

**М. С. Ткаченко**

*Студент кафедри інформаційних систем та технологій (ICT) Факультету інформатики та обчислювальної техніки (ФІОТ) Національного технічного університету України "Київський Політехнічний Інститут ім. Ігоря Сікорського", Київ, Політехнічна 41, корп. 18,  
appdev.max@gmail.com  
+380973743174, 0000-0002-5372-0883*

## **ЗАСТОСУВАННЯ R-CNN ПРИ АВТОМАТИЧНОМУ ПОЗИЦІОНУВАННІ ОБ'ЄКТІВ ЧЕРЕЗ НЕЙРОМЕРЕЖЕВИЙ АНАЛІЗ ГРАФІЧНИХ ДАНИХ**

*Визначено найбільш актуальні підходи, що використовуються у рамках побудови систем оптичного моніторингу та автоматичного контролю, на базі масивів даних відеореєстрації шляхом застосування алгоритмів нейромережевого аналізу. Зазначено переваги застосування при вирішенні даного класу задач згорткових нейромережевих архітектур, що здійснюють аналіз на основі визначення областей інтересу, а також методу опорних векторів та регресійних моделей. Метою розробки концепції оптимізації відповідних алгоритмів машинного аналізу стало зменшення навантаження на обчислювальний ресурс у процесі виділення і класифікації візуального об'єкту через застосування процедури вибіркового пошуку великого області інтересу для кожної матриці зображення та подальшого визначення вектора ознак для кожної області інтересу. Розроблений математичний апарат може бути ефективно використано при вирішенні широкого класу задач геолокації та надає можливість провести оцінку оптимізації системи машинного аналізу відповідно цільових показників точності, зменшення часу обробки вхідного запиту та зменшення навантаження на обчислювальний ресурс апаратного комплексу.*

*Ключові слова: система відеореєстрації, позиціонування візуального об'єкту, машинний аналіз, згорткова нейромережа, метод опорних векторів, регресійний аналіз, функція втрат.*

**Tkachenko Maksym**

## **APPLICATION OF R-CNN IN AUTOMATIC IN-DOOR POSITIONING THROUGH ANNS ANALYSIS OF GRAPHICAL DATA**

*The most relevant approaches used in the framework of the construction of optical monitoring and automatic control systems based on video registration data arrays by applying neural network analysis algorithms have been determined. The advantages of using convolutional neural network architectures for solving this class of problems, which perform analysis based on the definition of areas of interest, as well as the method of support vectors and regression models, are indicated. The purpose of the machine analysis relevant algorithms optimization concept development was to reduce the load on the computing resource in the process of selection and classification of the visual object through the application of the procedure of selective search of regions of interest for each image matrix and subsequent determination of the feature vector for each region of interest. The developed mathematical apparatus can be effectively used in solving a wide class of geolocation problems and provides an opportunity to evaluate the optimization of the machine analysis system in accordance with the target accuracy indicators, reducing the processing time of the incoming request and reducing the load on the computing resource of the hardware complex.*

*Keywords: video surveillance systems, visual object positioning, machine analysis, convolutional neural network, support vector method, regression analysis, loss function.*

**Вступ.** Розвиток транспортних систем, підвищення рівня життя громадян та впровадження заходів по збільшенню рівня інклюзивності міського простору для людей з особливими потребами, що є характерним для багатьох країн, призвели до необхідності розробки засобів оптичного моніторингу у рамках систем автоматизованого контролю, які надають можливість проводити координацію громадян, запобігати потенційним правопорушенням, тощо. На рівні інформаційних технологій зазначена проблема може бути вирішена через машинний аналіз у режимі реального часу масивів даних, що надходять від систем відеореєстрації та мобільних засобів зв'язку громадян.

