

doi: 10.17586/2226-1494-2024-24-2-190-197

Fast labeling pipeline approach for a huge aerial sensed dataset

Andrei M. Fedulin¹, Natalia V. Voloshina²✉

¹ “КТ — Unmanned Systems” JSC, Saint Petersburg, 199178, Russian Federation

² ITMO University, Saint Petersburg, 197101, Russian Federation

¹ af@kronshtadt.ru, <https://orcid.org/0000-0001-6951-4681>

² nvvoloshina@itmo.ru✉, <https://orcid.org/0000-0001-9435-9580>

Abstract

Modern neural network technologies are actively used for Unmanned Aerial Vehicles (UAVs). Convolutional Neural Networks (CNN), are mostly used for object detection, classification, and tracking tasks, for example, for such objects as fires, deforestations, buildings, cars, or people. However, to improve effectiveness of CNNs it is necessary to perform their fine-tuning on new flight data periodically. Such training data should be labeled, which increases total CNN fine-tuning time. Nowadays, the common approach to decrease labeling time is to apply auto-labeling and labeled objects tracking. These approaches are not effective enough for labeling of 8 hours’ huge aerial sensed datasets that are common for long-endurance USVs. Thus, reducing data labeling time is an actual task nowadays. In this research, we propose a fast aerial data labeling pipeline especially for videos gathered by long-endurance UAVs cameras. The standard labeling pipeline was supplemented with several steps such as overlapped frames pruning, final labeling spreading over video frames. The other additional step is to calculate a Potential Information Value (PIV) for each frame as a cumulative estimation of frame anomaly, frame quality, and auto-detected objects. Calculated PIVs are used than to sort out frames. As a result, an operator who labels video gets informative frames at the very beginning of the labeling process. The effectiveness of proposed approach was estimated on collected datasets of aerial sensed videos obtained by long-endurance UAVs. It was shown that it is possible to decrease labeling time by 50 % in average in comparison with other modern labeling tools. The percentage of average number of labeled objects was 80 %, with them being labeled for 40 % of total pre-ranged frames. Proposed approach allows us to decrease labeling time for a new long-endurance flight video data significantly. This makes it possible to speed up neural network fine-tuning process. As a result, it became possible to label new data during the inter-flight time that usually takes about two or three hours and is too short for other labeling instruments. Proposed approach is recommended to decrease UAVs operators working time and labeled dataset creating time that could positively influence on the time necessary for the fine-tuning a new effective CNN models.

Keywords

fast labeling pipeline, FLP, unmanned aerial vehicle, UAVs, long-endurance UAVs, adversarial attack, frames potential information value, PIV

For citation: Fedulin A.M., Voloshina N.V. Fast labeling pipeline approach for a huge aerial sensed dataset. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 2, pp. 190–197. doi: 10.17586/2226-1494-2024-24-2-190-197

УДК 004.032.26

Способ быстрой разметки сверхбольших данных аэросъемки

Андрей Михайлович Федюлин¹, Наталья Викторовна Волошина²✉

¹ АО «КТ — Беспилотные Системы», Санкт-Петербург, 199178, Российская Федерация

² Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ af@kronshtadt.ru, <https://orcid.org/0000-0001-6951-4681>

² nvvoloshina@itmo.ru✉, <https://orcid.org/0000-0001-9435-9580>

Аннотация

Введение. Современные нейросетевые технологии активно применяются для беспилотных летательных аппаратов. Для решения задач детектирования, классификации и сопровождения объектов, расположенных

© Fedulin A.M., Voloshina N.V., 2024

на поверхности Земли, используются сверточные нейронные сети. Для повышения эффективности работы сверточных нейронных сетей требуется периодическое дообучение применяемых моделей нейронных сетей на вновь поступающих полетных данных. Такие обучающие данные необходимо дополнительно размечать. Это приводит к увеличению общего времени подготовки дообученной модели нейронной сети. Задача сокращения времени разметки чаще всего решается путем применения процедуры авторазметки и трекинга размеченных объектов. Однако существующие подходы не являются эффективными при разметке сверхбольших данных аэросъемки со стандартной для беспилотных летательных аппаратов большой продолжительностью полета (более 8 ч). Таким образом, задача поиска дополнительных способов сокращения времени разметки является актуальной. В данной работе предложен способ быстрой разметки данных аэросъемки, собранных с видеокамер в процессе полетов беспилотных летательных аппаратов. **Метод.** Стандартная процедура разметки дополнена прореживанием сильно перекрывающихся кадров в сочетании с последующим переносом результирующей разметки на все кадры размечаемого видео. Для каждого оставшегося после прореживания кадра вычисляется значение его потенциальной информативности (Potential Information Value, PIV), как совокупная оценка аномалий кадра, его качества и количества автоматически детектированных объектов. Полученные значения PIV используются для ранжирования кадров по уровню значимости. Таким образом, оператору беспилотного летательного аппарата предоставляются в начале процедуры разметки наиболее значимые кадры. **Основные результаты.** Экспериментальное исследование эффективности предложенного подхода выполнено на подготовленных наборах данных аэросъемки, полученных с беспилотного летательного аппарата с продолжительностью съемки не менее 8 ч. Показано, что время разметки может быть уменьшено в среднем на 50 % относительно применения существующих программных средств. При этом первые 40 % отсортированных кадров содержат 80 % размеченных объектов. **Обсуждение.** Предложенный способ позволяет существенно уменьшить итоговое время разметки вновь поступающих полетных видеоданных для дальнейшего дообучения модели сверточных нейронных сетей. Это позволяет провести разметку непосредственно в межполетный интервал времени, составляющий в среднем 2–3 ч. Предлагаемый подход может быть применен для уменьшения загрузки операторов беспилотного летательного аппарата.

