

В.В. Кандиба¹, О.І. Кушнір¹, В.М. Бредіхін¹, І.О. Хорошилова²¹Харківський національний університет міського господарства імені О.М. Бекетова, Україна²Харківський національний автомобільно-дорожній університет, Україна

ДОСЛІДЖЕННЯ ІНСТРУМЕНТІВ ТА АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ ДЛЯ РОЗПІЗНАВАННЯ ТА ОЦИФРОВКИ ТОВАРНИХ ЧЕКІВ

Дана стаття досліджує питання розпізнавання текстів та оцифровки товарних чеків для ведення бухгалтерського обліку в компаніях та підприємствах. У статті розглянуто етапи виявлення та локалізації тексту, сегментації та розпізнавання тексту за допомогою поетапних та інтегрованих методів.

Ключові слова: датасет, нейронна мережа, цифрові технології, бінаризація, товарний чек, класифікація, OCR.

Постановка проблеми

За останні роки виявлення та розпізнавання тексту стало досить важливою задачею. Глибоке навчання досягло величезних успіхів у фундаментальних програмах комп'ютерного зору, таких як розпізнавання зображень, виявлення об'єктів та сегментація зображення [1]. Це привело до збільшення кількості додатків, заснованих на виявленні та розпізнаванні тексту.

Крім цього, глибоке навчання застосовується до таких областей, як виявлення тексту та розпізнавання тексту [2], а також їхнього подальшого використання для вилучення ключової інформації та розпізнавання іменованих сутностей [3].

Виявлення та розпізнавання тексту з паперових носіїв та веб-сторінок за допомогою різних методів оптичного розпізнавання символів привертає увагу науковців. Але досі проблема виявлення та розпізнавання тексту до кінця не вирішена. Сегментація та виділення тексту – це все ще дуже складна задача.

Аналіз останніх досліджень і публікацій

Вчені використовують різні підходи та різні навчальні набори даних, оптимізацію стратегії (наприклад, оптимізатори, графіки швидкості навчання, номери епох, попередньо навчені ваги та конвеєри збільшення даних), а також дизайн мережі (наприклад, мережева архітектура та втрати) для вирішення цієї задачі.

Серед алгоритмів виявлення тексту слід визначити Mask R-CNN [4], PANet [5], TextSnake [6], FCENet [7]. В них зазвичай використовуються мережі ResNet18, ResNet50 [8], ddrnet23-slim [9] в підходах до виявлення тексту.

Мета статті

Мета статті полягає у використанні отриманого досвіду з розпізнавання текстів для оцифровки товарних чеків для ведення бухгалтерського обліку в ком-

паніях та підприємствах. Дослідження засновано на використанні рекурентної мережі з довгостроковою пам'яттю, LSTM методах глибокого навчання та застосуванні алгоритму EAST для аналізу тексту чеків.

Виклад основного матеріалу дослідження

Отримання структурованої інформації полегшить майбутні дослідження виявлення тексту, визнання та його розуміння. В чеку слід виділяти «назва магазину», «товар» «вартість» та «загальна сума платежу» на зображеннях.

Процес виявлення та розпізнавання тексту поділяється на етап виявлення тексту та етап розпізнавання [10]. Виявлення тексту стосується пошуку текстової області у вхідному зображенні, тоді як розпізнавання має справу з перетворенням отриманого тексту в символи та слова. Методи, які використовуються для цієї мети, поділяються на поетапні методи та інтегровані методи.

Поетапні методи мають окремі етапи виявлення та локалізації тексту, класифікації, сегментації та розпізнавання і видаляють фонову частину з тексту на кожному етапі. Оскільки вони не залежать від розміру словника, їх можна використовувати для розпізнавання тексту із зображень незалежно від його обсягу.

Інтегровані методи мають обмін інформацією між етапами виявлення та розпізнавання, і вони спрямовані на розпізнавання слів із доступного тексту. На рис. 1 представлена блок-схема поетапного та комплексного методів.