На основі нейромережевих алгоритмів може бути здійснено відновлення та уніфікація відеоданих, що надходять від набору пристроїв фотореєстрації з різною якістю фотоматриці, а також виділено та класифіковано візуальні об'єкти і визначено їх положення та траєкторію руху у межах тривимірної сцени у рамках вирішення задачі позиціонування [1-4]. Слід зазначити, що вказана задача може бути спрощена через проведення додаткових процедур машинного аналізу аудіоданих, а також даних геолокації, але це додатково збільшує ресурсомісткість задачі позиціонування та, відповідно, час обробки вхідних даних.

*Аналіз сучасних досліджень і публікацій* присвячених проблемам нейромережевого аналізу відеоданих з метою позиціонування візуального об'єкту у межах тривимірної сцени вказав на пріоритет застосування алгоритмів на основі архітектури згорткової нейромережі (CNN) та,

зокрема методів вибіркового пошуку областей інтересу (RoI), що у рамках загальноприйнятої класифікації визначається як клас R-CNN [5-8]. Методи, що базуються на вибіркового пошуку областей інтересу дозволяють зменшити ресурсомісткість алгоритмів машинного аналізу на основі згорткових нейромереж, але за умов необхідності аналізу великих об'ємів даних у режимі реального часу при обмеженні обчислювального ресурсу апаратної платформи зазначений підхід не забезпечує достатній рівень оптимізації системи.

Тому на сьогоднішній день більш активно використовуються комбіновані моделі, що включають у себе метод опорних векторів (SVM) та регресійний аналіз, формуючі у рамках класу нейромережових архітектур R-CNN такі моделі, що є спеціалізованими для виділення візуальних об'єктів, як «Fast R-CNN» [9, 10], «Faster R-CNN» [11, 12], «R-CNN-JS» [13], «Mask R-CNN» [14, 15] і «YOLO: You Look Only Once» [16, 17]. Як показали експериментальні результати, наведені у відповідних публікаціях, алгоритми машинного аналізу на основі зазначених моделей нейромережової архітектури здатні суттєво зменшити час обробки графічних даних і відеоданих. Тим не менш, при роботі з масивом даних систем відеоспостереження суттєво зменшується точність машинного аналізу з метою виділення, класифікації та позиціонування візуального об'єкту, у зв'язку з меншою якістю набору зображень які представляють відеопотік, невідповідністю якості окремих зображень між собою, рівня освітлення, ракурсу візуальних об'єктів, тощо. Тому в рамках даного дослідження в якості *невирішеної частини загальної проблеми* розглядається саме адаптація алгоритмів на основі архітектури згорткової нейромережі при вирішенні задачі виділення, класифікації і позиціонування візуального об'єкту у межах тривимірної сцени.

Таким чином, **метою дослідження** стала розробка цілісної методології побудови і оптимізація відповідно цільових показників алгоритмів машинного аналізу відеоданих на основі згорткових нейромережових архітектур, що здійснюють аналіз на основі визначення областей інтересу, з метою вирішення позиціонування об'єктів у рамках приватних та громадських закладів міської інфраструктури.

### 1. Застосування нейромережових алгоритмів при вирішенні задачі позиціонування візуальних об'єктів

Узагальнена схема згідно якої здійснюється аналіз потокових даних систем відеоспостереження та мобільних пристроїв з метою проведення процедур виділення, класифікації і позиціонування візуальних об'єктів у межах тривимірної сцени включає у себе такі функціональні компоненти як складові системи відеоспостереження об'єкту цивільної інфраструктури, платформа машинного аналізу, база даних, що включає у себе відкриту інформацію про користувачів закладу та мобільні пристрої користувачів закладу з підсистемами відеореєстрації та прийому-передачі даних (рис. 1).

Відповідно до наведеної на рис. 1 схеми дані відеопотоку, що надходять від камер системи відеоспостереження та мобільних електронних пристроїв відвідувачів закладу поступають до системи нейромережового аналізу. Зазначена система виділяє візуальні об'єкти і співвідносить їх з інформацією представленою базою даних (по відношенню до користувачів закладу це включатиме біометричні дані), що дозволяє надійно ідентифікувати кожного окремого відвідувача, провести процедуру його позиціонування у інфраструктурі закладу, надати допоміжну інформацію, сформувати персональне рекламне оголошення, оцінити ризик здійснення правопорушення, тощо.

