

УДК 004.9

DOI: [10.26102/2310-6018/2025.49.2.010](https://doi.org/10.26102/2310-6018/2025.49.2.010)

Разработка модели компьютерного зрения для детекции областей в визуально насыщенных документах

П.В. Никитин, Р.И. Горохова✉

*Финансовый университет при Правительстве Российской Федерации, Москва,
Российская Федерация*

Резюме. Проблема эффективной автоматизации процесса обработки визуально насыщенных документов является важной частью исследований в области компьютерного зрения. Статья посвящена разработке модели компьютерного зрения для детекции областей в визуально насыщенных документах, с акцентом на обработку чеков с использованием обучения с подкреплением. В условиях растущего объема бумажной документации и необходимости автоматизации обработки данных, эффективная идентификация ключевых элементов чеков (таких как суммы, даты и наименования товаров) становится особенно актуальной. В статье представлена архитектура модели, основанная на сверточных нейронных сетях (CNN), которая обучена на разнообразных наборах данных, включающих изображения чеков различного формата и качества. Рассмотрены методы извлечения информации и алгоритм обучения с подкреплением, который использует обрезанную функцию потерь, цикл обучения с подкреплением, представленный в SpanIE-Recur. Описаны этапы предобработки данных, включая увеличение выборки и нормализацию изображений, что способствует повышению точности детекции. Результаты экспериментов показывают высокую эффективность предложенной модели, достигающую значительных показателей точности и полноты при идентификации областей интереса. Также обсуждаются возможные применения данной технологии в сферах автоматизации бухгалтерского учета, финансового анализа и электронного документооборота. В заключение подчеркивается важность дальнейших исследований в области улучшения алгоритмов обработки изображений и расширения функциональности модели для работы с другими типами документов.

Ключевые слова: визуально насыщенный документ, компьютерное зрение, обучение с подкреплением, детекция объектов, обработка чеков, автоматизация, области документа, предобработка данных, электронный документооборот.

Благодарности: Работа подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финуниверситета.

Для цитирования: Никитин П.В., Горохова Р.И. Разработка модели компьютерного зрения для детекции областей в визуально насыщенных документах. *Моделирование, оптимизация и информационные технологии.* 2025;13(2). URL: <https://moitvvt.ru/ru/journal/pdf?id=1858> DOI: 10.26102/2310-6018/2025.49.2.010

Developing a computer vision model for region detection in visually rich documents

P.V. Nikitin, R.I. Gorokhova✉

*Financial University under the Government of the Russian Federation, Moscow,
the Russian Federation*

Abstract. The problem of efficient automation of visually rich document processing is an important part of computer vision research. This paper is devoted to the development of a computer vision model for region detection in visually rich documents, with an emphasis on receipt processing using reinforcement

learning. In the context of the growing volume of paper documentation and the need to automate data processing, efficient identification of key elements of receipts (such as amounts, dates, and product names) is becoming especially relevant. The paper presents the architecture of the model based on convolutional neural networks (CNN), which is trained on a variety of datasets including receipt images of different formats and qualities. The methods of information extraction and the reinforcement learning algorithm are considered, which uses a trimmed loss function, a reinforcement learning loop presented in SpanIE-Recur. The stages of data preprocessing are described, including sample augmentation and image normalization, which contributes to increasing the detection accuracy. The experimental results show the high efficiency of the proposed model, achieving significant accuracy and recall in identifying regions of interest. Possible applications of this technology in the fields of accounting automation, financial analysis and electronic document management are also discussed. In conclusion, the importance of further research in the field of improving image processing algorithms and expanding the functionality of the model to work with other types of documents is emphasized.

Keywords: visually rich document, computer vision, reinforcement learning, object detection, receipt processing, automation, document areas, data preprocessing, electronic document management.

Acknowledgments: The work was prepared based on the results of research carried out at the expense of budgetary funds under a state assignment from the Financial University.