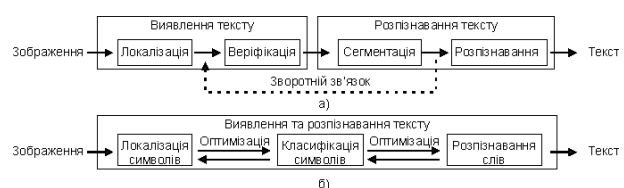


Рис. 1. Блок-схема:

а) поетапних методів; б) інтегрованих методів

Перший етап – виявлення та локалізація тексту – стосується виявлення присутності тексту у вхідному зображенні, тоді як локалізація тексту фіксує його розташування та формує групи текстових областей з максимальним усуненням фону. Ці кроки виконуються за допомогою компонентного аналізу (КА), де графік з'єднаних точок будується на основі кольорів або ребер із бінарного зображення [11].

На рис. 2 представлений спосіб виділення зв'язаних компонентів.

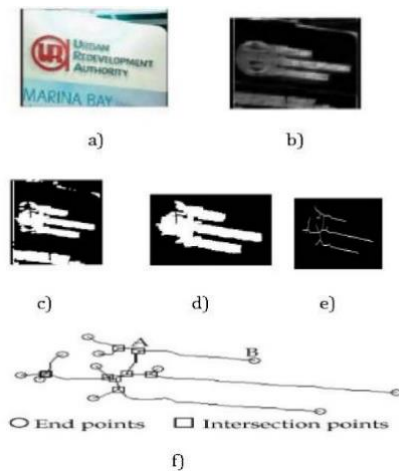


Рис. 2. Видалення зв'язаних компонентів:
а) оригінал зображення; б) карта максимальної різниці; в) кластер тексту; д) зв'язковий компонент (ЗК); е) скелет ЗК; ф) кінцеві точки та точки перетину ЗК [12]

Процес сегментації використовується для відокремлення тексту від фону та витягнення його із зображення.

Інтегровані методи, які зосереджуються на слові, часто заміняють складні стадії сегментації для точного виділення символів, які передаються на етап розпізнавання. Для цього спочатку виконують бінаризацію, яка перетворює кольорове або сіре зображення в чорно-біле. Для досягнення доброго результату сегментації незалежно від темного чи яскравого кольору тексту та фону для бінаризації використовується адаптивне порогове значення в алгоритмі кластеризації k -середніх ($k = 3$ або 4) як параметри кластера, що класифікує двійкове зображення на окремі тексти за допомогою імовірнісних моделей, як показано на рис. 3.

Для процесу виявлення тексту використовується алгоритм EAST (Efficient and Accurate Scene Text Detector), реалізований за допомогою згорткової нейронної мережі (CNN), яка складається з 4 рівнів:

- вхідний рівень;
- згортковий рівень;
- пуловий рівень;
- вихідний рівень.

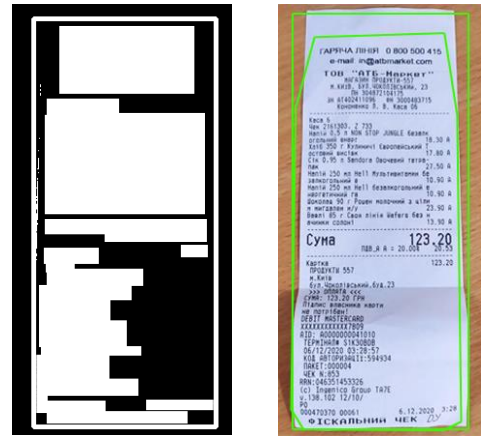


Рис. 3. Використання блоків для ідентифікації тексту рядків і слів

Попередньо оброблене зображення передається на вхідний шар. Зображення конвертується до розміру 128×32 . Потім змінене зображення передається на згортковий шар. Було використано 3 згорткові шари, які витягують відповідні функції із зображення. Згортка застосовує ядро фільтра з 25 нейронів для 1-го рівня та 9 нейронів для наступних двох рівнів. Потім дані передаються на шар пулу, який окреслює області зображення, що передається на вихід мережі. Шари об'єднання розміщуються паралельно згортковому шару. Уся карта функцій, отримана з шару об'єднання, об'єднується разом у повністю підключений рівень CNN. Кожен шар у CNN розходиться на повністю підключений рівень. Карта функцій із 8192 одиниць передається до наступного модуля для розпізнавання тексту.

