

Зінченко А. Ю., кандидат технічних наук, доцент,
доцент кафедри математичних методів системного аналізу
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0000-0003-1586-3645

Хайдуrow В. В., кандидат технічних наук, старший дослідник,
доцент кафедри математичного моделювання та аналізу даних
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0000-0002-4805-8880

ПОКРАЩЕННЯ РОЗПІЗНАВАННЯ СТРУКТУРОВАНОГО ТЕКСТУ НЕЙРОННОЮ МЕРЕЖЕЮ YOLO

У даній роботі розглядається підхід до покращення розпізнавання структурованого тексту за допомогою інтеграції нейронної мережі YOLO (You Only Look Once) та технології OCR (Optical Character Recognition). Метою роботи є дослідження можливостей підвищення ефективності розпізнавання структурованого тексту шляхом інтеграції моделі YOLO з технологією OCR, а також створення автоматизованої інформаційної системи для виявлення текстових об'єктів та їх подальшого розпізнавання, що дозволяє підвищити ефективність обробки структурованого тексту.

Авторами запропоновано архітектуру багатопотокової моделі, що використовує рекурентні та дво- та тривимірні згорткові нейронні мережі. Розроблене програмне забезпечення, що реалізує алгоритми обчислення оптичного потоку та частотного аналізу символів, написано на мові Python із використанням бібліотек Ultralytics, Pytesseract, Python Image Library та фреймворку веб-додатків Flask. Інтерфейс реалізовано за допомогою HTML, CSS та JavaScript, а для зберігання даних обрано базу даних MySQL. Головною особливістю системи є інтеграція моделі YOLO та OCR для забезпечення точного та швидкого розпізнавання текстових об'єктів на зображеннях.

Розроблена автоматизована інформаційна система є сервіс-орієнтованою, що дозволяє користувачам завантажувати інвойси, зберігати їхні дані та отримувати аналітику щодо витрат і взаємодії з клієнтами. Система включає нейронну мережу YOLOv10, натреновану на 500 зображеннях інвойсів, REST API для користувацької взаємодії, інтерфейс для завантаження інвойсів, а також об'єкти бази даних MySQL для зберігання інформації про користувачів та їхні інвойси.

Архітектура побудована за шаблоном MVC (Model-View-Controller): модель обробляє дані та бізнес-логіку, контролер зв'язує модель і представлення, а саме представлення демонструє дані користувачам. Така структура забезпечує чітке розмежування функцій між компонентами. Додатково система містить сервісні шари для бізнес-логіки та маршрутизації, а також використовує інструмент Blueprint фреймворку Flask для поділу додатка на менші компоненти та організації URL-адрес.

Аналіз результатів розпізнавання тексту показав високу точність OCR, зокрема для структурованого тексту, хоча зустрічаються недоліки, такі як порушення початкової структури тексту. Однак ці недоліки можуть бути мінімізовані шляхом використання мережі YOLO разом з технологією OCR.

Ключові слова: YOLO, OCR, розпізнавання тексту, нейронні мережі, обробка зображень.

Zinchenko A. Yu., Khaidurov V. V. Improving structured text recognition with YOLO neural network

This paper explores improving structured text recognition by integrating the YOLO (You Only Look Once) neural network and OCR (Optical Character Recognition) technology. This work aims to explore ways to enhance the efficiency of structured text recognition through the integration of the YOLO model with OCR technology, as well as to develop an automated information system for detecting and subsequently recognizing text objects, which improves the overall efficiency of structured text processing.

The developed system is a web application that enables users to upload invoices, store their data, and receive insights into expenses and client interactions. It includes a YOLOv10 neural network trained on a dataset of 500 invoice images, a REST API for user interaction, a user interface for invoice uploading, and a MySQL database to store information about users and their invoices.

The authors propose a multithreaded architecture model that utilizes recurrent and two- and three-dimensional convolutional neural networks. The software, which implements algorithms for optical flow calculation and frequency analysis of characters, is developed in Python using Ultralytics, Pytesseract, Python Image Library, and Flask libraries. HTML, CSS, and

JavaScript are used for the user interface, and MySQL is the chosen database model. The system's main feature is the integration of the YOLO and OCR models to ensure accurate and fast recognition of text objects in images.

The system architecture follows the MVC (Model-View-Controller) pattern, where the model handles data and logic, the controller acts as an intermediary between the view and model, and the view displays data to users. Each component has its respective roles and functions, making the system well-structured and easy to modify.

Additional service layers are used for business logic and routing, and Flask's Blueprint tool organizes the application into smaller components and URL structures. Overall, the system is well-structured, ensuring efficient data handling and a user-friendly interface for seamless interaction.

The analysis of text recognition results demonstrated high OCR accuracy, particularly with structured text, though some limitations were observed, such as disruptions in the original text structure. These drawbacks can be mitigated by combining the YOLO network with OCR technology.

Integrating YOLO with OCR enhances the system's text recognition efficiency, enabling more precise detection of text objects and their subsequent recognition. Despite these achievements, there is room for further improvement, specifically in refining object detection and text recognition algorithms to achieve even greater accuracy and processing speed.

Key words: YOLO, OCR, text recognition, neural networks, image processing.

