

УДК 004

DOI 10.31471/1993-9981-2024-1(52)-115-120

ОСОБЛИВОСТІ ПІДГОТОВКИ ДАТАСЕТІВ ДЛЯ НАВЧАННЯ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

І. С. Яковин¹, Р. Т. Красняк¹, А. А. Шкітов², Р. П. Томашівський¹

¹*Івано-Франківський національний технічний університет нафти і газу,
вул. Карпатська 15, м. Івано-Франківськ, 76019, Україна;
e-mail: igor.com.2009@gmail.com, roman.krasnyak@gmail.com*

²*Відкритий міжнародний університет розвитку людини «Україна»;
вул. Львівська 23, м. Київ, 03115, Україна; e-mail: office@uu.ua*

У статті проаналізовано ключові аспекти формування навчальних датасетів, включаючи зображення об'єктів, для підготовки штучних нейронних мереж, зокрема на платформі YOLO, яка широко використовується для задач розпізнавання. Предметом дослідження є процес ідентифікації та класифікації об'єктів, пов'язаних із дорожніми сценаріями, таких як трасові моделі авто, перепони та поверхня доріжки. У статті представлено підхід до створення двовимірних зображень із відповідною анотацією класів "моделі авто", "траса" та "перепони". Особливу увагу приділено методам аугментації даних, які дозволяють збільшити варіативність та покращити якість навчальних вибірок, що є критично важливим для ефективності навчання нейронних мереж. У дослідженні описано процес спеціалізації нейронної мережі на розпізнаванні об'єктів за специфічними характеристиками або в умовах з обмеженими ресурсами. Після навчання мережа використовується для ідентифікації та сегментації визначених об'єктів на зображеннях. Зібрані результати розпізнавання підлягають агрегуванню, що дозволяє інтегрувати інформацію з різних джерел, тим самим підвищуючи загальну точність системи. Запропоновані методи можуть слугувати основою для розробки та вдосконалення алгоритмів розпізнавання об'єктів у різних прикладних завданнях, таких як мобільна робототехніка, моніторинг та аналіз середовища. Актуальність роботи обумовлена потребою підвищення стабільності та точності роботи систем розпізнавання об'єктів у складних умовах, зокрема при наявності обмежених і низькоякісних навчальних даних. Отримані результати та підходи можуть бути застосовані для вирішення практичних завдань і підтримки подальших досліджень у цій галузі.

Ключові слова: розпізнавання об'єктів, нейронні мережі, датасети, обробка даних, анотація зображень.

The article analyzes key aspects of the formation of training datasets, including images of objects, for training artificial neural networks, in particular on the YOLO platform, which is widely used for recognition tasks. The subject of the study is the process of identifying and classifying objects associated with track scenarios, such as track robots, barriers and track surfaces. The article presents an approach to creating two-dimensional images with the appropriate annotation of the classes "robots", "track" and "barriers". Particular attention is paid to data augmentation methods that allow increasing the variability and improving the quality of training samples, which is critically important for the effectiveness of training neural networks. The study describes the process of specializing a neural network in recognizing objects by specific characteristics or in conditions with limited resources. After training, the network is used to identify and segment certain objects in images. The collected recognition results are subject to aggregation, which allows integrating information from different sources, thereby increasing the overall accuracy of the system. The proposed methods can serve as the basis for developing and improving object recognition algorithms in various applied tasks, such as mobile robotics, monitoring and environmental analysis. The relevance of the work is due to the need to increase the stability and accuracy of object recognition systems in difficult conditions, in particular in the presence of limited and low-quality training data. The obtained results and approaches can be applied to solve practical problems and support further research in this area.

Keywords: object recognition, neural networks, datasets, data processing, image annotation.

Вступ

Щоб моделі машинного навчання могли розпізнавати об'єкти категорій "automodel", "road" і "border", необхідно ретельно опрацьовувати процес створення навчальних датасетів, включаючи двовимірні зображення. Для цього

використовують фотографії та відео, які містять відповідні об'єкти, зняті за різних умов [1]. Важливим аспектом є забезпечення різноманітності даних: зображення повинні містити різні види автомоделей, елементи трас і перепон. Це дає змогу моделі пристосовуватися до змін

у параметрах освітлення, кутах огляду та навколишньому середовищі, що значно підвищує її ефективність у реальних сценаріях.