Ключевые слова

способ быстрой разметки, беспилотный летательный аппарат, БПЛА, продолжительность полета БПЛА, состязательные атаки, величина потенциальной информативности кадра, PIV

Ссылка для цитирования: Федюлин А.М., Волошина Н.В. Способ быстрой разметки сверхбольших данных аэросъемки // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 2. С. 190–197 (на англ. яз.). doi: 10.17586/2226-1494-2024-24-2-190-197

Introduction

Latest achievements in the object detection algorithms [1] relying on the Deep Neural Network (DNN) machine learning methods, such as Faster R-CNN [2–4], SSD [5], U-Net [6], YOLOv2 [7, 8], YOLOv3 [9], YOLOv5¹, and YOLOv7 [10], have contributed to a rapid development of the on-board High-Performance Intellectual Computing Systems (B-HPICS) for Unmanned Aerial Vehicles (UAV) capable to process data streams from sensors using DNN explicitly in-flight. Despite B-HPICS are always less powerful than ground ones, their on-board location provides a number of great advantages such as:

- Usage of the original data input instead of compressed one is often more confident;
- Sensor auto-control allows to scan area much faster and achieve higher search performance than manual one (Fig. 1);
- Sensor auto-control mode allows to remove significant part of the routine workload from UAV’s operators [11].

State-of-art UAVs equipped with B-HPICS [12] were designed as long-endurance UAVs to handle in real time radar station signals and infrared and optoelectronic camera video streams data allowing detecting and recognizing objects of the interest with the high accuracy. Such long-endurance UAVs are usually equipped with both high-quality sensors and powerful B-HPICS; thus, they may become a real competitor over conventional aerial sensing

solutions, such as manned aircrafts, Earth remote sensing satellites, and drones especially amidst huge areas [13]. B-HPICS should be flexible for the environmental changes and capable for self-improvement of its DNNs by training and fine-tuning processes on the new datasets collected and prepared explicitly from the recent flights. This task is still relevant for modern Convolutional Neural Networks (CNN) used by long-endurance UAVs.

Problem Statement

There have been developed several effective technologies [14–16] of the so called “few shot learning” to improve already trained CNN model by a small number of labeled images on which unrecognized objects of previous CNN model has been labeled. But there exists an open problem for such approach: how to label aerial sensed video (huge dataset of a million significantly overlapped frames) in a short inter-flight service time. Moreover, this operation should be usually done by UAVs’ operator on-site where there is neither high computer power no enough stuff available. The time line of such process is shown in Fig. 2.

Nowadays there exist labeling frameworks like SuperAnnotate (Fig. 3²), CVAT, V7 Darwin, and VGG Image Annotation [17] featured such powerful automation tools as auto-contour tool, automatic object classification, frames sampling and others.

¹ Available at: <https://zenodo.org/records/4679653> (accessed: 15.01.2024).

² SuperAnnotate labelling software observation on pipelines for AI website. Available at: <https://humansintheloop.org/tools-we-love-vol-3-superannotate> (accessed: 15.01.2024).

