UDC 004.352.242:[003.344+003.349 DOI: 10.31673/2412-4338.2025.048901

Ivanov Anatolii
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv
ORCID: 0009-0006-5370-445X

Onyshchenko Viktoriia
University of Warmia and Mazury in Olsztyn, Poland
ORCID: 0000-0002-3126-2260

TEXT EXTRACTION FROM SUPERMARKET PRICE TAGS IN ENGLISH AND CYRILLIC WITH OPEN-SOURCE OCR

Abstract: *Supermarket price tags often contain small, densely formatted text presented under challenging visual conditions such as glare, low contrast, irregular lighting, and the presence of both Latin and Cyrillic scripts. These factors significantly complicate text extraction, particularly for individuals with low vision, and reduce the effectiveness of general-purpose OCR systems commonly used in assistive applications. Reliable recognition of this text is crucial for tools intended to support independent navigation and price comprehension in retail environments. This article investigates the feasibility of using lightweight, open-source OCR engines to extract accurate and readable text from tightly cropped images of supermarket price tags in two scripts: English (Latin) and Cyrillic.*

This study evaluates three publicly available OCR frameworks—Tesseract, EasyOCR, and PaddleOCR—selected for their widespread accessibility, broad multilingual support, and compatibility with resource-constrained devices such as smartphones and embedded systems. To ensure a representative and diverse benchmark, a unified dataset was compiled from multiple open-source collections and organized into Latin-only, Cyrillic-only, and mixed-language subsets. Each OCR model was tested under both clean conditions and a variety of synthetic distortions designed to mimic real-world retail environments, including blur, contrast degradation, and perspective warp.

The article provides a detailed description of the dataset construction process, preprocessing techniques used to enhance image quality, evaluation methodology, and the metrics applied to measure recognition reliability. It also addresses practical challenges encountered during the study, such as inconsistent or noisy annotations and the difficulties posed by mixed-script content frequently found in bilingual retail contexts. By examining the strengths, limitations, and robustness of each OCR engine, this work offers guidance for developers and researchers creating OCR-based assistive technologies aimed at improving retail accessibility for users with visual impairments.

Keywords: *optical character recognition (OCR), price tag recognition, multilingual text, open-source models, text extraction, image analysis.*

Іванов Анатолій Ігорович
Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ
ORCID: 0009-0006-5370-445X

Онищенко Вікторія Валеріївна
Вармінсько-Мазурський університет в Ольштині, Польща
ORCID: 0000-0002-3126-2260

РОЗПІЗНАВАННЯ ТЕКСТУ НА СУПЕРМАРКЕТНИХ ЦІННИКАХ АНГЛІЙСЬКОЮ ТА КИРИЛИЦЕЮ ЗА ДОПОМОГОЮ ВІДКРИТИХ OCR-РІШЕНЬ

Анотація: *Цінники в супермаркетах часто містять дрібний, щільно форматований текст, представлений у складних візуальних умовах – з відблисками, низьким контрастом, нерівномірним освітленням. Такі фактори істотно ускладнюють читання тексту, особливо для людей із порушеннями зору, і знижують ефективність загальновживаних OCR-систем, які зазвичай застосовуються в допоміжних технологіях. У цій статті досліджується можливість використання легковагових, відкритих OCR моделей для отримання точного та читабельного тексту зі щільно обрізаних зображень цінників на двох мова: англійській (латиниці) та українській (кирилиці).*

У дослідженні оцінюються три публічно доступні OCR-фреймворки — Tesseract, EasyOCR та PaddleOCR — відібрані завдяки їх широкій доступності, підтримці багатьох мов і сумісності з пристроями, що мають обмежені обчислювальні ресурси, включно зі смартфонами та вбудованими системами. Для забезпечення *репрезентативної та різноманітної вибірки був сформований уніфікований датасет, зібраний із кількох відкритих джерел і організований у підмножини з латинським текстом, кириличним текстом та змішаним мовним наповненням. Кожну з OCR-моделей було протестовано як у «чистих» умовах, так і за різноманітних синтетичних спотворень, що імітують реальні роздрібні сценарії, зокрема розмиття, погіршення контрасту та перспективні деформації.*