Після виявлення тексту та етапу локалізації вихід може містити нетекстові області разом із текстовими, і їх потрібно помітити як хибні. Етап класифікації перевіряє області тексту та усуває нетекстові області за допомогою контрольованих або неконтрольованих алгоритмів.

Контрольовані алгоритми знають перед класифікацією такі властивості тексту, як колір, розмір, текстуру тощо. Ці алгоритми не мають попередніх знань про текстові особливості, тому потребують навчання перед класифікацією.

Неконтрольовані алгоритми класифікації не проходять навчання. Вони виділяють особливості тексту під час роботи і використовують ознаки, які виділені в попередній класифікації, для наступної.

Метод реалізовано за допомогою розділу зображення за допомогою вікна розміром 16×16 і отримання 36 функцій з кожного вікна зображення для класифікації текстових і нетекстових блоків. Для цього використовуються функції дисперсії ширини штриха, різниці між контрастністю тексту та фону, а також співвідношення сторін обмежувальної рамки для формування зв'язаних компонентів, які

класифікуються класифікатором на основі k -середніх. Ці глобальні особливості витягнено з усього зображення на відміну від контрольованого методу, де зображення поділено на під регіони.

Багато різних факторів можуть впливати на продуктивність моделей глибокого навчання для розпізнавання тексту. Тому було порівняно роботу алгоритмів ResNet18 і ResNet50, які зазвичай використовують в процесі виявлення тексту.

У табл. 1 представлено порівняння ResNet18, ResNet50 і ddrnet23-slim з точки зору розміру моделі і H -середнього значення для тестових зображень.

Як результат виявлено, що ddrnet23-slim працює трохи гірше, ніж ResNet18 і ResNet50, але має невеликий розмір.

Таблиця 1

Порівняння роботи моделей з точки зору їх розміру та H -середнього значення для тестових зображень

Модель	Метрики		
	FLOPs	Recall	Precision
ResNet18	37.1G	73.5	83.8
ResNet50	78.9G	78.4	83.1
ddrnet23-slim	16.7G	75.2	80.1

Запропонована архітектура системи складається з трьох модулів:

- попередньої обробки;
- виявлення тексту;
- розпізнавання тексту.

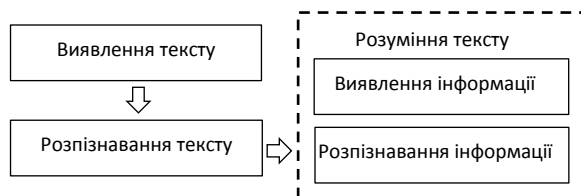


Рис. 4. Архітектура системи для розпізнавання чеків

Для розпізнавання тексту було використано рекурентну нейронну мережу (RNN) ddrnet23-slim. Послідовність ознак містить 256 ознак за один часовий крок, відповідна інформація передається через цю послідовність RNN. Тут працює двовимірний довготривала короткочасна пам'ять (2D-LSTM), яка передає інформацію на велику відстань і має потужні тренувальні характеристики. На виході створюється послідовність 32×80 . Для навчання використовувався набір даних IAM, який містить різну послідовність символів кожної літери. Для створення двонаправленої RNN мережі створюються 2 шари RNN довжиною 256 одиниць. Це допомагає проходити вхідну послідовність з обох кінців. Отримуємо 2 вихідні послідовності: вперед і назад довжиною 8192 нейронів. Цей вихід об'єднаний і має довжину

16384 нейронів. Ця послідовність зображена як розмір 2560 нейронів, значення яких передається на наступний шар. Набір даних складається з 79 унікальних символів.

Вхідні дані передаються як формат зображення будь-якого типу в запропоновану систему. Для розробки системи було використано мову Python, а для обробки зображень пакет OpenCV [13].

У результаті перевірки роботи на валідаційному зображенні чеку було отримано вихідну інформацію, представлену на рис. 5.

Встановлені категорії:	
Гігієна і косметика	22.10
Риба і морепродукти	19.90
Бакалія	12.08
Хлібобулочні вироби	13.98
Молочні продукти	22.98
Бакалія	19.30
Загальні витрати:	
Гігієна і косметика	22.10
Риба і морепродукти	19.90
Бакалія	31.38
Хлібобулочні вироби	13.98
Молочні продукти	22.98

Рис. 5. Результат роботи алгоритму

У результаті роботи було правильно розпізнано 5 з 6 категорій товарів. Також були допущені помилки розпізнавання цін товарів. Такі недоліки відбулися через неточне розпізнавання символів OCR-додатком.

