УДК 004.932.2:004.89

https://doi.org/10.15407/fmmit2026.42.158

Формалізація та первинна експериментальна перевірка адаптивного підходу до вибору OCR-послідовності для розпізнавання тексту на зображеннях.

Христина Грицай[1],Оксана Грицай[2], Ольга Терендій[3]

[1]студент, Національний університет «Львівська політехніка», вул. С. Бандери, 12, 79013, email:khrystyna.hrytsai.pz.2022@lpnu.ua, gryckris@gmail.com

[2]к.ф.-м.н., Національний університет «Львівська політехніка», вул. С. Бандери, 12, 79013, email:oksana.d.hrytsai@lpnu.ua

[3]к.т.н., Національний університет «Львівська політехніка», вул. С. Бандери, 12, 79013, email:olha.v.terendii@lpnu.ua , Інститут прикладних проблем механіки і математики ім. Я. С. Підстригача НАН України, вул. Наукова, 3-б, Львів, 79060, Україна, e-mail:ola_terendij@ukr.net

У статті розглянуто задачу вибору послідовності розпізнавання тексту на зображеннях із урахуванням методів попередньої обробки та особливостей сучасних OCR-моделей. На прикладі тестового зображення проілюстровано, що різні методи попередньої обробки можуть змінювати результат OCR-розпізнавання.У роботі запропоновано методику адаптивногоекспериментального підходу до вибору алгоритму послідовності розпізнавання тексту шляхом комбінування різних методів попередньої обробки зображень та сучасних OCR-моделей. У межах первинної експериментальної перевірки використано OCR-моделі: Tesseract, EasyOCR, PaddleOCR, RapidOCR та AmazonTextract.Запропонований підхід передбачає вибір конфігурації за схемою: тип зображення – метод попередньої обробки – OCR-модель – оцінювання результатів – вибір найкращої послідовності розпізнавання. Оцінювання ефективності виконується за інтегральною оцінкою, побудованою на основі метрик CER, WER, часу обробки, показника впевненості моделі та оцінки нечіткого зіставлення.Отримані результати мають попередній характер і розглядаються як основа для подальшого розширеного експериментального дослідження.

Ключові слова:*оптичне розпізнавання символів, OCR, попередня обробка зображень, Tesseract, EasyOCR, PaddleOCR, RapidOCR, AmazonTextract, CER, WER, інтегральна оцінка.*

Вступ.Перехід до цифрової економіки та широке впровадження електронного документообігу, де автоматизація і швидкість обробки даних визначає ефективність бізнесу, суттєво підвищили потребу в автоматизованих методах обробки інформації.У цьому контексті системи оптичного розпізнавання символів (OCR), що забезпечують перетворення зображень тексту у машинно-читабельний формат, відіграють ключову роль у процесах цифровізації даних [1, 2]. Актуальність задачі автоматичного розпізнавання тексту на зображеннях визначається тим, що значна частина даних зберігається у вигляді неструктурованих документів, таких як PDF-файли, скановані зображення та фотографії, що ускладнює їх аналіз, пошук та інтеграцію в інформаційні системи, зокрема у задачах машинного навчання та нейронних мереж[3, 4]. Крім того,

значним попитом користуються програмні рішення для автоматизації рутинних процесів,зокрема усунення ручного введення даних,а також застосунки, що забезпечують базові потреби доступності та мобільності, включаючи допоміжні технології для людей з обмеженими можливостями та системи розпізнавання тексту в реальному часі [5, 6, 7].

Водночас ефективність OCR-систем суттєво залежить від якості вхідних даних,які можуть містити шум, нерівномірне освітлення, геометричні викривлення та складні фони, що значно ускладнює процес розпізнавання [7, 8].

Підходи до автоматичного розпізнавання тексту можна умовно поділити на два основні напрями. Перший напрям представлений класичними модульними OCR-системами, у яких процес розпізнавання реалізується як послідовність етапів[6, 8]. Другий напрям охоплює інтегровані моделі аналізу документів, зокрема візуально-мовні та OCR-free архітектури, які розглядають документ як цілісний візуально-текстовий об’єкт[4, 9]. Вибір конкретного підходу залежить від типу документа, якості зображення, мови тексту, вимог до точності, швидкодії та доступних обчислювальних ресурсів.

Одним із ключових етапів модульних OCR-процесів є попередня обробка зображень [6, 8, 10].Але водночас ефективність попередньої обробки не є універсальною: метод, що покращує результат для одного типу зображення або однієї OCR-моделі, може не мати позитивного ефекту або навіть погіршувати результат в іншому випадку [11].

Незважаючи на значну кількість досліджень у галузі OCR, питання оптимального поєднання методів попередньої обробки та OCR-моделей залишається відкритим. Більшість наявних робіт зосереджені на вдосконаленні окремих етапів або на оцінюванні окремих моделей і не здійснюють системний аналіз їх взаємодії [3, 5, 8].

Метою цієї роботи є формалізація адаптивного підходу до вибору OCR-послідовності для розпізнавання тексту на зображеннях та його первинна експериментальна перевірка. Основне завдання полягає не в повномасштабному статистичному порівнянні OCR-моделей, а у перевірці логіки запропонованої методики, визначенні ключових параметрів експерименту та формуванні основи для подальшого розширеного дослідження.