Теоретичне обґрунтування

У найпростішому випадку доцільно обмежити наявну варіативність об'єктів розпізнавання трьома класами.

Клас "automodel" включає широкий спектр конфігурацій і типів автомобільних моделей, що можуть відрізнитися за розмірами, формами та кольорами. Для забезпечення ефективного навчання моделі необхідно включити до датасетів зображення з різноманітними варіаціями цих об'єктів, враховуючи їхній контекст і різні кути огляду. Категорія "road" охоплює зображення трас і шляхів, які характеризуються різними типами покриття та умовами освітлення. Важливо, щоб модель навчилася розпізнавати ці об'єкти навіть у ситуаціях, коли доріжка є частково прихованою або має нечіткі межі.

Клас "border" охоплює різні види перепон, які можуть бути виготовлені з таких матеріалів, як метал, пластик або дерево, і мати різноманітні форми.

Таким чином, датасети формуються за трьома основними напрямками: "automodel", "road" і "border". Кожна категорія поділяється на підгрупи, які містять детальний опис особливостей об'єктів та умов їхнього розташування.

У свою чергу, підкатегорія "автомодель" включає зображення автомобілів, розміщених на трасі.



Рисунок 1 – Розділення класу "automodel" на підкатегорії

Наступна ключова категорія, "road", включає різноманітні види зображень трас.



Рисунок 2 – Розділення класу "road" на підкатегорії

Категорія "перепона" включає об'єкти траси, що слугують для обмеження траєкторії руху автомобільних моделей.



Рисунок 3 – Розділення класу "border" на підкатегорії

Запропонована структура гарантує точність у процесі ідентифікації та класифікації зображень в датасеті, що сприяє покращенню аналізу та обробки даних для завдань розпізнавання й класифікації. Для полегшення розуміння, цей процес можна представити у вигляді схеми, де кожен рівень деталізації демонструє підкатегорії та їх взаємозв'язки.

Методика експерименту

Анотація є одним з основних етапів підготовки датасетів. Для кожного класу об'єктів необхідно точно позначити їх місце розташування на зображеннях, використовуючи рамки (bounding box),

полігональні контури або точкові маски в залежності від вимог задачі. Наприклад, для класу "automodel" доцільно застосовувати полігональні розмітки, оскільки це дозволяє точніше відобразити складні контури автомобільних моделей, тоді як для класів track та barrier можуть бути достатніми рамки, що визначають межі об'єктів.

Коли датасет обмежений у розмірі або має низьку якість, важливо використовувати техніки аугментації для поліпшення навчання моделі [1]. Аугментація дозволяє штучно збільшити кількість даних, застосовуючи різноманітні трансформації зображень, такі як зміна кута зйомки, коригування яскравості та контрасту, а також застосування інших геометричних і кольорових модифікацій. Для класів "automodel", "road" і "border" це може включати варіації в умовах та якості освітлення, додавання, спотворень, шумів або зміну фону.

Ще одним підходом, який застосовується для малих датасетів, є перенесене навчання [2,3]. Моделі, попередньо навчені на великих датасетах, таких як ImageNet, вже володіють загальними знаннями про ознаки об'єктів, що дозволяє адаптувати їх для конкретних задач, таких як розпізнавання автомобільних моделей, трас або перешкод. Цей метод значно зменшує потребу в нових даних і пришвидшує процес навчання.

Для поліпшення якості підготовки даних, зображення було завантажено на платформу RoboFlow [4], що значно полегшує їх обробку для завдань машинного навчання. Після цього було виконано детальну анотацію об'єктів на кожному зображенні, включаючи моделі авто, перешкоди та елементи дороги. Анотація полягала в точному визначенні меж (bounding boxes) для кожного об'єкта, приклад чого представлений на рис. 4, а також у присвоєнні відповідних класів.

З врахуванням ситуації, коли початкові дані були обмеженими і мали низьку якість, було ухвалено рішення застосувати

методи аугментації для збільшення кількості та різноманітності зображень у датасеті. Аугментація включала трансформації зображень, такі як обертання, масштабування, горизонтальні та вертикальні відображення, зсуви, а також зміну параметрів яскравості, контрасту та насиченості [5,6]. Ці методи дозволили нейронній мережі стати більш стійкою до змін у вхідних даних і покращити її здатність до узагальнення на обмеженому наборі прикладів. В результаті аугментації розмір датасету значно зріс, що позитивно вплинуло на ефективність навчання мережі.