Як було зазначено вище, обробка набору відеопотоків, що включає у себе реконструкцію вхідних даних значно збільшує навантаження на обчислювальний ресурс системи, що у багатьох випадках унеможливує роботу у режимі реального часу навіть при застосуванні оптимізованих алгоритмів, що базуються на згорткових нейромережових архітектурах з попереднім визначенням областей інтересу, стійких до зміни масштабу матриці зображення, зсуву і зміни ракурсу зйомки, наявності типових оптичних абераций, тощо.

Адаптація системи машинного аналізу на основі нейромережових алгоритмів відповідно конкретної задачі виділення, класифікації та позиціонування візуальних об'єктів в умовах обмеження на обчислювальний ресурс та час обробки даних вимагає розробити відповідний математичний апарат, на основі якого адекватно моделюються зазначені операції. Нехай відеопотік складається з набору  $i \in [1; L]$  зображень, представлених двовимірними матрицями розмірності  $M_i: X_{iM} \times Y_{iM}$ . При навчанні згорткової матриці формується ядро згортки розмірністю  $CK: X_{CK} \times Y_{CK}$ , причому слід зазначити, що  $X_{CK} \times Y_{CK} \ll X_{iM} \times Y_{iM}$  і для більшості завдань ядро

згортки формується як квадратне за формою, тобто  $X_{СК} = Y_{СК} = L_{СК}$  незалежно від співвідношення між  $X_{ІМ}$  і  $Y_{ІМ}$ .

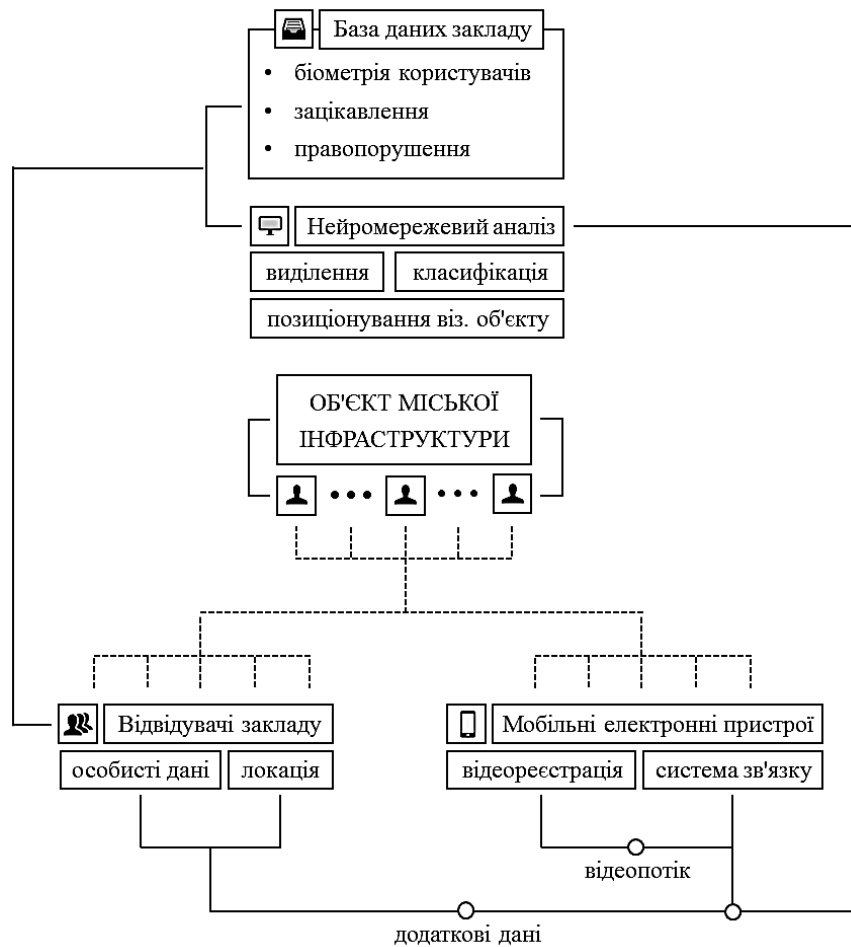


Рис. 1. Узагальнена схема машинного аналізу поточкових відеоданих з метою позиціонування відвідувачів закладу

