant improvements in object recognition and classification.*

Index Terms—Unmanned aerial vehicle; YOLO; feature maps extraction object detection; classification problem; hybrid neural networks.

I. INTRODUCTION

The scope of modern unmanned aerial vehicles (UAVs) affects both the civilian and special spheres: a reconnaissance of the terrain, aerial image analysis, security of protected objects, intelligent surveillance, route inspection, environmental monitoring, patrolling borders, traffic control, emergency assistance, etc [1] – [4]. In this article we focus on improving object detection. With help of image processing, based on video information from drones, and provide information on the numerous applications mentioned above.

In recent years, we have seen significant progress in object detection tasks using deep convolutional neural networks [5], [6]. Some well-known benchmark datasets, such as MS COCO [7] and PASCALVOC [8], contribute significantly to the development of object detection applications.

II. PROBLEM STATEMENT

The task of object selection on a static visual scene, on which there is a set of objects $O = \{o_m\}$, $M = 1, ..., M$ is understood to be the selection of a subset $O' \subset O$, each element $o'_{m'}$, where $m' = 1,..., M'$ is the index of an object from the set O' highlighted on the corresponding visual scene by a bounding box that covers this object in whole or in part.

However, most previous deep convolutional neural networks have been developed for images of natural scenes. The direct application of previous models to the problem of object detection in drone capture scenarios basically has three problems, which are intuitively illustrated by some cases in Fig. 1.

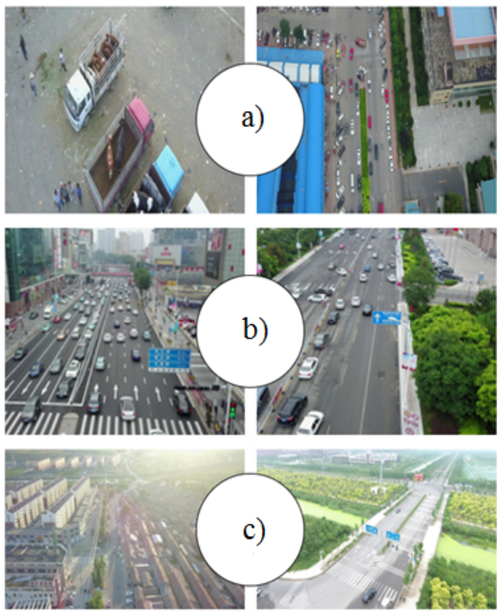

Fig. 1. Showcases to explain the three main problems in object detection on drone-captured images: (a) is the image size variation (large and small scale); (b) is the high-density of objects; (c) is the large surface coverage

Firstly, the scale of the object changes significantly, because the altitude of the drone flight varies greatly. Secondly, drone images contain high-density objects, resulting in overlap between objects. Thirdly, drone images always contain confusing geographical elements due to the fact that they cover large areas. The aforementioned three problems make it very difficult to detect objects in drone images.

III. Related Works

Consider combined method combined neural network methods for analyzing static visual scenes, which solve the problem of both selection and recognition of objects. A fairy fast method is YOLO (You Only Look Once) [10], which allows you to detect images on modern video cards withover 100 frames per second, but less accurate than Faster R-CNN. The method is based on the use of convolutional neural networks. The entire image is covered with a 7x7 grid. For each cell of the grid, two "anchors" are built. Then the class and the coordinate offset are predicted relative to the grid cell. Although the method is fast enough, it does not allow recognizing sufficiently small objects due to the small size of the grid.

To eliminate the disadvantage of YOLO accuracy, YOLOv2 was proposed [10]. The method, in comparison with YOLO, uses a number of additional "tricks" when training a convolutional neural network: data augmentation, training on images of higher dimensions, training on images of several scale. To form hypotheses, YOLOv2 uses anchors, the aspect ratio of which is formed on the basis of the k-means method. In addition, the work investigates the optimal number of such "anchors" for each cell of the map of signs of a convolutional neural network (5 "anchors" are used). Unlike YOLO, where the “anchors” are 98, the total number of “anchors” in YOLOv2 exceeds 1000. For each such “anchor”, as in YOLO, the class and the coordinate offset relative to the grid cell are predicted.

YOLOv5 has four different models including YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x. In general, YOLOv5 respectively uses the CSPDarknet53 architecture with SPP layer as the base, PANet as the neck and YOLO detection head [11, 14]. A Keras software package is provided to further optimise the entire architecture. Since it is the best known and most convenient single stage detector, we choose it as the base. When we train the model using the VisDrone2021 dataset [9] with a data augmentation strategy (Mosaic and MixUp), we find that YOLOv5x performs much better than YOLOv5s, YOLOv5m and YOLOv5l, with a gap in AP value of over 1.5%. Although the training costs of the YOLOv5x model are higher than the other three models, we still decided to use YOLOv5x to achieve the best detection performance. In addition, we tuned the parameters of commonly used photometric and geometric distortions to match the characteristics of the drone images.