Рисунок 4 – Приклад анотованого зображення

Після завершення анотації та застосування методів аугментації, проведено розділення датасету на три окремі фрагменти: 88% для тренувального набору, 8% для валідаційного та 4% для тестового. Така структура розподілу була обрана для максимального використання наявних даних для навчання мережі, водночас забезпечуючи достатню кількість даних для перевірки та тестування. Тренувальна частина використовувалася для безпосереднього навчання моделі, що дозволяло їй виявляти шаблони та характеристики об'єктів. Валідаційна частина слугувала для контролю процесу навчання та налаштування гіперпараметрів, що допомагало уникнути перенавчання та покращити здатність моделі до узагальнення. Тестова частина використовувалася для остаточної оцінки

ефективності моделі на нових даних, що дало змогу отримати точні показники її продуктивності.

Інформацію про розподіл даних, а також характеристики попередньої обробки та аугментації для створеного датасету можна побачити на рис. 5.

```

Augmentations Outputs per training example: 3
Crop: 0% Minimum Zoom, 20% Maximum Zoom
Rotation: Between -15° and +15°
Shear: ±10° Horizontal, ±10° Vertical
Grayscale: Apply to 15% of images
Saturation: Between -25% and +25%
Brightness: Between -15% and +15%
Blur: Up to 2.5px
Noise: Up to 0.1% of pixels
  
```

Рисунок 5 – Методи трансформації, застосовані до датасету

Для перевірки ефективності використаного датасету та оцінки результатів навчання нейронної мережі задіяно її тренування засобами платформи RoboFlow. Після завершення навчання були отримані статистичні показники, які відображають точність і якість нейронної мережі, що дозволило оцінити доцільність застосування датасету для вирішення поставленої задачі.

Зокрема, на рисунку 6 представлено графік, що демонструє зміни метрики середньої точності (mean Average Precision, mAP) в процесі навчання за послідовними епохами. Вісь абсцис показує кількість епох, на яких мережа проходила тренування на основному наборі даних, а вісь ординат відображає значення mAP, яке є оцінкою ефективності нейронної мережі щодо виявлення та класифікації об'єктів на основі отриманих двомірних зображень.

Графік відображає дві основні криві. Темно-фіолетова крива (mAP@50) демонструє швидке зростання середньої точності на перших 50 епохах, після чого вона стабільно утримується в межах 0.9–1.0, що свідчить про високу результативність мережі.

Світло-фіолетова крива (mAP@50:95) має менші значення, але також досягає

стабільності близько 0.7, що є позитивним показником.

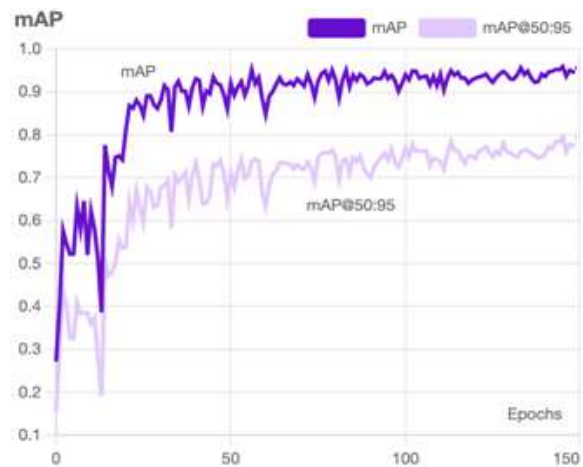


Рисунок 6 – Залежність середньої точності розпізнавання від тривалості навчання

Загальна картина вказує на те, що обидві метрики стабілізуються на високому рівні після початкового зростання. Значення функції втрат для об'єктних рамок (Box Loss) під час тренування для виявлення об'єктів наведено на рисунку 7.

