

**ЯКОВИН ІГОР**

Івано-Франківський національний технічний університет нафти і газу

<https://orcid.org/0009-0003-5844-8246>e-mail: [igor.com.2009@gmail.com](mailto:igor.com.2009@gmail.com)**МАЛІНОВСЬКИЙ РУСЛАН**

Івано-Франківський національний технічний університет нафти і газу

<https://orcid.org/0009-0002-3530-2271>e-mail: [ruslan1997malinovskyi@gmail.com](mailto:ruslan1997malinovskyi@gmail.com)**ЛАЦИК НАЗАРІЙ**

Івано-Франківський національний технічний університет нафти і газу

<https://orcid.org/0009-0004-4465-250X>e-mail: [nazarii.latsyk-a123-23@nung.edu.ua](mailto:nazarii.latsyk-a123-23@nung.edu.ua)**ВАТУЛЯК ТИМОФІЙ**

Івано-Франківський національний технічний університет нафти і газу

<https://orcid.org/0009-0003-8666-9484>-mail: [tymofiiivatuliak@gmail.com](mailto:tymofiiivatuliak@gmail.com)

## ОСОБЛИВОСТІ НЕЗАЛЕЖНОГО НАВЧАННЯ НЕЙРОМЕРЕЖ НА БАЗІ ПЛАТФОРМИ YOLO V8N

Імплементція нейромережесвих компонентів для вирішення спеціалізованих задач в межах невизначеності вхідних даних стає перспективним напрямком розвитку цифрових технологій. Одним з проблемних аспектів є обмеженість даних для організації навчання, що зумовлює необхідність маніпуляції даними, зокрема використання різних послідовностей даних наявного датасету. Такий підхід може сприяти кращому узагальненню, оскільки мережа отримує більше варіацій комбінацій даних. Це дозволяє моделі: навчитися працювати з варіативністю входів а також уникати перенавчання на статичному порядку даних.

Доцільно зазначити, що упередженість у даних порівняно часто виникає, якщо порядок входів залишається статичним. Якщо ж використати різні послідовності, то може зменшити ризик моделі залежати від шаблонів, що притаманні певній структурі даних а також сприяє формуванню більш стійких представлень інформації. У реальних сценаріях дані часто надходять неідеально впорядкованими або містять спотворення. В такій ситуації форкання різних послідовностей допомагає моделі бути більш стійкою до варіативності та менш залежною від специфіки ідеальних вхідних даних.

Фактично штучне створення різних навчальних послідовностей на основі одного датасету дозволяє розширити навчальну вибірку, не збільшуючи її реального розміру а також застосовувати техніки аугментації даних для задач, де розмір датасету обмежений.

Крім того, однією з проблем окремих нейронних мереж є надмірна впевненість у хибних результатах. Одним з варіантів вирішення згаданої проблеми може бути використання агрегації результатів функціонування кількох моделей, що дозволяє згладжувати згадані ситуації, оскільки декілька моделей зазвичай не роблять одну й ту ж помилку одночасно. Система розпізнавання, що реалізується на основі кількох нейронних мереж із методами агрегації буде більш адаптивною до нових типів даних або задач. Тож вона матиме кращу масштабованість, оскільки якщо додати нову модель, що спеціалізується на вузькій підзадачі, її результати буде нескладно інтегрувати у вже існуючу систему через механізм агрегації.

Ключові слова: нейронні мережі, розпізнавання, датасет, обробка даних, довірчий інтервал.

**YAKOVYN IGOR****MALINOVSKYI RUSLAN****LATSIK NAZARIY****VATULIAK TYMOFIY**

Ivano-Frankivsk National Technical University of Oil and Gas

## FEATURES OF INDEPENDENT LEARNING OF NEURAL NETWORKS BASED ON THE YOLO V8N PLATFORM

The implementation of neural network components for solving specialized tasks under conditions of input data uncertainty is becoming a promising direction in the development of digital technologies. One of the key challenges is the limitation of data available for training, necessitating data manipulation, particularly the use of different sequences of the existing dataset. This approach can contribute to better generalization, as the network is exposed to a greater variety of data combinations. As a result, the model learns to handle input variability and avoids overfitting to a static data order.