Об’єктом дослідження є процес автоматичного розпізнавання тексту на зображеннях і PDF-документах засобами оптичного розпізнавання символів. Предметом дослідження є методи попередньої обробки зображень, OCR-моделі та їхні комбінації, що впливають на значення метрик CharacterErrorRate (CER), Word ErrorRate (WER) і часу обробки.

Для досягнення поставленої мети необхідно проаналізувати сучасні підходи до OCR, охарактеризувати методи попередньої обробки зображень, розглянути особливості моделей Tesseract, EasyOCR, PaddleOCR, Amazon Textract, RapidOCR, формалізувати адаптивний підхід до вибору послідовності розпізнавання тексту, виконати первинну експериментальну перевірку та визначити напрями подальшої роботи.

Наукова новизна роботи полягає у формалізації процедури вибору OCR-послідовності, що базується на спільному аналізі характеристик вхідного зображення, методу його попередньої обробки, OCR-моделі та значень метрик CER, WER і часу обробки, показника впевненості та оцінки нечіткого зіставлення.

Практичне значення дослідження полягає у можливості використання отриманих результатів для побудови ефективних OCR-послідовностей у системах автоматизованої обробки документів, електронного документообігу, у цифрових архівах, інформаційно-пошукових системах і застосунках для розпізнавання тексту з фотографій.

1. Методологічні основи розпізнавання тексту на зображеннях.

1.1. Загальна схема процесу розпізнавання тексту на зображеннях.Процес розпізнавання тексту на зображеннях у класичних OCR-системах можна подати як послідовність взаємопов'язаних етапів. Загальна схема такого процесу охоплює отримання вхідного зображення, його попередню обробку, локалізацію текстових областей, сегментацію, безпосереднє розпізнавання тексту, постобробку результатів та оцінювання якості розпізнавання[3, 6, 8].

На першому етапі система отримує вхідне зображення, яке може бути представлене у вигляді сканованої сторінки, фотографії документа, PDF-сторінки або окремого фрагмента з текстом.

Другим етапом є попередня обробка зображення. Її мета полягає у покращенні візуальних характеристик вхідних даних перед передаванням їх до OCR-моделі.

Після підготовки зображення виконується локалізація текстових областей, тобто виявлення ділянок, які містять текст. Для документів зі складною структурою цей етап може включати визначення блоків тексту, таблиць, заголовків, колонок або окремих рядків. Неправильна локалізація тексту може призвести до втрати частини інформації або до включення у процес розпізнавання зайвих фрагментів зображення, що знижує точність результату.

Наступним етапом є розпізнавання тексту. OCR-модель перетворює візуальне представлення символів, слів або рядків у текстову послідовність. Результат може містити помилки заміни, вставки, пропуску символів або неправильного поділу слів.

Після розпізнавання може виконуватися постобробка результатів. Вона передбачає виправлення типових помилок, нормалізацію тексту, використання словників, мовних моделей або правил форматування.

Завершальним етапом є оцінювання якості розпізнавання. Для цього розпізнаний текст порівнюють з еталонним текстом, що дозволяє кількісно визначити рівень помилок. Найпоширенішими метриками для такого оцінювання є CharacterErrorRate (CER) та Word ErrorRate (WER), які відображають частку помилок відповідно на рівні символів і слів [11-12].

У загальному представленні цифрове зображення можна розглядати як тривимірниймасив піксельних значень:$I \in \mathbb{R}^{H \times W \times C}$, де H – висота зображення, W – ширина зображення, C – кількість каналів. Для RGB-зображенняC=3. Тобтодля кольорового зображення кожен піксель має вигляд:$I(x, y) = \big(R(x, y), G(x, y), B(x, y)\big)$. Усі подальші операції попередньої обробки виконуються над елементами цього масиву або над його похідними представленнями.

Певний метод попередньої обробкиможна подати як оператор$p: I \to I_p$, де I_p – зображення після застосування цього методу.

Формально процес OCR-розпізнавання можна описати таким чином. Нехай I – вхідне зображення, p – оператор попередньої обробки, m– OCR-модель,T– еталонний текст, а $\tilde{T}$– текст, отриманий у результаті розпізнавання. Тоді результат OCR-процесу можна подати у вигляді:$\tilde{T} = m\big(p(I)\big)$.

У цьому записі оператор pзмінює вхідне зображення відповідно до вибраного методу попередньої обробки, а модель m виконує розпізнавання підготовленого зображення. Якість отриманого результату визначається шляхом порівняння $\tilde{T}$з еталонним текстом T. Задача дослідження полягає у виборі такої комбінації оператора попередньої обробки та OCR-моделі, яка забезпечує найкраще значення обраних критеріїв якості для конкретного типу вхідного зображення.

1.2. Методи попередньої обробки зображень у задачах OCR. У практичних задачах зображення можуть містити шум, низький контраст, нерівномірне освітлення, розмиття, нахил тексту, складний фон або недостатню роздільну здатність, що ускладнює локалізацію текстових областей, сегментацію та безпосереднє розпізнавання.Попередня обробка зображення спрямована на зменшення впливу цих дефектів.

Перетворення зображення у відтінки сірого є базовим методом попередньої обробки, що передбачає вилучення кольорової інформації та збереження лише інтенсивності пікселів. Це зменшує обсяг даних через зменшення кількості каналів з трьох до одного:$p_{gray}: \mathbb{R}^{H \times W \times 3} \to \mathbb{R}^{H \times W}$,де значення інтенсивності кожного пікселя може бути обчислене як зважена сума RGB-компонент[13, 14].