Висновки

Запропонований додаток може бути використаний:

- для виявлення та розпізнавання символів, автоматично скануючи та оновлюючи поля рахунків і рахунків-фактур у базі даних;
- для вилучення тексту із зображення та автоматичного перетворення в цифровий формат та оновлення в базі даних;
- як інструмент для виявлення, розпізнавання та розуміння тексту.

Для забезпечення коректнішої роботи можуть бути виконані наступні дії:

- щоб уникнути помилкового віднесення товарів до найбільш поширених категорій, можливо провести збалансування навчальних даних;
- для підвищення точності розпізнавання цін товарів можливим рішенням є використання OCR-систем на комерційній основі, які надають точніші інструменти розпізнавання тексту.

Данні наробки можуть бути використані для промислового застосування.

Література

1. Borisyyuk F. Rosetta: Large scale system for text detection and recognition in images / F. Borisyyuk, A. Gordo, V. Sivakumar // *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, London (United Kingdom), August 19–23, 2018 yr. – New York (USA) : Association for Computing Machinery, 2018. – P. 71–79. – DOI: [10.1145/3219819.3219861](https://doi.org/10.1145/3219819.3219861).
2. Show, attend and read: A simple and strong baseline for irregular text recognition / H. Li, P. Wang, C. Shen, G. Zhang // *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, Honolulu, Hawaii (USA), 27 January – 1 February, 2019 yr. – AAAI Press, 2019. – P. 8610–8617. – DOI: [10.1609/aaai.v33i01.33018610](https://doi.org/10.1609/aaai.v33i01.33018610).
3. Kim W. A new approach for overlay text detection and extraction from complex video scene / W. Kim, C. Kim // *IEEE Transactions on Image Processing*. – 2009. – Vol. 18, Issue 2. – P. 401–411. – DOI: [10.1109/tip.2008.2008225](https://doi.org/10.1109/tip.2008.2008225).
4. Mask R-CNN / K. He, G. Gkioxari, P. Dollar, R. Girshick // *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice (Italy), October 22–29, 2017 yr. – New Jersey (USA) : Institute of Electrical and Electronics Engineers (IEEE), 2017. – P. 2961–2969. – DOI: [10.1109/iccv.2017.322](https://doi.org/10.1109/iccv.2017.322).
5. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network / W. Wang, E. Xie, X. Song, Y. Zang, W. Wang, T. Lu, G. Yu, C. Shen // *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul (South Korea), 27 October – 2 November, 2019 yr. – New Jersey (USA) : Institute of Electrical and Electronics Engineers (IEEE), 2019. – P. 8439–8448. – DOI: [10.1109/iccv.2019.00853](https://doi.org/10.1109/iccv.2019.00853).
6. TextSnake: A flexible representation for detecting text of arbitrary shapes / S. Long, J. Ruan, W. Zhang, X. He, W. Wu, C. Yao // *Proceedings of the 15th European Conference on Computer Vision (ECCV)*, Munich (Germany), September 8–14, 2018 yr. – Cham (Switzerland) : Springer, 2018. – P. 19–35. – DOI: [10.1007/978-3-030-01216-8_2](https://doi.org/10.1007/978-3-030-01216-8_2).
7. Fourier contour embedding for arbitrary-shaped text detection / Y. Zhu, J. Chen, L. Liang, Z. Kuang, L. Jin, W. Zhang // *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN (USA), June 20–25, 2021 yr. – New Jersey (USA) : Institute of Electrical and Electronics Engineers (IEEE), 2021. – P. 3122–3130. – DOI: [10.1109/CVPR46437.2021.00314](https://doi.org/10.1109/CVPR46437.2021.00314).
8. Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, J. Sun // *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV (USA), June 27–30, 2016 yr. – New Jersey (USA) : Institute of Electrical and Electronics Engineers (IEEE), 2016. – P. 770–778. – DOI: [10.1109/cvpr.2016.90](https://doi.org/10.1109/cvpr.2016.90).
9. Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes / Y. Hong, H. Pan, W. Sun, Y. Jia // *IEEE Transactions on Intelligent Transportation Systems*. – 2023. – Vol. 24, Issue 3. – P. 3448–3460. – DOI: [10.1109/TITS.2022.3228042](https://doi.org/10.1109/TITS.2022.3228042).
10. Elagouni K. A comprehensive neural-based approach for text recognition in videos using natural language processing / K. Elagouni, C. Garcia, P. Sbillot // *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, Trento (Italy), April 18–20, 2011 yr. – New York (USA) : Association for Computing Machinery, 2011. – DOI: [10.1145/1991996.1992019](https://doi.org/10.1145/1991996.1992019).
11. Мацуґа О. М. Навчальний посібник до вивчення курсу «Інформаційні технології розпізнавання образів» / О. М. Мацуґа, Ю. М. Архангельська, Н. М. Єрещенко. – Д. : ПБВ ДНУ, 2021. – 60 с.
12. Chaitanya R. K. Text Detection and Recognition: A Review / Chaitanya R. K., Ashwini B. B. // *International Research Journal of Engineering and Technology (IRJET)*. – 2019. – Vol. 4, Issue 6. – P. 179–185. – Regime of access: <https://www.irjet.net/archives/V4/i6/IRJET-V4I629.pdf>, free (date of the application: 05.09.2023).
13. OpenCV [Електрон. ресурс]. – Режим доступу: <https://opencv.org/>, вільний (дата звернення: 05.09.2023).