IV. Problem Solution

A. *Model modification overview*

In object detection task, YOLO series [10] play an important role in one-stage detectors. In this paper, we propose a modified model, based on YOLOv5 [12] to solve all the above-mentioned problems. The overview of the detection pipeline using our model modification is shown in Fig. 2. We follow the original CSPDarknet53 [13] and path aggregation network (PANet [14]) as the backbone and neck of our model.

At the head end, we add another head for the detection of tiny objects. To further improve the performance of our network, we use several "tweaks" (Fig. 2). In particular, we use data augmentation during training, which facilitates adaptation to abrupt changes in the size of objects in the images. We also add multiscale testing and multi-model ensemble strategies during output to obtain more convincing detection results.

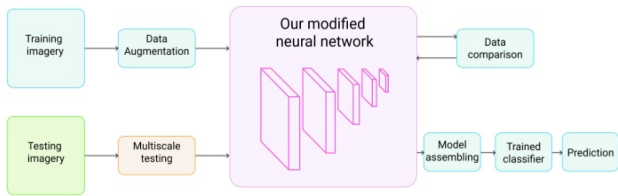

Fig. 2. The overview of working pipeline using our YOLO model modification

B. *Implementation details*

We implement our model on Pytorch 1.9.0. All of our models use a free GPU from Google Collab namely Tesla K80 for training, validation, and testing. In the training phase, we use part of pre-trained model from yolov5x, because our model modification uses most parts of backbone and some part of head from the original YOLOv5 model, by using these weights we can save a lot of training time.

We use VisDrone2021 training dataset for 70 epochs, and the first 3 epochs are used for warm-up. We use Adam optimizer for training, and 3e-5 as the initial learning rate. The learning rate of the last epoch decays to 0.11 of the initial learning rate.

Since the neural network only accepts photos with a size multiple of 32, each photo in the dataset has been pre-formatted to the required size of 416 x 416, and in some experiments 512 x 512 pixels, so that the neural network can learn from them. Each individual photo taken from the drone is also labeled in a separate file, which is necessary for the training phase of the network. The data is presented in PyTorch YOLOv5 format, where each drone photo corresponds to a text file with manually prepared data about the location of objects in the photo.

C. Experimental results

To assess the performance of the network, it was trained on 500 epochs, where improvements in object detection and recognition were assessed on each epoch. The following training metrics were obtained using the TensorBoard visualization tool (Fig. 3).

Our recognition results are as follows (Fig. 4). All of the data that came out of the network has been stored for future use and training on new samples of data to improve performance.

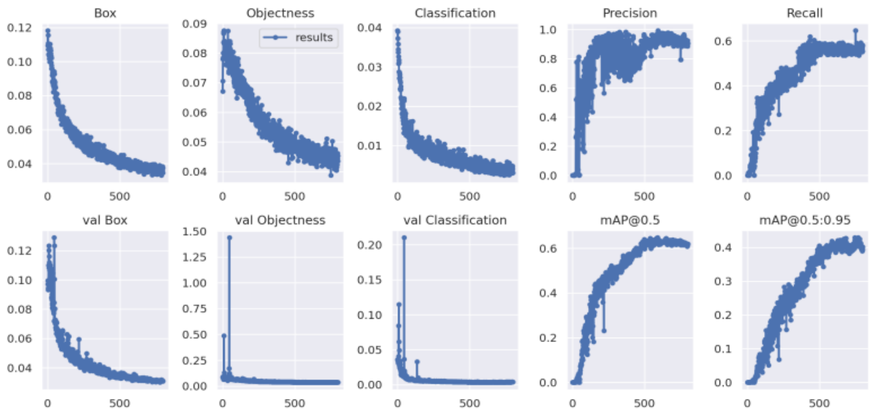

Fig. 3. Learning metrics of the neural network at 500 epochs

Fig. 4. Visualization of neural network detection and classification

D. Comparisons with the other models

Due to the limited number of entries on the VisDrone 2021 competition, we only received the results of 4 models in the testset-challenge and the final results of the ensemble of 5 models. We ended up with a good result of 37.2 points on the testset-challenge.

