

УДК 004.023

УЗАГАЛЬНЕНИЙ ПІДХІД ДО ВИБІРКОВОГО ПОШУКУ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ

Власенко В.О.,

vlasenko.vitaliy99@gmail.com, ORCID: 0009-0008-6951-8496

Дашкевич А.О., к.т.н.,

dashkewich.a@gmail.com, ORCID: 0000-0002-9963-0998

Воронцова Д.В., к.т.н.,

dvorontso@gmail.com, ORCID: 0000-0001-7868-0067

Охотська О.В.

lenaohotskaya@gmail.com, ORCID: 0000-0002-4810-2810

Національний технічний університет «Харківський політехнічний інститут» (Україна)

Роботу присвячено процесу дослідження та розробці власного підходу для розпізнавання об'єктів на зображеннях у випадках вибіркового пошуку. У сучасному світі у сфері комп'ютерного зору та обробки зображень, розпізнавання об'єктів є одним із найважливіших напрямків досліджень. Застосування нейронних мереж, таких як YOLO (You Only Look Once) та R-CNN (Region-based Convolutional Neural Network), виявилось дуже ефективним у вирішенні цієї задачі. Ці алгоритми здатні знаходити об'єкти на зображеннях та повертати обмежувальні рамки, які точно описують ці об'єкти. Однак, у деяких випадках, коли ми працюємо з інтерактивними програмами, наприклад, обираємо об'єкт, клацаючи або торкаючись розпізнаної області, виникає проблема вибору правильного об'єкта та його обмежувальної рамки. Це може вплинути на точність визначення обраного об'єкта в контексті вибіркового пошуку. Виникає необхідність знайти таку область пошуку, яка дозволить нам належним чином визначити обраний об'єкт, особливо у випадку перетину обмежувальних рамок. Ефективний підхід до визначення розміру області пошуку та візуалізація процесу дослідження можуть покращити точність і швидкість вибору об'єктів, забезпечуючи більш зручний та ефективний пошук об'єктів на зображеннях. Ми пропонуємо вирішення проблеми перетину обмежувальних рамок, що виникає при роботі нейронних мереж типу YOLO та R-CNN, розробивши метод оцінки оптимального розміру області пошуку, який дозволить знайти відповідний об'єкт та його обмежувальну рамку та пропонуємо узагальнений підхід до візуалізації процесу дослідження, що дозволить наочно представити перекриття обмежувальних рамок та полегшить вибір оптимального об'єкта. Для підтвердження ефективності запропонованого методу ми проводимо експерименти на відповідному наборі даних та порівнюємо їх. Результати таких досліджень можуть мати значний практичний вплив

на розробку систем розпізнавання об'єктів і покращення їх функціональності в цілому. Майбутні дослідження можуть фокусуватися на розширенні набору даних для випадків вибіркового пошуку, включаючи різні сценарії перекриття обмежувальних рамок та об'єктів з різною формою та розмірами.

Ключові слова: Дрони, вибіркового пошуку об'єктів, обмежувальні рамки, нейронні мережі, YOLO, R-CNN, алгоритм розпізнавання.

Постановка проблеми. У процесі пошуку об'єктів на зображеннях нейронні мережі типу YOLO [1] або R-CNN [2] повертають області у вигляді обмежувальних рамок об'єктів (рис. 1). Проблема полягає в тому, що обмежувальні рамки можуть накладатися одна на одну, а в деяких інтерактивних програмах, наприклад обираючи об'єкт, клацаючи або торкаючись розпізнаної області, нам потрібно знайти відповідний об'єкт і його послідовну обмежувальну рамку, яка найкраще підходить до місця клацання.



Рис. 1. Результат виявлення об'єктів за допомогою YOLO v4

Дано квадратне вікно W з розміром $s_w \times s_w$, набір знайдених обмежувальних прямокутників $B = \{b_1, \dots, b_N\}$, $b_i = \{x, y, w, h\}$, де x, y – координати лівої верхньої точки прямокутника з шириною w та висотою h , а також прямокутна область пошуку b_s з розміром $s_b \times s_b$, що центрується навколо точки $(\frac{s_w}{2}, \frac{s_w}{2})$. Процес вибіркового пошуку відбувається наступним чином: треба знайти таку підмножину $B' \in B$, щоб кожен $b_i \in B'$ перетинався з b_s :

$$b_i \cap b_s \neq \emptyset, \quad (1)$$

Тому нам потрібно знайти значення розміру області пошуку s_b , що дасть нам можливість належно визначити обраний об'єкт, особливо у випадку перекриття обмежувальних рамок.