Через квадрат цього значення  $(L_{СК})^2$  визначається вага ядра згортки. Відповідно, процедура згортки здійснюється через накладання матриці  $СК$  на всі області матриці зображення  $M_i$  з частковим перетином ядер згортки, що дозволяє виділити окремі ознаки і, отже, сформувати карту ознак у багатоканальному режимі для зменшення часу обробки вхідних даних. На математичному рівні часткове перекриття визначається через крок сканування  $S_{СК} < L_{СК}$ . Навантаження на обчислювальний ресурс системи машинного аналізу поточкових відеоданих на основі неймережевих згорткових алгоритмів розраховується через визначення загальної кількості параметрів  $N_{\Sigma}$ , що використовуються у рамках виконання даної процедури, як добуток ваги згортки  $(L_{СК})^2$ , кількості каналів згортки  $n_{СК} \in [1; N_{СК}]$  і кількості каналів виділених ознак згортки  $n_{CF} \in [1; N_{CF}]$ , отже, відповідно —  $N_{\Sigma} = N_{СК} \cdot N_{CF} \cdot (L_{СК})^2$ .

Крім того математичне моделювання процесу машинного аналізу за допомогою алгоритму на основі згорткової неймережі включає у себе формалізацію процедури субдискретизації. На етапі субдискретизації зменшується розмірність карти ознак матриці зображення, що виділяються неймережевим алгоритмом.

Це безпосередньо впливає як на точність машинного аналізу, так і на загальний час обробки системою вхідного запиту. У загальному вигляді неймережевий аналіз може бути представлений у вигляді сандвіч-структури, де процедури згортки чергуються з процедурами субдискретизації, як це показано на рис. 2. Оптимізація відповідних неймережевих алгоритмів має базуватись на схемі, що включає у себе оцінку параметрів процедур згортки та субдискретизації, а також схему навчання неймережі.