Дана робота містить опис процесу побудови датасету, методів попередньої обробки, застосованих для підвищення якості зображень, методології оцінювання та метрик, використаних для вимірювання надійності розпізнавання. Також розглянуто практичні виклики, що виникли під час роботи, зокрема неконсистентні анотації, а також труднощі, пов'язані зі змішаними скриптами, які часто зустрічаються у двомовних роздрібних середовищах. Завдяки аналізу сильних сторін, обмежень та стійкості кожного OCR-двигуна це дослідження пропонує рекомендації для розробників і науковців, що створюють OCR-інструменти для асистивних технологій, спрямованих на підвищення доступності роздрібної торгівлі для користувачів із порушеннями зору.

Ключові слова: *оптичне розпізнавання символів (OCR), розпізнавання цінників, багатомовний текст, моделі з відкритим вихідним кодом, аналіз зображень.*

1. Introduction

Reading price tags on supermarket shelves poses a significant challenge for individuals with low vision, especially in visually dense and cluttered shelf environments. Furthermore, locating and interpreting small or irregularly printed text in poorly lit or crowded retail spaces can be a non-trivial task even for fully sighted individuals. So, for individuals with impaired vision, the difficulty magnified, potentially leading to dependence on others, difficulty with price comparison and loss of independence during shopping.

Advances in mobile computing, computer vision, and speech synthesis technologies have created opportunities for developing assistive systems capable of being some sort of translator between visual and auditory feelings. At the core of such systems is the ability to extract semantic content from visual scenes. For example, in the retail domain, this possibility involves reading product information, prices, and promotional details directly from shelf labels or tags. Optical character recognition (OCR) serves as one of the foundational steps in this pipeline, converting images of printed text into machine-readable characters.

At the same time, existing assistive technologies frequently rely on general-purpose OCR solutions that were not designed for the specific visual challenges posed by supermarket environments. Price tags often use small fonts, stylized layouts, non-standard character placement, and may be obstructed by shelf edges, have lighting artefacts, and products may have text around on photo. Furthermore, multilingual content is common in international or bilingual retail contexts, which introduces additional complexity when OCR models are not adequately trained for mixed-script scenarios. These factors can degrade recognition accuracy and undermine the reliability of assistive systems in real-world use.

To address these limitations, the current work investigates the feasibility of using lightweight, open-source OCR systems to extract readable text from tightly cropped price tags. This task is considered a critical subcomponent of broader assistive applications, where the goal is to convert shelf information into accessible formats such as speech or haptics in near real-time. The focus of this article is exclusively on the OCR component: recognizing the content of a price tag once it has been visually isolated, assuming that detection and localization are handled separately by upstream modules.

Three widely available OCR engines (Tesseract, EasyOCR, and PaddleOCR) are evaluated under controlled experiments using real-world price tag datasets containing text in English and Cyrillic scripts. These tools represent a diverse set of architectural paradigms and resource requirements, from classical LSTM-based recognition (Tesseract) to modern deep learning pipelines (EasyOCR and PaddleOCR). The primary goals are to benchmark their baseline performance and to assess their robustness under realistic conditions such as blur, low contrast, skewed perspective, and image compression. These are the main artefacts that commonly affect mobile-captured images in uncontrolled environments.

2. Related work

Optical character recognition (OCR) remains a fundamental component in computer vision systems, connecting image data and semantic understanding, particularly in assistive technologies, document processing, and retail automation. Modern OCR systems have evolved from template-matching and handcrafted pipelines to deep learning architectures capable of recognizing text under varied conditions. Yet, small font sizes, complex layouts, low contrast, and mixed-language content continue to pose challenges, particularly in real-world settings such as price tags, receipts, and medicine packaging.

