УДК 004.932:629.7

АДАПТИВНА ПОПЕРЕДНЯ ОБРОБКА ВІДЕОПОТОКУ БПЛА З ПЕРЕВІРКОЮ ЗА ВІДГУКОМ ДЕТЕКТОРА

Гладченко А.С.
Національний університет кораблебудування імені адмірала Макарова
Україна, Миколаїв
251244@nuos.edu.ua

Анотація. *Запропоновано адаптивний вибір попередньої обробки кадру за відгуком детектора: алгоритм застосовується, лише коли детектор знаходить із ним більше стійких рамок, ніж на кадрі без обробки. За розміткою UAVDT і VisDrone вибір не погіршує розпізнавання в жодній умові, а в тумані підвищує MOTA на 0,015. На темних відео без розмітки правильних рамок стає на 31 % більше за майже незмінної їх частки, а обробка працює в реальному часі на платформі 15 Вт.*

Ключові слова: *комп'ютерний зір; безпілотні літальні апарати; попередня обробка зображень; детекція об'єктів; YOLO.*

Вступна частина. Виявлення об'єктів на відео з безпілотних літальних апаратів (БПЛА) ускладнюють малий розмір об'єктів, рух камери й обмежені обчислення на борту [1]. За слабкого освітлення задача стає ще складнішою. Глибокі методи покращення зображень дають якісний результат, але додають окрему нейромережу до бюджету кадру [2, 3], а спільне навчання покращення з детектором потребує зміни самої моделі [4]. Дешеві фільтри на основі таблиць підстановки й підвищення різкості працюють у реальному часі, проте жоден фіксований фільтр не підходить для всіх умов. Тому доцільним є вибір фільтра за ознаками кадру. Його вада в тому, що ознаки кадру не показують, чи допоможе фільтр детектору саме на цьому відео.

Метою дослідження є адаптивна попередня обробка відеопотоку БПЛА, яка в реальному часі збільшує кількість правильно виявлених об'єктів і не погіршує розпізнавання на нових даних.

Основна частина. У початковій схемі конвеєр обробляє кадр 960×540 у такій послідовності: вибір алгоритму, фільтрація, детекція моделлю YOLO26n [5], навченою на VisDrone [6], супровід ByteTrack [7]. Сітка містить 78 обчислювально дешевих алгоритмів, а дерево рішень обирає алгоритм за чотирма ознаками: середньою яскравістю, контрастом, часткою тіней і щільністю контурів.

Дію вибору оцінено сліпою перевіркою рамок, у яких два варіанти розходяться: відношення перетину до об'єднання (IoU) з усіма рамками іншого варіанта менше 0,3. З кожного набору випадково взято по 25–30 рамок, які автор оцінював без назви варіанта. Чистий приріст правильних рамок на кадр обчислено за формулою (1):

$$G = \frac{D_1 \cdot p_1 - D_2 \cdot p_2}{F} \tag{1}$$

де D_1, D_2 – кількість рамок лише першого й лише другого варіанта; p_1, p_2 – частки правильних рамок у вибірках; F – кількість кадрів. Межі отримано підстановкою 95 % інтервалів Вілсона в найгіршому й найкращому поєднанні.

На власному дуже темному відео вибір додав 4,28 правильної рамки на кадр, але на 77 темних послідовностях відкритого набору нічних відео з БПЛА NAT2021 [8] він спричинив

втрату 0,67 рамки на кадр з межами [−0,77; −0,50]. Отже, ознаки кадру не відрізняють відео, на яких підсвічування допомагає, від відео, на яких воно шкодить.