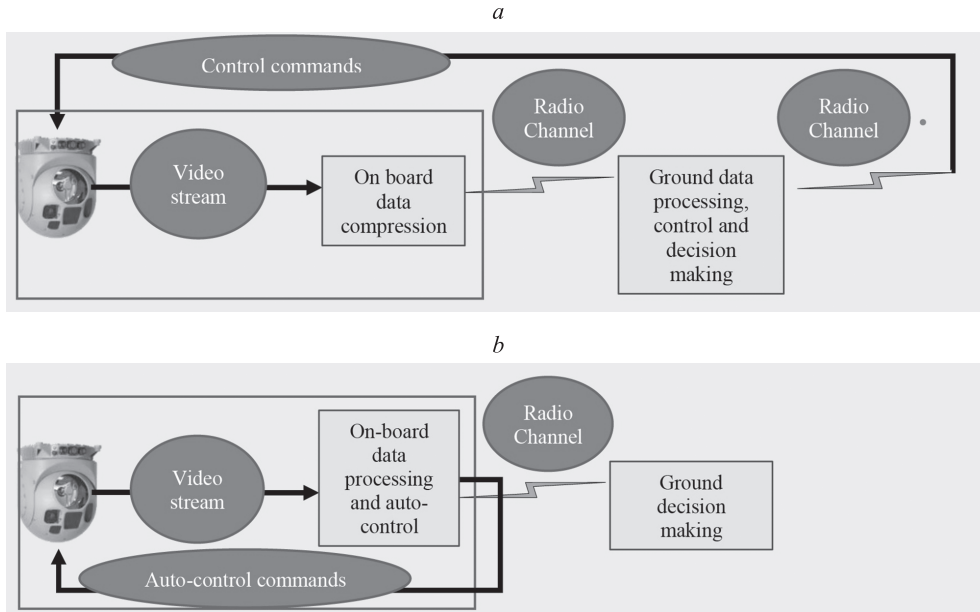


Fig. 1. Sensor control schemas: ground data processing with manual sensor control (a), on-board data processing by B-HPICS with sensor auto-control (b)

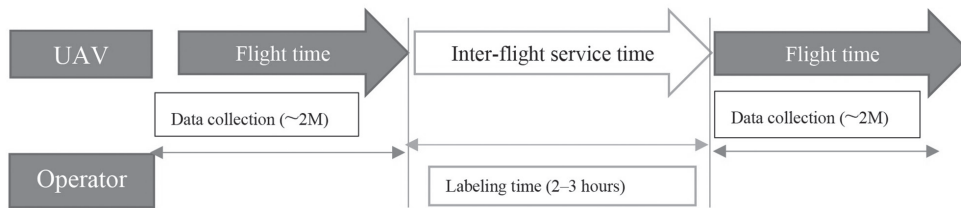


Fig. 2. Time line of flight cycle

Our experiments showed that all these tools are usually failed to process 8 hours video in a 2–3 hours of inter-flight service time. Thus, our goal was to create a fast labeling automation pipeline that allows operator to identify

and annotate as much as possible objects of interest in an aerial sensed video with a computing power limit up to 10 TFlops involving only one operator in a required inter-flight service time limitation.

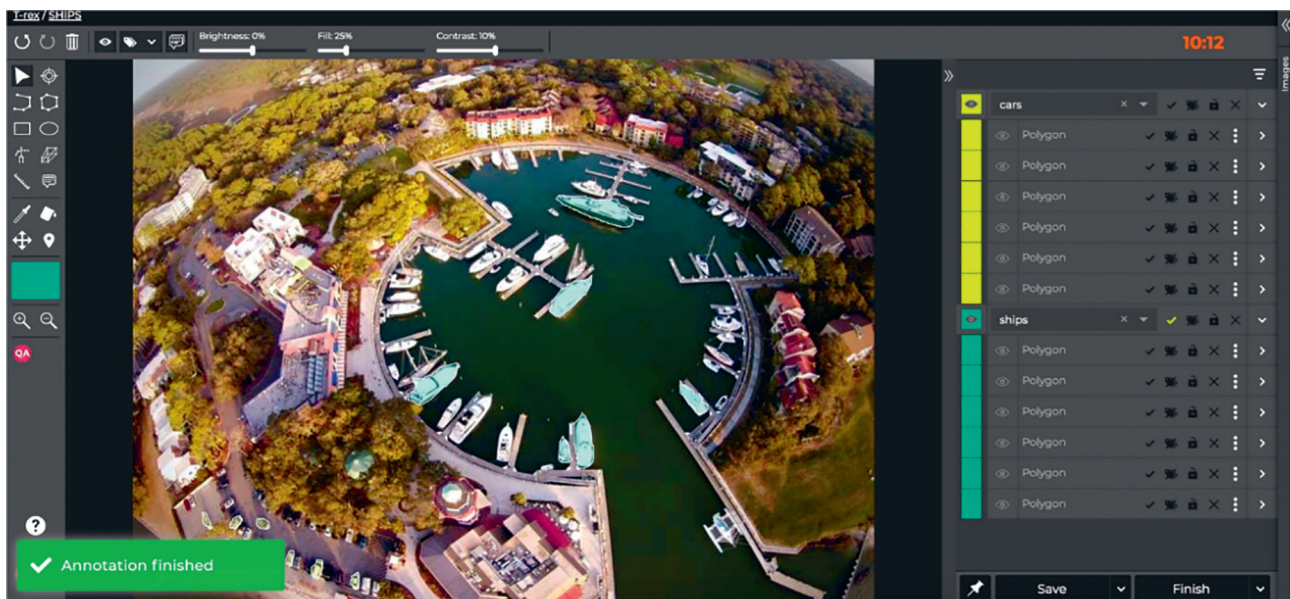


Fig. 3. An example of SuperAnnotate user interface

Proposed approach

As long as there is not enough time to process all long-duration aerial video in a short inter-flight service time, we propose several automated processes. One is to yield frames for labeling not in the sequential order but by its informativity metric defined as Potential Information Value (PIV). PIV should demonstrate not only visual quality of the frame, but if there are any objects of interest and some animalities and how many of them are in the frame at the same time. These parameters were chosen as the most important video frame characteristics for labeling process.