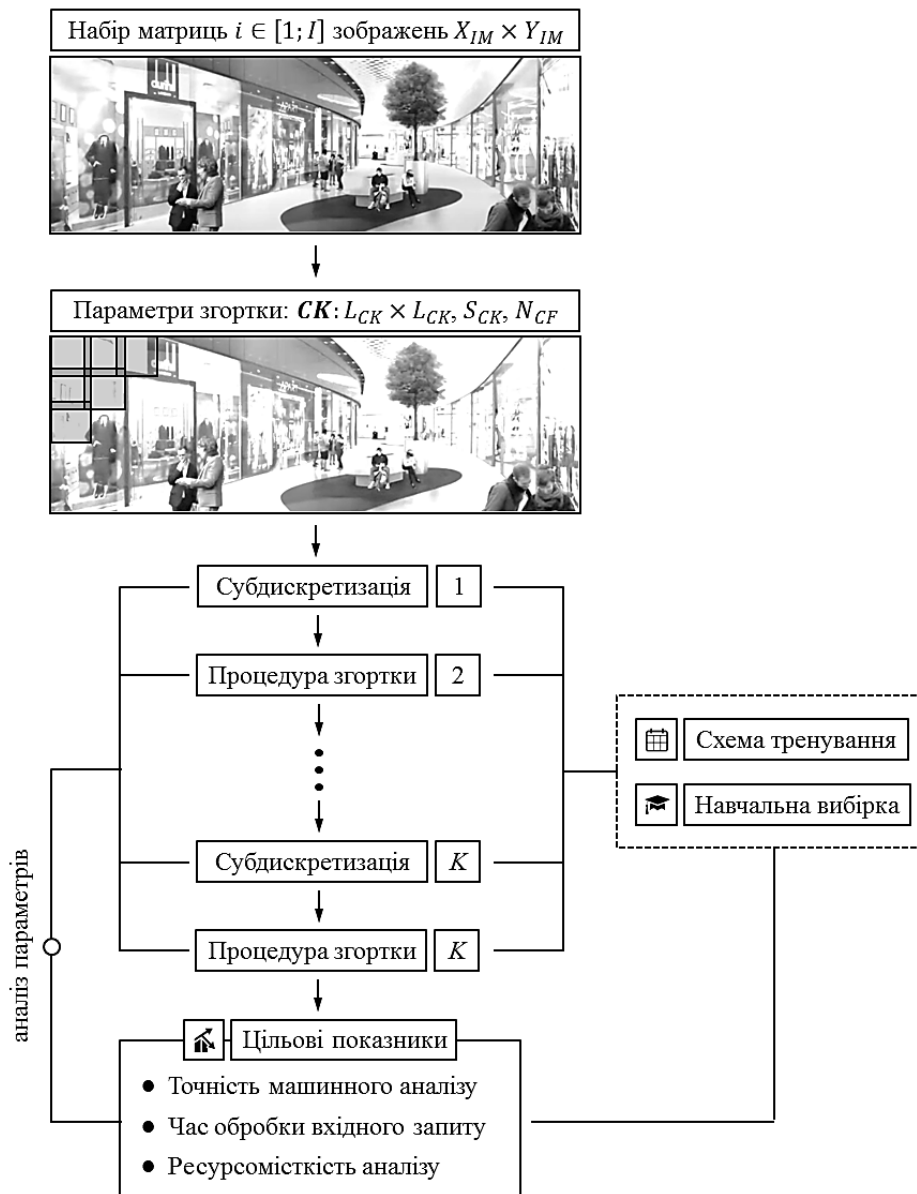


Рис. 2. Базова схема визначення цільових показників нейромережевого аналізу поточкових відеоданих

## 2. Нейромережевий аналіз поточкових відеоданих через вибіркового пошуку областей інтересу

Як було вказано вище основним підходом по оптимізації роботи згорткових нейромереж при виділенні, класифікації та позиціонуванні візуальних об'єктів є вибіркового пошуку областей інтересу, на які поділяються матриці вхідного набору даних. У рамках даного дослідження пропонується розглянути спрощену схему нейромережевого аналізу, що обмежується поділом матриці зображення на області інтересу, та подальше застосування для кожної області класичного набору процедур згортки та субдискретизації. Це дозволяє запобігти подальшого ускладнення методики машинного аналізу, а також спростити оптимізації нейромережевих алгоритмів відповідно до визначення глобальних екстремумів цільових показників.

Відповідна схема роботи алгоритму машинного аналізу, що базується на нейромережеві архітектурі класу R-CNN у загальному випадку складається з наступних етапів (рис. 3):

1. Підготовка навчальної вибірки виділення, класифікації та позиціонування візуальних об'єктів. Застосування отриманої навчальної вибірки та додаткових алгоритмів для тренування базової моделі згорткової нейромережі.

2. Проведення процедури вибіркового пошуку областей інтересу, що потенційно можуть включати у себе візуальні об'єкти, які підлягають виділенню, класифікації та позиціонуванню.

3. Проведення процедури уніфікації розмірності областей інтересу відповідно до актуальних параметрів нейромережевої архітектури. Визначення областей інтересу, що не містять візуальні об'єкти, які підлягають виділенню, класифікації та позиціонуванню. Видалення відповідних областей інтересу.

4. Налаштування параметрів нейромережевої архітектури відповідно параметрів областей інтересу, що пройшли процедуру уніфікації та містять візуальні об'єкти, які підлягають виділенню, класифікації та позиціонуванню.