It is worth noting that data bias often arises when the input sequence remains static. Utilizing different sequences can reduce the risk of the model becoming dependent on patterns inherent in a specific data structure, thereby promoting the formation of more robust data representations. In real-world scenarios, data is often imperfectly ordered or contains distortions. In such cases, varying the input sequences helps the model become more resilient to variability and less dependent on ideal input data structures.

Artificially generating different training sequences from a single dataset expands the training sample without increasing its actual size. Additionally, data augmentation techniques can be applied to tasks with limited dataset sizes.

Another challenge in neural networks is their excessive confidence in incorrect predictions. One potential solution is the aggregation of multiple model outputs, which helps mitigate this issue, as different models rarely make the same mistake simultaneously. A recognition system based on multiple neural networks with aggregation methods becomes more adaptable to new data types and tasks. Consequently, such a system offers improved scalability, as adding a new model specializing in a narrow subtask allows for seamless integration into the existing system through the aggregation mechanism.

Keywords: neural networks, recognition, dataset, data processing, confidence interval.

## Вступ

У випадках обмеженого за обсягом чи зі значними спотвореннями датасету, що зустрічаються порівняно часто, доцільно використовувати методів аугментації, що є важливим інструментом для покращення результатів навчання моделі [1]. Для задач, де контекст визначає значення, навчання на різних послідовностях дозволяє нейронній мережі покращити розуміння локального та глобального контекстів, розширити можливості у випадках, коли важливі як попередні, так і наступні дані.

Використання методів аугментації даних може дозволити суттєво збільшити розмір датасету за рахунок застосування різноманітних трансформацій зображень, таких як зміна кутів огляду, регулювання яскравості чи інших кольорних або геометричних змін, а також зміна послідовності подачі даних датасету. Для обраних класів "automodel", "road" і "border" це може включати зміну освітлення, адитивне накладання шуму або використання різних фонових імплементацій.

Одним із дієвих підходів для роботи з невеликими датасетами може бути використання перенесеного навчання [1,2]. Заздалегідь натреновані моделі на великих наборах даних, зокрема таких як ImageNet, вже містять знання про основні характеристики об'єктів. Це дає змогу їх нескладної адаптації до нішових задач, зокрема розпізнавання об'єктів роботів, треків і бар'єрів. Такий підхід значно знижує потребу в нових даних та може скоротити час навчання мережі.

Коли збір реальних даних є ускладненим, вдаються до використання синтетичних даних [3]. Їх генерація (формування датасетів) за допомогою 3D-моделювання або генеративно-змагальних мереж (GANs) дозволяє розширювати набори додаткових прикладів для навчання, змінюючи характеристики об'єктів а також фонові умови. Такий підхід найбільш доцільний для класів "automodel", "road" і "border", коли обсяг реальних даних обмежений.

Зображення низької якості або з високим рівнем зашумлення потребують попередньої обробки, що включає їх нормалізацію [4]. Також ця процедура може включати методи шумозаглушення, покращення контрастності чи використання алгоритмів суперрезолюції для підвищення деталізації, що в свою чергу може покращити точність моделі.

Балансування даних між класами також відіграє важливу роль. Якщо один клас домінує за кількістю зображень, це може призвести до упередженості моделі, знижуючи загальну ефективність. Для вирішення задач такого плану використовують методи oversampling і undersampling, що забезпечують рівномірний розподіл даних між класами [5], що як і в попередньому випадку актуально для класів "automodel", "road" і "border".

Іншим способом збільшення обсягу анотованих даних є техніка псевдо-міток [6]. Цей метод передбачає автоматичну анотацію оригінальних зображень за допомогою використання попередньо натренованої моделі, що дозволяє отримати більше прикладів для тренування, зокрема для розмитих/нечітких зображень.

З огляду на обмежену кількість та якість вихідних експериментальних даних, використано аугментацію для збільшення обсягу і різноманітності датасету. Проведені трансформації передбачали обертання під різними кутами, зміну масштабу, відображення, зсуви та корекцію параметрів: яскравість, контрастність, насиченість. Крім того, при навчання кожної моделі послідовність прикладів була різною. Такий підхід дав змогу моделі стати стійкішою до змін у вхідних даних і краще узагальнювати знання з обмеженого набору прикладів. У наслідок аугментації розмір датасету став більшим і різним для кожної моделі.