Тому вибір доповнено зворотним зв'язком. Дерево лише пропонує алгоритм, а раз на 60 кадрів той самий кадр подається на детектор двічі: із запропонованим і з базовим алгоритмом Orig+Shp1.4, який лише підвищує різкість. Рахуються рамки зі значенням впевненості не нижче 0,45. Результат проби k обчислюється за формулою (2) і згладжується за формулою (3):

$$r_k = \ln \frac{N_P + 1}{N_B + 1} \tag{2}$$

$$s_k = (1 - \alpha) \cdot s_{k-1} + \alpha \cdot r_k \tag{3}$$

де N_P, N_B – кількість рамок із запропонованим і базовим алгоритмом; $\alpha = 0{,}35$ – вага нової проби. Пропозиція приймається, коли $s_k > 0{,}10$, і відхиляється, коли $s_k < 0$, а до прийняття застосовується базовий алгоритм.

Правило, параметри й критерії зафіксовано до перевірки. Основний критерій: нижня межа G проти варіанта без фільтрів більша за нуль. Основний тест виконано на 238 темних послідовностях навчальної частини NAT2021 (46 301 кадр), які не брали участі в налаштуванні, додатковий – на 103 світліших послідовностях. Результати наведено в табл. 1.

Таблиця 1 – Чистий приріст рамок на кадр, у дужках консервативні межі

Порівняння	Набір	Правильні рамки	Хибні
З перевіркою проти без фільтрів	238 темних, нові	+0,096 [+0,041; +0,141]	+0,002
Без перевірки проти без фільтрів	238 темних, нові	−0,013 [−0,104; +0,076]	−0,012
З перевіркою проти без перевірки	238 темних, нові	+0,034 [−0,030; +0,104]	+0,090
З перевіркою проти Orig+Shp1.4	238 темних, нові	+0,007 [−0,024; +0,038]	+0,005
З перевіркою проти без фільтрів	103 світліші, нові	+0,094 [−0,044; +0,223]	+0,024
З перевіркою проти без фільтрів	77 темних, налаштування	+0,001 [−0,071; +0,079]	+0,059

Основний критерій виконано: приріст становить близько 5 % за 2,05 рамки на кадр без фільтрів, а кількість хибних рамок майже не змінюється. Вибір без перевірки на цьому наборі значущої різниці не дав, а втрату на послідовностях налаштування усунуто. На світліших послідовностях шкоди не виявлено, приросту також не доведено.

Перевагу над фіксованим Orig+Shp1.4 не підтверджено: цим алгоритмом оброблено 71 % кадрів. Порівняно з вибором без перевірки хибних рамок стало більше на 0,090 на кадр з межами [+0,020; +0,154].

На вбудованій платформі Ryzen Z1 Extreme за ліміту 15 Вт обробка кадру займає 23,75 мс проти 23,55 мс без перевірки, відкинуто 0,9 % кадрів потоку 30 кадрів за секунду.

Сліпа перевірка не враховує об'єктів, пропущених обома варіантами, і перемикань ідентичності. За показником точності супроводу MOTA [9], що обчислюється за розміткою, схема з деревом поступається обробці без фільтрів (табл. 2), тому схему перебудовано (версія 4; дві проміжні версії перевірки за розміткою не пройшли).

Кадр зменшується до ширини 640 пікселів зі збереженням пропорцій (замість 960×540), бо в такій роздільній здатності його бачить детектор, і обробляється вже зменшеним. Базою замість Orig+Shp1.4 стає кадр без обробки, кандидатами – 11 алгоритмів на основі підвищення різкості, підсвічування таблицями підстановки, обмеженого за контрастом адаптивного вирівнювання гістограми (CLAHE) [10] і білатерального фільтра [11]. Дерева рішень немає: раз на 60 кадрів детектор у фоновому потоці оцінює базу й усіх кандидатів на кадрах f і $f + 3$. У формулі (2) рахуються лише стійкі рамки – із впевненістю понад 0,45 на кадрі f, які той

самий алгоритм знаходить знову на кадрі $f + 3$ з IoU ≥ 0,5, щоб одиничні хибні рамки не підвищували оцінку. Кандидат приймається, коли його s_k перевищує 0,20 або на 0,20 перевищує значення поточного, і відхиляється, коли $s_k < 0$; на кадрах із середньою яскравістю понад 160 дозволено лише підвищення різкості.