Постановка проблеми. Автоматичне виявлення та розпізнавання тексту на зображеннях і відео є однією з найважливіших задач у галузі комп'ютерного зору та обробки зображень. Незважаючи на значний прогрес у цій сфері, ідентифікація структурованого тексту залишається складним завданням. Це пов'язано з різноманіттям шрифтів, розмірів та орієнтацій текстових елементів, а також із зовнішніми факторами, такими як умови освітлення та низька якість зображень. Ці обставини ускладнюють точне розпізнавання текстових об'єктів, що в свою чергу впливає на ефективність автоматизації в таких секторах, як фінанси, медицина, логістика, транспорт та інші.

Структурований текст зустрічається в багатьох сферах: у документах, табличних даних, на номерних знаках, плакатах тощо. З огляду на це, точне і швидке розпізнавання таких текстових об'єктів є критично важливим для автоматизації процесів обробки інформації. Використання технології OCR (Optical Character Recognition) стало стандартом для розпізнавання тексту на зображеннях, проте її традиційні методи є недостатньо ефективними у випадках, коли текст має складну структуру або розміщений на низькоякісних зображеннях.

Щоб подолати ці труднощі, в даній роботі пропонується новий підхід – інтеграцію нейронної мережі YOLO, яку використано для виявлення текстових об'єктів, з технології OCR для подальшого їх розпізнавання. Нейронна мережа YOLO відзначається високою швидкістю роботи і точністю виявлення об'єктів, що робить її перспективною для застосування у розпізнаванні тексту, особливо структурованого. Поєднання можливостей YOLO з OCR дозволяє суттєво підвищити ефективність обробки текстових даних, розширюючи функціональність таких систем для обробки великого обсягу інформації у вигляді текстових зображень, зокрема інвойсів, рахунків та інших документів.

Таким чином, розроблення інтегрованої моделі, яка об'єднує YOLO та OCR, спрямована на подолання недоліків традиційних методів та створення автоматизованої інформаційної системи для високоточного і швидкого розпізнавання тексту на зображеннях. Важливість цього підходу зростає в умовах цифрової трансформації, коли швидкість та якість обробки даних є критичними для багатьох сфер, таких як фінанси, охорона здоров'я та транспорт. Саме тому розроблення та застосування ефективних методів та алгоритмів, які б забезпечували точне та швидке розпізнавання структурованого тексту на зображеннях, а також створення автоматизованої інформаційної системи (AIC) для виявлення текстових об'єктів та їх подальше розпізнавання, є темою актуальною та перспективною.

Аналіз останніх досліджень і публікацій. Аналіз останніх досліджень показав, що в останні роки значна увага приділяється вдосконаленню методів обробки зображень для розпізнавання та класифікації об'єктів. Хоча основи оптичного розпізнавання символів (OCR) були закладені десятиліття тому, проте і донині дана технологія є однією з ключових для перетворення тексту з зображень у цифровий формат. Так, в роботі Supriyanto та ін. [1] розглядається використання мультискейлового підходу Retinex для аналізу розпізнавання облич, що показує перспективу підходу до поліпшення якості обробки зображень з неоднорідним освітленням. Інші дослідження, зокрема Patel та ін. [2], зосереджуються на розпізнаванні патологій за допомогою обробки зображень, зокрема раку легень на комп'ютерних томографіях, використовуючи різні методи обробки зображень для покращення точності діагностики. У контексті загальних задач комп'ютерного зору цікаві результати представлені в роботі Vardhan та ін. [3] щодо застосування методів обробки зображень для розпізнавання емоцій.

У статті [4] розглядається еволюція алгоритмів YOLO для розпізнавання об'єктів, починаючи з першої версії і до YOLOv8. Автори аналізують покращення, які з'явилися на кожному етапі розвитку моделі, та нові технічні особливості YOLOv8. В статті запропоновані модифікації моделі, що вплинули на точність і продуктивність мережі в реальних умовах. Зокрема, зроблено акцент на вдосконаленнях архітектури. В роботі [5] запропоновано поліпшений алгоритм виявлення об'єктів на дорогах за допомогою інфрачервоного зображення, який базується на механізмі уваги (attention mechanism), що інтегрований у YOLOv8.

Дослідники адаптували YOLOv8 для роботи з інфрачервоними зображеннями, що забезпечує ефективніше розпізнавання об'єктів у складних умовах, таких як нічна зйомка або низька видимість. Це дозволило фокусуватися на найважливіших частинах зображення, що дозволяє досягати високої точності та стабільності алгоритму навіть у складних дорожніх умовах [5].

Основні принципи розроблення та проектування автоматизованих інформаційних систем, зокрема для багатовимірної обробки зображень, описані в роботах [6,7].

Мета статті: дослідження можливостей підвищення ефективності розпізнавання структурованого тексту шляхом інтеграції моделі YOLO з технологією OCR, а також створення автоматизованої інформаційної системи для виявлення текстових об'єктів та їх подальшого розпізнавання, що дозволяє підвищити ефективність обробки структурованого тексту.