5. Виділення набору ознак для кожної області інтересу. Визначення для кожної області інтересу функціональних векторів за допомогою метода опорних векторів. Оптимізація точності проведення процедури визначення областей інтересу на основі регресійної моделі.

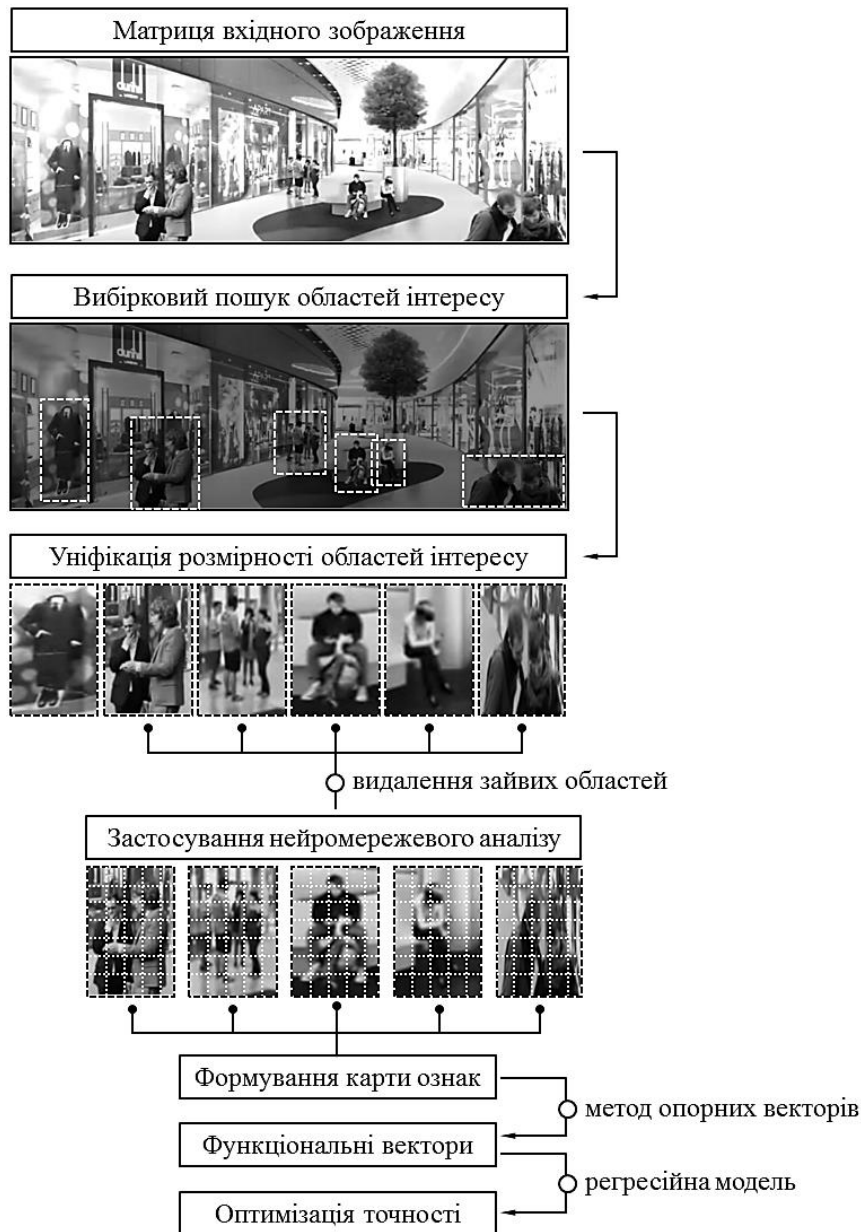


Рис. 3. Алгоритм аналізу поточкових відеоданих через вибірковий пошук областей інтересу