For citation: Nikitin P.V., Gorokhova R.I. Developing a computer vision model for region detection in visually rich documents. *Modeling, Optimization and Information Technology*. 2025;13(2). (In Russ.). URL: <https://moitvvt.ru/ru/journal/pdf?id=1858> DOI: 10.26102/2310-6018/2025.49.2.010

Введение

В повседневной деятельности предприятий возникает необходимость обработки большого количества визуально насыщенных документов (счета, договоры, заявки, чеки и отчеты) для извлечения структурированной информации [1]. Обработка визуально насыщенных документов остается сложной задачей из-за необходимости учитывать не только текст, но и его положение на странице. Это часто затрудняет даже опытным операторам извлечение нужной информации и создает текущие проблемы обработки [2]. Многие компании продолжают полагаться на ручную обработку, что приводит к ошибкам и увеличенным затратам. В среднем, работник способен обрабатывать около 80 чеков в день, что замедляет процесс и создает негативные последствия ручного труда.

В последнее время все активнее используется внедрение электронного документооборота (ЭДО), однако многие компании сталкиваются с проблемами при передаче и обработке документов, что не решает основную проблему извлечения информации в силу неэффективности действующих систем [3]. Для решения проблемы существует высокая потребность в разработке автоматизированных систем для извлечения данных из визуально насыщенных документов, что может снизить затраты, уменьшить время обработки и минимизировать количество ошибок [4]. Автоматизация может оказать влияние на производственные процессы и привести к значительному сокращению численности сотрудников, заменяя до четырех работников, а в перспективе даже целые отделы.

Процесс извлечения данных из сложных документов может быть улучшен с помощью современных методов Intelligent Document Processing (IDP) и метрики Visually Rich Document Understanding (VRDU), которые учитывают разнообразие макетов и вложенных сущностей [5, 6]. Также возможно использование готовых инструментов. Среди доступных можно выделить такие инструменты, как OCRMiner и CloudScan, которые предлагают различные подходы к автоматизации извлечения данных, применяя машинное обучение и логику, основанную на атрибутах макета [7]. Данные инструменты реализуют различные технологические подходы. Так, OCRMiner [8] использует символичный вывод и системы последовательного извлечения знаний, в то время как

CloudScan [9] применяет рекуррентные нейронные сети для захвата контекста и улучшения качества обработки.

Более интересным методом является дополнение традиционного обучения кросс-энтропии различными обучающими целями, лучше соответствующими критериям оценки. Авторы формулируют задачу извлечения информации как проблему обучения с подкреплением (RL), где информационный экстрактор является политической сетью, а его выход соответствует действиям [10].

Еще один предлагаемый метод – объединенная каркасная система извлечения информации из визуально богатых документов, в которой чтение текста и извлечение информации могут усиливать друг друга через хорошо спроектированный много-модальный блок контекста. В частности, часть чтения текста предоставляет много-модальные функции, такие как визуальные, текстовые и макетные функции [11]. Метод FieldSwap является самым новым и заключается в замене ключевых фраз поля источника на ключевые фразы поля-назначения для генерации новых синтетических примеров поля-назначения для использования в обучении [12].

Целью данного исследования является создание программы, способной извлекать структурированные данные визуально-насыщенных документов, в частности, из чеков, с точностью более 75 %.

Для достижения цели исследования необходимо выполнить подготовку и очистку данных, создание алгоритмов для извлечения текста, оценку конечного результата.

Материалы и методы

Метод «Основа извлечения информации» заключается в использовании такого модуля как SpanIE-Rescur, который извлекает информацию с помощью формулировки «вопроса» [13]. А именно, он заменяет последовательность *labeled* данных, используемых в LayoutLM [14] предсказаниями начальных и конечных позиций выданных ответов.