2.1 OCR Techniques for Non-Standard Text

Contemporary OCR architectures commonly rely on convolutional neural networks (CNNs) combined with sequence models for text line decoding. The CRNN (Convolutional Recurrent Neural Network) architecture remains a widely used baseline that incorporates CNN features with bi-directional LSTMs and CTC loss to transcribe sequences [1-2]. Despite its efficiency and moderate data requirements, CRNN-based systems struggle when input images exhibit severe perspective distortion, curved baselines, or stylized text. To address these limitations, attention-based models and encoder-decoder architectures have become prominent.

At the same time, preprocessing remains a critical enhancement layer for OCR on degraded inputs. Spatial Transformer Networks (STNs) can perform different transformations to normalize skewed or distorted text regions before recognition, improving performance on curved or rotated text lines [3]. Similarly, photometric normalization and adaptive thresholding have been used to handle regions with low contrast or poor illumination. In addition, recent research indicates that end-to-end models trained directly on down sampled inputs (e.g., 60–75 DPI) can surpass super-resolution OCR pipelines, achieving near-perfect character recognition by implicitly learning robust features [4].

Multilingual OCR introduces additional complexity, particularly when scripts with similar glyphs (e.g., Latin and Cyrillic) co-occur. Script-agnostic training, character-level embedding, and hybrid decoding (with script detection modules) have been proposed to mitigate errors caused by script confusion. Yet, recognition accuracy remains uneven across scripts, and models trained primarily on English or Latin alphabets tend to degrade significantly when confronted with low-resource or cursive scripts [5]. These gaps are especially relevant for regions where supermarket tags and product packaging frequently mix multiple scripts.

2.2 OCR Usage in Real-World Applications

Real-world OCR scenarios evolve significantly from clean document scans. Applications such as receipt understanding, food labelling, and pharmaceutical inventory management involve document-like layouts with non-standard spacing, varied fonts, and substantial noise. The ReceiptSense benchmark introduced a large-scale, multilingual dataset of grocery receipts featuring Arabic-English content [6]. It revealed that even advanced neural OCR models struggle with column misalignment, font variability, and faded thermal printing. While Tesseract performed best on printed English regions, it faltered on mixed-script zones. Transformer models were more flexible in structure but required GPU resources and lacked accuracy on small or low-contrast fonts.

A similar benchmark focused on South African food packaging highlighted how packaging-specific distortions, like reflections, text curvature, and stylized layouts, impact OCR performance [7]. In this domain, Tesseract surprisingly outperformed newer Transformer models on small print and layout-dense regions like nutrition labels. These findings underline that classic OCR engines, when paired with proper page segmentation or layout hints, can still offer strong baselines in structured domains.

Medicine packaging poses a unique OCR challenge due to the presence of small, embossed, or printed on curved surface identifiers such as expiry dates and batch codes. A photometric preprocessing pipeline introduced in [8] used directional lighting and edge map fusion to isolate embossed characters before applying OCR, showing that image acquisition conditions can be as important as the recognition model itself. Similarly, studies in pharmaceutical automation suggest that OCR pipelines must be designed to handle reflective surfaces, minute alphanumeric strings, and near-invisible ink markings.

2.3 Price Tag OCR: Challenges and Specialized Approaches

Supermarket price tag recognition is a focused yet impactful area of OCR research. Unlike general text or document OCR, shelf tags pose specific challenges: small, dense text, visual clutter, angled or curved surfaces, and frequent occlusions. At the same time, they are often photographed under poor conditions, like glare, blur, or weak focus. Still, price tag OCR remains underexplored due to limited public datasets and diverse scripts in retail settings.

The Unitail benchmark [9] introduced a pipeline combining text recognition and product identification by matching printed descriptors (e.g., “500g”, “chocolate”) with a detailed product database. This showed that small packaging text helps distinguish similar SKUs. OCR was used for both text detection and product matching, highlighting its retail value.

Laptev et al. [10] focused on shelf tags using a YOLOv4-tiny detector and EasyOCR for text extraction. They found that lightweight detectors can localize tags in cluttered scenes, and pretrained OCR models reached about 95% accuracy on well-cropped tags. Their use of Cyrillic data exposed weaknesses in OCR systems trained on Latin text.