Особливість адаптації архітектури згорткової нейромережі з вибіркоким пошуком областей інтересу до аналіз поточкових відеоданих у режимі реального часу полягає побудові алгоритму, що включає у себе застосування метод опорних векторів та регресійну модель з метою зменшення навантаження на обчислювальний ресурс. Вхідні дані відеопотоку представлені як набір двовимірних розмірності  $X_{IM} \times Y_{IM}$  поділяються відповідно областей уніфікованого розміру  $X_U \times Y_U$  на множину  $(X_{IM}/X_U) \times (Y_{IM}/Y_U)$ . Виконання відповідного нейромережевого алгоритму,

таким чином, включає у себе (i) проведення вибіркового пошуку областей інтересу; (ii) перехід від пулінгового шару класичної згорткової нейромережі до пулінгового шару, що визначається через область інтересу; (iii) заміна останнього повнозв'язного шару на повнозв'язний шар фонового класу; (iv) заміна логістичної функції Softmax на функцію Softmax фонового класу; (v) визначення дискретної функції розподілу ймовірності для кожної з областей інтересу; (vi) застосування для областей інтересу регресійної моделі.

Якщо у рамках розширення математичного апарату ввести такі показники як набір міток класу  $m \in [1; M]$  для кожного з яких за допомогою нормованої експоненційної функції визначається функція розподілу ймовірності як набір  $\{P_1, P_2, \dots, P_m, \dots, P_M\}$ , а також алгоритми мінімальної обмежувальної коробки для значень  $B: \{B_{XA}, B_{YA}, B_{XB}, B_{YB}\}$  та відповідні прогнозовані набори для значень  $B_m: \{B_{XA}(m), B_{YA}(m), B_{XB}(m), B_{YB}(m)\}$ , функцію втрат  $LF$  як кількісний цільовий показник точності машинного аналізу можна визначити через розрахунок суми функції  $CL$ , на основі якої визначаються втрати та функції  $BB$ , на основі якої прогнозуються втрати:

$$LF = CL(P_m) + BB(B_m), \text{ де } \begin{cases} P_m \in [P_1; P_M] \\ B_m: \{B_{XA}(m), B_{YA}(m), B_{XB}(m), B_{YB}(m)\} \end{cases} \quad (1)$$

Функція  $CL(P_m)$  визначається через функцію логарифму, функція  $BB(B_m)$  через робастну функцію втрат  $RL$  прогумовану для всіх значень міток класу  $m \in [1; M]$ :

$$CL = -\log(P_m), \quad (2)$$

$$F_{BB}(B_m) = \sum_{m=1}^M (RL(\Delta B_m)). \quad (3)$$

Таким чином, відповідно до побудованої моделі математичний апарат надає можливість перевести задачу оптимізації нейромережевого алгоритму до математичної задачі пошуку глобальних мінімумів функції втрат  $LF(B_m, P_m)$  та середнього часу обробки вхідного запиту.

### Висновки

У результаті проведеної роботи було розглянуто основні методи побудови нейромережевих алгоритмів машинного аналізу, що базуються на архітектурі згорткової нейромережі з вибірковою аналізом областей інтересу, з метою виділення, класифікації та позиціонування візуальних об'єктів у рамках вирішення задачі геолокації у міському просторі у режимі реального часу та за умов обмеження обчислювального ресурсу апаратної платформи.

У рамках проведеного дослідження для вирішення задачі розробки цілісної методології побудови нейромережевих алгоритмів машинного аналізу було розроблено:

- узагальнену схему машинного аналізу потокових відеоданих з метою позиціонування відвідувачів об'єкту міської інфраструктури на основі потокових даних системи відеоспостереження та мобільних електронних пристроїв відвідувачів;
- базову схему визначення цільових показників нейромережевого аналізу потокових відеоданих на основі класичної архітектури згорткової нейромережі;
- узагальнений алгоритм аналізу потокових відеоданих через вибіркового пошук областей інтересу матриць зображень через які представлено відеодані;
- алгоритм вирішення задачі оптимізації нейромережевого алгоритму через пошук глобальних мінімумів функції втрат та середнього часу обробки вхідного запиту.