В математическом плане можно представить, что $D = \{w_0, w_1, \dots, w_n\}$ – это входные данные, состоящие из n входных токенов. Предварительно обученная языковая модель преобразует их в набор скрытых представлений $H = \{h_0, h_1, \dots, h_n\}$. Вопрос под номером $t - q_t$ представлен вектором e_{qt} . Модуль взаимодействия запроса с данными g , который реализуется с помощью слоев внимания, выводит начальную и конечную позицию соответствующего диапазона ответа:

$$\text{start}_t, \text{end}_t = g(H, e_{qt}). \quad (1)$$

Метод «Proximal Policy Optimization» представляет собой алгоритм обучения с подкреплением, который использует обрезанную функцию потерь [15]. Применяется для того, чтобы избежать больших изменений в обновлении политики и при этом гарантировать улучшения. Пусть s_t и a_t обозначают положение агента (state) и действие (action) на временном шаге t , а π_0 – стратегия (policy). При таком раскладе цель PPO определяется следующим образом:

$$J^{\text{PPO}}(\theta) = \mathbb{E} [\min\{\text{ra}(\theta)\hat{A}_{\theta_{\text{old}}}(s_t, a_t), \text{clip}(\text{ra}(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_{\theta_{\text{old}}}(s_t, a_t)\}], \quad (2)$$

где θ – параметры стратегии, $\hat{A}_{\theta_{\text{old}}}(s_t, a_t)$ – преимущество, вычисленное на старых параметрах политики θ_{old} перед каждым обновлением политической итерации, с использованием любого алгоритма оценки преимущества для преобразования вознаграждений [16], а $\text{ra}(\theta) = \frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$ – это отношение между старой и новой политикой. В случае, если отношение $\text{ra}()$ выходит за пределы диапазона $1 - \epsilon$ и $1 + \epsilon$, функция преимущества будет обрезана [17].

В исследовании применим адаптацию оригинальной модели SpanIE-Recur, как стратегию агента и ее доработку с помощью обучения с подкреплением. Единственное отличие заключается в замене обучаемого вектора вопроса e_{qt} на вектор, полученный из предобученного многопредметного текстового кодировщика, который принимает текст вопроса в качестве входных данных. Это может быть полезным для настройки передачи обучения. Рассматривается процесс извлечения информации как последовательный процесс принятия решений, чтобы была возможность использовать обучение с подкреплением, где каждый вопрос к документу соответствует одному шагу в эпизоде. На каждом шаге t модель получает информацию документа D и t -й вопрос q_t в качестве входных данных, то есть агент находится в состоянии $s_t = (D, q_t)$. Модель предсказывает начальную позицию $start_t$ и конечную позицию end_t для ответа, таким образом, агент принимает действие $a_t = (start_t, end_t)$, которое соответствует извлечению строки ответа $\hat{y} = D[start_t: end_t + 1]$ из контекста документа. Затем агент получает вознаграждение r на основе функции вознаграждения, которая описана в следующей секции, и переходит к следующему шагу $t + 1$, пока агент не пройдет все вопросы [18]. Цикл обучения с подкреплением приведен на Рисунке 1 [17].

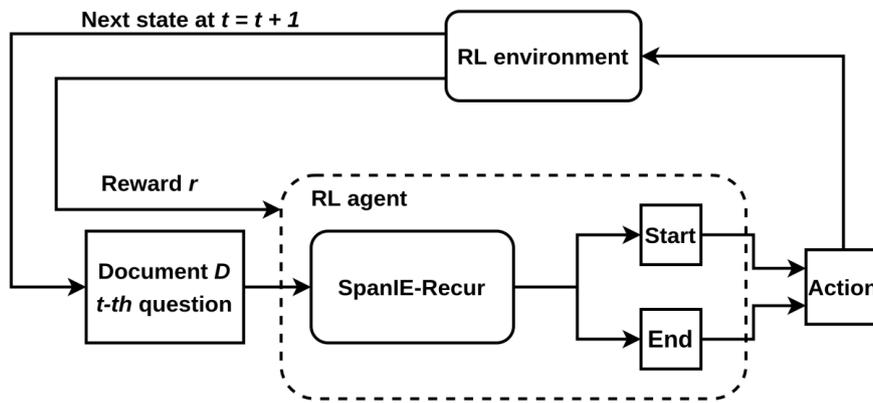


Рисунок 1 – Цикл обучения с подкреплением, представленный в SpanIE-Recur
 Figure 1 – Reinforcement learning loop presented in SpanIE-Recur

Для улучшения модели SpanIE-Recur используются новые функции вознаграждения, которые сводятся в одну, чтобы максимально точно дать оценку действию агента.

