

ЗАСТОСУВАННЯ АЛГОРИТМУ YOLO В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ НА ОСНОВІ ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ

Вінницький національний технічний університет

Анотація. *Останніми роками комп'ютерні технології, зокрема технології комп'ютерного зору, відіграють усе важливішу роль у медичній діагностиці, включаючи отоскопію — процес візуалізації зовнішнього слухового ходу та барабанної перетинки для виявлення патологій вуха. Однак отоскопічні зображення мають низку особливостей, які ускладнюють автоматичну класифікацію, зокрема варіації в освітленні, орієнтації та наявності артефактів. Ці фактори вимагають розробки надійних моделей глибокого навчання для точної класифікації захворювань. У цій роботі розглянуто сучасні підходи до вирішення цієї проблеми. Розглянуто використання можливостей згорткових нейронних мереж (CNN) для виявлення об'єктів у медичних зображеннях, зокрема фотографій і відео, отриманих за допомогою отоскопічних досліджень, з метою полегшення діагностичних процедур. Проведено аналіз архітектури OTONet, спеціально розроблену для класифікації отоскопічних зображень, робіт з використанням квантового машинного навчання (QML). Було розглянуто гібридну квантово-класичну архітектуру CNN, яка використовує квантові та класичні шари згортки для вилучення прихованих функцій із зображень. Розглянуто використання архітектур глибокого навчання, таких як CNN, з імплементацією алгоритму YOLO, що є перспективним підходом до автоматизації діагностики за допомогою отоскопічних зображень. Описано процес навчання моделі, який передбачає апроксимацію функції на основі прикладів і використання унітарного кодування для зручної роботи з категоріями. Мережа прогнозує об'єкти, ділячи зображення на сітку клітинок і призначаючи рамки для кожного об'єкта, з урахуванням їх положення, розмірів та достовірності. Такий підхід дозволяє вирішувати складні завдання класифікації та виявлення об'єктів з високою точністю. Цей напрямок досліджень має великий потенціал для подальшого розвитку в медичній сфері, забезпечуючи ефективні рішення для обробки медичних зображень та покращуючи якість діагностики.*

Ключові слова: алгоритм, нейронна мережа, аналіз, метод, зображення, модель.

Вступ

Останнім часом спостерігається все більш інтенсивна інтеграція комп'ютерних технологій, зокрема технологій комп'ютерного зору, у процес діагностики різноманітних захворювань. У випадку цільового діагностичного методу даного дослідження – отоскопії, яка полягає у візуалізації зовнішнього слухового ходу і барабанної перетинки, допомагаючи у виявленні патологічних станів у вусі, існує низка проблем через відмінні характеристики даних медичних зображень.

Отоскопічні зображення часто демонструють варіації в освітленні, орієнтації та наявності артефактів, що робить важливим розробку надійних і оптимізованих моделей глибокого навчання для точної класифікації. В області комп'ютерного зору останнім часом все більшої популярності набувають штучні нейронні мережі для вирішення завдань виявлення об'єктів. Однак, виникає низка труднощів у ході реалізації даної концепції. Постає необхідність знайти відповідний метод розпізнавання об'єктів на зображенні, а також відповідну мережеву архітектуру та прийнятні гіперпараметри задля інтеграції технології виявлення об'єктів у архітектуру нейронної мережі. Наукові діячі сьогодення внесли значний вклад у розробку методологій по створенню біотехнічної системи для обробки та аналізу отоскопічних зображень з використанням технології виявлення об'єктів. Був проведений ряд досліджень для вирішення даної проблеми.

У роботі [1] було представлено, структуру OTONet, спеціально розроблену для класифікації отоскопічних зображень. Дана архітектура використовує октавну 3D-конволюцію та комбінацію функцій і модулів фокусування на регіоні для створення точної системи класифікації, здатної розрізняти різні отоскопічні стани. Ця архітектура розроблена для ефективного захоплення та обробки просторової та функціональної інформації, наявної на отоскопічних зображеннях. Використовуючи загальнодоступний набір даних отоскопії, OTONet досягла точності класифікації 99,3% і оцінки F1 99,4% для 11 класів захворювань вуха.