Аналіз останніх досліджень і публікацій. Аналіз зображень земної поверхні – дуже актуальна тема, яка сьогодні стає все більш популярною. Сучасні дрони оснащені різноманітними датчиками, за допомогою яких можна збирати дані про земну поверхню, включаючи дані про рельєф, водні ресурси, ґрунт, рослинність тощо. Існує багато науково-технічних публікацій, які досліджують можливості використання дронів для аналізу земної поверхні. Вони охоплюють широкий спектр тем, включаючи методи збору даних, обробку інформації, проектування алгоритмів і багато іншого. У статті [3] автори наводять приклад використання дронів і фотограмметрії для збору та аналізу даних у будівельній галузі. В ній описано методи, інструменти та технології, що використовуються для створення 3D-моделей та картографічних продуктів. Наприклад, наступна стаття [4] представляє можливості використання дронів для картографування та моніторингу пасовищ. Вона описує методи, що використовуються для збору даних про рослинність і ґрунт, а також розробку алгоритмів для їх аналізу. В інших роботах [5] автори описують процес моніторингу, який включає фото- та відеозйомку з БПЛА, а також обробку даних за допомогою комп'ютерного зору та машинного навчання для виявлення хвороб і шкідників на рослинах.

Також розглядається використання дронів для збору та аналізу ландшафтних зображень у контексті прийняття рішень щодо землекористування. Автори статті зазначають, що використання безпілотників дає змогу отримувати якісні зображення та дані про земельні ресурси у більш швидкому та доступному форматі, що може покращити прийняття рішень у різних галузях, таких як сільське, лісове та інші середовища [5]. Дослідники розглядають можливість використання малих дронів для дистанційного зондування стихійних лих і пов'язаних з ними катастроф. Автори відзначають, що малі дрони надають нові можливості для збору даних і забезпечення доступу до віддалених або важкодоступних місць, що може покращити моніторинг і реагування на стихійні лиха [6]. Навіть виявлення пошкоджень залізниці тепер здійснюється за допомогою відеоданих [7]. У статті [8] представлено дослідження, проведене з використанням БПЛА для інспекції мостової інфраструктури. Автори статті описують процес зйомки, який включає зйомку фото і відео з БПЛА в різних ракурсах. Потім дані обробляються програмним забезпеченням для створення точної 3D-моделі інфраструктури мосту. Загалом, дослідження в області аналізу земної поверхні за допомогою дронів продовжують активно розвиватися і застосовуватися в різних галузях. Нові технології, алгоритми та методи обробки даних дозволяють отримувати все більш точну та детальну інформацію про земну поверхню, яка може бути корисною для прийняття рішень у різних галузях, включаючи сільське господарство, геологію, екологію тощо. На сьогоднішній день штучні нейронні мережі досягають надзвичайного успіху у вирішенні завдань, таких як розпізнавання об'єктів [9] або сегментація зображень [10], крім

того, вони можуть працювати на мобільних пристроях, таких як смартфони або планшетні ПК.

Формулювання цілей статті. Розробка узагальненого підходу до оцінки та візуалізації процесу дослідження алгоритмів розпізнавання об'єктів у випадках вибіркового пошуку.

Основна частина. У нашому підході ми використовуємо наступну схему, щоб дослідити вплив розміру області пошуку для належного виявлення об'єктів:

1. Ми отримуємо блок пікселів P під плаваючим вікном W з вхідного зображення.
2. Передаємо блок пікселів P в детектор об'єктів, щоб створити вихідний набір даних B .
3. Нехай $s_b = s_{start}$.
4. Тоді для кожного $b_i \in B$ ми перевіряємо умову (1), щоб отримати B' .
5. У разі перетину блоків, якщо $|B'| > 1$, де $|\cdot|$ є числом блоків у B' , тоді ми застосовуємо додатковий тест, щоб знайти блок, який має максимальну площу перетину з b_s .
6. Збільшуємо s_b та переходимо до кроку 4.
7. Етап оцінювання.