Функция вознаграждения на основе соотношения текста. Для сверки схожести текста используется расстояние Левенштейна между выходным y и искомым \hat{y} текстом. Чем больше схожесть, тем выше награда.

$$r_{\text{string}} = 1 - \text{Levenshtein_distance}(y, \hat{y}). \quad (3)$$

Функция вознаграждения на основе соотношения положения. Для сверки положения используются координаты выходного и искомого текста. Пусть $a = (start, end)$ – выходные координаты, а $a_{gt} = (start_{gt}, end_{gt})$ – искомые координаты. Полную формулу, где $interArea$ – расстояние между начальной и конечной позицией, можно увидеть ниже:

$$r_{\text{location}} = \frac{interArea}{(end - start) + (end_{gt} - start_{gt}) - interArea}. \quad (4)$$

Функция вознаграждения на основе соотношения маркировки. Для сверки маркировки, как и в случаях, описанных выше, проводится сверка между выходным и искомым значением. В частности, при вводе вопроса q_t , если модель выдает ответ

$\hat{y} = D[\text{start: end} + 1]$, то модель также должна выдать маркировку, которая совпадает с q_t , то есть:

$$r_{\text{label}} = \begin{cases} 1, & \text{если } \text{label}(\hat{y}) = q_t \\ 0, & \text{если } \text{label}(\hat{y}) \neq q_t \text{ и } \text{label}(\hat{y}) = \text{other} \\ -1, & \text{если } \text{label}(\hat{y}) \neq q_t \text{ и } \text{label}(\hat{y}) \neq \text{other} \end{cases} \quad (5)$$

В данном случае $\text{label}()$ возвращает маркировку при ответе \hat{y} . Other – специальная маркировка, используемая, в случаях, когда невозможно определить правильную. Данное вознаграждение наказывает неправильно указанные маркировки больше, чем использование other.

Конечная формула для вознаграждения агента может быть записана следующим образом:

$$r = a_1 \cdot r_{\text{string}} + a_2 \cdot r_{\text{location}} + a_3 \cdot r_{\text{label}} + a_4 \cdot r_{\text{semantic}}, \quad (6)$$

где a_i – соответствующие веса для каждого компонента вознаграждения [19].

Результаты и обсуждение

Начиная с модели с предварительно обученными весами, происходит дообучение предварительно обученной модели на обучающем наборе. На практике этот обучающий набор разбивается на мелкие поднаборы заданного размера (2 %, 5 %, 10 % или 100 % от всей обучающей выборки). Далее модель дообучается с помощью обучения с подкреплением до $n_{\text{max}} = 100,000$ итераций. Во время этого процесса используются предоставленные каждым поднабором истинные значения, чтобы вычислить функцию вознаграждения. Цель заключается в том, чтобы определить, помогает ли RL-настройка улучшить производительность обычного обучения [17].

Значение метрик качества алгоритма относительно размера обучающего датасета представлены в Таблице 1.

Таблица 1 – Значение метрик качества
Table 1 – The importance of quality metrics

Dataset		Original SL		
		Precision	Recall	F1
CORD	2 %	21,69	2,77	3,04
	5 %	64,33	47,83	51,48
	10 %	79,45	72,98	75,32
	100 %	96,31	94,91	95,39

Исходя из представленных значений метрик качества алгоритма (Precision, Recall и F1) для различных размеров обучающего датасета (CORD), можно сделать следующие выводы:

1. Анализ метрик по размерам датасета:

– при использовании лишь 2 % данных точность (Precision) составила 21,69, что говорит о низком уровне корректности извлечения информации. Показатель полноты (Recall) оказался на уровне 2,77, что указывает на то, что модель практически не охватывает все значимые элементы в данных. F1-мера, как гармоническое среднее Precision и Recall, составила всего 3,04, что свидетельствует о крайне низкой эффективности модели при таком небольшом размере обучающего датасета;