Виклад основного матеріалу. Особливості застосування моделі YOLO для інтелектуального аналізу структурованого тексту. YOLO або «You Only Look Once» – це алгоритм виявлення об'єктів, що значно покращив можливості комп'ютерного зору в реальному часі, дозволяючи швидко і точно визначати об'єкти на зображеннях і відео. На відміну від традиційних алгоритмів виявлення об'єктів, які працюють у кілька етапів і потребують значних обчислювальних ресурсів, YOLO вирішує задачу виявлення об'єктів як проблему регресії, обчислюючи ймовірності класу і координати обмежувальних рамок у один прохід, безпосередньо з необроблених пікселів зображення. Для визначення координат обмежувальних рамок YOLO ділить вхідне зображення на сітку, де кожна комірка відповідає за об'єкти, що знаходяться в межах цієї комірки, з розрахунком їх класових ймовірностей. Такий підхід дозволяє YOLO виявляти кілька об'єктів різних класів за один прохід, забезпечуючи обробку в реальному часі, що робить його ідеальним для таких застосувань, як автономне водіння, відеоспостереження та робототехніка.

Структурований текст – це тип тексту, який має чітко визначену, логічно організовану структуру, як правило, у вигляді таблиць, списків або впорядкованих полів. Прикладами структурованого тексту є таблиці даних, інвойси, формуляри або будь-яка текстова інформація, що містить чіткі блоки, що легко виділяються та розпізнаються за певними правилами. Обробка структурованого тексту є важливою для автоматизації документів, оскільки така структура дозволяє легко витягувати й аналізувати конкретні дані з документів.

Основою багатьох сучасних алгоритмів глибокого навчання, включаючи модель YOLO, є згорткові нейронні мережі (CNN). CNN завдяки своїй здатності виявляти та ідентифікувати візуальні ознаки, дозволяють YOLO розпізнавати об'єкти у реальному часі. Ця можливість критично важлива для обробки структурованого тексту, де потрібно визначити, локалізувати та інтерпретувати текстові області в складних умовах.

Основна особливість CNN полягає у використанні локальних рецептивних полів, завдяки яким кожен нейрон обробляє тільки частину зображення, що дає змогу зменшити кількість обчислень. В YOLO цей підхід реалізується через поділ зображення на сітку, де кожна комірка відповідає за виявлення об'єктів, включаючи текстові блоки, в межах своїх областей. На перших етапах CNN знаходить прості елементи, такі як контури, тоді як на наступних шарах виділяються складніші структури, що дозволяє YOLO ідентифікувати текстові області різних форматів, шрифтів та розмірів.

Застосування CNN у YOLO також передбачає використання згорткових, об'єднувальних та повноз'єднаних шарів. Згортковий шар виділяє ознаки, які є важливими для аналізу тексту, а об'єднувальні шари зменшують просторовий розмір даних, що запобігає перенаванчання моделі та зменшує обсяг обчислень. Завдяки повноз'єднаним шарам, які інтегрують виявлені ознаки, YOLO може точно класифікувати знайдені області як текстові, а OCR може виконувати подальше детальне розпізнавання символів у цих областях.

Однією з основних метрик для оцінки точності моделей виявлення об'єктів є коефіцієнт перетину та об'єднання (IoU), який показує, наскільки точно обмежувальні рамки, передбачені моделлю, відповідають реальним об'єктам на зображенні. Для оцінки продуктивності моделей також використовується середня точність (AP), що визначається за допомогою кривої точності та відкликання. Відмінність YOLO від інших алгоритмів виявлення об'єктів полягає у використанні однієї повнозв'язної нейронної мережі для прогнозування об'єктів на зображенні, що спрощує процес і підвищує продуктивність. Важливим компонентом YOLO є метод неакумулятивного стиснення (NMS), який підвищує точність і ефективність, видаляючи зайві обмежувальні рамки.

Оптичне розпізнавання тексту. Оптичне розпізнавання тексту (OCR) – це процес механічного або електронного перетворення зображень рукописного, друкованого або машинописного тексту в послідовність кодів, які використовуються для відображення в текстовому редакторі. Ця технологія широко застосовується для перетворення книг і документів в електронний формат, автоматизації облікових систем у бізнесі та публікації текстів на вебсайтах. OCR дозволяє редагувати текст, виконувати пошук слів чи фраз, зберігати інформацію в компактнішому вигляді, аналізувати дані, а також здійснювати електронний переклад, форматування чи перетворення тексту в мовлення.

Розглянемо, як працює OCR на конкретному прикладі. Для цього використаємо рушій tesseract-ocr і подамо на вхід файл зі структурованим текстом (рис. 1), у нашому випадку – інвойс. Для взаємодії з рушієм створимо простий скрипт на Python, скориставшись бібліотекою Pytesseract, яка забезпечує інтерфейс для роботи з рушієм.