Після завершення анотації та аугментації датасет поділено на три частини: тренувальну, валідаційну та тестову. Пропорції розподілу датасету: 84.0% - тренувальний, 8.0% - валідаційний та 8.0% - перевірочний/тестовий.

Тренувальний набір слугував для навчання мережі, дозволяючи їй виявляти характерні особливості об'єктів. Валідаційний набір використовувався для контролю за навчанням і налаштування гіперпараметрів, що сприяло зменшенню ризику перенавчання і підвищенню агрегаційних властивостей моделі.

Тестовий набір забезпечував фінальну оцінку продуктивності моделі на нових даних, що дозволило отримати емпіричні показники її ефективності.

Такий підхід дозволяє ефективно використати доступні дані, забезпечивши достатній обсяг для навчання, валідації та подальшої перевірки/тестування.

### Навчання нейронних мереж на одному датасеті з різною послідовністю їх подачі

Після створення набору даних наступним етапом є тренування нейронних мереж. У рамках натурного експерименту використано три окремі моделі, побудовані на основі архітектури YOLOv8n(nano) [7,8]. Після тренування кожної з таких мереж проведено аналіз із використанням нормалізованої матриці похибок, міри F1 (Confidence Curve) та довірчого інтервалу.

Процес навчання здійснювався за допомогою стандартної команди yolodetect:

```
yolo detect train
data=robots.yaml model=yolov8n3.pt
epochs=110 imgsz=640 device=mps
```

На рисунку 1а зображено нормалізовану матрицю похибок, яка ілюструє роботу моделі при класифікації даних на чотири категорії: "automodel", "road", "border" та "background". Уздовж

горизонтальної осі показано реальні мітки класів (True Labels), тоді як вертикальна вісь відображає передбачувані моделі мітки (Predicted Labels).

Модель виявила порівняно високу ефективність у розпізнаванні класу "border", правильно класифікуючи 92.0% прикладів, хоча 7.0% помилково віднесено до класу "background". Щодо класу "automodel", точність досягла 90.0%, із 10.0% зразків, помилково визначених як "background". Найвищу продуктивність модель продемонструвала для класу "road", де 95.0% об'єктів були коректно класифіковані, а лише 5.0% потрапили до "background" через помилки.

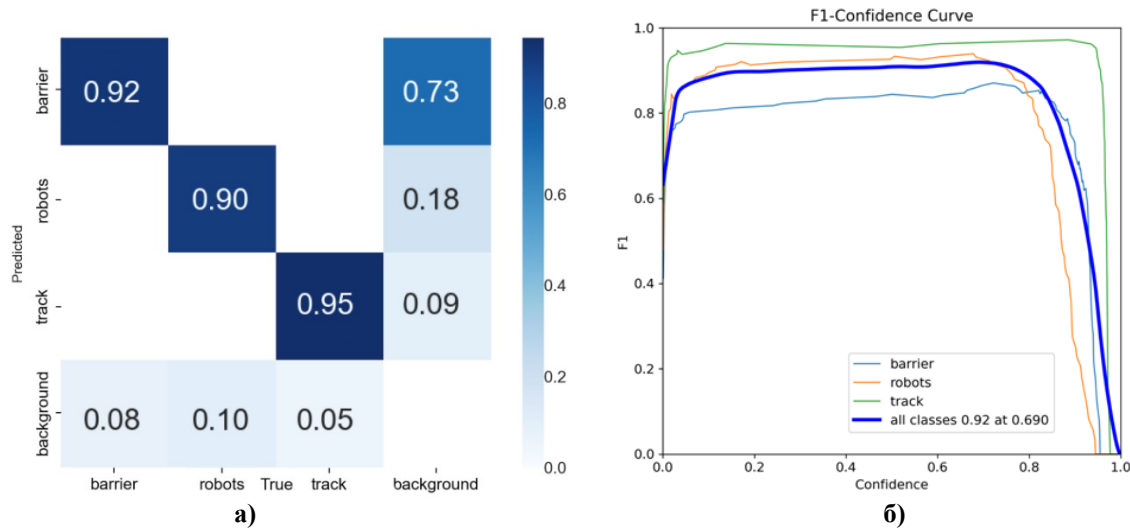


Рис. 1. Результати контролю навчання першої нейромережі: а) нормалізована матриця похибок, б) оцінки міри F1 та довірного інтервалу

На рисунку 1б представлені графіки, що показують залежність міри F1 від рівня певності моделі (Confidence Curve) для трьох категорій: "automodel", "road" і "border", а також для всіх категорій разом.