Нормалізація яскравості використовується для зменшення впливу нерівномірного освітлення,що передбачає приведення значень інтенсивності пікселів до заданого діапазону. Її застосовують тоді, коли одна частина документа є затемненою, а інша – надмірно освітленою.

Підвищення контрасту спрямоване на збільшення різниці між текстом і фономчерез збільшення різниці між яскравими та темними ділянками. Цей метод є корисним для низькоконтрастних документів, старих сканів, фотографій із блідими символами або зображень, де текст недостатньо чітко відділений від фону.

Формально оператор підвищення контрасту можна подати як перетворення вхідного зображення Iу нове зображення $I_{contrast}$:$p_{contrast}: \mathbb{R}^{H \times W} \to \mathbb{R}^{H \times W}$.

Для зображення у відтінках сірого один із простих способів підвищення контрасту можна описати лінійним перетворенням інтенсивності [13-15]. Окремим методом локального підвищення контрасту є CLAHE–ContrastLimitedAdaptiveHistogramEqualization[16].

Масштабування зображення або застосування методів покращення роздільної здатності використовують для покращення читабельності тексту малого розміру або документів із низькою роздільною здатністю. Застосовуючи коефіцієнт масштабування до матриці інтенсивності, отримаємо нову матрицю, де для визначення інтенсивності значення пікселів використовуються методи інтерполяції. Найпоширенішими методами інтерполяції є:найближчий сусід, білінійна інтерполяція, бікубічна інтерполяція, інтерполяція Ланцоша.

Глобальна бінаризація полягає у перетворенні зображення у двоколірне представлення: пікселі поділяються на текстові та фонові на основі одного порогового значення. Метод Оцу є автоматичним способом вибору порогового значення для бінаризації. Він ґрунтується на аналізі гістограми яскравості та намагається знайти такий поріг, який найкраще розділяє пікселі на два класи: фон і текст.

Фільтрація шуму використовується для усунення випадкових піксельних завад, плям, артефактів сканування або цифрового шуму. Водночас надмірне шумозаглушення може розмити контури символів і погіршити OCR-результат.

Морфологічні операції використовуються для структурної обробки бінарних або напівбінарних зображень. У задачах OCR вони можуть застосовуватися для з’єднання розірваних елементів символів, видалення дрібних шумових об’єктів або покращення форми текстових компонентів.

Корекція нахилу тексту спрямована на вирівнювання документа або текстових рядків відносно горизонтальної осі. Ризиком є помилкове визначення кута нахилу, що може додатково спотворити зображення.

Жоден із методів попередньої обробки не є універсальним. Метод, який покращує результат для одного типу зображення або однієї OCR-моделі, може не мати позитивного ефекту або навіть погіршити результат в іншому випадку.

Для подальшого експериментального дослідження доцільно розглядати не окремі методи попередньої обробки ізольовано, а їхні комбінації з конкретними OCR-моделями. Це дозволяє оцінити, які методи є ефективними для Tesseract, EasyOCR, PaddleOCR, RapidOCR та AmazonTextract у різних умовах якості вхідних зображень.

1.3. Сучасні OCR-моделі та інструменти розпізнавання тексту. Сучасні системи оптичного розпізнавання символів істотно відрізняються за архітектурою, способом розгортання, вимогами до обчислювальних ресурсів, підтримкою мов і стійкістю до дефектів вхідного зображення. У межах цього дослідження розглянуто п’ять інструментів:Tesseract OCR, EasyOCR, PaddleOCR, RapidOCR та AmazonTextract.

Tesseract OCR є одним із найвідоміших OCR-рушіїв з відкритим вихідним кодом[2]. Його використовують як базову модель для порівняння, оскільки він є

доступним, локальним, підтримує багато мов і не потребує підключення до хмарних сервісів. У сучасних версіях Tesseract використовується LSTM-модель, яка розглядає текст як послідовність і прогнозує символи з урахуванням контексту сусідніх елементів. Після розпізнавання результат може уточнюватися за допомогою словників, мовних правил і параметрів сегментації сторінки. Якість роботи Tesseract істотно залежить від попередньої обробки[2].

EasyOCR є бібліотекою з відкритим вихідним кодом, орієнтованою на розпізнавання тексту на зображеннях різного типу, зокрема на фотографіях, документах і тексті в природних сценах[17]. EasyOCR працює як двоетапна система: спочатку виконується виявлення текстових областей на зображенні, після чого знайдені фрагменти передаються до модуля розпізнавання. Такий підхід робить модель придатною для зображень зі складнішим фоном, хоча результат усе одно залежить від контрасту, розміру тексту та якості локалізації.

PaddleOCRє комплексною OCR-системою з відкритим вихідним кодом, орієнтованою на розпізнавання тексту, обробку документів і вилучення структурованої інформації з PDF-файлів та зображень[18]. У типовій OCR-послідовності PaddleOCR спочатку визначає координати текстових областей, після чого ці області передаються до модуля розпізнавання. Додатково можуть використовуватися модулі класифікації напряму тексту та структурного аналізу документа.

RapidOCR є відкритим OCR-інструментом, орієнтованим на швидке локальне розгортання та ефективне виконання OCR-задач у прикладних системах[19]. Його основна перевага полягає у швидкодії, мультиплатформності та можливості автономного використання без звернення до хмарних сервісів. RapidOCR зазвичай працює як оптимізована OCR-послідовність, у якій окремо виконуються виявлення текстових областей, за потреби класифікація орієнтації тексту та подальше розпізнавання.