Thus, the PIV of a frame is defined as a combination of three parameters that we named as: image quality (IQ), the number of auto-detected objects (DO), and the number of found anomalies (AN) (frames zones with potential objects of interest):

$$PIV_i = F(AN_i, DO_i, IQ_i), \quad (1)$$

where AN_i, DO_i, IQ_i are corresponding coefficients that are calculated for current frame and i — index of current video frame that should be analyzed and labeled. F is a function with corresponding parameters that combine them to represent informativeness of the frame.

The main hypothesis is that the more is the value of of the frame the more is probability that it contains the objects of interest that should be labeled and are able to be noticed and recognized by the operator. Thus, if the hypothesis is true, it allows to review just 30–40 % of the most informative frames to be label a significant number of objects of interest.

Besides of frames ranging by PIV and state-of-art labeling automation we also propose another two features to be included in the standard labeling pipeline:

- 1) to use auto-spreading of labeling results because aerial sensed video often contains significantly overlapped frames;
- 2) to sample video dynamically by dense optical flow threshold exceedance instead of fixed time interval.

In addition, it is proposed to use image modification detection step at the beginning of the labeling process to make labeled data more reliable with aspect of information security. Thus, resulting training, fine-tuning processes, and CNN models could become more stable to adversarial attacks [18, 19] that make all labeling process more safe and secure.

In accordance with proposed hypothesis, the typical labeling pipeline have been modified. The proposed Fast Labeling Pipeline (FLP) is represented in Fig. 4.

The proposed new labeling pipeline steps are marked with gray color in Fig. 4.

Potential information value

When realizing FLP it is necessary to calculate the (1). In our research we propose to calculate it according the equation:

$$PIV = 1/2 C_{IQ}(C_{AN} + C_{DO}),$$

where C_{IQ} — image quality coefficient, $C_{IQ} \in [0, 1]$; C_{AN} — anomaly coefficient is calculated as:

$$C_{AN} = AN/AN_{all}, C_{AN} \in [0, 1],$$

where AN — number of anomalies that were found on current frame; AN_{all} — number of anomalies that were found on all frames; C_{DO} — meaningfulness coefficient is calculated as:

$$C_{DO} = DO/DO_{all}, C_{DO} \in [0, 1],$$

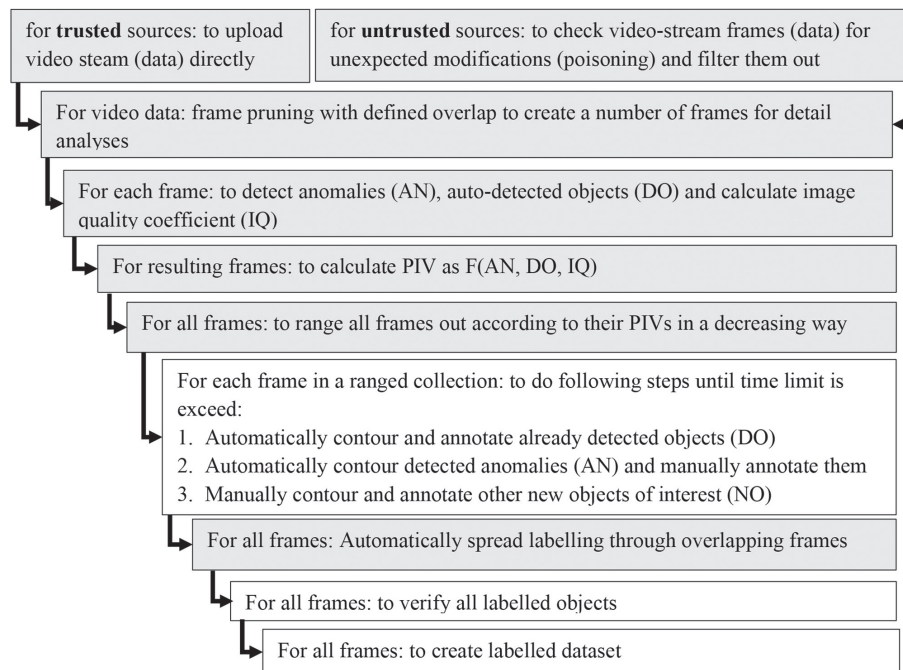


Fig. 4. Proposed fast labeling pipeline

where DO — number of objects that were detected by the current pretrained model of CNN; DO_{all} — number of objects that were detected by the current pretrained model of CNN on all frames of video.