Вісь абсцис відображає рівень впевненості моделі в класифікації, який змінюється від 0 до 1, де більш високі значення вказують на більшу впевненість у передбаченнях мережі. Вісь ординат показує значення міри F1, яка являє собою гармонійне середнє між точністю та повнотою, варіюючись від 0 до 1, де 1 вказує на максимальні значення обох цих показників.

Як можна побачити, для класу "border" динаміка міри F1 знаходиться між показниками класів "automodel" та "road". Міра F1 для "border" поступово збільшується до рівня впевненості 0.7, після чого повільно знижується, досягаючи приблизно 0.8 на максимальних рівнях певності. Для класу "automodel" результати є дещо гіршими: високі значення міри F1 спостерігаються лише при рівнях впевненості понад 0.4, а після досягнення 0.7 міра F1 швидко падає.

Навчання другої нейронної мережі виконувалося за допомогою типової команди:

```
yolo detect train
data=robots.yaml model=yolov8n3.pt
epochs=120 imgsz=640 device=mps
```

На рисунку 2а проілюстровано нормалізовану матрицю похибок для цієї моделі. З неї можна побачити, що мережа досягає високої точності у виявленні класу "border": 95.0% зразків коректно ідентифіковано, а решта 5.0% помилково віднесено до "background". Це може свідчити про наявність спільних ознак між бар'єрами та елементами фону.

Водночас для класу "automodel" модель правильно класифікує лише 54.0% прикладів; ще 36.0% зразків хибно віднесено до класу "road", а 10.0% – до "background".

Такий розподіл може бути спричинений частковим збігом візуальних характеристик між роботами та треком (наприклад, схожістю форм чи кольорів), а також впливом шуму або обмеженою якістю навчальних даних.

У випадку класу "road" точність сягає 98.0%, і лише 2.0% зображень опиняються у "background" через помилки. Це демонструє чітку відмінність між треком та іншими класами й вказує на здатність моделі ефективно розпізнавати "road".

На рисунку 2б наведено взаємозв'язок між мірою F1 (гармонійним середнім точності та повноти) і порогом певності моделі.

По вертикальній осі відкладено значення міри F1 у межах від 0 до 1, де 1 вказує на ідеальне співвідношення точності та повноти, а 0 – на найнижчі показники. Горизонтальна вісь відображає поріг певності (Confidence Threshold), від 0 до 1, що визначає рівень довіри моделі до власних прогнозів. Для класу "border" F1-міра лишається стабільно високою (близько 0.9) у широкому діапазоні порогів від 0.2 до 0.9.