– увеличение объема обучающего датасета до 5 % значительно улучшает показатели: точность повысилась до 64,33, а полнота – до 47,83. F1-мера также

улучшилась, составив 51,48, что указывает на более сбалансированную работу модели, хотя все еще остается ниже желаемых значений;

– при использовании 10 % объема данных результаты продолжают улучшаться: Precision достигает 79,45, Recall – 72,98, а F1-мера – 75,32. Эти показатели демонстрируют более надежную работу алгоритма и говорят о достаточно высоком качестве извлечения данных;

– при использовании полного объема обучающего датасета (100 %) метрики достигают своих наивысших значений: Precision – 96,31, Recall – 94,91, F1 – 95,39. Это подтверждает, что увеличение объема данных позволяет значительно повысить эффективность модели, обеспечивая высокую точность и полноту извлечения информации.

Выявляется общая тенденция значения метрик и размера датасета. Анализ показывает, что качество работы алгоритма прямо пропорционально зависит от объема обучающего датасета. С увеличением процента данных модель становится все более точной и полноценной. Это соответствует общепринятым ожиданиям в области машинного обучения, где более обширные и качественные данные приводят к лучшему обучению модели и, как следствие, к более высоким показателям метрик.

Модель, описанная выше, смогла произвести корректное извлечения данных в 75 % чеков, при использовании 10 % данных. За корректное извлечение считается полностью правильная пометка необходимых регионов, при наличии ошибки хотя бы в одном из полей, результат не засчитывается.

Заключение

Результаты статьи подчеркивают значимость разработки эффективных решений для автоматизации обработки визуально насыщенных документов с использованием современных методов компьютерного зрения. Представленная модель, основанная на сверточных нейронных сетях и методах обучения с подкреплением, демонстрирует высокие результаты в идентификации ключевых элементов, что открывает новые горизонты для автоматизации в таких областях, как бухгалтерский учет и финансовый анализ.

Результаты проведенных экспериментов подтверждают эффективность предложенного подхода и его потенциал для практического применения. Тем не менее, дальнейшие исследования остаются актуальными для улучшения алгоритмов обработки изображений и расширения функциональности модели, что позволит адаптировать технологии к разнообразным типам документов и повысить общую производительность систем автоматизации документооборота. Важно продолжать работу в этом направлении, чтобы обеспечить более высокую точность и надежность в обработке данных, что, в свою очередь, будет способствовать оптимизации бизнес-процессов и повышению их эффективности.

Для улучшения результата планируется внедрение LayoutLM3 [20], модель, выпущенная в 2022 году, специализирующаяся на извлечении структурированных данных из визуально насыщенных документов. Проблема ее использования – необходимость большого количества качественных данных для обучения.