AmazonTextract є хмарним сервісом інтелектуальної обробки документів, який виходить за межі класичного OCR і призначений не лише для вилучення тексту, а й для аналізу структури документа[20-21]. На відміну від локальних OCR-бібліотек, AmazonTextract обробляє документ через API та повертає не лише розпізнані слова і рядки та інші структурні компоненти документа.

Вибір саме цих моделей для дослідження є методологічно доцільним, оскільки вони охоплюють різні класи OCR-рішень: класичний локальний OCR-рушій, нейромережеві бібліотеки з відкритим кодом, оптимізовані інструменти для швидкого інференсу та хмарний сервіс інтелектуальної обробки документів.

2. Адаптивний підхід до вибору послідовності розпізнавання тексту на зображеннях

Використання одного фіксованого методу попередньої обробки або однієї OCR-моделі для всіх вхідних даних не завжди забезпечує оптимальний результат.У межах цієї роботи пропонується адаптивний підхід до вибору послідовності розпізнавання тексту, який передбачає формування кількох варіантів попередньої обробки для одного вхідного зображення, подальше OCR-

розпізнавання кожного з них, аналіз координат знайдених текстових областей, уточнення результатів на вирізаних фрагментах і вибір найкращої конфігурації на основі інтегральної оцінки якості.

Нехай I– вхідне зображення, яке містить текстову інформацію. Це може бути сканована сторінка, PDF-сторінка, фотографія документа або зображення з текстом у природному середовищі.Формально цифрове зображення можна подати як тривимірний масив піксельних значень.Усі подальші операції попередньої обробки виконуються над елементами цього масиву або над його похідними представленнями.

Для зображення I задається множина методів попередньої обробки:$P = \{p_i | i = 1, \dots, n\}$,де кожен p_iє окремим методом попередньої обробки або наперед визначеною комбінацією таких методів, аn– кількість розглянутих конфігурацій. Для кожної конфігурації $p_i \in P$ формується окрема версія вхідного зображення:$I_i = p_i(I)$, де I_i– зображення після застосування i-їконфігурації попередньої обробки. Таким чином, замість одного варіанта вхідних даних система отримує набір підготовлених зображень, кожне з яких може бути більш придатним для певної OCR-моделі або певного типу дефекту.

Далі кожне попередньо оброблене зображення I_iпередається на вхід OCR-моделі. У загальному випадку можна розглядати множину OCR-моделей:$M = \{m_j | j = 1, \dots, k\}$,де кожен m_jє окремою OCR-моделлю, наприклад Tesseract, EasyOCR, PaddleOCR, RapidOCR або AmazonTextract, аk– кількість моделей у множині M.

Результатом роботи OCR-моделі m_jдля зображення I_iє проміжний розпізнаний текст$\tilde{t}_{ij}$, множина координат знайдених текстових областейB_{ij}та показник впевненості OCR-моделіc_{ij} для відповідної конфігурації.

Координати текстових областей мають важливе значення, оскільки вони дають змогу не лише отримати розпізнаний текст, а й виділити фрагменти зображення, на основі яких цей текст було сформовано.У загальному випадку координати текстових областей можна подати як множину прямокутних областей$B_{ij} = \left\{ b_{ij}^{(r)} \middle| r = 1, \dots, q \right\}$, де кожна область $b_{ij}^{(r)} = (x_{min}, y_{min}, x_{max}, y_{max})$, задається координатами лівого верхнього та правого нижнього кутів.Для кожної конфігурації виконується вирізання областей інтересу. Операцію вирізання можна записати так:$R_{ij}^{(r)} = crop\left(I_i, b_{ij}^{(r)}\right)$, де $R_{ij}^{(r)}$ – вирізаний фрагмент зображення, що міститьr-ту текстову область.

На наступному етапі для отриманих фрагментів виконується ще одна перевірка для уточнення. Вона передбачає повторне OCR-розпізнавання окремих текстових областей на початковому або обробленому зображенні, порівняння результатів між різними конфігураціями попередньої обробки, застосування нечіткого пошуку та зіставлення з очікуваними словами, словником, шаблонами або доменною термінологією. Такий етап є доцільним у випадках, коли первинне OCR-розпізнавання дає кілька близьких за якістю результатів або коли окремі слова мають низький показник впевненості.Уточнений результат для

конфігурації (p_i, m_j) можна подати у вигляді:$T_{ij} = V\left(R_{ij}^{(r)}, \tilde{t}_{ij}\right)$, де V – оператор перевіркита уточнення результату, що може включати повторне розпізнавання вирізаних текстових фрагментів, зіставлення з результатами інших конфігурацій і нечітке порівняння з очікуваними значеннями.

Для додаткової перевірки результату OCR використовується нечітке зіставлення розпізнаного тексту з множиною очікуваних значень, словником або набором шаблонів.Позначимо таку множину через$D = \{d_l | l = 1, \dots, s\}$, де d_l – слово, фраза, шаблон або текстове значення, з яким може порівнюватися результат розпізнавання.Оцінку нечіткого зіставлення для конфігурації (p_i, m_j)позначимо якF_{ij}. ЯкщоT_{ij} – текст, отриманий після перевірки та уточнення результату OCR

для цієї конфігурації, тоді F_{ij}можна визначити як максимальну подібність міжT_{ij}та елементами множиниD: $F_{ij} = \max_{d \in D} \text{sim}(T_{ij}, d)$, де $\text{sim}(T_{ij}, d)$ – функція подібності між двома текстовими рядками.Як функцію подібності можна використати нормалізовану відстань редагування між рядками, що ґрунтується на мінімальній кількості операцій вставки, видалення та заміни символів [22]:

$$\text{sim}(T_{ij}, d) = 1 - \frac{\text{EditDist}(T_{ij}, d)}{max(|T_{ij}|, |d|)}.$$

Тут $\text{EditDist}(T_{ij}, d)$ – відстань редагування між рядками, тобто мінімальна кількість операцій вставки, видалення або заміни символів, необхідних для перетворення одного рядка в інший.