Робота [2] була направлена на розробку форми машинного навчання, квантового машинного

навчання (Quantum machine learning – QML), як нової стратегії для подолання проблеми використання спеціальних високопродуктивних обчислювальних машин, які є дуже дорогими та важкодоступними. Гібридна квантово-класична CNN використовує як квантовий, так і класичний шари згортки, призначені для використання параметризованої квантової схеми. Запропонована модель використовує підхід квантових схем для побудови алгоритмів QML, які потім використовуються для перетворення квантового стану для вилучення прихованих функцій зображення.

Результати дослідження

Вирішальними для використання алгоритму YOLO[3] є набори даних, що містять зображення, на яких «рішення» проблеми вже анотовано мітками Ground Truth. У контексті розпізнавання об'єктів це передбачає позначення об'єктів на зображеннях у наборі даних за допомогою обмежувальних рамок. Кероване навчання в основному передбачає апроксимацію функції за допомогою прикладів. У контексті розпізнавання об'єктів це передбачає абстрагування інформації з прикладів зображень і висновок про критерії рішення для таких завдань, як наявність об'єкта, локалізація та класифікація в межах зображень. Успішне виконання цих критеріїв передбачає отримання моделі розпізнавання об'єктів, здатної обробляти нові, незнайомі зображення, ефективно локалізувати та класифікувати відповідні об'єкти в них.

У процесі навчання нейронної мережі набори даних позначаються мітками, і в таких випадках доцільно використовувати метод унітарного кодування. Унітарне кодування передбачає призначення вектора нулів і одиниць кожній категорії, причому розмірність вектора відповідає різним можливим категоріям. Кожній категорії присвоюється певний індекс у цьому векторі, в результаті чого вектор категорії повністю складається з нулів. Відповідно до цього принципу сума значень у векторі дорівнює одиниці, що робить вектор чітко визначеним розподілом ймовірної відповідної категорії. Перевагою цього підходу є те, що якщо пізніше буде введено нову категорію, до існуючих векторів потрібно буде додати лише один вимір із нульовим значенням. Це передбачає можливість розглядати нормалізовані значення вихідного рівня як вектор розподілу ймовірностей.

При використанні алгоритму YOLO нейронна мережа використовується для прогнозування об'єктів у вхідному зображенні. Розуміння роботи даного алгоритму вимагає розуміння структури базових міток істини та передбачень, які становлять вихід нейронної мережі: вхідне зображення розбивається на сітку клітинок $S \times S$. Комірка, що охоплює центр об'єкта на зображенні, відповідає за розпізнавання об'єктів, в ідеалі приписуючи кожному об'єкту одну клітинку. У кожній із цих клітинок виділено B обмежувальних рамок. Вибір кількості комірок (S) і обмежувальних рамок (B) є довільним у конструкції моделі виявлення об'єктів YOLO.

Обмежувальна рамка характеризується координатами (x, y, w, h) і достовірною оцінкою.[4] Координати (x, y) представляють центр обмежувальної рамки як відносне зміщення до відповідної комірки. Координати w і h позначають висоту і ширину обмежувальної рамки відносно всього зображення. Обмежувальні прямокутники служать для визначення положення та розмірів об'єктів на зображеннях. Оцінка достовірності, кінцеве значення, що інкапсулює обмежувальну рамку, визначає, чи містить вона об'єкт і ступінь відповідності цьому об'єкту. Формально оцінка надійності обмежувальної рамки визначається наступним чином:

$$Confidence = Pr(Object) \times IoU_{pred}^{truth} \quad (1)$$

$Pr(Object)$ – показник об'єктності, який визначає, наскільки ймовірно, що рамка містить об'єкт. IoU_{pred}^{truth} визначає, наскільки точні розміри та положення коробки порівнюються з розмірами наявного об'єкта.

Система розпізнавання об'єктів будується шляхом моделювання CNN, здатної генерувати тензор, схожий на передбачення. Протягом фази навчання модель піддається оптимізації, та поступовій коригуванню своїх ваг. Мета полягає в тому, щоб звести до мінімуму розбіжності між передбаченнями та їхніми відповідними основними мітками істини для кожного вхідного зображення. Коли модель навчається за допомогою численних зображень, вона вчиться дистилювати інформацію та формулювати критерії прийняття рішень, які ідеально можна узагальнити на невидимі зображення за межами навчального набору даних. CNN поступово зменшує роздільну здатність вхідного зображення до розміру $S \times S$, одночасно збільшуючи канали, наприклад, з трьох кольорових каналів до $(B \times 5 + C)$. [5] Таким чином, мережа генерує вихідний тензор у формі $S \times S \times (B \times 5 + C)$.