Данную модель не планируется выводить в производство, по объективным причинам: она не создана для реального мира, а существует лишь в целях изучения проблемы. Для вывода необходимо использовать более современные методы и большее количество обучающих данных.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Хуан Д. Алгоритмы извлечения информации из текстов, парсинг веб-страниц с использованием языка программирования Python. *Актуальные исследования*. 2022;(30):21–24.
Huang D. Algorithms for Extracting Information from Texts, Parsing Web Pages Using the Python Programming Language. *Aktual'nye issledovaniya*. 2022;(30):21–24. (In Russ.).
2. Шестакова М.В., Головнина А.А., Головнин О.К. Извлечение знаний из графических и текстовых документов средствами искусственного интеллекта. В сборнике: *Проблемы информатики в образовании, управлении, экономике и технике: сборник статей XXIV Международной научно-технической конференции, 22–23 ноября 2024 года, Пенза, Россия*. Пенза: Пензенский государственный университет; 2024. С. 273–278.
Shestakova M.V., Golovnina A.A., Golovnin O.K. Knowledge Mining from Graphic and Text Documents Based on Artificial Intelligence. In: *Problems of Computer Science in Education, Management, Economy and Technology: Proceedings of 24th International and Technical Conference, 22–23 November 2024, Penza, Russia*. Penza: Penza State University; 2024. P. 273–278. (In Russ.).
3. Чиняков О.Е. Электронный документооборот: свойства и проблемы внедрения *Гуманитарные и политико-правовые исследования*. 2023;(1):43–50.
Chinyakov O.E. Electronic Document Management: Properties and Problems of Implementation. *Humanitarian and Political-Law Studies*. 2023;(1):43–50. (In Russ.).
4. Исачкова Л.Н., Асанова Н.А., Хут С.Ю., Ешугова Ф.Р. Обеспечение экономической безопасности в системе электронного документооборота в условиях цифровой трансформации бизнеса. *Вестник Академии знаний*. 2021;(45):113–117. <https://doi.org/10.24412/2304-6139-2021-11342>
Isachkova L.N., Asanova N.A., Huth S.J., Yeshugova F.R. Ensuring Economic Security in the Electronic Document Management System in the Context of Digital Business Transformation. *Vestnik Akademii znanii*. 2021;(45):113–117. (In Russ.). <https://doi.org/10.24412/2304-6139-2021-11342>
5. Mandvikar Sh. Augmenting Intelligent Document Processing (IDP) Workflows with Contemporary Large Language Models (LLMs). *International Journal of Computer Trends and Technology*. 2023;71(10):80–91. <https://doi.org/10.14445/22312803/IJCTT-V71I10P110>
6. Nicolaieff L., Kandi M.M., Zegaoui Yo., Bortolaso Ch. Intelligent Document Processing with Small and Relevant Training Dataset. In: *2022 International Conference on Intelligent Systems and Computer Vision (ISCV), 18–20 May 2022, Fez, Morocco*. IEEE; 2022. P. 1–7. <https://doi.org/10.1109/ISCV54655.2022.9806100>
7. Wang Z., Zhou Y., Wei W., Lee Ch.-Yu, Tata S. VRDU: A Benchmark for Visually-rich Document Understanding. In: *KDD '23: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 06–10 August 2023, Long Beach, CA, USA*. New York: Association for Computing Machinery; 2023. P. 5184–5193. <https://doi.org/10.1145/3580305.3599929>
8. Ha H.T., Horák A. Information Extraction from Scanned Invoice Images using Text Analysis and Layout Features. *Signal Processing: Image Communication*. 2022;102. <https://doi.org/10.1016/j.image.2021.116601>
9. Aggarwal N., Patra S., Sinha S., Jaiman A., Ghosh D. Data Extraction from Scanned Invoice Documents in Multiple Languages. In: *International Workshop on Signal*