TABLE I. THE PERFORMANCE COMPARISON OF VISDRONE 2021 DATASET PARTICIPANTS' MODELS

Methods	mAP (%)	AP50 (%)
RetinaNet[15]	11.81	21.37
DetNet59[16]	15.26	29.23
FPN[17]	16.51	32.20
Light-RCNN[18]	16.53	32.78
RRNet (2019 2 nd) [19]	29.13	55.82
SMPNet (2020 2 nd) [20]	35.98	59.53
Our model modification	37.2	62.3

V. CONCLUSIONS

In this paper, we add to YOLOv5 some advanced technologies such as transformer coding block and some experimental techniques and create a state-of-the-art detector that is especially good for object detection in drone shooting scenarios. We have updated the record of the VisDrone2021 dataset, our experiments have shown that our modification of the model has achieved good performance on the VisDrone2021 dataset. We have tried a large number of features and used some of them to improve the accuracy of the object detector. We hope that this paper will help other artificial intelligence researchers to achieve good results in object detection with drones.

REFERENCES

[1] V. Radetsky, I. Rusnak, and Yu. Danyk, Unmanned Aviation in Modern Armed Forces: Monograph, Kyiv: NAOU (2008), p. 224.

[2] R. U. Biard and T. U. MacLjejn, *Light Unmanned Aerial Vehicles: Theory and Practice*, Moscow, Technosfera, 2015, p. 312. [in Russian]

[3] *Countering unmanned aerial systems: Methodical manual*, V. Tiurin, O. Martyniuk, et all, Kyiv: NUOU 2016, p. 30. [in Ukrainian]

[4] *Application of UAV in the conflicts of the present*, Yu. Ziatdinov, M. Kuklinsky, S. Mosov, A. Feschenko, et all.; ed. S. Mosov, Kyiv: 2013, p. 248. [in Ukrainian]

[5] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, 28:91–99, 2015.

[6] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You only look once: Unified, real-time object detection," *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016. https://doi.org/10.1109/CVPR.2016.91

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, "Microsoft coco: Common objects in context," *In European conference on computer vision*, pp. 740–755. Springer, 2014. https://doi.org/10.1007/978-3-319-10602-1_48

[8] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John M. Winn, and Andrew Zisserman, "The pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, 88(2):303–338, 2010. https://doi.org/10.1007/s11263-009-0275-4

[9] *Visdrone Team*, Visdrone 2020 leaderboard. Website, 2020. http://aiskyeye.com/visdrone-2020-leaderboard/

[10] *A Brief History of YOLO Object Detection Models From YOLOv1 to YOLOv5*, Website, 2021. https://machinelearningknowledge.ai/a-brief-history-of-yolo-object-detection-models/

[11] K. Wang, J. H. Liew, Y. Zou, D. Zhou, & J. Feng, "Panet: Few-shot image semantic segmentation with prototype alignment," *in Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9197–9206. https://doi.org/10.1109/ICCV.2019.00929

[12] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," *in Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2778–2788. https://doi.org/10.1109/ICCVW54120.2021.00312

[13] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, & I. H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," *in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 390–391. https://doi.org/10.1109/CVPRW50498.2020.00203

[14] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia, "Path aggregation network for instance segmentation," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8759–8768, 2018. https://doi.org/10.1109/CVPR.2018.00913

[15] T. Y. Lin, P. Goyal, R. Girshick, K. He, & P. Dollár, "Focal loss for dense object detection," *in Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988. https://doi.org/10.1109/ICCV.2017.324

[16] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng and J. Sun, "DetNet: Design backbone for object detection", *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 334–350. https://doi.org/10.1007/978-3-030-01240-3_21

[17] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, & S. Belongie, "Feature pyramid networks for object detection," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125. https://doi.org/10.1109/CVPR.2017.106

[18] Z. Li,, C. Peng,, G. Yu., X. Zhang,, Y. Deng,, & J. Sun, *Light-head r-cnn: In defense of two-stage object detector*. 2017. arXiv preprint arXiv:1711.07264.

[19] S. Oh, H. J. S. Kim, J. Lee, & J. Kim, "RRNet: Repetition-Reduction Network for Energy Efficient Depth Estimation," *IEEE Access*, 2020, 8, 106097–106108. https://doi.org/10.1109/ACCESS.2020.3000773

[20] Visdrone Team. Visdrone 2020 leaderboard. Website, 2020. http://aiskyeye.com/ visdrone-2020-leaderboard/.

Received September 02, 2021

Sineglazov Victor. ORCID 0000-0002-3297-9060. Doctor of Engineering Science. Professor. Head of the Department.
Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine, (1973).
Research area: Air Navigation, Air Traffic Control, Identification of Complex Systems, Wind/Solar power plant, artificial intelligence.
Publications: more than 660 papers.
E-mail: svm@nau.edu.ua

Kalmykov Vadym. Post-graduate Student.
Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation, Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (2020).
Research interests: artificial neural networks, artificial intelligence, programming.
Publications: 12.
E-mail: kvvad@ukr.net