Правило налаштовано на VisDrone-MOT val і семи місцях зйомки UAVDT [12] і зафіксовано разом із критеріями. Тест за розміткою виконано на інших семи місцях UAVDT (22 послідовності, 16 472 кадри: день, ніч, туман) і на VisDrone2019-MOT test-dev (17 послідовностей, 6 635 кадрів); варіант без фільтрів обробляється тим самим шляхом. За окремою фіксацією версію 4 запущено також на NAT2021-test, де розмічено одну ціль на кадр і рахується частка знайдених цілей, і на 28 відео з БПЛА без розмітки (204 790 кадрів), не використаних для налаштування версії 4. Для них введено перевірений бал (ПБ), формула (4):

$$\text{ПБ} = 100 \cdot \frac{T - E - I}{F} \tag{4}$$

де T, E – кількість правильних і хибних рамок; I – перемикань ідентичності. На розмічених даних ПБ = 100·MOTA·(об'єктів на кадр). Без розмітки T і E оцінено сліпою перевіркою 300 рамок, по 50 із наборів «лише версія 4», «лише без фільтрів» і «спільні» для темних (яскравість менше 30) і решти відео, I не враховується.

Таблиця 2 – Результати версії 4, у дужках частка правильних рамок

Набір	Показник	Версія 4	Без фільтрів	З деревом
UAVDT, 22 посл.	MOTA	0,573 (77,2 %)	0,567 (77,5 %)	0,565 (77,0 %)
у тому числі туман, 2	MOTA	0,425 (79,3 %)	0,410 (81,0 %)	0,417 (80,2 %)
VisDrone test-dev, 17	MOTA	0,380 (84,3 %)	0,380 (84,3 %)	0,375 (83,8 %)
NAT2021-test, 77 темних	Цілі	45,0 %	44,1 %	–
NAT2021-test, 103 світлі	Цілі	57,9 %	58,1 %	–
Без розмітки, 6 темних	ПБ	853 (95,6 %)	660 (96,5 %)	–
Без розмітки, 22 інші	ПБ	1 295 (97,2 %)	1 222 (97,9 %)	–

Усі критерії тесту за розміткою виконано: версія 4 не гірша за обробку без фільтрів у кожній умові UAVDT і перевершує схему з деревом на обох наборах (MOTA вища на 13 з 22 і 11 з 17 послідовностей). Приріст понад обробку без фільтрів помітний лише в тумані, де правило обирало CLAHE; на VisDrone test-dev результати практично однакові, тож перевагу над схемою з деревом здебільшого дає база без обробки.

На темних відео без розмітки ПБ зріс на 193,5 бала на кадр (29 %) з консервативними межами [+120,4; +238,6]: 9,36 рамки на кадр проти 7,10, з них правильних 8,95 проти 6,85. На інших відео ПБ зріс на 73,6 [+29,8; +106,8]. На темних послідовностях NAT2021-test частка знайдених цілей зросла на 0,9 п. п. з 95 % бутстреп-інтервалом [−0,2; +2,5], приріст не доведено; на світлих погіршення немає.

На платформі 15 Вт обробка кадру з детекцією займає 19,0 і 24,7 мс на нічному й денному тестових відео, відкинуто 0,3 і 0,5 % кадрів, стільки ж, як без проб.

Висновки. Перевірка вибору деревом за відгуком детектора дала на нових темних відео приріст правильних рамок, але за розміткою поступилася обробці без фільтрів. Версія 4, у якій детектор обирає серед 11 алгоритмів за стійкими рамками від кадру без обробки, за розміткою не погіршує розпізнавання в жодній умові, перевершує схему з деревом і допомагає в тумані, а на темних відео без розмітки знаходить на 31 % більше правильних рамок за майже незмінної їх частки в реальному часі на платформі 15 Вт. Обмеження: один оцінювач сліпої перевірки й недоведений приріст на нічній розмітці NAT2021. Далі планується донавчання детектора на реальних нічних кадрах і розмітка нічних відео з багатьма об'єктами.

