UDC 004.93 (045)
DOI:10.18372/1990-5548.72.16935

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

GENERATION OF UAV-BASED TRAINING DATASET USING SEMI-SUPERVISED LEARNING

[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**—***The paper considers the problem of constructing a training sample based on the use of semi-supervised learning a teacher. The problem statement related to the problem posed is substantiated. It is shown that obtaining a training sample in some cases is a difficult task that requires significant computational and financial costs. The use of semi-supervised learning made it possible to label unlabeled data and thus ensure the creation of a labeled sample of sufficient size. The paper gives examples of generating a training sample, as well as its use for training neural networks, which are used to solve the problem of multiclass classification. Using this approach, you can get a robust data set consisting of a small amount of manually labeled images and a huge amount of pseudo-labeled or augmented data. Using this approach, one can train a classifier to detect and classify any objects in images with bounding boxes and label them accordingly.*

Index Terms—Dataset formation; semi-supervised learning; pseudo-labeling; unmanned aerial vehicle; YOLOv5; object detection; classification problem.

I. INTRODUCTION

Traditionally, supervised learning tasks formulate like this: we are given an ordered collection of l marked data points $D_L = ((x_i, y_i))_{i=1}^{l}$. Each data point (x_i, y_i) consists of an object $x_i \in X$ from a given input space X, and has a label y_i associated with it, where y_i is a real value in regression problems and a categorical value in classification problems. Based on a collection of these data points (training data), supervised learning techniques attempt to derive a function that can successfully identify the label y^* for some previously unseen input x^*.

In many classification problems we also have access to a collection of data points u, $D_U = (x_i)_{i=1}^{l+u}$, whose labels are unknown. Where l is a labelled data point. The data points for which we want to make predictions, commonly referred to as test data, are, by definition, unlabeled.

Semi-supervised classification methods attempt to use unlabeled data points to construct a learner model whose performance exceeds learners obtained using only labelled data.

The proposed training dataset can be helpful in a combat environment where we have many drone photos and videos that would take an average person several days of thorough work to process. The proposed model of training sampling construction allows increasing the amount of training data which will consequently improve the efficiency of object recognition for already existing models using inference and creation of new ones.

II. PROBLEM STATEMENT

A. Dataset formation problems

The problem of generating the right samples for training a neural network has been described earlier [2], [3]. In a situation where there are huge amounts of data from UAVs, there is a need to automatically process this data and use it to identify a potential enemy. The data source can be either UAV-based or satellite-based. A satellite transmits data in several channels at once, which complicates the construction of neural network models, but at the same time gives additional context for the detection of enemy troops and vehicles. The biggest drawback of any supervised learning algorithm is that the data set must be manually tagged by either a machine learning engineer or a data analyst. This is a very expensive process, especially when dealing with large amounts of data. The biggest disadvantage of any non-self-learning is that its range of applications is limited. To deal with these disadvantages, the concept of semi-supervised learning (SSL) was introduced [4]. In this type of learning, the algorithm is trained on a combination of labeled and unlabeled data. Typically, this combination contains a very small amount of labeled data and a very large amount of unlabeled data. The basic procedure is that the programmer first clusters similar data using an unsupervised learning algorithm and then uses the existing labeled data to label the remaining unlabeled data. Typical use cases for this type of algorithm have in common that it is relatively cheap to obtain unlabeled data while labeling this data is very expensive.

B. UAV-based object detection problems

There are three main problems for object detection on UAV-captured images:

1) image size variation (large and small scale);
2) high-density of objects;
3) large surface coverage

First of all, the scale of the object changes drastically as the altitude of the drone flight varies. Secondly, UAV images contain high-density objects, resulting in overlapping objects. Thirdly, UAV-based images always contain confusing geographical elements due to the fact that they cover large areas. The above-mentioned three problems make the detection of objects in UAV imagery very difficult.

III. Related Works

A. Object detection

It is a fundamental task of computer vision and has been widely studied in the literature [6] – [8]. Popular object detection systems include Region-based CNN (RCNN) [6], YOLO [9], SSD [10] and others. The progress made in existing works is mainly to train a stronger or faster object detector when sufficient annotated data is available. There is growing interest in improving detectors using unlabeled training data using a semi-observable object detection system. Recently, a consistency-based semi-supervised object detection method has been proposed in [11], which provides consistent prediction of the unlabeled image and its inverted counterpart. Their method requires a more complex Jensen–Shannon divergence to compute the consistency regularization.