Тоді оцінка нечіткого зіставлення набуває вигляду:

$$F_{ij} = \max_{d \in D}\left(1 - \frac{\text{EditDist}(T_{ij}, d)}{max(|T_{ij}|, |d|)}\right).$$

Значення F_{ij}належить інтервалу[0; 1], де 1 відповідає повному збігу, а значення, близькі до 0, свідчать про низьку подібність між розпізнаним текстом та очікуваними значеннями.Множина Dформується залежно від типу задачі: на основі доменного словника, шаблону документа, переліку очікуваних значень або попередньо визначених фраз, характерних для певного типу зображень чи документів.

Остаточний вибір конфігурації не повинен ґрунтуватися лише на одному критерії, оскільки найнижчий рівень помилок не завжди означає найкраще практичне рішення. Наприклад, одна модель може забезпечувати вищу точність, але мати значно більший час обробки; інша –працювати швидше, але давати нестабільний результат на шумних зображеннях. Тому доцільно використовувати інтегральну функцію якості, побудовану на основі методу зваженої суми для вибору найкращої альтернативи за кількома критеріями[23]:

$$Q(p_i, m_j) = \alpha \times CER_{ij} + \beta \times WER_{ij} + \gamma \times t_{ij}^{norm} - \delta \times c_{ij}^{norm} - \lambda \times F_{ij}^{norm},$$

де CER_{ij} – помилка розпізнавання на рівні символів для конфігурації (p_i, m_j); WER_{ij} – помилка розпізнавання на рівні слів; t_{ij}^{norm} – нормалізований час обробки; c_{ij}^{norm} – нормалізований показник впевненості OCR-моделі; F_{ij}^{norm} –

нормалізована оцінка нечіткого зіставлення; α, β, γ, δ, λ –вагові коефіцієнти, що визначають відносну важливість відповідних критеріїв.

Для моделювання адаптивного підходу до вибору послідовності розпізнавання тексту на зображеннях запропоновано такі вагові коефіцієнти: $\alpha = 0.3$, $\beta = 0.3$, $\gamma = 0.2$, $\delta = 0.1$, $\lambda = 0.1$.Ці значення не розглядаються як універсальні та можуть бути змінені залежно від прикладної задачі. Наприклад, для систем реального часу більшої ваги може набувати критерій часу обробки, тоді як для архівного розпізнавання документів пріоритетом може бути точність.

Для коректного порівняння всі складові інтегральної оцінки доцільно нормалізувати до спільного діапазону, наприклад[0; 1]. У такому випадку найкращою вважається конфігурація, для якої значенняінтегральної оцінки$Q(p_i, m_j)$є мінімальним:$(p^*, m^*) = arg \min_{p_i \in P,\, m_j \in M,} Q(p_i, m_j)$, де$p^*$ – оптимальний метод або конфігурація попередньої обробки, а m^* – OCR-модель, що в поєднанні з цією конфігурацієюзабезпечує найкраще значення інтегральної оцінки.Тоді фінальний розпізнаний текст визначається як результат, отриманий для найкращої конфігурації:$T_{best} = T_{ij}$, для $(p_i, m_j) = (p^*, m^*)$.

Узагальнена схемаалгоритму адаптивного вибору OCR-послідовності:

Вхідні дані:

I– вхідне зображення;

$P = \{p_i | i = 1, \ldots, n\}$– множина конфігурацій попередньої обробки;

$M = \{m_j | j = 1, \ldots, k\}$– множина OCR-моделей;

D – множина очікуваних значень, доменний словник або набір шаблонів для нечіткого зіставлення.

Вихідні дані:

T_{best}– найкращий розпізнаний текст;

(p^*, m^*)–конфігурація попередньої обробки та OCR-модель, для яких отримано найкраще значення інтегральної оцінки.

Кроки алгоритму:

1. Для кожної конфігурації попередньої обробки $p_i \in P$ сформувати відповідне оброблене зображення I_i.
2. Для кожної OCR-моделі $m_j \in M$виконати первинне OCR-розпізнавання зображенняI_iта отримати проміжний розпізнаний текст$\tilde{t}_{ij}$, координати знайдених текстових областейB_{ij} і показник впевненості моделіc_{ij}.
3. На основі координат текстових областей виконати вирізання фрагментів зображенняR_{i_j}, що містять текст.
4. Сформувати або задати множинуDочікуваних значень, доменний словник чи набір шаблонів для нечіткого зіставлення.
5. Виконати уточнення результату і отримати значенняуточнений текстT_{ij}.
6. Для кожної конфігурації (p_i, m_j)обчислити метрики якості розпізнавання: CER, WER, час обробки, показник впевненості OCR-моделі та оцінку нечіткого зіставленняF_{ij}.

7. На основі обчислених метрик визначити інтегральну оцінку якості$Q(p_i, m_j)$.
8. Вибрати конфігурацію(p^*, m^*),для якої значення інтегральної оцінки є мінімальним.
9. Як фінальний результат повернути текстT_{best},отриманий для конфігурації(p^*, m^*).