Invoice no: 12060439

Date of issue:

08/29/2020

Seller:

Brown-Johnson
310 Amanda Corner Suite 472
North William, MN 33119

Tax Id: 981-94-7235
IBAN: GB78IXYE58273701690538

Client:

Calderon-Duran
9884 Roberts Tunnel
North Lindaside, VA 02674

Tax Id: 996-81-8911

ITEMS

No.	Description	Qty	UM	Net price	Net worth	VAT [%]	Gross worth
1.	Rectangle Area Rug Wood Grain Carpet Mat for Living Room Bedroom - 180x60cm	4,00	each	32,81	131,24	10%	144,36
2.	Rug for living room-Minion And Banana Area Rugs Living Room Carpet, Christmas...	2,00	each	29,96	59,92	10%	65,91
3.	Yilong 5'x7.5' Hand Knotted Silk Area Rug Medium Size Classic Floor Carpets 0186	3,00	each	2 800,00	8 400,00	10%	9 240,00

SUMMARY

	VAT [%]	Net worth	VAT	Gross worth
	10%	8 591,16	859,12	9 450,28
Total		\$ 8 591,16	\$ 859,12	\$ 9 450,28

Рис. 1. Інвойс сканований за допомогою OCR

Виконаємо наступний скрипт:

- import pytesseract
- from PIL import Image
- print(pytesseract.image_to_string(Image.open("invoice_0_charspace_1.png")))

Результатом виконання буде набір рядків тексту зчитаних із зображення:

Date of issue: 08/29/2020

Seller: Client:

Brown-Johnson Calderon-Duran
310 Amanda Corner Suite 472 9884 Roberts Tunnel
Morth Willlarm, MN 33119 Maorth Lindaside, WA 028674
Tax Id: 981-94-7235 Tax Id: 986-81-8911
IBAN: GB78IXYE58273701690538

ITEMS

No. Description Qty UM Net price Net worth VAT [%] Gross worth
1. Rectangle Area Rug Wood Grain 4,00 each 32,81 131,24 10% 144,36
Carpet Mat for Living Room
Bedroom – 180x60cm
2. Rug for living room-Minion And 2,00 each 29,96 59,92 10% 65,91
Banana Area Rugs Living Room
Carpet, Christmas...
3. Yilong 5'x7.5' Hand Knotted Silk 3,00 each 2 800,00 8 400,00 10% 9 240,00
Area Rug Medium Size Classic
Floor Carpets 0186

SUMMARY

VAT [%] Net worth VAT Gross worth
10% 8591,16 859,12 9 450,28
Total \$ 8 591,16 \$ 859,12 \$ 9 450,28

Недоліком цього підходу є втрата початкової структури тексту, що може ускладнити подальшу обробку даних, наприклад, конвертацію у формат JSON. Оскільки програма, використовуючи лише OCR, не здатна класифікувати кожне слово за його значенням, знадобилося б розробити алгоритм, який вручну розділяє та сортує всі слова і рядки для відновлення оригінальної структури й коректної обробки, що може бути складним завданням.

Хоча OCR є ефективним інструментом для перетворення тексту в цифровий формат, його використання має свої обмеження та потребує додаткової обробки для відновлення початкової структури тексту й досягнення більшої точності у складних випадках. Як показало проведене дослідження, оптимальним рішенням могла б стати інтеграція методів оптичного розпізнавання символів із додатковими технологіями, наприклад, згортковими нейронними мережами.

Запропонована архітектура та результати тестування. Авторами запропонована архітектура багатопотокової моделі для розпізнавання структурованого тексту (рис. 2), яка використовує рекурентні, а також дво- та тривимірні згорткові нейронні мережі. Розроблене програмне забезпечення створено на мові програмування Python з використанням бібліотек Ultralytics, pytesseract, Python Image Library та Flask. Основною характеристикою цієї архітектури є застосування алгоритмів обчислення оптичного потоку та частотного аналізу для розпізнавання символів, завдяки інтеграції технології OCR у систему.