In our approach all coefficients should be defined in a way that the higher is its value the better is the frame informativeness. In this case the higher is frames PIV the more informative is the frame.

For the C_{IQ} we propose to calculate its value as a complex quality coefficient according to the equation:

$$C_{IQ} = F1(N, H, Sh), \quad (2)$$

where $F1$ is a function with corresponding parameters that combine them to represent frame quality. It was proposed to define such quality parameters as: N — complex frame noisiness parameter, H — complex frame histogram quality parameter, Sh — complex frame sharpness parameter. Complex frame noisiness parameter should show how noisy is a frame because it is well known that the higher is frame noise the harder to detect any object in it. Complex frame histogram parameter should show how close is frame histogram to the normal one. That means that it should not contain many peaks or be concentrated in local part of all histogram. Complex frame sharpness parameter should show if it contains abnormal sharpen elements.

In addition, a C_{IQ} value of each frame could be transformed from quantitative to qualitative form to be shown for an operator who labels frames of a flight video to make it easier to understand if it has good quality or not.

Experimental part

The proposed fast labeling pipeline was implemented as a cloud-based web-service featuring GPGPU support for fast performance of image processing operations.

During the experiment, there used several pretrained neural networks: YOLOv7 [20] with predefined classifier for an object detection and classification, WideResnet-50 [21] as an anomaly detector, and FlowNet2S [22] as an optical flow definer.

To calculate quality coefficient C_{IQ} (2) we choose following approaches to define quality parameters. Complex frame noisiness parameter N is obtained by analyzing results of smoothing filters such as Gaussian, Wiener, mean, and median filters with 3×3 filtering windows size in combination with analysis of one- and two-dimensional Fourier spectrum in an aspect of higher spectrum coefficients. The mean square distortion of elements was taken to estimate noisiness. Complex frame histogram quality parameter H is obtained by analyzing average offset of normalized brightness, average offset of normalized contrast, and histogram density. Complex frame sharpness parameter Sh is obtained based on analyzes of both Gaussian and Laplacian filters output with a frame as an input of the first one.

In our realization the OpenCV libraries¹ were used to implement all applied filters and transforms.

¹ OpenCv website. Available at: <https://opencv.org/home/> (accessed: 15.01.2024).

Thus, in proposed method the frame quality is calculated as quality coefficient C_{IQ} :

$$C_{IQ} = \frac{\sum_{i=1}^3 C_i}{3} \prod_{i=1}^3 Sigm_{C_i} \quad (3)$$

where C_i — quality parameter, C_1 — frame histogram quality parameter (equal to H in (2)), C_2 — frame sharpness parameter (equal to Sh in (2)), C_3 — frame noisiness parameter (equal to N in (2)), and $Sigm_{C_i}$ is a quality parameters weighted coefficient that is calculated for each corresponding quality parameter C_i :

$$Sigm_{C_i} = 1 - \frac{k5C_i}{\sqrt{1 + (5C_i)^2}}, \quad k = \sqrt{26/25}.$$

This quality parameter weighted coefficients $Sigm_{C_i}$ are proposed to balance influence of low values of quality parameters in nonlinear way. So that value of quality coefficient C_{IQ} strongly drops down if any of quality parameters C_i become extremely low.

The frame quality is estimated by the resulting value of $C_{IQ} \in [0, 1]$ (3). For better visualization (Fig. 5, *b*) in our experiment the transformation thresholds were expertly defined as:

$$\begin{aligned} C_{IQ} \geq 0.55 & \text{ — good quality,} \\ 0.3 \leq C_{IQ} < 0.55 & \text{ — medium quality,} \\ C_{IQ} < 0.3 & \text{ — low quality.} \end{aligned}$$

The examples of proposed fast labeling pipeline results are shown in Fig. 5 and Fig. 6.

Fig. 5 and Fig. 6 show that the operator is provided with the most informative frames and its parameters at the very beginning of labeling process. Auto-labeling process matches existing anomalies is they are proper for predefined classifier. Additional parameters of the auto-labeling results are also shown for the operator to check labeling correctness and to find and label unlabeled objects of interest in much shorter time.

For the experimental part there was chosen six types of 8-hour FullHD aerial sensed video in a day time over countryside area. Total number of objects lays in a range from 5 to 100. The number of object classes is 3 (3 types of vehicles). One of the testing videos has low quality (smoothed view with a block structure on it). There are videos that have concentration of objects of interest at a short part of the video (at the beginning or end or somewhere in the middle) and the others have approximately uniform distribution of objects over the video frames. Three experienced specialists of labeling took part in the experiment.