References

1. Borisyyuk, F., Gordo, A., & Sivakumar, V. (2018). Rosetta: Large scale system for text detection and recognition in images. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)* (pp. 71–79). Association for Computing Machinery. DOI: [10.1145/3219819.3219861](https://doi.org/10.1145/3219819.3219861)
2. Li, H., Wang, P., Shen, C., & Zhang, G. (2019). Show, attend and read: A simple and strong baseline for irregular text recognition. *Proceedings of the 33rd AAAI Conference on Artificial Intelligence* (pp. 8610–8617). AAAI Press. DOI: [10.1609/aaai.v33i01.33018610](https://doi.org/10.1609/aaai.v33i01.33018610)
3. Kim, W., & Kim, C. (2009). A new approach for overlay text detection and extraction from complex video scene. *IEEE Transactions on Image Processing*, 18(2), 401–411. DOI: [10.1109/tip.2008.2008225](https://doi.org/10.1109/tip.2008.2008225)
4. He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 2961–2969). Institute of Electrical and Electronics Engineers (IEEE). DOI: [10.1109/iccv.2017.322](https://doi.org/10.1109/iccv.2017.322)
5. Wang, W., Xie, E., Song, X., Zang, Y., Wang, W., Lu, T., Yu, G., & Shen, C. (2019). Efficient and accurate arbitrary-shaped text detection with pixel aggregation network. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 8439–8448). Institute of Electrical and Electronics Engineers (IEEE). DOI: [10.1109/iccv.2019.00853](https://doi.org/10.1109/iccv.2019.00853)
6. Long, S., Ruan, J., Zhang, W., He, X., Wu, W., & Yao, C. (2018). TextSnake: A flexible representation for detecting text of arbitrary shapes. *Proceedings of the 15th European Conference on Computer Vision (ECCV)* (pp. 19–35). Springer. DOI: [10.1007/978-3-030-01216-8_2](https://doi.org/10.1007/978-3-030-01216-8_2)
7. Zhu, Y., Chen, J., Liang, L., Kuang, Z., Jin, L., & Zhang, W. (2021). Fourier contour embedding for arbitrary-shaped text detection. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3122–3130). Institute of Electrical and Electronics Engineers (IEEE). DOI: [10.1109/CVPR46437.2021.00314](https://doi.org/10.1109/CVPR46437.2021.00314)
8. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). Institute of Electrical and Electronics Engineers (IEEE). DOI: [10.1109/cvpr.2016.90](https://doi.org/10.1109/cvpr.2016.90)
9. Hong, Y., Pan, H., Sun, W., & Jia, Y. (2023). Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes. *IEEE Transactions on Intelligent Transportation Systems*, 24(3), 3448–3460. DOI: [10.1109/TITS.2022.3228042](https://doi.org/10.1109/TITS.2022.3228042)
10. Elagouni, K., Garcia, C., & Sbillot, P. (2011). A comprehensive neural-based approach for text recognition in videos using natural language processing. *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*. Association for Computing Machinery. DOI: [10.1145/1991996.1992019](https://doi.org/10.1145/1991996.1992019)
11. Matsuha, O. M. (2021). *Navchalnyi posibnyk do vvyvchennia kursu «Informatsiini tekhnologii rozpoznavannia obraziv»*. RVV DNU.
12. Chaitanya, R. K., & Ashwini, B. B. (2019). Text Detection and Recognition: A Review. *International Research Journal of Engineering and Technology (IRJET)*, 4(6), 179–185.