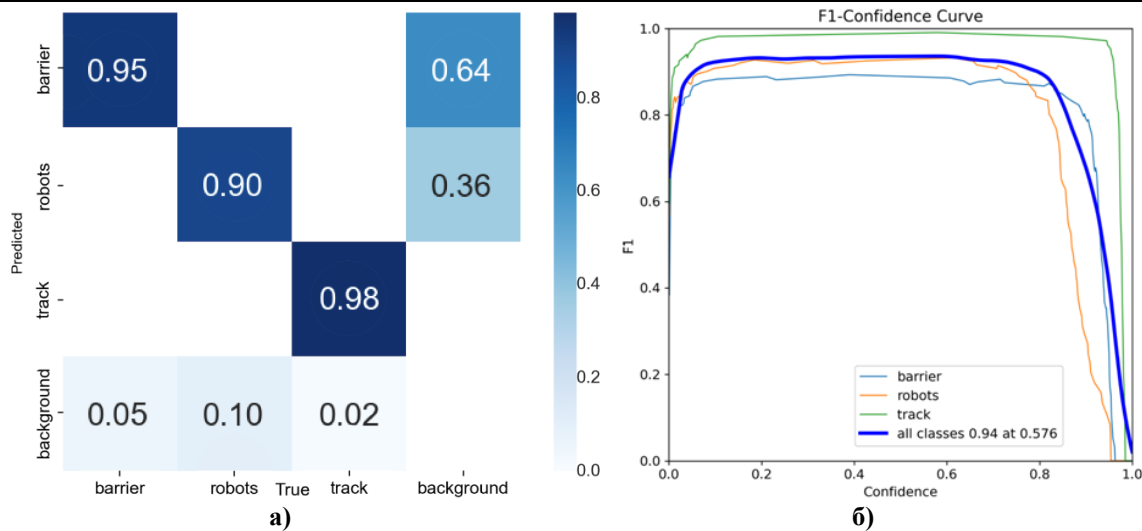


Рис. 2. Результати контролю навчання другої нейронної мережі: а) нормалізована матриця помилок, б) оцінки міри F1 та довірчого інтервалу

Відчутне зниження міри F1 для цього класу спостерігається лише за дуже високих порогів певності (близько 0.95 і вище). Така поведінка вказує на те, що мережа загалом добре визначає цей клас, утримуючи баланс між точністю та повнотою навіть при високих рівнях упевненості. Значна частина зразків класифікується коректно, що свідчить про якісне відображення характерних ознак у навчальному наборі.

Щодо класу "automodel", максимальне значення міри F1 (приблизно 0.9) досягається при порогах певності в діапазоні 0.1–0.4. Проте з ростом порогу певності міра F1 поступово знижується і стає майже нульовою при значеннях понад 0.8.

Це може означати, що мережа краще розпізнає цей клас за менш жорстких критеріїв певності, тоді як підвищення порогу призводить до плутанини з іншими класами.

Для класу "road" міра F1 залишається дуже стабільною та високою (понад 0.95) у широкому інтервалі порогів від 0.1 до 0.95. Різке падіння цього показника відбувається лише за порогів упевненості, що перевищують 0.95.

Така картина свідчить про здатність мережі відмінно розпізнавати характерні риси класу "road" та водночас підтримувати гармонійне співвідношення точності й повноти.

Навчання третьої нейронної мережі проводилося за допомогою такої типової команди:

```
yolo detect train
data=robots.yaml model=yolov8n3.pt
epochs=130 imgsz=640 device=mps
```

На рисунку 3а проілюстровано нормалізовану матрицю помилок для цієї моделі. Аналіз показує, що мережа коректно класифікує 92.0% зразків класу "border", демонструючи високий рівень точності. Водночас 8.0% прикладів віднесено до "background" помилково, що може свідчити про схожість певних рис між бар'єрами та фоном.

У випадку класу "automodel" мережа правильно класифікує 77.0% зразків. Проте 11.0% було хибно віднесено до "background", а 12.0% – до "road".

Така ситуація може пояснюватися схожими властивостями об'єктів класів "automodel" і "road" або ж наявністю шуму та неповної сегментації в навчальних даних, що призводить до неправильної класифікації.

Клас "road" виявляється найкраще розпізнаваним: 96.0% зразків ідентифіковано безпомилково, а лише 4.0% віднесено до "background". Цей результат зрештою може свідчити про чітку вираженість характеристик класу "road" і здатність моделі розпізнавати їх із високою точністю.

Натомість клас "background", як можна побачити, демонструє помітний рівень плутанини з "border" та "automodel". Однією з причин цього може бути обумовлено наявністю фонового оточення, що нагадує риси інших об'єктів, а також неоднозначністю або неповнотою сегментації фонів під час навчання.

На рисунку 3б подано криві залежності міри F1 від рівня певності моделі для кожного класу. Для класу "border" міра F1 стартує приблизно з 0.6 на найнижчих порогах (порядку 0.1) і досягає максимуму (приблизно 0.85–0.9) у діапазоні порогів 0.2–0.8. Після перевищення порогу 0.8 спостерігається різке падіння F1-міри, яка на рівні впевненості 1.0 стає мінімальною.

Така поведінка свідчить, що мережа якісно класифікує цей клас на середніх значеннях порога, але при надто високих рівнях впевненості зростає кількість пропущених зразків.