Recent work on Korean shelf tags [11] and Roboflow datasets underscores the variation in tag design, font, and language, demanding flexible or fine-tuned OCR models. Some researchers trained digit-specific price readers, treating prices as structured numbers and detecting digits and separators separately, improving numeric accuracy under difficult conditions [12].

Yet, key issues persist: OCR performance drops sharply with blur or glare, and models often misread Cyrillic or other non-Latin scripts [6][10]. There is also minimal research on structured field extraction from tags (e.g., identifying which part is price, product, or unit), leaving room for layout-aware OCR or multimodal learning techniques.

3. Purpose of the study

The purpose of this study is to investigate the viability and effectiveness of open-source OCR engines in recognizing text from tightly cropped supermarket price tag images. This research addresses a critical sub problem in assistive technologies focused on individuals with low vision, for whom accurate and timely interpretation of price tag information is essential to fostering independence and equitable access in retail environments.

Supermarket price tags present a unique OCR challenge due to their small physical size, dense layout, varied typography, and exposure to real-world distortions such as blur, glare, and background clutter. These issues are escalated even more when the text appears in non-Latin scripts, such as Cyrillic, or when multiple languages are present, which is common in our country. Despite the growing availability of OCR frameworks, there is a lack of focused evaluation on their ability to handle these specific conditions - particularly within lightweight systems suitable for mobile or embedded devices.

This study focuses on evaluating three publicly available OCR systems: Tesseract, EasyOCR, and PaddleOCR, as candidate solutions for integrating into assistive applications. These models were selected based on their open-source availability, multilingual capabilities, and support for deployment in resource-constrained environments. The primary goal is to benchmark these systems in terms of their accuracy, robustness, and responsiveness when applied to price tag OCR, with emphasis on English and Cyrillic text recognition.

4. Methodology and Experiments

4.1 Objective

The primary objective of this study is to evaluate the performance of open-source OCR systems on small, tightly cropped supermarket price tags. The experiments aim to assess not only baseline recognition accuracy but also the robustness of each system under realistic distortions and multilingual settings, with a particular emphasis on Cyrillic and English scripts. The goal is to identify lightweight models that are viable for assistive applications where speed, accuracy, and language coverage are critical.

4.2 OCR Models Selection

Three open-source OCR systems were selected based on their accessibility, multilingual support, and deployment feasibility:

– Tesseract OCR, being one of the most established open-source OCR engines, often used as a baseline in many academic and industrial settings. It is based on LSTM networks for character sequence prediction and supports over 100 languages. Its architecture optimized for character segmentation followed by sequence decoding, making it effective for well-aligned and clean text.

– EasyOCR, which combines modern convolutional backbones with attention-based sequence decoders. It supports more than 80 languages out of the box and is known for ease of use and fast inference. Its modularity allows experimentation with custom detection and recognition models.

– PaddleOCR provides an end-to-end deep learning pipeline including text detection, classification, and recognition. It supports multilingual OCR and includes lightweight versions suitable for mobile deployment. Its flexible design and broad model zoo make it a strong candidate for domain adaptation.

Each model was also evaluated without fine-tuning to assess practical readiness and generalization to price tag domains.

4.3 Used datasets

To ensure a good foundation for evaluating the price tag text extraction system, data were collected from multiple Roboflow datasets directly related to price tag OCR technology. All selected datasets focus on retail environments appearing under diverse real-world conditions. These datasets were merged to create a more comprehensive and representative source of labelled data. Some of the used price tags can be seen in Fig. 1. After collecting and analyzing several relevant datasets, all images and annotations were unified into a consistent structure and label format. Based on the content and language characteristics of the text within the images, the final dataset was organized into three distinct subsets:

- Latin dataset, containing price tags written in Latin characters (e.g., English).
- Cyrillic dataset, featuring price tags with Cyrillic text (e.g., Ukrainian).
- Mixed dataset, which includes images containing both Latin and Cyrillic text, reflecting realistic multilingual shelf scenarios often present in international or border-region retail settings.

Fig. 1. Some price tags presented in a compiled dataset.