B. Semi-supervised learning

Semi-supervised learning for image classification has recently improved considerably. Consistency regularization is becoming one of the popular approaches among recent methods and is inspiring in object detection. The idea is to force the model to generate consistent predictions when data is augmented with label preservation. Examples include Mean-Teacher [12], UDA [13], and MixMatch. Another popular class of SSL is pseudo-labeling, which can be seen as a hard version of consistency regularization: the model performs self-training to generate pseudo labels of unlabeled data and thereby trains randomly augmented unlabeled data to match the corresponding pseudo labels.

C. Data augmentation

Data augmentation is crucial for improving generalization and model stability [14], especially gradually they become the main stimulus for semi-supervised learning.

Finding appropriate color and geometric transformations of input spaces has been shown to be crucial for improving generalization [15]. However, most augmentations are mainly studied in image classification. The complexity of data augmentation for object detection is much higher than for image classification, since the global geometric transformations of the data affect the annotations of the bounding boxes.

D. Modified YOLOv5 detector

Our modification follows [1] (Fig. 1) the original CSPDarknet53 [17] and path aggregation network (PANet [18]) as the backbone and neck of our model.

At the head end, we add another head for the detection of tiny objects. To improve the performance and accuracy of our network, we use several "tweaks" (Fig. 1). We use data augmentation during training, it provides adaptation to abrupt changes in the size of objects in the images. We also use multi-model ensemble strategies and multiscale testing for the output to obtain more accurate detection results.

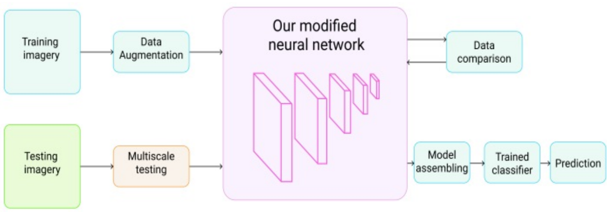

Fig. 1. The overview of working pipeline using our YOLOv5 model modification [1]

The main advantage of our model is that it can work efficiently with different UAV photo sizes. Its loss-counting function considers the differences between heights and the number of elements in the photo.

IV. Problem Solution

A. Implementation details

In order to obtain the best UAV object recognition results, it was decided to choose the YOLOv5 neural network model with a modification [1], which we have described before. We did not aim to achieve the best recognition results in a short period of time (the network was trained for a few days only). Our goal was to generate a fully labeled sample from a huge amount of unlabeled data.

The essence of our sampling approach is that we take and manually label approximately 150 photos and then train the detector to find the planes in the UAV photo. We then use this detector to mark the remainder of the photos, in our case 650 more photos. After that we need to refine the neural network using the new data, having previously prepared it.

We implement our model on Pytorch 1.12.1. All of our models use a free graphics processing unit (GPU) from Google Collab for training, validation, and testing. For training, we use our modified backbone with only one class to detect – airplanes. As an initial dataset, we chose Roboflow aerial dataset [16]. Our dataset creation procedure consists of several stages:

1) data selection;
2) data normalization;
3) manual labeling;
4) data augmentation;
5) initial model warm-up with 150 images;
6) image labeling (650 images);
7) further inference.

Unsupervised loss formulation that leverages unlabeled data is the key to SSL. Many advancements in SSL for classification rely on some forms of consistency regularization. We use a K-way classification. The formula for consistency regularization is defined as follows:

$$l_u = \sum_{x \in X} w(x) l\big(q(x), p(x;\theta)\big), \qquad (1)$$

here $x \in X$ is an image, p, $q : X \to [0, 1]K$ map x into a $(K-1)$-simplex, and $w : X \to \{0, 1\}$ maps x into a binary value. $\ell(\cdot, \cdot)$ measures a distance between two vectors. Typical choices include cross entropy and L2 distance. p represents the prediction of the model that is parameterized by θ. q is the prediction target, and w is the weight. It determines the contribution of x to the model loss.

B. Pseudo-labeling

Pseudo-labeling can be defined as follows:

$$\begin{aligned} &q(x) = \text{ONE\_HOT}\big(\arg\max\big(p(x;\theta)\big)\big), \\ &w(x) = 1 \quad \text{if} \quad \max\big(p(x;\theta)\big) \geq \tau. \end{aligned} \qquad (2)$$

C. Experimental results

To evaluate the performance of the network, it was trained on 450 epochs, where at each epoch improvements in object detection and recognition were evaluated. The following training metrics were obtained using the matplotlib visualization tool (Fig. 2). Our recognition results are as follows (Fig. 3). All data obtained by the network were stored for further use and training on new data samples to improve performance.

D. Comparison with the other models

According to the result of our previous research, we can compare our model modification with other popular models of neural networks. We haven't changed the model described in [1] (Fig. 4), so the results are accurate.