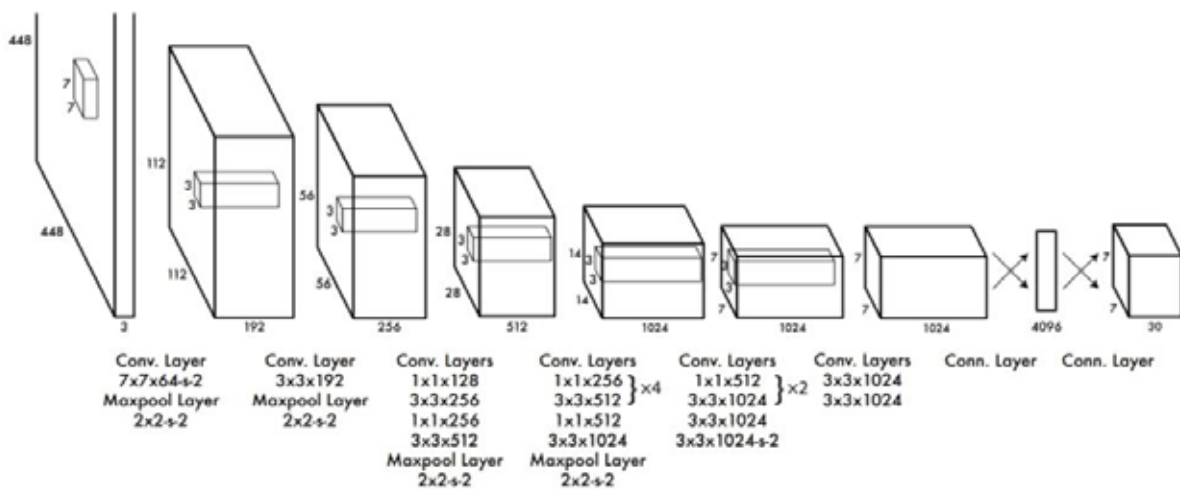


Рис. 2. Архітектура багатопотокової моделі розпізнавання структурованого тексту

Перші 20 згорткових шарів моделі пройшли попереднє навчання з використанням бази даних ImageNet, при цьому були підключені тимчасові шари Average Pooling та повнозв'язні шари Fully Connected. Після цього попередньо навчена модель адаптується для виконання завдання розпізнавання. Результати попередніх досліджень показали, що додавання згорткових і повнозв'язних шарів до попередньо навченої мережі підвищує її продуктивність. Останній повнозв'язний шар YOLO прогнозує як ймовірності класу, так і координати обмежувальної рамки.

Одним із ключових методів, які використовуються в моделях YOLO, є немаксимальне стиснення (NMS). NMS – це етап постобробки, що забезпечує підвищення точності та ефективності виявлення об'єктів. Під час виявлення об'єктів для одного об'єкта на зображенні зазвичай генерується кілька обмежувальних рамок. Ці рамки можуть перекриватися або розташовуватися в різних позиціях, але всі вони представляють один і той же об'єкт. NMS допомагає ідентифікувати та усунути зайві або неправильно побудовані обмежувальні рамки, виводячи єдину обмежувальну рамку для кожного об'єкта на зображенні.

Для подальшого використання та інтеграції в розроблену систему модель YOLOv10 була навчена на вибірці з 200 зображень інвойсів. Після 1500 ітерацій навчання було досягнуто 99% точності в розпізнаванні текстових даних на інвойсах. Серед можливих дефектів було виявлено некоректне накладання міток «generalText» на інші мітки. Для навчання нейромережі використовувалася мова програмування Python та бібліотека ultralytics, яка забезпечує доступ до різних типів моделей YOLO і дозволяє тренувати їх на власних даних. Алгоритм YOLO приймає зображення як вхідні дані, а потім використовує просту глибоку згорточну нейронну мережу для виявлення об'єктів на зображенні.

Ultralytics – це універсальна платформа для навчання моделей з виявлення об'єктів, сегментації екземплярів та класифікації зображень.

Виконаємо наступний скрипт:

1. `from ultralytics import YOLO`
2. `model = YOLO('yolov10x.yaml')`

3. results = model.train(data='config.yaml', epochs=5000, patience=0)
4. metrics = model.val()
5. model.predict('OCRdata/images/test', show=True, save=True)
6. model.export()

Процес і результати навчання представлені на рисунках 3 і 4.

Для оптичного розпізнавання тексту в розробленій архітектурі нейронної мережі був використаний рушій tesseract-ocr – Tesseract Open Source OCR Engine. Для інтеграції tesseract-ocr в інформаційну систему використовувалася бібліотека Python Pytesseract.

Epoch	GPU_mem	box_loss	cls_loss	dfl_loss	Instances	Size
1609/5400	14.5G	0.4724	0.2737	0.7985	1679	640: 100% ██████████ 11/11 [00:12<00:00, 1.14s/it]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 1/1 [00:00<00:00, 4.14it/s]
	all	10	286	0.46	0.458	0.44 0.255
1610/5400	15G	0.4501	0.2679	0.802	1689	640: 100% ██████████ 11/11 [00:12<00:00, 1.14s/it]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 1/1 [00:00<00:00, 4.48it/s]
	all	10	286	0.445	0.467	0.431 0.251
1611/5400	14.1G	0.4494	0.2666	0.8081	1359	640: 100% ██████████ 11/11 [00:12<00:00, 1.13s/it]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 1/1 [00:00<00:00, 4.13it/s]
	all	10	286	0.523	0.449	0.459 0.274
1612/5400	15G	0.4482	0.2666	0.8085	913	640: 100% ██████████ 11/11 [00:12<00:00, 1.13s/it]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 1/1 [00:00<00:00, 4.46it/s]
	all	10	286	0.475	0.489	0.45 0.250
1613/5400	14.3G	0.4557	0.2688	0.8017	1776	640: 100% ██████████ 11/11 [00:12<00:00, 1.14s/it]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 1/1 [00:00<00:00, 4.41it/s]
	all	10	286	0.475	0.489	0.45 0.250