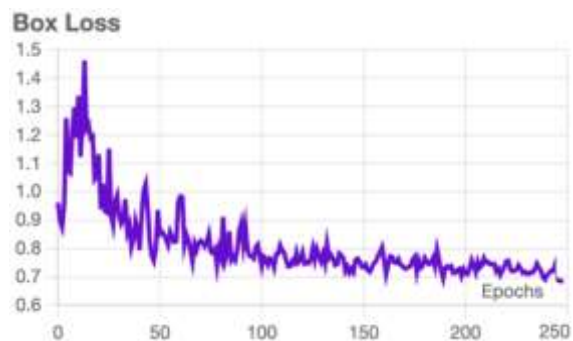


Рисунок 7 – Значення функції втрат для рамок об'єктів (Box Loss) при тренуванні мережі

На графіку вісь абсцис відповідає за кількість епох навчання, в той час як вісь ординат відображає Box Loss, який оцінює похибки у прогнозуванні координат обмежувальних рамок; чим менше значення, тим вища точність моделі. У перші 50 епох значення Box Loss знижується з 1.5 до 0.9–1.0, що свідчить про помітне покращення. Після 50 епох цей показник стабілізується та поступово знижується до рівня 0.7–0.8. Коли досягається 200 епох, коливання

практично зникають, і значення Box Loss стабільно тримаються на рівні 0.7, що вказує на зрілість процесу навчання. Графік свідчить про успішний тренінг, а різке зниження і стабілізація після 100 епох підтверджують, що мережа здатна точно передбачати рамки об'єктів.

На рисунку 8 наведено графік втрат класифікації (Class Loss) під час навчання.

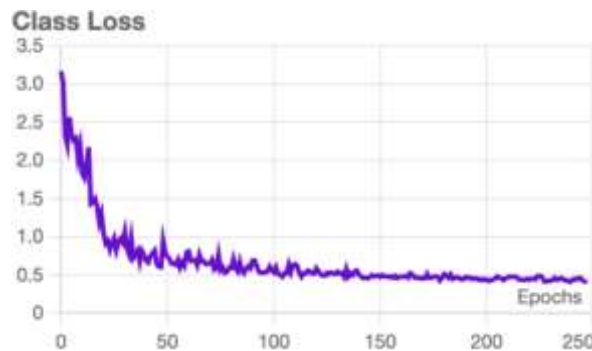


Рисунок 8 – Значення функції втрат для класифікації об'єктів (Class Loss) при тренуванні мережі

Показник Class Loss відображає ефективність процесу класифікації об'єктів на різних етапах навчання: чим менше значення, тим вища точність класифікації. Як і у попередньому графіку, вісь X демонструє кількість епох навчання, а вісь Y — рівень втрат класифікації. На початковому етапі навчання значення Class Loss починає з високих показників (приблизно 3.0), що є типовим для початку тренування.

Протягом перших 50 епох цей показник швидко знижується, досягнувши значення нижче 1.0, що вказує на суттєве покращення здатності мережі до класифікації.

Після 50 епох спостерігається стабільність, коли значення Class Loss стабільно коливається близько 0.5 і не змінюється помітно після 100 епох. Швидке зниження втрат у початковий період (0–50 епох) свідчить про ефективне освоєння класифікації об'єктів з початковими втратами 3.0, які зменшуються до приблизно 0.5. Стабілізація після 100 епох підтверджує, що мережа досягла стабільних результатів і ефективно класифікує об'єкти.

На рисунку 9 наведено графіки, що показують зміни в метриках точності (Precision) та повноти (Recall) протягом 300 епох навчання мережі. Обидві ці метрики є важливими для оцінки продуктивності мереж у задачах класифікації та виявлення об'єктів.

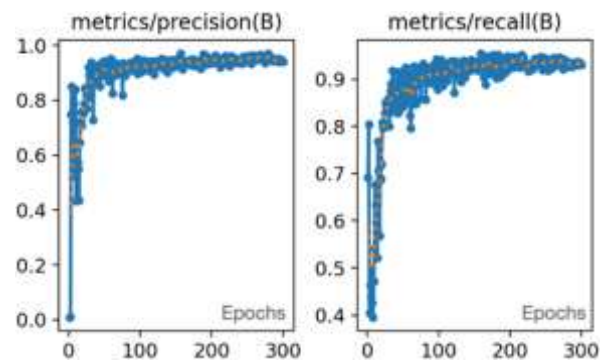


Рисунок 9 – Значення точності (Precision) та повноти (Recall) при тренуванні мережі

На першому графіку, як можна побачити, відображено зміни точності мережі: на осі абсцис — кількість епох, що характеризує процес навчання, а на осі ординат — рівень точності, який розраховується як частка правильно передбачених позитивних класів серед усіх передбачених позитивних класів.