Experimental results are shown in Fig. 7. It shows dependence between the average percentage of labeled objects $Nobj$ and the average percentage of labeled video frames Nfr .

Graphics in Fig. 7 show that with proposed Fast Labeling Pipeline (FLP) 80 % of objects are labeled in 40 % of first pre-ranged frames in average in comparison with CVAT [23]. It was shown experimentally that



Fig. 5. An example of proposed software interface with sorted by PIV frames: an interface view (a), sorting parameters description (b)



Fig. 6. An example of frame with anomalies with result of auto-labeling with current CNN of proposed software: an interface view (a), auto-labeling parameters description (b)

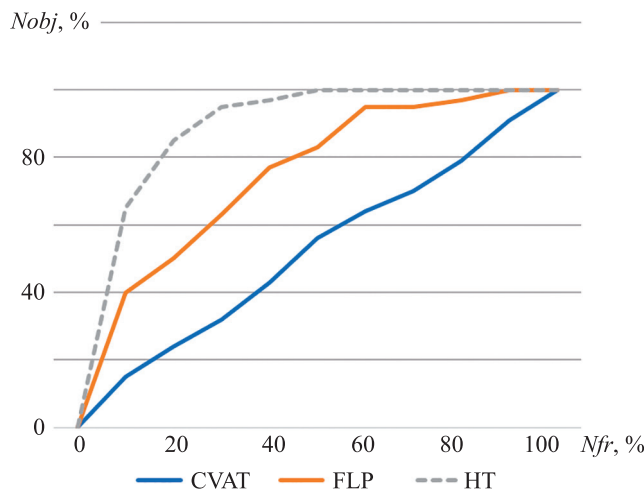


Fig. 7. Experimental results of dependence between the average percentage of labeled objects N_{obj} and the average percentage of labeled video frames N_{fr} : CVAT — Computer Vision Annotation Tool, FLP — proposed Fast Labelling Pipeline, HT — hypothetical effectiveness threshold

resulting labeling time become 50 % less in average due to the applied auto-labeling and proposed PIV ranging and label spreading methods. HT is a hypothetical effectiveness threshold that is based on practical needs of real labeling process. HT could be looked at as a goal for future optimization.

Conclusion

In the presented research, a new pipeline was proposed allowing us to fit labeling time into a short inter-flight period. Such effect was achieved by combining both state-of-art automation tools (such as object detection and auto-contour tools) and proposed tools: ranging frames by its PIV, dynamic labels spreading through overlapped frames and smart frames sampling. A cloud-based labeling web-service was developed and it was shown experimentally that proposed pipeline allows labeling 80 % objects of interest of all existing objects just by processing 40 % of pre-ranged frames in average that fits 2–3 hours of inter-flight service time. Future research will be aimed to find optimal PIV parameters calculation algorithm (close to hypothetical effectiveness threshold) by input video analysis.