Для класу "automodel" початкова F1-міра становить близько 0.5, після чого вона зростає до приблизно 0.85 у діапазоні порогів від 0.2 до 0.4. Далі показник різко знижується і майже досягає нуля при значеннях порога вище за 0.9.

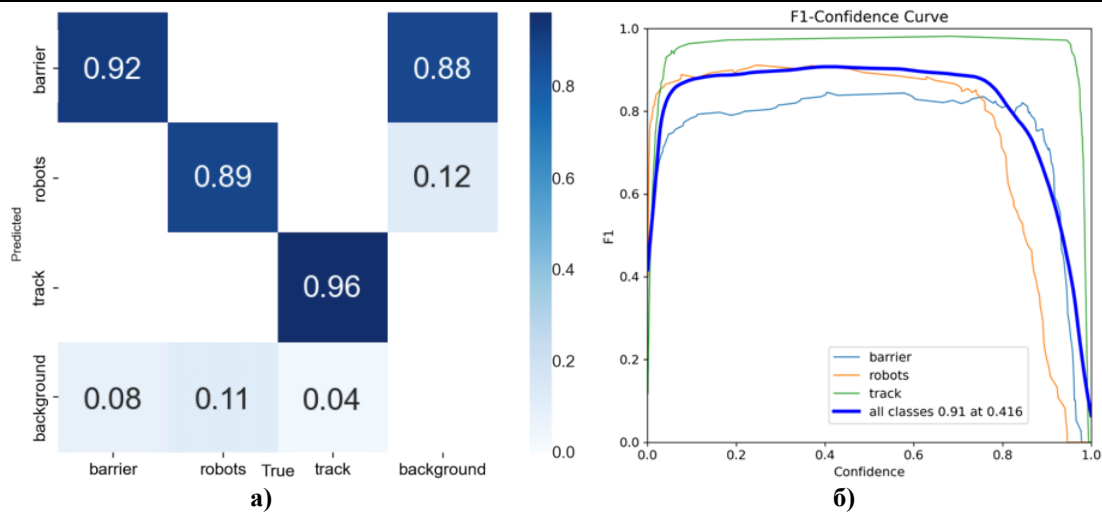


Рис. 3. Результати контролю навчання третьої нейромережі: а) нормалізована матриця похибок, б) оцінки міри F1 та довірчого інтервалу

Це може свідчити про труднощі мережі з розпізнаванням цього класу за високих порогів впевненості, ймовірно через схожість із іншими класами або недостатньо виражені характеристики. При цьому оптимальна продуктивність мережі для "automodel" спостерігається на нижчих рівнях впевненості, тож можна розглянути способи поліпшити показники на високих порогах.

Щодо класу "road", він характеризується високою й дуже стабільною мірою F1 (понад 0.95%) в межах порогів від 0.1 до 0.95. Лише при порогах, більших за 0.95, спостерігається поступове зниження, але воно залишається несуттєвим. Така стабільність свідчить про здатність моделі ефективно розпізнавати об'єкти класу "road" навіть при високих рівнях впевненості, зберігаючи відмінний баланс між точністю та повнотою.

Таким чином, використання методів агрегації результатів розпізнавання вище розглянутих нейронних мереж обґрунтовується кількома ключовими аспектами, які сприяють підвищенню якості результатів і надійності системи в цілому, зокрема: компенсувати вплив випадкових помилок, покращити адаптивність до нових типів даних або задач, підвищити стійкість і надійність роботи тощо.

### Висновки

Використання кількох нейронних мереж для детектування об'єктів з подальшою агрегацією результатів їх функціонування може дозволити зменшити ймовірність помилок, оскільки рішення кількох моделей можуть компенсувати недоліки однієї. Використання кількох нейромереж, які працюють за різними алгоритмами, зокрема проходили навчання за різними послідовностями датасету, дає можливість уникнути упередженості однієї конкретної моделі, покращити загальну точність завдяки об'єднанню сильних сторін різних підходів. Зокрема, якщо одна модель має перевагу в розпізнаванні статичних об'єктів, а інша краще розпізнає динамічні то їх комбіновані результати дадуть більш повну інформацію.