Використання інструментів ШІ. Claude Opus 5.5 (Anthropic) у середовищі Claude Code використано для написання й налагодження коду конвеєра та скриптів оцінювання й обробки результатів експериментів (основна частина). Сліпу перевірку рамок і вимірювання на вбудованій платформі виконав автор, який перевірив текст і відповідає за результати.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Wu, X., Li, W., Hong, D., Tao, R., & Du, Q. (2022). Deep learning for unmanned aerial vehicle-based object detection and tracking: A survey. *IEEE Geoscience and Remote Sensing Magazine*, *10*(1), 91–124. https://doi.org/10.1109/MGRS.2021.3115137

[2] Li, C., Guo, C., Han, L., Jiang, J., Cheng, M.-M., Gu, J., & Loy, C. C. (2022). Low-light image and video enhancement using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(12), 9396–9416. https://doi.org/10.1109/TPAMI.2021.3126387

[3] Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1777–1786). https://doi.org/10.1109/CVPR42600.2020.00185

[4] Liu, W., Ren, G., Yu, R., Guo, S., Zhu, J., & Zhang, L. (2022). Image-adaptive YOLO for object detection in adverse weather conditions. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(2), 1792–1800. https://doi.org/10.1609/aaai.v36i2.20072

[5] Sapkota, R., Cheppally, R. H., Sharda, A., & Karkee, M. (2025). *YOLO26: Key architectural enhancements and performance benchmarking for real-time object detection*. arXiv. https://doi.org/10.48550/arXiv.2509.25164

[6] Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., & Ling, H. (2022). Detection and tracking meet drones challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(11), 7380–7399. https://doi.org/10.1109/TPAMI.2021.3119563

[7] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. In *Computer Vision – ECCV 2022* (pp. 1–21). Springer. https://doi.org/10.1007/978-3-031-20047-2_1

[8] Ye, J., Fu, C., Zheng, G., Paudel, D. P., & Chen, G. (2022). Unsupervised domain adaptation for nighttime aerial tracking. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8886–8895). https://doi.org/10.1109/CVPR52688.2022.00869

[9] Bernardin, K., & Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: The CLEAR MOT metrics. *EURASIP Journal on Image and Video Processing*, *2008*, Article 246309. https://doi.org/10.1155/2008/246309

[10] Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In P. S. Heckbert (Ed.), *Graphics gems IV* (pp. 474–485). Academic Press. https://doi.org/10.1016/B978-0-12-336156-1.50061-6

[11] Tomasi, C., & Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision* (pp. 839–846). IEEE. https://doi.org/10.1109/ICCV.1998.710815

[12] Du, D., Qi, Y., Yu, H., Yang, Y., Duan, K., Li, G., Zhang, W., Huang, Q., & Tian, Q. (2018). The unmanned aerial vehicle benchmark: Object detection and tracking. In *Computer Vision – ECCV 2018* (pp. 375–391). Springer. https://doi.org/10.1007/978-3-030-01249-6_23

Hladchenko A.S.

Adaptive UAV video stream preprocessing with detector feedback verification

Abstract. *Adaptive selection of frame preprocessing based on detector feedback is proposed: an algorithm is applied only when the detector finds more persistent boxes with it than on the unprocessed frame. On annotated UAVDT and VisDrone data the selection does not degrade detection in any condition and raises MOTA by 0.015 in fog. On dark unannotated videos it yields 31% more correct boxes with almost unchanged precision and runs in real time on a 15 W platform.*

Keywords: *computer vision; unmanned aerial vehicles; image preprocessing; object detection; YOLO.*