References

Литература

1. Zhao Z., Zheng P., Xu S., Wu X. Object detection with deep learning: A review. *arXiv*, 2019, arXiv:1807.05511. <https://doi.org/10.48550/arXiv.1807.05511>
2. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 2015, vol. 28, pp. 91–99.
3. Liu S., Liu Z. Multi-channel CNN-based object detection for enhanced situation awareness. *Sensors & Electronics Technology (SET) panel Symposium SET-241 on 9th NATO Military Sensing Symposium*, 2017.
4. Mahalanobis A., McIntosh B. A comparison of target detection algorithms using DSIAC ATR algorithm development data set. *Proceedings of SPIE*, 2019, vol. 10988, pp. 1098808. <https://doi.org/10.1117/12.2517423>
5. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A. SSD: Single shot multibox detector. *Lecture Notes in Computer Science*, 2016, vol. 9905, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
6. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, 2015, vol. 9351, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
7. Redmon J., Farhadi A. YOLO9000: better, faster, stronger. *Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6517–6525. <https://doi.org/10.1109/CVPR.2017.690>
8. Chen H.-W., Reyes M., Marquand B., Robie D. Advanced automated target recognition (ATR) and multi-target tracker (MTT) with electro-optical (EO) sensors. *Proceedings of SPIE*, 2020, vol. 11511, pp. 115110V. <https://doi.org/10.1117/12.2567178>
9. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. *arXiv*, 2018, arXiv:1804.02767v1. <https://doi.org/10.48550/arXiv.1804.02767>
10. Wang C.Y., Bochkovskiy A., Liao M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proc. of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. <https://doi.org/10.1109/CVPR52729.2023.00721>
11. Fedulin A.M., Evstaf'ev D.V., Kondrashova G.L., Artemenko N.V. Human-autonomy teaming interface design for multiple-UAV control. *Russian Aeronautics*, 2022, vol. 65, no. 2, pp. 419–424. <https://doi.org/10.3103/S1068799822020222>
12. Barnell M., Raymond C., Capraro Ch., Isereau D., Cicotta Ch., Stokes N. High-performance computing (HPC) and machine learning demonstrated in flight using agile condor. *Proc. of the 2018 IEEE High Performance extreme Computing Conference (HPEC)*, 2018, pp. 1–4. <https://doi.org/10.1109/HPEC.2018.8547797>
13. Fedulin A.M., Driagin D.M. Prospects of MALE-class UAVs using for the huge territories aerial survey. *Izvestiya SFedU. Engineering Sciences*, 2021, no. 1(218), pp. 271–281. (in Russian). <https://doi.org/10.18522/2311-3103-2021-1-271-281>
14. Fei-Fei L., Fergus R., Perona P. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, vol. 28, no. 4, pp. 594–611. <https://doi.org/10.1109/tpami.2006.79>
15. Fink M. Object classification from a single example utilizing class relevance metrics. *Advances in Neural Information Processing Systems*, 2004, vol. 17, pp. 449–456.
16. Alajaji D., Alhichri H.S., Ammour N., Alajlan N. Few-shot learning for remote sensing scene classification. *Proc. of the 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*, pp. 81–84. <https://doi.org/10.1109/M2GARSS47143.2020.9105154>
17. Sager Ch., Janiesch Ch., Zschech P. A survey of image labelling for computer vision applications. *Journal of Business Analytics*, 2021, vol. 4, no. 2, pp. 91–110. <https://doi.org/10.1080/2573234X.2021.1908861>
18. Oprea A., Vassilev A., Fordyce A., Anderson H. Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations: Report NIST AI100-2E2023. 107 p. <https://doi.org/10.6028/NIST.AI.100-2e2023>
19. Choi J.I., Tian Q. Adversarial attack and defense of YOLO detectors in autonomous driving scenarios. *Proc. of the 2022 IEEE Intelligent*
1. Zhao Z., Zheng P., Xu S., Wu X. Object detection with deep learning: A review // *arXiv*. 2019. arXiv:1807.05511. <https://doi.org/10.48550/arXiv.1807.05511>
2. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks // *Advances in Neural Information Processing Systems*. 2015. V. 28. P. 91–99.
3. Liu S., Liu Z. Multi-channel CNN-based object detection for enhanced situation awareness // *Sensors & Electronics Technology (SET) panel Symposium SET-241 on 9th NATO Military Sensing Symposium*. 2017.
4. Mahalanobis A., McIntosh B. A comparison of target detection algorithms using DSIAC ATR algorithm development data set // *Proceedings of SPIE*. 2019. V. 10988. P. 1098808. <https://doi.org/10.1117/12.2517423>
5. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A. SSD: Single shot multibox detector // *Lecture Notes in Computer Science*. 2016. V. 9905. P. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
6. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation // *Lecture Notes in Computer Science*. 2015. V. 9351. P. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
7. Redmon J., Farhadi A. YOLO9000: better, faster, stronger // *Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. P. 6517–6525. <https://doi.org/10.1109/CVPR.2017.690>
8. Chen H.-W., Reyes M., Marquand B., Robie D. Advanced automated target recognition (ATR) and multi-target tracker (MTT) with electro-optical (EO) sensors // *Proceedings of SPIE*. 2020. V. 11511. P. 115110V. <https://doi.org/10.1117/12.2567178>
9. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement // *arXiv*. 2018. arXiv:1804.02767v1. <https://doi.org/10.48550/arXiv.1804.02767>
10. Wang C.Y., Bochkovskiy A., Liao M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors // *Proc. of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023. <https://doi.org/10.1109/CVPR52729.2023.00721>
11. Fedulin A.M., Evstaf'ev D.V., Kondrashova G.L., Artemenko N.V. Human-autonomy teaming interface design for multiple-UAV control // *Russian Aeronautics*. 2022. V. 65. N 2. P. 419–424. <https://doi.org/10.3103/S1068799822020222>
12. Barnell M., Raymond C., Capraro Ch., Isereau D., Cicotta Ch., Stokes N. High-performance computing (HPC) and machine learning demonstrated in flight using agile condor // *Proc. of the 2018 IEEE High Performance extreme Computing Conference (HPEC)*. 2018. P. 1–4. <https://doi.org/10.1109/HPEC.2018.8547797>
13. Федулин А.М., Дрягин Д.М. Перспективы применения крупно-размерных БПЛА при решении задач комплексного обследования территорий // *Известия ЮФУ. Технические науки*. 2021. № 1(218). С. 271–281. <https://doi.org/10.18522/2311-3103-2021-1-271-281>
14. Fei-Fei L., Fergus R., Perona P. One-shot learning of object categories // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006. V. 28. N 4. P. 594–611. <https://doi.org/10.1109/tpami.2006.79>
15. Fink M. Object classification from a single example utilizing class relevance metrics // *Advances in Neural Information Processing Systems*. 2004. V. 17. P. 449–456.
16. Alajaji D., Alhichri H.S., Ammour N., Alajlan N. Few-shot learning for remote sensing scene classification // *Proc. of the 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*. P. 81–84. <https://doi.org/10.1109/M2GARSS47143.2020.9105154>
17. Sager Ch., Janiesch Ch., Zschech P. A survey of image labelling for computer vision applications // *Journal of Business Analytics*. 2021. V. 4. N 2. P. 91–110. <https://doi.org/10.1080/2573234X.2021.1908861>
18. Oprea A., Vassilev A., Fordyce A., Anderson H. Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations: Report NIST AI100-2E2023. 107 p. <https://doi.org/10.6028/NIST.AI.100-2e2023>
19. Choi J.I., Tian Q. Adversarial attack and defense of YOLO detectors in autonomous driving scenarios // *Proc. of the 2022 IEEE Intelligent Vehicles Symposium (IV)*. 2022. P. 1011–1017. <https://doi.org/10.1109/IV51971.2022.9827222>