У межах цієї роботи алгоритм подано як формалізовану методику первинного експериментального оцінювання конфігурацій OCR-послідовності. Його призначення полягає не в остаточному ранжуванні OCR-моделей, а у формуванні відтворюваної процедури оцінювання конфігурацій попередньої обробки та розпізнавання.Особливістю запропонованої послідовності є використання координат текстових областей для повторного аналізу фрагментів зображення. Це дає змогу уточнювати результат розпізнавання не на всьому зображенні, а на окремих ділянках, які були ідентифіковані OCR-моделлю як текстові.

3. Первинна експериментальна перевірка адаптивного підходу

Запропонований у роботі адаптивний підхід до вибору OCR-послідовності потребує експериментальної перевірки, однак у межах цієї статті експериментальна частина розглядається не як повномасштабне порівняння OCR-моделей, а як первиннаекспериментальна перевірка методики. Її основне призначення полягає у перевірці логіки запропонованої послідовності. Це не дає підстав для остаточного ранжування OCR-моделей, проте дозволяє перевірити працездатність запропонованого підходу, визначити параметри, які доцільно фіксувати в подальших експериментах, і підготувати основу для розширеного дослідження на більшому наборі зображень різної якості.

3.1. Тестові зображення. Для перевірки запропонованого підходу використано зображення інформаційної таблички з текстом NOTICE(рис. 1)[24].

Рис. 1. Тестове зображення інформаційної таблички з текстом.

Це зображення належить до типу зображень із текстом у природному середовищі, оскільки містить об'єкт із текстовою інформацією на складному фоновому зображенні. Такий тип даних є складнішим для OCR, ніж чистий скан документа, оскільки система має спочатку локалізувати текстову область, відокремити її від фону, а потім виконати розпізнавання.

Експериментальна перевіркавиконується за єдиною послідовністю дій, описаною вище. Весь процес автоматизовано мовоюPython з використанням бібліотек OpenCV.Фінальний результат експериментальної перевірки для зображення подається як найкращий розпізнаний текст.

Для експериментальної перевірки використано п'ять конфігурацій: без попередньої обробки, перетворення у відтінки сірого, підвищення контрасту, шумозаглушення з масштабуванням і бінаризацію методом Оцу (рис.2).

Рис.2.Результати застосування різних методів попередньої обробки до тестового зображення.

Для кожної конфігурації виконано OCR-розпізнавання за допомогою моделі Tesseract.Такий підхід дозволяє не перевантажувати експериментальну частину, але водночас показати, як різні варіанти підготовки зображення можуть впливати на результат OCR-розпізнавання.Результати експерименту подані у таблиці 1.

Таблиця 1.Результати розпізнавання тексту на рис. 1 за допомогою моделі Tesseract.

Опис	CER	WER	Час, мс	Впевненість моделі	Оцінка нечіткого зіставлення	Інтегральна оцінка

без попередньої обробки	0.56	0.71	450	0.55	0.65	0.411
відтінки сірого	0.28	0.38	360	0.78	0.86	0.154
підвищення контрасту	0.34	0.44	390	0.72	0.78	0.214
шумозаглушення + масштабування	0.18	0.25	510	0.9	0.9	0.119
бінаризаціяОцу	0.25	0.34	420	0.82	0.68	0.167

У результаті визначено локалізацію тексту і наявний на зображенні текст (рис.3). Основна помилка: модель не розпізнала слово NOTICE у верхній частині таблички. Решта ключового тексту розпізнана досить точно.

Рис. 3. Результат локалізації та розпізнавання тексту за допомогою моделі Tesseract.

За таким самим принципом визначено оцінки і текст для решти моделей. Зведені результати інтегральних оцінок відносно попередньої обробки зображення подано у таблиці 2. Значення інтегральної оцінки мінімізується: чим менше тим краще. Найкращі значення є від'ємними завдяки високому відсотку впевненості моделі. Навіть без адаптивного підходу моделі AmazonTextract і RapidOCRпродемонструвалинижчі значення інтегральної оцінки за вибраних вагових коефіцієнтів. Однак, застосування попередньої обробки ще більше покращує інтегральну оцінку. Зокрема, у межах використаного тестового зображення та за вибраних вагових коефіцієнтів найнижче значення інтегральної оцінки отримано для поєднання RapidOCR із шумозаглушенням та масштабуванням.Такий результат може бути пов'язаний з особливостями реалізації RapidOCR та його орієнтацією на ефективне локальне виконання. Якщо більшої ваги надати точності і меншій кількості помилок, то лідером є поєднання моделі AmazonTextractішумозаглушення та масштабування.

Таблиця 2. Інтегральні оцінки для різних OCR-моделей та конфігурацій попередньої обробки

Опис	Tesseract	EasyOCR	PaddleOCR	RapidOCR	Amazon Textract
без попередньої	0.411	0.759	0.62	-0.087	-0.012

обробки					
відтінки сірого	0.154	0.397	0.276	-0.135	-0.073
підвищення контрасту	0.214	0.477	0.354	-0.122	-0.059
шумозаглушення + масштабування	0.119	0.391	0.248	-0.174	-0.1
бінаризаціяОцу	0.167	0.402	0.295	-0.143	-0.09

У результаті було визначено наявний на зображенні текст:
NOTICE
LEVATOR OCCUPANCY LIMIT
1 PERSON OR FAMILY
LIMIT 1 PERSON
LIMIT 1 FAMILY.
THANK YOU FOR PRACTICING SOCIAL DISTANCING.

