УДК 004.932.72’1:004.912

АНАЛІЗ ТА НАВЧАННЯ YOLO-МОДЕЛЕЙ ДЛЯ ДЕТЕКЦІЇ ТА НОРМАЛІЗАЦІЇ ЧЕКІВ З МЕТОЮ ПІДВИЩЕННЯ ТОЧНОСТІ РОЗПІЗНАВАННЯ ТЕКСТУ

Зуйко І.В.
e-mail: iryna.zuiko@nure.ua
Науковий керівник – канд. техн. наук, доц. Любченко В.А.
Харківський національний університет радіоелектроніки, каф. ІНФ
м. Харків, Україна

This work explores the problem of OCR models with text recognition when the text is at an angled position. To solve the problem, it is proposed to normalize the image. To do this, you need to detect the corners of the document and correct the image in perspective. The goal is to analyze and train YOLO models for finding receipt boundaries using OBB or segmentation. Both methods successfully reduce the angle deviation of the text on the receipt. OBB is slightly faster than segmentation, but segmentation better highlights the boundaries of the receipt and prospectively corrects the image.

Останніми роками все більше помітна тенденція оцифрування даних, зокрема тексту: документів, книг, чеків та іншого. Оскільки в цифровому форматі набагато зручнішим та легшим є зберігання, редагування даних, здійснення пошуку необхідного слова або фрази серед усієї інформації. Tesseract, TrOCR Large, CRNN, ASTER – усе це одні з програм, що застосовуються для розпізнавання та конвертації тексту в електронний формат. OCR спрощує і пришвидшує рутинну роботу, перенесення даних, залишаючи людям час на більш важливі завдання.

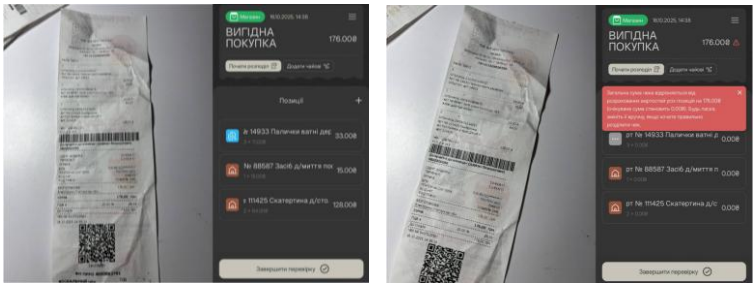
Однак, незважаючи на значні переваги, при деяких умовах OCR стикається з проблемами. Складністю для систем розпізнавання тексту є ситуації, де текст знаходиться не вертикально або горизонтально вирівняним, а викривленим в площині, тоді інформація з документа може передаватися некоректно. Текст може переноситься зі зсувом, неточно або не розпізнаватися взагалі. На рис. 1 представлені приклади вирівняного і спотвореного зображення.

Для покращення точності роботи систем OCR і мінімізування куту нахилу тексту, необхідно нормалізувати зображення: виявити межі документу та перспективно скорегувати зображення, де буде лише вирівняний документ.

Метою даного дослідження є застосування та аналіз різних моделей YOLO для виявлення чеків. З самого початку моделі не розпізнають чеки, іноді сприймають і виділяють їх як книги, але разом з ними виділяють інші нерелевантні об’єкти, такі як руки, стіл та інше. Тому для початку було зібрано якісний датасет в Roboflow з 2694 зображень чеків, їх було розмічено і натреновано ними моделі.

Тепер треба обрати спосіб, що найкраще виділить чек. Звичайна детекція не підходить, так як проблемою є фото чеків під кутом, а бокс детекції не

обертається (рис. 2, а). Моделі орієнтованих обмежувальних боксів (OBB) можуть підлаштувати бокс під обертання чеку, значно зменшуючи нахил тексту, однак кут боксу не завжди точний, через що може залишатися незначний нахил (рис. 2, б). Сегментація найкраще виділяє межі чеку не залежно від кута фото, що сприятиме найкращому вирівнюванню фото (рис. 2, в).



а) б)

Рисунок 1 – Розпізнавання тексту OCR на різних чеках:
а) рівний чек; б) нахилений чек

а) б) в)

Рисунок 2 – Виділення чеку: а) детекція; б) OBB; в) сегментація

Для різних моделей YOLO було проведено аналіз статистичних даних (табл. 1). Точність у моделей приблизно однакова, більше 98%, що каже про якісне навчання. Сегментація трохи повільніша за OBB, тому що виконує більш складну задачу. DLF втрати найбільші в OBB і менші при сегментації. Спад свідчить про точніше виявлення координат меж чеків.

Таблиця 1 – Порівняння характеристик різний моделей YOLO

Модель	Втрати боксів	DFL втрати	Втрати класу	Втрати сегментації	Точність тестування (%)	Кількість параметрів	Розмір моделі (MB)	Час на передбачення (с)
11n-obb	0.3	0.986	0.29	–	98.98	2653918	5.7	0.19
26s-obb	0.37	0.202	0.36	–	98.69	9751554	20.8	0.48
11n-seg	0.2	0.705	0.21	0.19	98.33	2834763	5.7	0.27
11s-seg	0.19	0.715	0.19	0.27	98.39	10067203	19.5	0.66
26n-seg	0.22	0.008	0.14	0.34	98.25	2689079	6.2	0.25
26s-seg	0.25	0.009	0.19	0.28	98.49	10365727	22.2	0.67
26m-seg	0.23	0.008	0.17	0.26	98.27	23508239	51.9	2.02

Наступним кроком нормалізації чека є знаходження 4 кутів і їх впорядкування. В сегментації для отримання 4 кутів з маски виділяється контур чеку, який апроксимується полігоном. OBB одразу надає 4 кути прямокутника. Далі йде впорядковування точок відповідно до їх розташування. За цими 4 точками відбувається проєктивне перетворення площини, що вирівнює перспективні спотворення зображення (рис 3).

Рисунок 3 – Перспективне вирівнювання чеків при сегментації

Отже, сегментація найкраще виділяє межі і найбільше зменшує кут нахилу чеку, хоча витрачає більше часу ніж OBB на передбачення. Розмір моделі (nano, small, medium) несуттєво впливає на збільшення точності, але значно збільшує час витрачений на передбачення та займає більше місця, тому вибір залежить від потужності пристрою.

Список використаних джерел:

1. Харченко В.В., Любченко В.А. Аналіз сучасних підходів детекції та розпізнавання тексту на зображеннях. Technologies, theories and developments : Modern scientific teaching. Proceedings of the IV International Scientific and Practical Conference (23–26 вересня 2025 р.). Valencia, 2025. С. 54–59.

2. Ultralytics. Model training with YOLO : вебсайт. URL: https://docs.ultralytics.com/models/ (дата звернення: 10.03.2026).

3. Tesseract OCR. Tesseract Open Source OCR Engine : вебсайт. URL: https://github.com/tesseract-ocr/tesseract (дата звернення: 10.03.2026).