Рис. 3. Процес тренування моделі YOLO

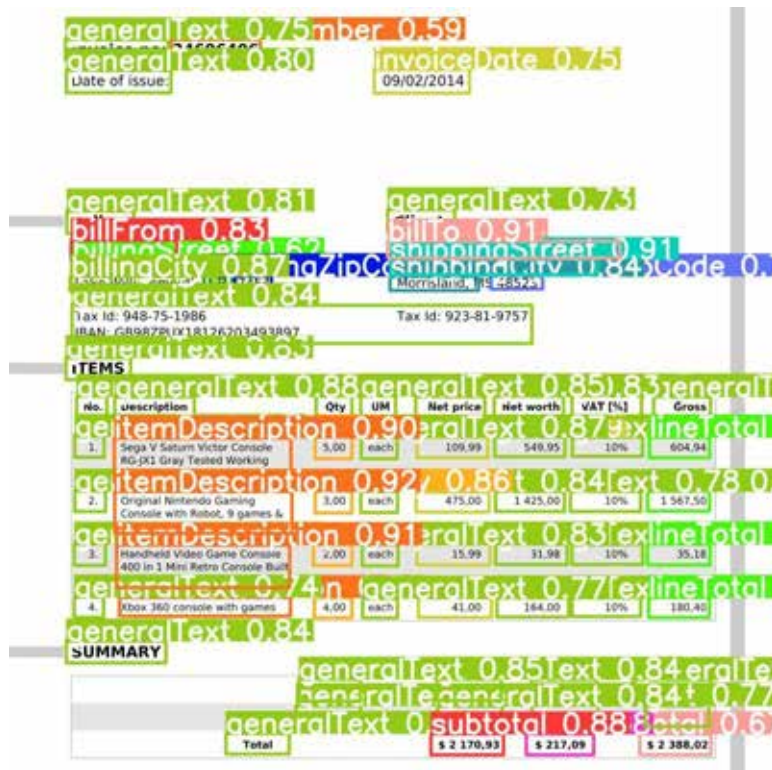


Рис. 4. Результат тренування YOLO.

Розроблення автоматизованої інформаційної системи розпізнавання структурованого тексту. Розроблена система проектувалася по сервіс-орієнтованій архітектурі (реалізація патерну MVC) і представляє собою веб-додаток, який дозволяє користувачам завантажувати інвойси, зберігати дані про них у своєму обліковому записі та переглядати аналітику витрат і взаємодії з клієнтами. Додаток включає такі компоненти:

1. Нейронну мережу YOLOv10, навчена на вибірці з 500 зображень інвойсів.
2. REST API, що забезпечує взаємодію користувача із системою.
3. Інтерфейс користувача для завантаження інвойсів та перегляду необхідної інформації про них.
4. Базу даних MySQL для збереження даних про зареєстрованих користувачів і завантажені ними інвойси.