Висновки

При використанні моделей архітектури CNN з імплементованим алгоритмом YOLO завдання виявлення об'єктів оформлюється як регресійна задача, в якій об'єкти ідентифікуються всередині зображення за допомогою аналізу нейронної мережі піксельних даних. Аналіз кривих втрат навчання та перевірки продемонстрував високу ефективність навчання та перевірки. Невелика різниця між кривими втрат під час навчання та підтвердження вказує на здатність моделі мінімізувати перенавчання, забезпечуючи надійне узагальнення для реальних зображень отоскопії. Зменшення значень втрат із плином епох означає швидку конвергенцію моделі та її здатність швидко покращувати свою продуктивність.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Rao D., Singh R., Kamath S., Pendekanti S., Pai D., Kolekar S., Holla R., Pathan S. OTONet: Deep Neural Network for Precise Otoscopy Image Classification. P. 1-1. 2024. DOI:10.1109/ACCESS.2024.3351668.
- [2] Ragab M., Jadid Abdulkadir S., Muneer A., Alqushaibi A., Sumiea E., Qureshi R., Al-Selwi S., Alhussian H. A Comprehensive Systematic Review of YOLO for Medical Object Detection. 2024. P. 1-1. DOI:10.1109/ACCESS.2024.3386826.
- [3] Binol H., Niazi M., Elmaraghy C. OtoXNet-automated identification of eardrum diseases from otoscope videos: a deep learning study for video-representing images. *Neural Comput & Applic.* 2022 №34. DOI:10.1007/s00521-022-07107-6
- [4] Quan H., Lai H., Gao G., Ma J., Li J., Chen Dongji. Pairwise CNN-Transformer Features for Human-Object Interaction Detection. *Entropy.* 2024. №26. 205 p. DOI:10.3390/e26030205.
- [5] Wang J., Kang S., Wang J., Zhang X., Ma Zh. Mirror R-CNN: Object Detection with Flipped Image. 2024. P.154-166. DOI:10.1007/978-981-97-0855-0_16.

Марчук Андрій Юрійович - аспірант кафедри біомедичної інженерії та оптико-електронних систем, Вінницький національний технічний університет, м. Вінниця, andriu4934@gmail.com.

A. Y. Marchuk

YOLO ALGORITHM USING IN DECISION SUPPORT SYSTEMS BASED ON THE USE OF NEURAL NETWORKS

Vinnitsia National Technical University

Abstract. *In recent years, computer technologies, particularly computer vision technologies, have played an increasingly important role in medical diagnostics, including otoscopy—the process of visualizing the external auditory canal and tympanic membrane to detect ear pathologies. However, otoscopic images have several characteristics that complicate automatic classification, such as variations in lighting, orientation, and the presence of artifacts. These factors necessitate the development of reliable deep learning models for accurate disease classification. This paper examines modern approaches to address this issue. It discusses the use of convolutional neural networks (CNNs) for object detection in medical images, specifically photographs and videos obtained through otoscopic examinations, to facilitate diagnostic procedures. An analysis of the OTONet architecture, specifically designed for classifying otoscopic images, as well as research utilizing quantum machine learning (QML), is conducted. The paper explores a hybrid quantum-classical CNN architecture that employs both quantum and classical convolutional layers for extracting hidden features from images. The use of deep learning architectures, such as CNNs, integrated with the YOLO algorithm, is also discussed as a promising approach to automating diagnostics using otoscopic images. The training process of the model is described, which involves function approximation based on examples and the use of one-hot encoding for convenient handling of categories. The network predicts objects by dividing the image into a grid of cells and assigning bounding boxes to each object, taking into account their position, size, and confidence level. This approach allows for the effective resolution of complex classification and object detection tasks with high accuracy. This line of research holds significant potential for further development in the medical field, providing effective solutions for medical image processing and improving diagnostic quality.*

Key words: *algorithm, neural network, analysis, method, image, model.*

Marchuk Andrii Yuriyovych - Postgraduate student, Department of Biomedical Engineering, Vinnitsia National Technical University, Vinnitsia, andriu4934@gmail.com.