### Список використаних джерел:

21. Rego, M., & Fonseca, P. (2021). OCC based indoor positioning system using a smartphone camera. *2021 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. <https://doi.org/10.1109/icarsc52212.2021.9429782>.
22. Jackermeier, R., & Ludwig, B. (2019). Door transition detection for long-term stability in pedestrian indoor positioning. *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. <https://doi.org/10.1109/ipin.2019.8911810>.



23. Zhu, C., Joerger, M., & Meurer, M. (2020). Quantifying Feature Association error in camera-based positioning. *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*. <https://doi.org/10.1109/plans46316.2020.9109919>.
24. Xiong, Q., Yang, L., & Chen, Q. (2021). Research on positioning algorithm of binocular camera based on multi-media. *The 2nd International Conference on Computing and Data Science*. <https://doi.org/10.1145/3448734.3450476>.
25. Cartwright, H. M. (2021). *Artificial Neural Networks*. Humana Press.
26. Rajchakit, G., Agarwal, P., & Ramalingam, S. (2021). *Stability Analysis of Neural Networks*. Springer.
27. Gajja, M. (2020). Brain tumor detection using mask R-CNN. *Journal of Advanced Research in Dynamical and Control Systems*, 12 (SP8), 101–108. <https://doi.org/10.5373/jardcs/v12sp8/20202506>.
28. Rizki, Y., Medikawati Taufiq, R., Mukhtar, H., Apri Wenando, F., & Al Amien, J. (2020). Comparison between faster R-CNN and CNN in recognizing weaving patterns. *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*. <https://doi.org/10.1109/icimcis51567.2020.9354324>.
29. Yang, R. (2022). Retraction note: UAV landmark detection on Fast Region-based CNN. *Arabian Journal of Geosciences*, 15 (8). <https://doi.org/10.1007/s12517-022-10069-7>.
30. Chaudhuri, A. (2021). Hierarchical modified fast R-CNN for object detection. *Informatica*, 45 (7). <https://doi.org/10.31449/inf.v45i7.3732>.
31. Ye, M., & Li, J. (2022). Abalone counting based on improved faster R-CNN. *2022 2nd International Conference on Bioinformatics and Intelligent Computing*. <https://doi.org/10.1145/3523286.3524542>.
32. Farhat, H. J., Sakr, G. E., Kilany, R., Smayra, T., & Mallak, I. (2022). Pneumonia and COVID-19 classification in chest X-rays using faster region-based convolutional neural networks (Faster R-CNN). <https://doi.org/10.21203/rs.3.rs-1285679/v1>.
33. Kızılluluk, S., & Sert, E. (2022). Hurricane-faster R-CNN-JS: Hurricane Detection with faster R-CNN using artificial Jellyfish Search (JS) optimizer. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-022-13156-9>.
34. Zhu, H., Wang, Y., & Fan, J. (2022). IA-mask R-CNN: Improved Anchor Design Mask R-CNN for surface defect detection of automotive engine parts. *Applied Sciences*, 12(13), 6633. <https://doi.org/10.3390/app12136633>.
35. Zhu, H., Wang, Y., & Fan, J. (2022). IA-mask R-CNN: Improved Anchor Design Mask R-CNN for surface defect detection of automotive engine parts. *Applied Sciences*, 12(13), 6633. <https://doi.org/10.3390/app12136633>.
36. Levy, C., Čiprijanović, A., Drlica-Wagner, A., Mutlu-Pakdil, B., Nord, B., & Tanoglidis, D. (2021). Detecting low surface brightness galaxies with mask R-CNN. *Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021)*. <https://doi.org/10.2172/1825283>.
37. Yan, W., Liu, T., & Fu, Y. (2021). Yolo-tight: An efficient dynamic compression method for Yolo Object Detection Networks. *2021 13th International Conference on Machine Learning and Computing*. <https://doi.org/10.1145/3457682.3457740>.