This structure allows targeted evaluation of OCR model performance under different linguistic and visual conditions while maintaining consistency across annotation formats and image preprocessing procedures.

4.4 Experimental Pipeline

The evaluation process was designed to simulate real-world OCR deployment on previously isolated price tags. The following pipeline was adopted in the following way:

- Preprocessing. In this step, input images were preprocessed to normalize lighting, improve contrast, or enhance sharpness. Techniques such as grayscale conversion, adaptive thresholding, and unsharp masking were used.
- OCR Inference. Cropped and optionally enhanced images processed by each OCR model using its default settings, with no language fine-tuning or domain-specific retraining applied.
- Output Matching. The raw OCR outputs were compared with correct annotations on a per-tag basis, allowing error calculation at both character and word levels.
- Metric Logging. For each sample and model, error metrics were computed and logged, facilitating later statistical analysis.

This pipeline was repeated under both clean and distorted image conditions to measure resilience and reliability across realistic inputs.

4.5 Image Distortion Tests

To evaluate the robustness of OCR models under non-ideal imaging conditions, a series of synthetic distortions was applied to clean images. These distortions mimic common issues encountered during mobile capture in supermarkets:

- Gaussian Blur, simulating camera shake or slight defocus.
- Low Contrast, which reflects poor lighting.
- Perspective Warp, which represents images captured at oblique angles relative to the shelf surface.
- JPEG Compression, including down sampling or lossy storage formats used by handheld devices.
- Synthetic Glare Overlay: Simulates reflections from plastic label covers or glossy surfaces.

Each transformation was applied using standard Python libraries with parameters selected to represent typical user-captured image conditions. By applying these perturbations, the study captures the practical performance boundaries of the OCR models in everyday retail use cases.

4.6 Evaluation Metrics

Performance was measured using the following metrics:

– Character Error Rate (CER), measuring the edit distance between predicted and reference text, normalized by the total number of characters. Lower values indicate fewer recognition errors on the character level.
– Word Accuracy (%), representing the percentage of correctly recognized words out of all words in the ground truth. It captures recognition precision at the lexical level.
– Exact Line Match (%): Indicates the proportion of text lines that are perfectly recognized without any errors. It reflects the model's ability to reproduce complete lines exactly as in the reference.

These metrics reflect both fine-grained (character-level) and task-level (line transcription) accuracy, relevant for speech-based assistive output.

4.7 Results and Analysis

Two sets of evaluations were conducted. First, OCR accuracy was measured on clean, tightly cropped price tag images across three language settings: Cyrillic-only, Latin-only, and mixed scripts. Table 1 summarizes the Character Error Rate (CER), Word Accuracy, and Exact Line Match metrics.

Table 1

Baseline performance (CER, Word Accuracy) for each model on clean input

Model	Dataset	Character Error Rate (CER)	Word Accuracy (%)	Exact Line Match (%)
Tesseract	Cyrillic Only	9.2	86.4	77.3
	Latin Only	5.4	92.5	85.2
	Mixed (Cyr+Latin)	11.1	80.2	77.3
EasyOCR	Cyrillic Only	5.5	87.4	76.5
	Latin Only	3.3	95.7	94.2
	Mixed (Cyr+Latin)	7.7	85.8	80.1
PaddleOCR	Cyrillic Only	16.1	65.2	52.2
	Latin Only	2.2	97.3	92.6
	Mixed (Cyr+Latin)	14.7	72.4	59.3

In the second set of experiments, robustness was tested under six synthetic distortions, applied to the mixed-language tag dataset. Word accuracy was recorded in each scenario (Table 2).

Table 2

OCR Word Accuracy (%) under different distortions (Mixed Language Dataset)

Model	Clean image	Blur	Low Contrast	Warp	JPEG Compression	Glare
Tesseract	80	72	70	68	74	64
EasyOCR	86	80	77	76	82	68
PaddleOCR	72	63	65	60	67	55

5. Results and Discussion

The experimental results reveal several important trends in OCR model performance on supermarket price tags. Results from Table 1 are also presented in Fig. 2 for easier analysis.