Як детектор об'єктів на кроці 2 ми використовуємо штучні нейронні мережі YOLOv4 та Faster R-CNN, але є можливість використовувати будь-які інші моделі виявлення об'єктів, що повертають на виході набір обмежувальних прямокутників.

Ми пропонуємо оцінити площу перетину $b_i \cap b_s$ як кількість пікселів перетину b_s з відповідним b_i :

$$b_i \in B | s_i > 0, \quad (2)$$

де s_i – кількість пікселів, що потрапили до перетину.

Після того, як ми знайдемо набір усіх обмежувальних рамок B' , які перетинаються з областю пошуку b_s , ми проводимо попарний тест між усіма парами областей у B' , щоб знайти найкраще підібраний обмежувальний прямокутник b_F з найбільшою площею перетину (2):

$$b_F = b_{i|i=\operatorname{argmax}(S)} \quad (3)$$

де $S = \{S_1, \dots, S_{N'_B}\}$ – набір площ областей у B' , N'_B – загальна кількість блоків, що перетинаються.

Щоб оцінити точність вибору обмежувальної рамки, ми використовуємо відношення кількості областей, які перетинають b_s заданого розміру відносно загальної кількості блоків, знайдених у P як кількісний показник:

$$R = \frac{|N_{B'}|}{|N_B|}, \quad (4)$$

За умови (4) шукаємо s_B , які дають мінімальне R . Для оцінки умови (4) ми провели серію експериментів.

Під час експериментів ми оцінюємо наступні параметри посилаючись на s_B :

- середнє значення N'_B ;
- співвідношення S_1/S_2 ;
- середнє значення R .

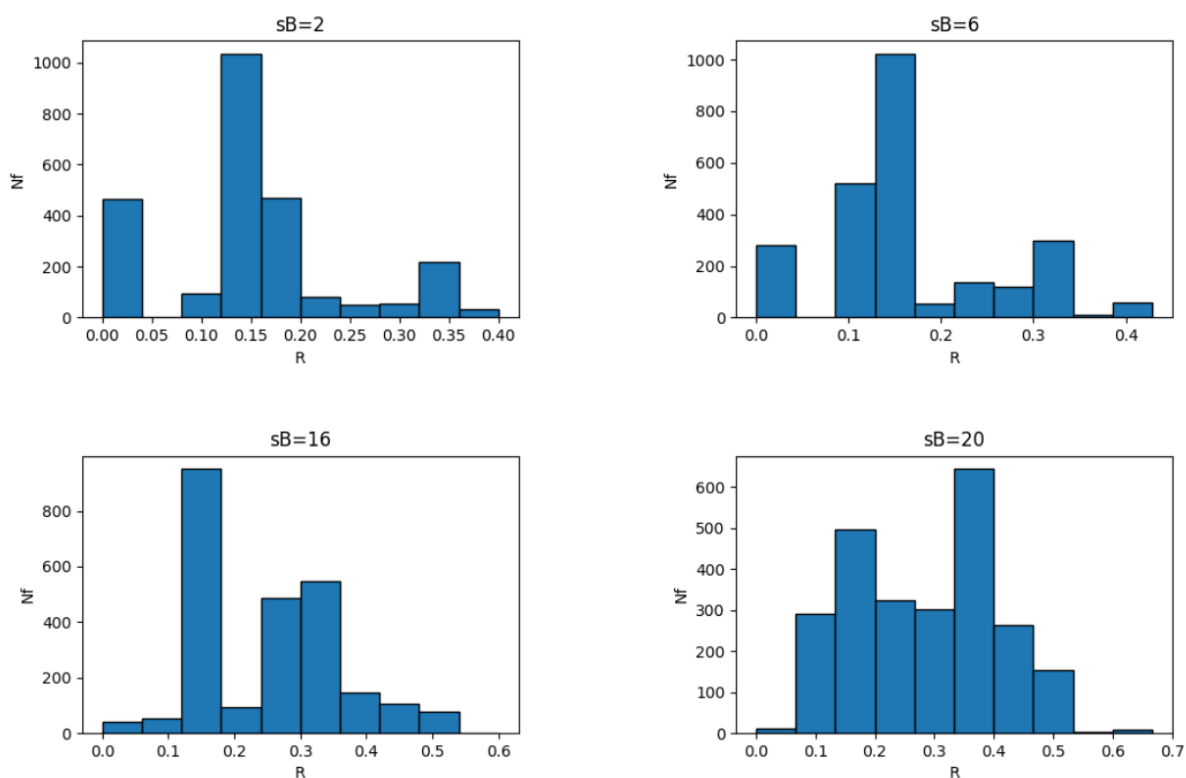
У таблиці 1 наведено параметри кожного з експериментів, загальна кількість використаних блоків з пікселями становить $N = 2500$, середня кількість обмежувальних прямокутників, знайдених в результаті експериментів, $N_B = 6.9928$.