- Vehicles Symposium (IV)*, 2022, pp. 1011–1017. <https://doi.org/10.1109/IV51971.2022.9827222>
20. Wang C.Y., Bochkovskiy A., Liao H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 7464–7475. <https://doi.org/10.1109/cvpr52729.2023.00721>
 21. Defard T., Setkov A., Loesch A., Audigier R. PaDiM: A Patch distribution modeling framework for anomaly detection and localization. *Lecture Notes in Computer Science*, 2021, vol. 12664, pp. 475–489. https://doi.org/10.1007/978-3-030-68799-1_35
 22. Ilg E., Mayer N., Saikia T., Keuper M., Dosovitskiy A., Brox T. FlowNet 2.0: Evolution of optical flow estimation with deep networks. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1647–1655. <https://doi.org/10.1109/cvpr.2017.179>
 23. Guillermo M., Billones R.K., Bandala A., Vicerra R.R., Sybingco E., Dadios E.P., Fillone A. Implementation of automated annotation through mask RCNN Object Detection Model in CVAT using AWS EC2 Instance. *Proc. of the 2020 IEEE REGION 10 CONFERENCE (TENCON)*, 2020, pp. 708–713. <https://doi.org/10.1109/tencon50793.2020.9293906>
 20. Wang C.Y., Bochkovskiy A., Liao H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors // *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023. P. 7464–7475. <https://doi.org/10.1109/cvpr52729.2023.00721>
 21. Defard T., Setkov A., Loesch A., Audigier R. PaDiM: A Patch distribution modeling framework for anomaly detection and localization // *Lecture Notes in Computer Science*. 2021. V. 12664. P. 475–489. https://doi.org/10.1007/978-3-030-68799-1_35
 22. Ilg E., Mayer N., Saikia T., Keuper M., Dosovitskiy A., Brox T. FlowNet 2.0: Evolution of optical flow estimation with deep networks // *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. P. 1647–1655. <https://doi.org/10.1109/cvpr.2017.179>
 23. Guillermo M., Billones R.K., Bandala A., Vicerra R.R., Sybingco E., Dadios E.P., Fillone A. Implementation of automated annotation through mask RCNN Object Detection Model in CVAT using AWS EC2 Instance // *Proc. of the 2020 IEEE REGION 10 CONFERENCE (TENCON)*. 2020. P. 708–713. <https://doi.org/10.1109/tencon50793.2020.9293906>

Authors

Andrei M. Fedulin — Director for Software Development, “KT — Unmanned Systems” JSC, Saint Petersburg, 199178, Russian Federation, [sc 57514263700](https://orcid.org/0000-0001-6951-4681), <https://orcid.org/0000-0001-6951-4681>, af@kronshtadt.ru
Natalia V. Voloshina — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 55511854200](https://orcid.org/0000-0001-9435-9580), <https://orcid.org/0000-0001-9435-9580>, nvvoloshina@itmo.ru

Received 06.02.2024

Approved after reviewing 14.03.2024

Accepted 28.03.2024

Авторы

Федулин Андрей Михайлович — директор центра разработки программного обеспечения, АО «КТ — Беспилотные Системы», Санкт-Петербург, 199178, Российская Федерация, [sc 57514263700](https://orcid.org/0000-0001-6951-4681), <https://orcid.org/0000-0001-6951-4681>, af@kronshtadt.ru
Волошина Наталья Викторовна — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 55511854200](https://orcid.org/0000-0001-9435-9580), <https://orcid.org/0000-0001-9435-9580>, nvvoloshina@itmo.ru

Статья поступила в редакцию 06.02.2024

Одобрена после рецензирования 14.03.2024

Принята к печати 28.03.2024



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»