First of all, on clean, tightly cropped tag images, EasyOCR consistently outperformed Tesseract and PaddleOCR, achieving the highest word accuracy across all script conditions. Its performance was particularly notable on Latin-only tags, with 95.7% word accuracy and 94.2% exact line match. Even in mixed-script scenarios, it maintained a strong balance between Cyrillic and Latin recognition.

Tesseract, though older and more widely used, produced competitive results—particularly for Cyrillic and mixed-language tags, thanks to its well-established support for Cyrillic and strong segmentation accuracy. However, its performance dropped noticeably under distortion, especially with glare and warping, reflecting its reliance on structured, well-lit inputs.

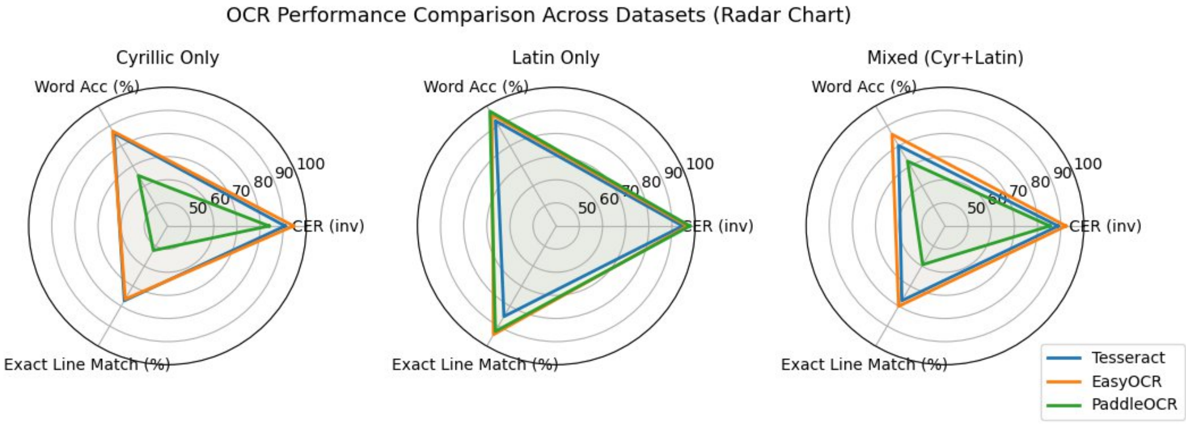

Fig. 2. Qualitative examples of success and failure cases per model

PaddleOCR, while demonstrating state-of-the-art accuracy on Latin text (97.3% word accuracy), suffered significant degradation in Cyrillic and mixed-script recognition. Its limited support for Cyrillic scripts, likely due to model training priorities, contributed to lower baseline and distorted performance, especially on multilingual tags.

When distortion was introduced (Table 2), all models experienced reduced accuracy. EasyOCR proved the most robust, maintaining over 76% word accuracy under moderate blur, low contrast, and warping. Its resilience under JPEG compression (82%) also makes it a viable candidate for mobile or edge applications where storage or bandwidth may be limited.

Tesseract and PaddleOCR were more susceptible to quality degradation. Notably, PaddleOCR dropped to 55% accuracy under glare, reinforcing the importance of pre-capture filtering or hardware-level lighting adjustments for real-world use.

In several cases, particularly when feeding full-price-tag images into OCR engines without prior field segmentation, the models returned extraneous characters or symbols. Some of such results can be seen on Fig. 3. This issue was most noticeable in tags containing barcodes, QR codes, or condensed numeric blocks, where OCR engines attempted to interpret these non-text elements as alphanumeric content. Such false positives were especially common with Tesseract and EasyOCR in Cyrillic tags containing dense barcodes. This highlights a potential need for preprocessing or region-specific masking to exclude visual elements that are not intended for OCR, such as graphics or encoded patterns.

Fig. 3. Artefacts during price tag text recognition.