We also compared our model with YOLOv5 itself. There are several varieties of YOLOv5 models, namely YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x. Each of them differs in their performance for different tasks. Some are better suited for recognizing small objects in large photos, others are better at recognizing large objects in close-up photos, while others can recognize both with varying success. We modified YOLOv5s model so that it can work with UAV-based data and detect small objects in real time with high accuracy.

We also did comparison of different YOLOv5 modifications in Table I.

TABLE I. THE PERFORMANCE COMPARISON OF YOLOV5 MODEL MODIFICATIONS

Model	Size	mAP 0.5:0.95	mAP 0.5	Speed (ms)	FLOPs
YOLOv5s	416	37.2	56	0.9	16.5
YOLOv5m	416	45.2	63.9	1.7	49
YOLOv5l	416	48.8	67.2	2.7	109.1
YOLOv5x	416	50.7	68.9	4.8	205.7
Our model	416	53.2	70.1	5.2	234.3

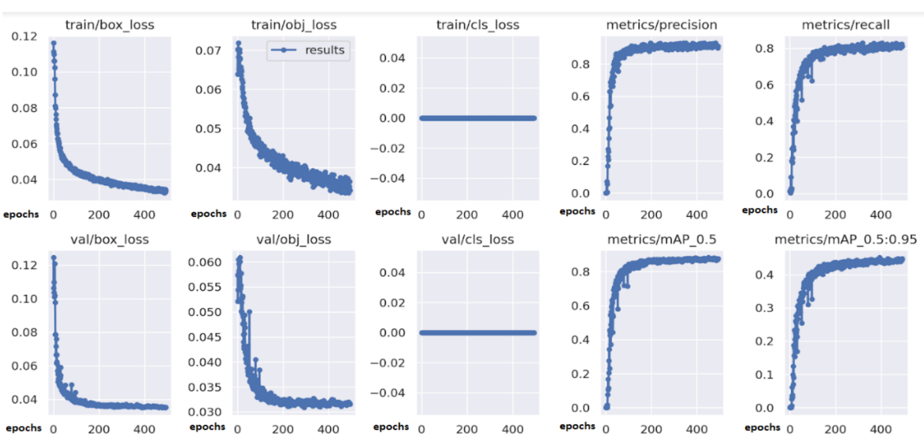

Fig. 2. Learning metrics of the neural network at 450 epochs

Fig. 3. Visualization of airplane detection and classification

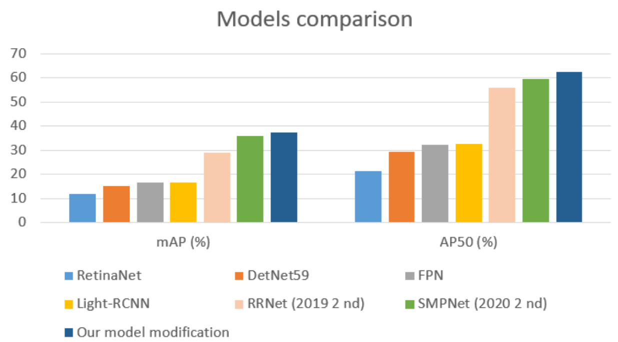

Fig. 4. Performance comparison with other popular models

V. CONCLUSIONS

In this paper, we propose an algorithm to form a dataset using a semi-supervised learning model. To do so, we suggest using our YOLOv5 model modification for classification model training. For demonstration purposes, we used Roboflow's open aerial dataset. To prove the concept of semi-supervised learning, we manually labelled about two hundred images. After that, we trained our modified YOLOv5 model to tag other five hundred pictures. Our modified model showed increased detection accuracy in a short-term classification task compared to other known models. The proposed approach can be utilized to expand the dataset with more data samples and for further inference of already trained object classifiers.

REFERENCES

[1] V. M. Sineglazov and V. V. Kalmykov, “Image Processing from Unmanned Aerial Vehicle Using Modified YOLO Detector,” Electronics and control systems, NAU Kyiv: vol. 3, no. 69, pp.37–42, 2021. https://doi.org/10.18372/1990-5548.69.16425

[2] J. H. Kim, J. Kim, S. J. Oh, S. Yun, H. Song, J. Jeong, & H. O. Song, (2022). Dataset Condensation via Efficient Synthetic-Data Parameterization. arXiv preprint arXiv:2205.14959.

[3] M. Maranghi, A. Anagnostopoulos, I. Cannistraci, I. Chatzigiannakis, F. Croce, G. Di Teodoro, & P. Velardi, (2022). AI-based Data Preparation and Data Analytics in Healthcare: The Case of Diabetes. arXiv preprint arXiv:2206.06182.