Варто зазначити, результати отримано відповідно до вагових коефіцієнтів зі значеннями ваги для часу 0.2 і впевненості моделі 0.1, нормалізацію часу обрано як $t/600$, де 600 мс прийнято як умовне верхнє значення для тестового набору конфігурацій. Оскільки показник впевненості моделейформуються різними OCR-системами за різними внутрішніми механізмами, у межах цієї первинної експериментальної перевірки цей показник використано лише як допоміжний критерій. Його безпосереднє порівняння між моделями потребує додаткового калібрування у подальших дослідженнях.Такий спосіб нормалізації використано лише для демонстрації роботи методики і потребує уточнення в розширеному експерименті.Від'ємні значення інтегральної оцінки можливі через входження показників, що максимізуються, зі знаком мінус.Залежно від задачі оптимізації за часом, оптимізації за точністю чи якістю, вагові коефіцієнти можуть бути змінені, а отже, і інтегральна оцінка також може змінюватися. Отриманий результат ілюструє потенційну доцільність використання попередньої обробки перед OCR-розпізнаванням для зображень із текстом у природному середовищі.

Такий обсяг даних є недостатнім для статистично обґрунтованого порівняння OCR-моделей, однак є прийнятним для первинної перевірки послідовності експериментальних дій.

Висновки.У роботі розглянуто задачу адаптивного вибору послідовності розпізнавання тексту на зображеннях з урахуванням методів попередньої обробки та особливостей OCR-моделей. Запропоновано підхід, у межах якого OCR-процес розглядається не як застосування однієї моделі, а як конфігурація, що поєднує попередню обробку зображення, розпізнавання тексту та багатокритеріальне оцінювання результату.

Формалізовано процедуру вибору OCR-послідовності на основі множини методів попередньої обробки, множини OCR-моделей і набору метрик якості.

Для оцінювання конфігурацій використано CharacterErrorRate, Word ErrorRate, час обробки, показник впевненості OCR-моделі та оцінку нечіткого зіставлення. Запропоновано інтегральну оцінку, яка дозволяє порівнювати різні конфігурації залежно від вибраних вагових коефіцієнтів.

Проведена первинна експериментальна перевірка показала працездатність запропонованої методики на прикладі зображення з текстом у природному середовищі. Отримані результати мають попередній характер і не можуть розглядатися як остаточне ранжування OCR-моделей, оскільки експериментальна частина виконана на обмеженій кількості зображень. Водночас експериментальна перевірка дала змогу перевірити логіку запропонованої методики, визначити основні параметри для фіксації під час експерименту та показати, що результати розпізнавання можуть суттєво змінюватися залежно від обраної конфігурації попередньої обробки.Основна цінність роботи полягає у формалізації методики, перевірці логіки експериментальної послідовності та визначенні параметрів, які потрібно враховувати у подальших дослідженнях.

Практичне значення роботи полягає в тому, що запропонований підхід може бути використаний як методична основа для побудови OCR-пайплайнів у системах автоматизованої обробки документів, цифровізації архівів, електронного документообігу, інформаційно-пошукових системах і застосунках для розпізнавання тексту з фотографій.

Подальші дослідження доцільно спрямувати на розширення експериментальної бази, зокрема на використання більшої кількості зображень різних типів: чистих сканів, PDF-сторінок, шумних зображень, низькоконтрастних документів, фотографій із нерівномірним освітленням, зображень із нахиленим текстом, а також текстів українською мовою та змішаномовних текстів. Також доцільно виконати аналіз чутливості інтегральної оцінки до зміни вагових коефіцієнтів, калібрування показників впевненості різних OCR-моделей і порівняння запропонованого підходу з базовими OCR-пайплайнами.

Література

1. Wang X.-F., He Z.-H., Wang K., Wang Y.-F., Zou L., Wu Z.-Z. A survey of text detection and recognition algorithms based on deeplearning technology. *Neurocomputing*. 2023. Vol. 556. Article 126702. DOI: 10.1016/j.neucom.2023.126702.
2. Smith R. An Over view of the Tesseract OCR Engine. *Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007).* Curitiba, Brazil, 2007. P. 629–633. DOI: 10.1109/ICDAR.2007.4376991.
3. Cui L., Xu Y., Lv T., Wei F. Document AI: Benchmarks, Models and Applications. Arxiv preprint. 2021. DOI: 10.48550/arXiv.2111.08609.URL: https://arxiv.org/abs/2111.08609.
4. Appalaraju S., Jasani B., Kota B. U., Xie Y., Manmatha R. DocFormer: End-to-End Transformer for Document Understanding. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).* 2021. P. 993–1003.
5. Baviskar D., Ahirrao S., Potdar V., Kotecha K. Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and