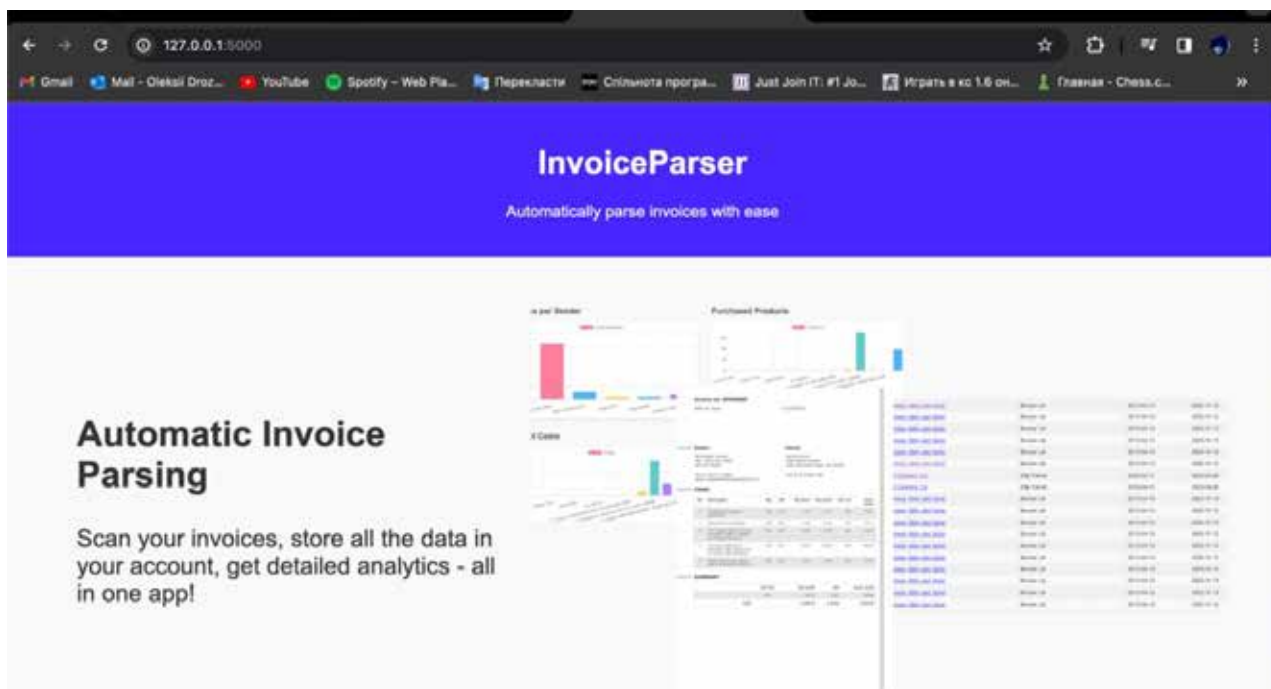


Рис. 5. Головна сторінка додатку

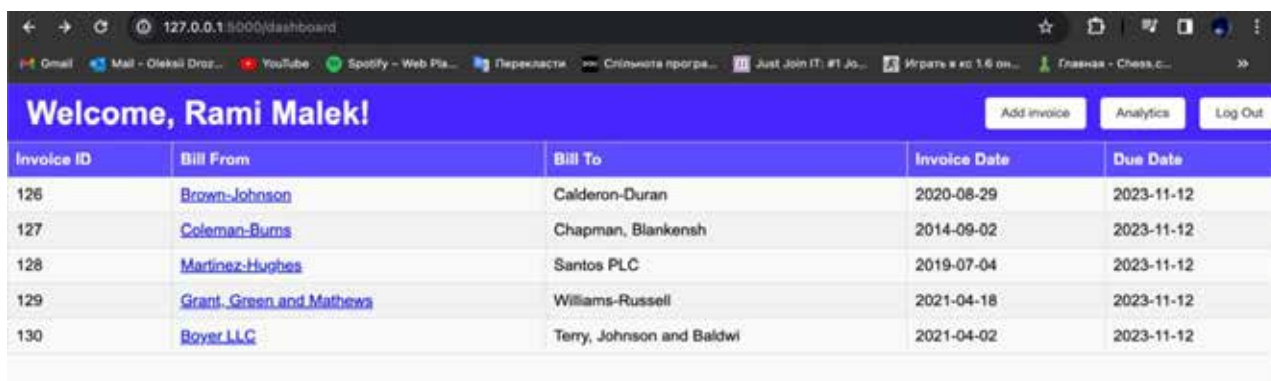


Рис. 6. Особистий кабінет користувача з усіма інвойсами

Серверний код, який відповідає за завантаження інвойсів, аналітику, роботу нейронної мережі та обробку її результатів, написаний на Python з використанням об'єктно-орієнтованого підходу, бібліотеки Flask для розробки API та Ultralytics для навчання й використання моделі YOLOv10. Інтерфейс користувача реалізований на HTML, CSS та JavaScript.

На рис. 5–7 представлені головний інтерфейс програми, особистий кабінет та можливості аналітики – звітність розпізнавання структурованого тексту нейронною мережею YOLO.

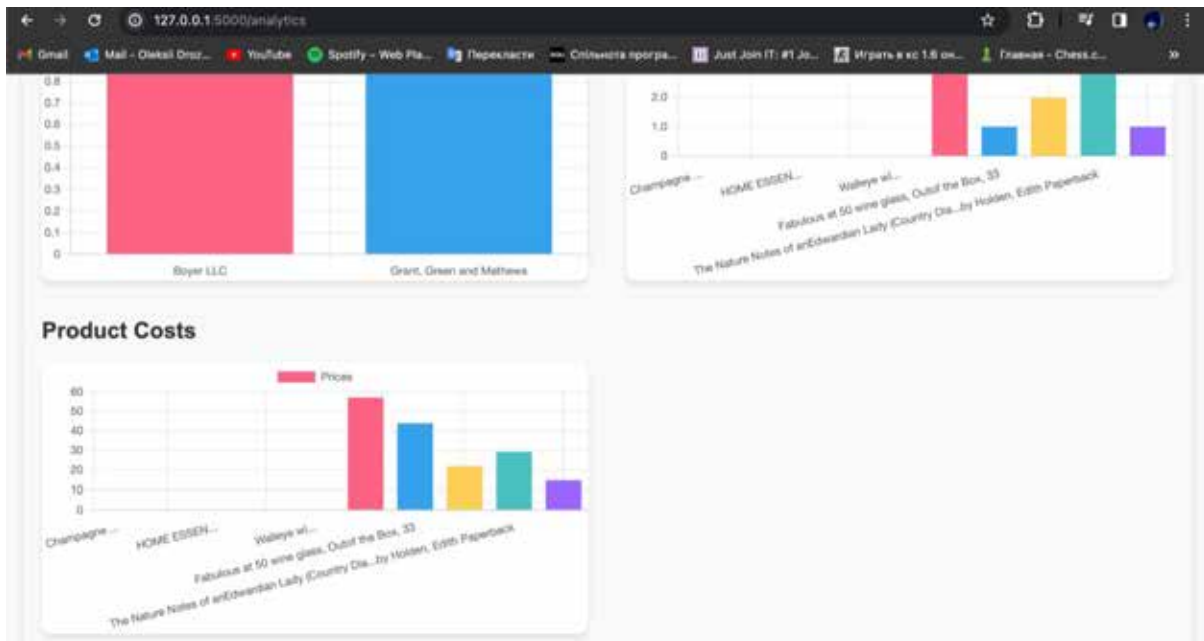


Рис. 7. Сторінка для відображення звітів

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі. В даній роботі розроблене програмне забезпечення може сканувати зображення інвойсів, зберігати витягнуті дані в базі даних та надавати аналітику за всіма інвойсами, завантаженими користувачем. Унікальність АІС полягає в тому, що, на відміну від існуючих рішень, які лише сканують інвойси й конвертують їх у такі формати як JSON, вона забезпечує глибший аналіз і зберігання даних.

Результати розпізнавання структурованого тексту за допомогою розробленої архітектури, що лежить в основі програмного забезпечення, показали високу точність при використанні технології OCR. Можливі недоліки, наприклад втрата початкової структури тексту, можуть бути компенсовані комбінацією методів, зокрема нейронної мережі YOLO. Інтеграція YOLO з OCR значно підвищує ефективність системи розпізнавання тексту, забезпечуючи більш точне виявлення текстових об'єктів для подальшого їх розпізнавання. Подальші дослідження можуть зосередитися на вдосконаленні алгоритмів виявлення об'єктів для ще більшої точності та швидкості роботи системи.