Retrieved from <https://www.irjet.net/archives/V4/i6/IRJET-V4I629.pdf>
13. OpenCV team. (2023). OpenCV. Retrieved from <https://opencv.org/>

Рецензент: д-р техн. наук, проф. А.Л. Литвинов, Харківський національний університет міського господарства імені О.М. Бекетова, Україна.

Автор: КАНДИБА Валентина Валентинівна
здобувач вищої освіти 2-го курсу магістратури
навчально-наукового інституту енергетичної,
інформаційної та транспортної інфраструктури
Харківський національний університет міського
господарства імені О.М. Бекетова
E-mail – valentyyna.kandyba@kname.edu.ua

Автор: КУШНІР Олег Ігорович
здобувач вищої освіти 2-го курсу магістратури
навчально-наукового інституту енергетичної,
інформаційної та транспортної інфраструктури
Харківський національний університет міського
господарства імені О.М. Бекетова
E-mail – oleg.kushnir@kname.edu.ua

Автор: БРЕДІХІН Володимир Михайлович
кандидат технічних наук, доцент, доцент кафедри
комп'ютерних наук та інформаційних технологій
Харківський національний університет міського
господарства імені О.М. Бекетова
E-mail – bredixinv@gmail.com
ID ORCID: <https://orcid.org/0000-0002-6063-5046>

Автор: ХОРОШИЛОВА Ірина Олександрівна
кандидат економічних наук, доцент, доцент
кафедри обліку і оподаткування
Харківський національний автомобільно-дорожній
університет
E-mail – verbytska67@gmail.com
ID ORCID: <https://orcid.org/0000-0001-5343-5161>

STUDY OF MACHINE LEARNING TOOLS AND ALGORITHMS FOR RECOGNITION AND DIGITALISATION OF SALES RECEIPTS

V. Kandyba¹, O. Kushnir¹, V. Bredikhin¹, I. Khoroshylova²

¹O.M. Beketov National University of Urban Economy in Kharkiv, Ukraine

²Kharkiv National Automobile and Highway University, Ukraine

This article discusses the issue of processing images of sales receipts for subsequent text information extraction using OCR methods. This application is helpful for maintaining a family budget or for conducting accounting in small companies. The main problem with recognising receipts is the low quality of ink and printing paper, which is why it wrinkles and tears easily, and printed letters quickly fade. The study is based on a series of algorithms based on stepwise methods and integrated image transformation methods that can significantly improve the resulting character recognition. The step-by-step methods localise the text, carry out classification, segmentation, and text recognition, and remove the background part at each algorithm stage. Since they do not depend on the size of the dictionary, they can be used to recognise text from images regardless of its size. To solve the problem, we proposed a unique algorithm for image normalisation, which includes finding a receipt in the image, processing the resulting image area, removing shooting defects and media defects, and using a neural network to process and restore characters. We used the EAST (Efficient and Accurate Scene Text Detector) algorithm implemented using a convolutional neural network (CNN) for the text-finding process. Based on a comparison of the performance of the models in terms of their size and H-mean value, we selected the ddrnet23-slim neural network for the test images. The developed application can significantly increase the accuracy of text information recognition and, simultaneously, is small in size. The developed system recognises characters with reasonably high accuracy and shows the accuracy of the recognition result at a level of 97% and higher. The proposed system can be used: to detect and recognise characters by automatically scanning and updating invoice fields in the database; to extract text from an image and automatically convert it to digital format and update it in the database; as a tool for detecting, recognising, and understanding texts.

Keywords: dataset, neural network, digital technologies, binarisation, sales receipt, classification, OCR.