Future Directions. *IEEE Access*. 2021. Vol. 9. P. 72894–72936. DOI: 10.1109/ACCESS.2021.3072900.
6. Subramani N., Matton A., Greaves M., Lam A. A Survey of Deep Learning Approaches for OCR and Document Understanding. arXivpreprint. 2020. DOI: 10.48550/arXiv.2011.13534.
7. Long S., He X., Yao C. Scene Text Detection and Recognition: TheDeep Learning Era. *International Journal of Computer Vision*. 2021. Vol. 129. P. 161–184. DOI: 10.1007/s11263-020-01369-0.
8. Raisi Z., Naiel M. A., Fieguth P., Wardell S., Zelek J. TextDetectionandRecognitionintheWild: A Review. arXivpreprint. 2020. DOI: 10.48550/arXiv.2006.04305.
9. Kim G., Hong T., Yim M., Nam J., Park J., Yim J., Hwang W., Yun S., Han D., Park S. OCR-Free Document Understanding Transformer. *Computer Vision – ECCV 2022*. *Lecture Notesin Computer Science*. Cham : Springer, 2022. Vol. 13688. P. 498–517. DOI: 10.1007/978-3-031-19815-1_29.
10. Kshetry R. L. ImagePreprocessingandModifiedAdaptiveThresholdingforImproving OCR. arXivpreprint. 2021. DOI: 10.48550/arXiv.2111.14075. URL: https://arxiv.org/abs/2111.14075.
11. Otsu N. A Thre shold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*. 1979. Vol. 9, No. 1. P. 62–66. DOI: 10.1109/TSMC.1979.4310076.
12. Quality Assurancein OCR-D: Evaluation Specification. OCR-D Documentation. 2022. URL: https://ocr-d.de/en/spec/ocrd_eval.html.
13. Recommendation ITU-R BT.601-7. Studioen coding parameters of digital television for standard 4:3 andwide-screen 16:9 aspectratios. Geneva : International Telecommunication Union, 2011. 20 p. URL: https://www.itu.int/rec/R-REC-BT.601.
14. Gonzalez R. C., Woods R. E. DigitalImageProcessing. 4th ed. NewYork : Pearson, 2018. 1168 p.
15. Tekalp A. M. Digital Video Processing. 2nd ed. Hoboken : Prentice Hall Press, 2015. 624 p.
16. Zuiderveld K. Contrast Limited Adaptive Histogram Equalization. Graphics Gems IV / ed. by P. S. Heckbert. SanDiego : Academic Press, 1994. P. 474–485. DOI: 10.1016/B978-0-12-336156-1.50061-6.
17. Easy OCR: Ready-to-use OCR with 80+ supportedlanguages. Git Hubrepository. URL: https://github.com/JaidedAI/EasyOCR
18. PaddleOCR: Turnany PDF orimagedocumentintostructureddataforyour AI. GitHubrepository. URL: https://github.com/PaddlePaddle/PaddleOCR.
19. RapidOCR: Opensource OCR toolformulti-platformand of flined eployment. Git Hubrepository. URL: https://github.com/RapidAI/RapidOCR
20. Amazon Textract Developer Guide. Amazon Web Services Documentation. URL: https://docs.aws.amazon.com/textract/latest/dg/what-is.html
21. Analyze Document — Amazon Textract API Reference. Amazon Web Services Documentation.URL: https://docs.aws.amazon.com/textract/latest/dg/API_AnalyzeDocument.html
22. Wagner R. A., Fischer M. J. The String-to-String Correction Problem. Journal of the ACM. 1974. Vol. 21, No. 1. P. 168–173. DOI: 10.1145/321796.321811.
23. Hwang C.-L., Yoon K. Multiple Attribute Decision Making: Methods and Applications. A State-of-the-ArtSurvey. Berlin ; Heidelberg ; NewYork : Springer-Verlag, 1981. 259 p. DOI: 10.1007/978-3-642-48318-9.

24. Shop for Elevator Limit Sign.Banner Buzz.URL :https://www.bannerbuzz.com/elevator-occupancy-limit-1-person-aluminum-sign-non-reflective/p.

Formalization and Initial Experimental Evaluationof an Adaptive Approachto OCR Pipeline Selection for Text RecognitioninImages

Khrystyna Hrytsai, Oksana Hrytsai, Olha Terendii

The paper addresses the problem of selecting an appropriate text recognition pipeline for images by considering image preprocessing methods and the specific features of modern optical character recognition (OCR) models. The relevance of the study is determined by the fact that OCR quality depends not only on the selected recognition model but also on the characteristics of the input image, including noise, contrast, illumination, resolution, text skew, and background complexity.

The aim of the paper is to formalize an adaptive approach to OCR pipeline selection and to perform its initial experimental evaluation. The proposed approach is based on generating several preprocessed versions of the same input image, applying OCR models to each version, obtaining recognized text, text region coordinates, confidence scores, and processing time, and then evaluating the obtained results using a multi-criteria quality score. The study considers the following OCR tools: Tesseract, EasyOCR, PaddleOCR, RapidOCR, and Amazon Textract. The preprocessing configurations include the original image without preprocessing, grayscale conversion, contrast enhancement, denoising with scaling, and Otsu binarization. The quality assessment is based on Character Error Rate (CER), Word Error Rate (WER), processing time, model confidence score, and fuzzy matching score. The experimental part is considered as an initial experimental evaluation rather than a full-scale statistical comparison of OCR models. Its purpose is to verify the logic of the proposed methodology, identify the main parameters that should be fixed in further experiments, and prepare a basis for extended research on a larger dataset of images of different quality. The obtained results demonstrate that the quality of OCR recognition may vary depending on the selected combination of preprocessing method and OCR model. However, the results should be interpreted as preliminary and cannot be considered a final ranking of OCR models. The practical value of the proposed approach lies in its potential use as a methodological basis for building OCR pipelines in automated document processing systems, digital archives, electronic document management systems, information retrieval systems, and applications for text recognition from images.

Отримано 12.05.2026