Слід розуміти, що процес детектування потребує відповідні параметри, зокрема: локацію зображення, поріг певності, що дозволяє локалізувати зображення для опрацювання та за рівнем певності задається прийнятний поріг для ідентифікації об'єктів. Крім того, додаткова конфігурація повинна містити інформацію про назви класів об'єктів для розпізнавання а також, для забезпечення коректного відображення результатів детектування на зображенні, маркерні кольори, що присвоюють кожному класу. Використання допоміжних функцій для виконання проміжних операцій, зокрема: обчислення показника перекриття між фреймами, об'єднання/агрегація результатів від кількох моделей, зберігання опрацьованого зображення а також кольорів для кожного класу, дозволяє вирішити завдання основних етапів детектування, полегшити операції підготовки зображень.

Одним з варіантів покращення показників виявлення об'єктів на двомірних зображеннях може бути використання мультимодельного підходу. Такий підхід передбачає використання кількох, в нашому випадку трьох, моделей YOLO де остаточний результат визначається як медіанне значення серед результатів детектування таких моделей. Тобто, після зчитування вхідного зображення кожна з моделей детектує наявні об'єкти, після чого відбувається агрегація отриманих результатів шляхом порівняння фреїв від різних моделей.

Такий підхід потребує більше обчислювальних ресурсів, однак дозволяє уникнути дублювання результатів агрегуючи кілька, в нашому випадку три, результатів детектування одного об'єкта на двомірному зображенні.

Доцільно зазначити, що підвищення точності детектування можна використати алгоритм Non-Maximum Suppression, що реалізує процедуру вилучення зайвих фреїв за релевантністю. Фактично згадана процедура дозволяє уникнути повторень детектування одного і того ж об'єкта на зображенні.

Результатом залучення кількох нейронних мереж до опрацювання одного зображення буде маркування об'єктів фреймами з відповідними кольорами, тобто кожен колір фрейму матиме позначення відповідного класу та рівень певності.

### References

1. A Survey on Data Augmentation Techniques / K. Nanthini та ін. *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, м. Erode, India, 23–25 лют. 2023 р. 2023. URL: <https://doi.org/10.1109/iccmc56507.2023.10084010> (дата звернення: 23.11.2024).
2. A transfer learning approach for predictive modeling of bioprocesses using small data / A. W. Rogers та ін. *Biotechnology and Bioengineering*. 2021. Т. 119, № 2. С. 411–422. URL: <https://doi.org/10.1002/bit.27980> (дата звернення: 23.11.2024).
3. Gürsakal N., Çelik S., Birişçi E. Synthetic Data Generation with Python. *Synthetic Data for Deep Learning*. Berkeley, CA, 2022. С. 159–214. URL: [https://doi.org/10.1007/978-1-4842-8587-9\\_5](https://doi.org/10.1007/978-1-4842-8587-9_5) (дата звернення: 23.11.2024).
4. Huang L. Normalization Techniques in Deep Learning. Cham : Springer International Publishing, 2022. URL: <https://doi.org/10.1007/978-3-031-14595-7> (дата звернення: 23.11.2024).
5. 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. URL: <https://doi.org/10.23883/ijrter.2017.3168.0uwxm> (дата звернення: 23.11.2024).
6. Curriculum Labeling: Revisiting Pseudo-Labeling for Semi-Supervised Learning / P. Cascante-Bonilla та ін. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021. Т. 35, № 8. С. 6912–6920. URL: <https://doi.org/10.1609/aaai.v35i8.16852> (дата звернення: 23.11.2024).
7. Mupparaju Sohan, Thotakura SaiRam, Ch Venkata RamiReddy. A Review on YOLOv8 and its Advancements. ResearchGate. URL: [https://www.researchgate.net/publication/377216968\\_A\\_Review\\_on\\_YOLOv8\\_and\\_Its\\_Advancements](https://www.researchgate.net/publication/377216968_A_Review_on_YOLOv8_and_Its_Advancements) (дата звернення: 21.11.2024).
8. Investigations of Object Detection in Images/Videos Using Various Deep Learning Techniques and Embedded Platforms—A Comprehensive Review / C. B. Murthy та ін. *Applied Sciences*. 2020. Т. 10, № 9. С. 3280. URL: <https://doi.org/10.3390/app10093280> (дата звернення: 21.11.2024).