Таблиця 1

Результати експериментів

№ експеримента	s_B	Середнє значення N'_B	S_1/S_2	Середнє значення R
1	2	0.9604	1.0342	0.1432
2	6	1.1492	1.4601	0.1703
3	10	1.3304	2.0067	0.1958
4	16	1.6724	2.2235	0.2460
5	20	1.8800	3.4423	0.2758

На рис. 2 наведено гістограми загальної кількості оброблених областей зображення зі знайденим R .

Рис. 2. Гістограми зі знайденим R

На рис. 3 показано діаграми розсіювання площ пар блоків, що перекриваються (точки позначено випадковими кольорами для візуалізації).

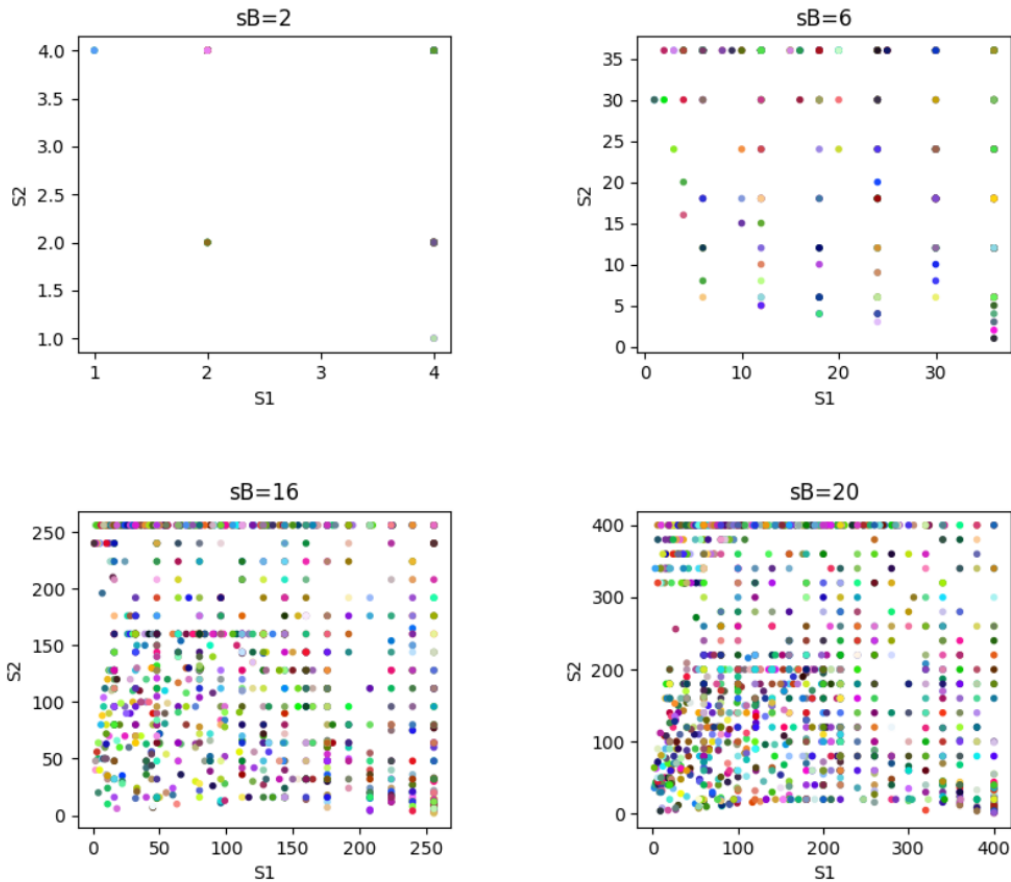


Рис. 3. Діаграми розсіювання S_1 та S_2

Деякі приклади областей зображення показані на рис. 4. Можна побачити, що збільшення s_b збільшує кількість блоків, що перетинаються, і коефіцієнт R , який повинен мати мінімальне (але більше нуля) значення у випадку найкращого розміру вікна пошуку s_b . Також слід зазначити, що при збільшенні значень S_1 та S_2 збільшується часова складність алгоритму, тому найкращі рішення були отримані з невеликими значеннями s_b .