Список використаних джерел:

1. Supriyanto S., Maisevli H., Maya S. R., Diena R. R. Multiscale Retinex Application to Analyze Face Recognition. *Jurnal Online Informatika*. 2020. Vol 5. No. 2. P. 217. DOI: <http://dx.doi.org/10.15575/join.v5i2.668> (дата звернення: 01.11.2024)
2. Patel B., Pankaj K. M., Amit K. Lung Cancer Detection on CT Images by using Image Processing. *International Journal of Trend in Scientific Research and Development*. 2018. Vol. 2. Issue 3. P. 2525-2531. DOI: <http://dx.doi.org/10.31142/ijtsrd11674>. (дата звернення: 01.11.2024)
3. Bardhan Y., Tejas A. F., Prabhat R., Shekhar U., Bharate V.D. Emotion Recognition using Image Processing. *International Journal of Trend in Scientific Research and Development*. 2018. Vol. 2. Issue 3. P. 1523-1526. DOI: <http://dx.doi.org/10.31142/ijtsrd10995> (дата звернення: 01.11.2024)
4. Jagan Mohan R. N. V., Vasamsetty C. S., Gupta V. M. N. S. S. V. K. R. Algorithms in Advanced Artificial Intelligence // Prakash I.V., Palanivelan M. A Study of YOLO (You Only Look Once) to YOLOv8. 2024. London. CRC Press. p. 257–266. DOI: <https://doi.org/10.1201/9781003529231> (дата звернення: 01.11.2024)
5. Luo Z., Tian Y. Improved Infrared Road Object Detection Algorithm Based on Attention Mechanism in YOLOv8. *IAENG International Journal of Computer Science*. 2024. Vol. 51, p. 673 – 680. URL: https://www.iaeng.org/IJCS/issues_v51/issue_6/IJCS_51_6_12.pdf (дата звернення: 01.11.2024)
6. Legland D., Marie-Françoise D. ImageM: a user-friendly interface for the processing of multi-dimensional images with Matlab. 2021. *F1000Research*. p. 10-33. DOI: <http://dx.doi.org/10.12688/f1000research.51732.1> (дата звернення: 01.11.2024)
7. Зінченко А. Ю. Проектування розподілених інформаційних систем на основі використання технології слабозв'язаних компонентів. *Системи та технології*. 2023. 63(1), с. 5-14. DOI: <https://doi.org/10.32782/2521-6643-2022.1-63.1> (дата звернення: 01.11.2024)

References:

1. Supriyanto S., Maisevli H., Maya S. R., Diena R. R. Multiscale Retinex Application to Analyze Face Recognition. *Jurnal Online Informatika*. 2020. Vol 5. No. 2. P. 217. DOI: <http://dx.doi.org/10.15575/join.v5i2.668> (accessed on: 01.11.2024)
2. Patel B., Pankaj K. M., Amit K. Lung Cancer Detection on CT Images by using Image Processing. *International Journal of Trend in Scientific Research and Development*. 2018. Vol. 2. Issue 3. P. 2525-2531. DOI: <http://dx.doi.org/10.31142/ijtsrd11674>. (accessed on: 01.11.2024)
3. Bardhan Y., Tejas A. F., Prabhat R., Shekhar U., Bharate V.D. Emotion Recognition using Image Processing. *International Journal of Trend in Scientific Research and Development*. 2018. Vol. 2. Issue 3. P. 1523-1526. DOI: <http://dx.doi.org/10.31142/ijtsrd10995> (accessed on: 01.11.2024)
4. Jagan Mohan R. N. V., Vasamsetty C. S., Gupta V. M. N. S. S. V. K. R. Algorithms in Advanced Artificial Intelligence // Prakash I.V., Palanivelan M. A Study of YOLO (You Only Look Once) to YOLOv8. 2024. London. *CRC Press*. p. 257–266. DOI: <https://doi.org/10.1201/9781003529231> (accessed on: 01.11.2024)
5. Luo Z., Tian Y. Improved Infrared Road Object Detection Algorithm Based on Attention Mechanism in YOLOv8. *IAENG International Journal of Computer Science*. 2024. Vol. 51, p. 673 – 680. URL: https://www.iaeng.org/IJCS/issues_v51/issue_6/IJCS_51_6_12.pdf (accessed on: 01.11.2024)
6. Legland D., Marie-Françoise D. ImageM: a user-friendly interface for the processing of multi-dimensional images with Matlab. 2021. *F1000Research*. p. 10-33. DOI: <http://dx.doi.org/10.12688/f1000research.51732.1> (accessed on: 01.11.2024)
7. Zinchenko A. Yu. Proektuvannia rozpodilennykh informatsiinykh system na osnovi vykorystannia tekhnolohii slabozviazanykh komponentiv. *Systemy ta tekhnolohii*. 2023. 63(1), s. 5–14. DOI: <https://doi.org/10.32782/2521-6643-2022.1-63.1> (accessed on: 01.11.2024).