Overall, these results suggest that EasyOCR is best suited for general-purpose retail OCR tasks, especially in assistive applications, where multilingual support and tolerance to real-world noise are required. Tesseract remains a viable alternative for Cyrillic-dominant environments with limited compute resources, while PaddleOCR may excel in high-contrast, Latin-script deployments such as Western supermarkets.

5. Conclusion

This study evaluated the performance of open-source OCR engines—Tesseract, EasyOCR, and PaddleOCR for recognizing text on supermarket price tags in English, Cyrillic, and mixed-script settings. EasyOCR consistently achieved the best overall results, demonstrating strong multilingual capabilities and robustness to common distortions like blur, glare, and low contrast. PaddleOCR excelled on Latin-only inputs but underperformed in Cyrillic and mixed scenarios. Tesseract, while older, showed solid results on Cyrillic but struggled with noise and glare.

A recurring issue was the presence of extra characters when models processed full price tag images containing barcodes or dense numeric fields. These misreadings suggest a need for preprocessing methods, such as targeted masking, to suppress irrelevant visual content.

One more major limitation was the lack of high-quality Cyrillic price tag datasets. While Latin datasets were readily available, assembling Cyrillic data required merging smaller, inconsistent sources. This likely introduced label noise that may have impacted recognition accuracy. Future work would benefit from creating a high-quality Cyrillic dataset representing diverse fonts and layouts.

In conclusion, EasyOCR offers a strong baseline for assistive retail tools, particularly in multilingual environments. However, reliable deployment, especially for Cyrillic, will require better data, robust preprocessing, and continued domain-specific adaptation.

References

1. Shi B., Bai X., Yao C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. arXiv preprint arXiv:1507.05717. 2015.
2. Li M., Lv T., Chen J., Cui L., Lu Y., Florêncio D., Zhang C., Li Z., Wei F. TrOCR: Transformer-based optical character recognition with pre-trained models. Proceedings of the AAAI Conference on Artificial Intelligence. 2023. Vol. 37 (11). P. 13094–13102.
3. Jaderberg M., Simonyan K., Zisserman A., Kavukcuoglu K. Spatial transformer networks. Advances in Neural Information Processing Systems 28 (NeurIPS 2015). 2015. P. 2017–2025.
4. Gilbey J. D., Schönlieb C.-B. An end-to-end optical character recognition approach for ultra-low-resolution printed text images. arXiv preprint arXiv:2105.04515. 2021.
5. Peng X., Cao H., Setlur S., Govindaraju V., Natarajan P. Multilingual OCR research and applications: An overview. Proceedings of the 4th International Workshop on Multilingual OCR (MOCR 2013). Washington, DC: ACM. 2013.
6. Abdallah A., Mounis M., Abdalla M., Kasem M. S., Mahmoud M., Abdelhalim I., Elkasaby M., ElBendary Y., Jatowt A. ReceiptSense: Beyond traditional OCR – A dataset for receipt understanding. arXiv preprint arXiv:2406.04493. 2024.
7. Nagayi M., Khan A. S., Frank T., Swart R., Nyirenda C. Evaluating OCR performance on food packaging labels in South Africa. arXiv preprint arXiv:2510.03570. 2025.
8. Koponen J., Haataja K., Toivanen P. Text recognition of cardboard pharmaceutical packages by utilizing machine vision. IS&T Electronic Imaging 2021 – Image Processing: Algorithms and Systems XIX. 2021. P. 235-1–235-7.
9. Chen F., Zhang Y., Ahmed U., Zhu C., Savvides M. Unitail: Detecting, reading, and matching in retail scenes. Proceedings of ECCV 2022. Vol. 13677. Springer. 2022. P. 705–722.
10. Laptev P., Litovkin S., Davydenko S., Konev A., Kostyuchenko E., Shelupanov A. Neural network-based price tag data analysis. Future Internet. 2022. Vol. 14 (3).
11. Kratos-AI. Korean Price Tags Image Dataset [Data set]. Hugging Face. n.d.
12. Netzer Y., Wang T., Coates A., Bissacco A., Wu B., Ng A. Street View House Numbers (SVHN) Dataset [Data set]. UFLDL, Stanford University. 2011.