В. М. Синєглазов В. В. Калмиков. Обробка зображень безпілотних літальних апаратів з використанням модифікованої архітектури детектора YOLO
Ідентифікація об’єктів із зображенням дронів – одне з найсучасніших завдань для штучних нейронних мереж. Оскільки дрони завжди переміщуються на різній висоті, масштаб об’єкта сильно варіюється, що ускладнює оптимізацію мереж. Більше того, політ на великих швидкостях і малих висотах призводить до нечіткого зображення густонаселених об’єктів під час руху, що є проблемою під час вирішення завдання розпізнавання та класифікації невеликих за розміром об’єктів. У статті розглядаються розв’язання вищевказаних проблем і вони вирішуються шляхом застосування додаткової моделі прогнозування для ідентифікації об’єктів різного масштабу. Ми також модифікуємо функцію втрат, щоб більші об'єкти ставити в невигідне становище і навпаки, щоб стимулювати розпізнавання дрібніших об'єктів. Щоб досягти покращень, використовуємо передові методи, такі як багатомасштабне тестування, розмиття зображення, поворот об’єкта та спотворення даних. Експерименти з великим набором даних показують, що розглянута модель добре працює на зображеннях дронів. Порівняно з базовою моделлю (YOLOv5) розглянута модель демонструє значні покращення у розпізнаванні та класифікації об’єктів.
Ключові слова: безпілотні літальні апарати; нейронна мережа YOLO; вилучення карт об’єктів; розпізнавання об’єктів; класифікація об’єктів; гібридні нейронні мережі.

Синєглазов Віктор Михайлович. ORCID 0000-0002-3297-9060.
Доктор технічних наук. Професор. Завідувач кафедрою.
Кафедра авіаційних комп’ютерно-інтегрованих комплексів, Факультет аеронавігації, електроніки і телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Київський політехнічний інститут, Київ, Україна, (1973).
Напрям наукової діяльності: аеронавігація, управління повітряним рухом, ідентифікація складних систем, вітроенергетичні установки, штучний інтелект.
Кількість публікацій: більше 660 наукових робіт.
E-mail: svm@nau.edu.ua

Калмиков Вадим Віталійович. Аспірант.
Кафедра авіаційних комп’ютерно-інтегрованих комплексів, Факультет аеронавігації, електроніки та телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (2020).
Напрям наукової діяльності: штучні нейронні мережі, штучний інтелект, програмування.
Публікації: 12.
E-mail: kvvad@ukr.net

В. М. Синеглазов, В. В. Калмыков. Обработка изображений беспилотных летательных аппаратов с использованием модифицированной архитектуры детектора YOLO
Идентификация объектов по изображениям дронов – одна из самых современных задач для искусственных нейронных сетей. Поскольку дроны всегда перемещаются на разной высоте, масштаб объекта сильно различается, что затрудняет оптимизацию сетей. Более того, полет на больших скоростях и малых высотах приводит к нечеткому изображению густонаселенных объектов во время движения, что является проблемой

при решении задачи распознавания и классификации небольших по размеру объектов. В данной статье рассматриваются решения вышеуказанных проблем и они решаются путем применения дополнительной модели прогнозирования для идентификации объектов разного масштаба. Мы также модифицируем функцию потерь, чтобы более крупные объекты ставить в невыгодное положение и наоборот, чтобы стимулировать распознавание более мелких объектов. Чтобы добиться улучшений, используются передовые методы, такие как многомасштабное тестирование, размытие изображения, поворот объекта и искажение данных. Эксперименты с большим набором данных показывают, что рассмотренная модель хорошо работает на изображениях дронов. По сравнению с базовой моделью (YOLOv5) рассмотренная модель демонстрирует значительные улучшения в распознавании и классификации объектов.
Ключевые слова: беспилотные летательные аппараты; нейронная сеть YOLO; извлечение карт объектов; распознавание объектов; классификация объектов; гибридные нейронные сети.

Синеглазов Виктор Михайлович. ORCID 0000-0002-3297-9060.
Доктор технических наук. Профессор. Заведующий кафедрой.
Кафедра авиационных компьютерно-интегрированных комплексов, Факультет аэронавигации, электроники и телекоммуникаций, Национальный авиационный университет, Киев, Украина.
Образование: Киевский политехнический институт, Киев, Украина, (1973).
Направление научной деятельности: аэронавигация, управление воздушным движением, идентификация сложных систем, ветроэнергетические установки, искусственный интеллект.
Количество публикаций: более 660 научных работ.
E-mail: svm@nau.edu.ua

Калмыков Вадим Витальевич. Аспирант.
Кафедра авиационных компьютерно-интегрированных комплексов, Факультет аэронавигации, электроники и телекоммуникаций, Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина, (2020).
Направление научной деятельности: искусственные нейронные сети, искусственный интеллект, программирование.
Публикации: 12.
E-mail: kvvad@ukr.net