[4] C. Shneider, A. Hu, A. K. Tiwari, M. G. Bobra, K. Battams, J. Teunissen, & E. Camporeale, (2021). A Machine-Learning-Ready Dataset Prepared from the Solar and Heliospheric Observatory Mission. arXiv preprint arXiv:2108.06394.

[5] Semi-supervised learning, 2022, https://en.wikipedia.org/wiki/Semi-supervised_learning

[6] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR 2014. https://doi.org/10.1109/CVPR.2014.81

[7] Kaiming He, Georgia Gkioxari, Piotr Doll´ar, and Ross Gir-shick. Mask r-cnn. In ICCV 2017.

[8] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In CVPR 2018.

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In CVPR 2016. https://doi.org/10.1109/CVPR.2016.91

[10] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In ECCV 2016. https://doi.org/10.1007/978-3-319-46448-0_2

[11] Jisoo Jeong, Seungeui Lee, Jeesoo Kim, and Nojun Kwak. Consistency-based semi-supervised learning for object detection. In NeurIPS, 2019.

[12] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In NeurIPS, 2017.

[13] Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V. Le. Unsupervised data augmentation for consistency training. arXiv preprint arXiv:1904.12848, 2019.

[14] Daniel Ho, Eric Liang, Ion Stoica, Pieter Abbeel, and Xi Chen. Population based augmentation: Efficient learning of augmentation policy schedules. arXiv preprint arXiv:1905.05393, 2019.

[15] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation strategies from data. In CVPR, 2019. https://doi.org/10.1109/CVPR.2019.00020

[16] Aerial dataset. Website, 2022. https://universe.roboflow.com/gdit/aerial-airport/dataset/1.

[17] Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). 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 (pp. 390–391).

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

Received April 24, 2022

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 670 papers.
E-mail: svm@nau.edu.ua

Kalmykov Vadym. Post-graduate student.
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: 14.
E-mail: kvvad@ukr.net

В. М. Синєглазов В. В. Калмиков. Формування набору даних для навчання на базі БПЛА з використанням напівкерованого навчання

У роботі розглянуто завдання побудови навчальної вибірки на основі використання навчання із частковим залученням вчителя. Обґрунтовано постановку завдання, пов'язану з поставленою проблемою. Показано, що отримання навчальної вибірки у ряді випадків є складним завданням, яке потребує значних обчислювальних та фінансових витрат. Використання машинного навчання із частковим залученням вчителя дозволило розмітити немарковані дані і цим забезпечити створення маркованої вибірки достатнього обсягу. В роботі наведено приклади генерації навчальної вибірки, а також її використання для навчання нейронних мереж, які застосовуються для вирішення багатокласової задачі класифікації. Використовуючи цей підхід, можна отримати надійний набір даних, що складається з невеликої кількості ручних зображень і величезної кількості псевдорозмічених даних або доповнених даних. Використовуючи цей підхід, можна навчити класифікатор виявляти та класифікувати будь-які об'єкти на зображеннях з обмежувальними рамками та маркувати їх відповідним чином.

Ключові слова: формування набору даних; напівкероване навчання; псевдомаркування; безпілотний літальний апарат; YOLOv5; виявлення об’єктів; задача класифікації.

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.
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: 14.
E-mail: kvvad@ukr.net

В. М. Синеглазов, В. В. Калмыков. Создание обучающих наборов данных на основе БПЛА с использованием полууправляемого обучения

В работе рассмотрена задача построения обучающей выборки на основе использования обучения с частичным привлечением учителя. Обоснована постановка задачи, связанная с поставленной проблемой. Показано, что получение учебной выборки в ряде случаев является сложной задачей, которая требует значительных вычислительных и финансовых затрат. Использование машинного обучения с частичным привлечением учителя позволило разметить немаркированные данные и тем самым обеспечить создание маркированной выборки достаточного объема. В работе приведены примеры генерации обучающей выборки, а также ее использование для обучения нейронных сетей, которые применяются для решения задачи многоклассовой классификации. Используя этот подход, можно получить надежный набор данных, состоящий из небольшого количества размеченных вручную изображений и огромного количества псевдоразмеченных данных или дополненных данных. Используя этот подход, можно обучить классификатор обнаруживать и классифицировать любые объекты на изображениях с ограничительными рамками и маркировать их соответствующим образом.

Ключевые слова: формирование массива данных; полууправляемое обучение; псевдоразметка; беспилотный летательный аппарат; YOLOv5; обнаружение объектов; проблема классификации.

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.
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: 14.
E-mail: kvvad@ukr.net