Рис. 4. Приклади блоків, що перетинаються ($s_b = 20$)

Висновки. В ході виконання роботи було запропоновано підхід до оцінки та візуалізації процесу дослідження алгоритмів розпізнавання об'єктів у випадках вибіркового пошуку. Досліджено вплив параметрів запропонованого алгоритму на якість процесу пошуку. Встановлено, що для якісного вибору потрібного об'єкта на зображенні достатньо використовувати області пошуку невеликого розміру. Підхід реалізовано у вигляді програмного додатку мовою програмування Python.

Обмеження нашого підходу полягає в тому, що він не може автоматично знаходити параметри, і в майбутній роботі ми зосередимося на автоматичному оцінюванні параметрів підходу.

Література

1. Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y.M., 2021. Scaled-YOLOv4: Scaling Cross Stage Partial Network, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 13029–13038.
2. Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. 'Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation'. arXiv. <https://doi.org/10.48550/arXiv.1311.2524>
3. Asnafi, M., Dastgheibifard, S., 2018. A Review on Potential Applications of Unmanned Aerial Vehicle for Construction Industry. *Sustainable Structures and Materials, An International Journal* 1, 44–53. <https://doi.org/10.26392/SSM.2018.01.02.044>
4. Hafeez, A., Husain, M.A., Singh, S.P., Chauhan, Anurag, Khan, Mohd.T., Kumar, N., Chauhan, Abhishek, Soni, S.K., 2023. Implementation of drone technology for farm monitoring & pesticide spraying: A review. *Information Processing in Agriculture* 10, 192–203. <https://doi.org/10.1016/j.inpa.2022.02.002>
5. Kleinschroth, F., Banda, K., Zimba, H., Dondeyne, S., Nyambe, I., Spratley, S., Winton, R.S., 2022. Drone imagery to create a common understanding of landscapes. *Landscape and Urban Planning* 228, 104571. <https://doi.org/10.1016/j.landurbplan.2022.104571>
6. Kucharczyk, M., Hugenholtz, C.H., 2021. Remote sensing of natural hazard-related disasters with small drones: Global trends, biases, and research opportunities. *Remote Sensing of Environment* 264, 112577. <https://doi.org/10.1016/j.rse.2021.112577>
7. Mahmoud, A., Mohamed, M.G., El Shazly, A., 2022. Low-cost framework for 3D reconstruction and track detection of the railway network using video data. *The Egyptian Journal of Remote Sensing and Space Science* 25, 1001–1012. <https://doi.org/10.1016/j.ejrs.2022.11.001>
8. Mandirola, M., Casarotti, C., Peloso, S., Lanese, I., Brunesi, E., Senaldi, I., 2022. Use of UAS for damage inspection and assessment of bridge

- infrastructures. *International Journal of Disaster Risk Reduction* 72, 102824. <https://doi.org/10.1016/j.ijdr.2022.102824>
9. Liu, Chengxin, Kewei Wang, Hao Lu, Zhiguo Cao, and Ziming Zhang. 2022. ‘Robust Object Detection With Inaccurate Bounding Boxes’. arXiv. <https://doi.org/10.48550/arXiv.2207.09697>.
 10. Liu, Xiaolong, Zhidong Deng, and Yuhan Yang. 2019. ‘Recent Progress in Semantic Image Segmentation’. *Artificial Intelligence Review* 52 (2): 1089–1106. <https://doi.org/10.1007/s10462-018-9641-3>.