Спочатку точність низька, але швидко підвищується, досягнувши значення близько 0.9 після перших 50 епох, і залишається стабільною в діапазоні 0.9–1.0 до завершення навчання.

Синя лінія з точками представляє фактичні показники точності на кожній епосі, тоді як помаранчева пунктирна лінія візуалізує згладжене значення для відображення загальної тенденції.

На другому графіку, як можна побачити, демонструється зміна повноти мережі: на осі X — кількість епох, а на осі Y — значення повноти, яке визначається як частка правильно класифікованих позитивних класів серед усіх реальних позитивних класів.

Як і в першому графіку, на початкових етапах повнота низька, але швидко зростає протягом перших 50 епох. Після 100 епох значення Recall стабілізується, коливаючись близько 0.9, що свідчить про

здатність мережі правильно класифікувати більшість позитивних класів.

Синя лінія з точками відображає фактичні показники, а помаранчева пунктирна лінія відображає згладжене значення для візуалізації загальної тенденції.

Висновки

Згідно з проведеним аналізом, мережа відображає значне покращення в обох метриках на початкових етапах тренування, швидко освоюючи класифікацію об'єктів і зменшуючи кількість помилкових прогнозів. Після 100 епох точність і повнота досягають значень близько 0.9–1.0, що свідчить про стабільну і ефективну роботу моделі. Малі флуктуації можуть бути пов'язані з особливостями складності даних або з більш складними зображеннями для класифікації.

У процесі підготовки датасету було важливо звернути увагу на якість даних. Початково зображення мали низьку якість, але завдяки детальній анотації та аугментації вдалося створити більш збалансований і різноманітний датасет.

Також важливо стежити за рівномірним розподілом класів об'єктів серед всіх частин датасету, щоб уникнути зміщення моделі на конкретні класи.

Список використаних джерел / References

1. Nanthini K., Sivabalaselvamani D. and oth. A Survey on Data Augmentation. 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023. pp. 913-920. DOI: [10.1109/ICCMC56507.2023.10084010](https://doi.org/10.1109/ICCMC56507.2023.10084010).

2. Rogers A. W. and oth. A transfer learning approach for predictive modeling of bioprocesses using small data. *Biotechnology and Bioengineering*. 2021. Vol. 119, № 2. P. 411–422. DOI: [10.1002/bit.27980](https://doi.org/10.1002/bit.27980)

3. Gürsakal N., Çelik S., Birişçi E. Synthetic Data Generation with Python. *Synthetic Data for Deep Learning*.

Berkeley, CA, 2022. С. 159–214. DOI: [10.1007/978-1-4842-8587-9_5](https://doi.org/10.1007/978-1-4842-8587-9_5)

4. Roboflow: Computer vision tools for developers and enterprises. *Roboflow: Computer vision tools for developers and enterprises*. URL: <https://roboflow.com/>

5. Huang L. Normalization Techniques in Deep Learning. Cham: Springer International Publishing, 2022. DOI: [10.1007/978-3-031-14595-7](https://doi.org/10.1007/978-3-031-14595-7)

6. A Review on Imbalanced Data Handling Using Undersampling and Oversampling Technique. *International Journal of Recent Trends in Engineering and Research*. 2017. Т. 3, № 4. С. 444–449. DOI: [10.23883/ijrter.2017.3168.0uwxm](https://doi.org/10.23883/ijrter.2017.3168.0uwxm)

DATASETS PREPARATION FEATURES FOR TRAINING ARTIFICIAL NEURAL NETWORKS

I. S. Yakovyn¹, R. T. Krasnyak¹,
A. A. Shkitov², R. P. Tomashivskiy¹

¹Ivano-Frankivsk National Technical University of Oil and Gas, 15 Karpatska Str., Ivano-Frankivsk, 76019, Ukraine; e-mail: igor.com.2009@gmail.com, roman.krasnyak@gmail.com

²Open International University of Human Development "Ukraine"; 23 Lvivska St., Kyiv, 03115, Ukraine; e-mail: office@uu.ua