GENERALIZED APPROACH FOR OBJECT SELECTIVE SEARCH IN IMAGES

Vlasenko Vitalii, Dashkevych Andrii, Vorontsova Darya, Okhotska Olena

The work is devoted to the process of research and development of an own approach for object recognition in images in cases of selective search. In the modern world in the field of computer vision and image processing, object recognition is one of the most important areas of research. The use of neural networks, such as YOLO (You Only Look Once) and R-CNN (Region-based Convolutional Neural Network), has proven to be very effective in solving this problem. These algorithms are able to find objects in images and return bounding boxes that accurately describe those objects. However, in some cases when we work with interactive programs, for example, selecting an object by clicking or touching a recognized area, there is a problem of selecting the correct object and its bounding box. This can affect the accuracy of the selected object in the context of a selective search. It is necessary to find such a search area that will allow us to properly identify the selected object, especially in the case of crossing bounding boxes. An effective approach to the size of the search area and the visualization of the research process can improve the accuracy and speed of object selection, providing a more convenient and efficient object search in images. We offer a solution to the problem of crossing the bounding frames that arises in the operation of YOLO and R-CNN type neural networks by developing a method for estimating the optimal size of the search area, which will allow finding the appropriate object and its bounding frame, and offer a generalized approach to the visualization of the research process, which will allow to visually represent the overlap of the bounding frames and will facilitate the selection of the optimal object. To confirm the effectiveness of the proposed method, we conduct experiments on the appropriate data set and compare them. The results of such research can have a significant practical impact on the development of object recognition systems and improving their functionality as a whole. Future research may focus on expanding the dataset for selective search cases, including different bounding box overlap scenarios and objects with different shapes and sizes.

Keywords: Drones, selective object search, bounding boxes, neural networks, YOLO, R-CNN, recognition algorithm.

References

1. Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y.M. (2021) Scaled-YOLOv4: Scaling Cross Stage Partial Network, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 13029–13038.
2. Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. (2014) ‘Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation’. arXiv. <https://doi.org/10.48550/arXiv.1311.2524>.
3. Asnafi, M., Dastgheibifard, S. (2018). A Review on Potential Applications of Unmanned Aerial Vehicle for Construction Industry. *Sustainable Structures and Materials, An International Journal*, 1, 44–53. <https://doi.org/10.26392/SSM.2018.01.02.044>.
4. Hafeez, A., Husain, M.A., Singh, S.P., Chauhan, Anurag, Khan, Mohd.T., Kumar, N., Chauhan, Abhishek, Soni, S.K. (2023). Implementation of drone technology for farm monitoring & pesticide spraying: A review. *Information Processing in Agriculture*, 10, 192–203. <https://doi.org/10.1016/j.inpa.2022.02.002>.
5. Kleinschroth, F., Banda, K., Zimba, H., Dondeyne, S., Nyambe, I., Spratley, S., Winton, R.S. (2022) Drone imagery to create a common understanding of landscapes. *Landscape and Urban Planning* 228, 104571. <https://doi.org/10.1016/j.landurbplan.2022.104571>.
6. Kucharczyk, M., Hugenholtz, C.H., 2021. Remote sensing of natural hazard-related disasters with small drones: Global trends, biases, and research opportunities. *Remote Sensing of Environment* 264, 112577. <https://doi.org/10.1016/j.rse.2021.112577>.
7. Mahmoud, A., Mohamed, M.G., El Shazly, A., 2022. Low-cost framework for 3D reconstruction and track detection of the railway network using video data. *The Egyptian Journal of Remote Sensing and Space Science* 25, 1001–1012. <https://doi.org/10.1016/j.ejrs.2022.11.001>.
8. Mandirola, M., Casarotti, C., Peloso, S., Lanese, I., Brunesi, E., Senaldi, I., 2022. Use of UAS for damage inspection and assessment of bridge infrastructures. *International Journal of Disaster Risk Reduction* 72, 102824. <https://doi.org/10.1016/j.ijdrr.2022.102824>.
9. Liu, Chengxin, Kewei Wang, Hao Lu, Zhiguo Cao, and Ziming Zhang. 2022. Robust Object Detection With Inaccurate Bounding Boxes. arXiv. <https://doi.org/10.48550/arXiv.2207.09697>.
10. Liu, Xiaolong, Zhidong Deng, and Yuhan Yang. (2019) Recent Progress in Semantic Image Segmentation. *Artificial Intelligence Review* 52 (2): 1089–1106. <https://doi.org/10.1007/s10462-018-9641-3>.