- Processing and Machine Learning (WSPML 2023): Proceedings: Volume 12943, 22–24 September 2023, Hangzhou, China.* 2023. <https://doi.org/10.1117/12.3019910>
10. Чуб В.С. Извлечение признаков рекуррентными нейронными сетями из больших объёмов данных. В сборнике: *За нами будущее: взгляд молодых ученых на инновационное развитие общества: сборник научных статей 3-й Всероссийской молодежной научной конференции, 03 июня 2022 года, Курск, Россия.* Курск: Юго-Западный государственный университет; 2022. С. 536–539.
 11. Хань И., Чэнь Ч., Хэ Х. Искусственный интеллект и технологии языкового анализа. *Столыпинский вестник.* 2024;6(10). URL: <https://elibrary.ru/item.asp?id=75102830>
Han Y., Chen Zh., He H. Artificial Intelligence and Language Analysis Technologies. *Stolypin Annals.* 2024;6(10). (In Russ.). URL: <https://elibrary.ru/item.asp?id=75102830>
 12. Бурнашев Р.Ф., Анварова Л.А. Применение нейронных сетей в автоматическом переводе и обработке естественного языка. *Universum: технические науки.* 2024;(4–1):39–43.
Burnashev R., Anvarova L. Application of Neural Networks in Automatic Translation and Natural Language Processing. *Universum: tekhnicheskie nauki.* 2024;(4–1):39–43. (In Russ.).
 13. Xie J., Wendt J.B., Zhou Y., Ebner S., Tata S. FieldSwap: Data Augmentation for Effective Form-Like Document Extraction. In: *2024 IEEE 40th International Conference on Data Engineering (ICDE), 13–16 May 2024, Utrecht, Netherlands.* IEEE; 2024. P. 4722–4732. <https://doi.org/10.1109/ICDE60146.2024.00359>
 14. Xu Y., Li M., Cui L., Huang Sh., Wei F., Zhou M. LayoutLM: Pre-training of Text and Layout for Document Image Understanding. In: *KDD '20: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 06–10 July 2020, Virtual Event, CA, USA.* New York: Association for Computing Machinery; 2020. P. 1192–1200. <https://doi.org/10.1145/3394486.3403172>
 15. Новиков Н.П., Виноградов В.И. Опыт использования архитектуры сети Transformer для аппроксимации политики агента в обучении с подкреплением. *Моделирование и анализ данных.* 2024;14(2):7–22. <https://doi.org/10.17759/mda.2024140201>
Novikov N.P., Vinogradov V.I. Experience in Using the Transformer Network Architecture to Approximate Agent's Policy in Reinforcement Learning. *Modelirovanie i analiz dannykh.* 2024;14(2):7–22. (In Russ.). <https://doi.org/10.17759/mda.2024140201>
 16. Alves E.L.G., Carvalho C., De Lima P.M., Pinheiro V., Furtado V. Information Extraction from Financial Statements Based on Visually Rich Document Models. In: *Proceedings of the 20th National Meeting on Artificial and Computational Intelligence (ENIAC 2023), 25–29 September 2023, Belo Horizonte, Brazil.* Porto Alegre: Sociedade Brasileira de Computação; 2023. P. 894–908. <https://doi.org/10.5753/eniac.2023.234520>
 17. Huang Yu., Lv T., Cui L., Lu Yu., Wei F. LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking. In: *MM '22: Proceedings of the 30th ACM International Conference on Multimedia, 10–14 October 2022, Lisboa, Portugal.* New York: Association for Computing Machinery; 2022. P. 4083–4091. <https://doi.org/10.1145/3503161.3548112>
 18. Mistry J., Arzeno N.M. Document Understanding for Healthcare Referrals. In: *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI 2023), 26–29 June 2023, Houston, TX, USA.* IEEE; 2023. P. 460–464. <https://doi.org/10.1109/ICHI57859.2023.00067>
 19. Ding Y., Vaiani L., Han C., et al. 3MVRD: Multimodal Multi-task Multi-teacher Visually-Rich Form Document Understanding. In: *Findings of the Association for Computational Linguistics, ACL 2024, 11–16 August 2024, Bangkok, Thailand.* New

- York: Association for Computational Linguistics; 2024. P. 15233–15244.
<https://doi.org/10.18653/v1/2024.findings-acl.903>
20. Tchuitcheu W.C., Lu T., Doods A. Table Representation Learning Using Heterogeneous Graph Embedding. *Pattern Recognition*. 2024;156. <https://doi.org/10.1016/j.patcog.2024.110734>

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Никитин Петр Владимирович, кандидат педагогических наук, доцент, доцент кафедры искусственного интеллекта, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация.
e-mail: pnikitin@fa.ru
ORCID: [0000-0001-8866-5610](https://orcid.org/0000-0001-8866-5610)

Pyotr V. Nikitin, Candidate of Pedagogical Sciences, Docent, Associate Professor at the Department of Artificial Intelligence, Financial University under the Government of the Russian Federation, Moscow, the Russian Federation.

Горохова Римма Ивановна, кандидат педагогических наук, доцент, доцент кафедры информационных технологий, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация.
e-mail: rigorokhova@fa.ru
ORCID: [0000-0001-7818-8013](https://orcid.org/0000-0001-7818-8013)

Rimma I. Gorokhova, Candidate of Pedagogical Sciences, Docent, Associate Professor at the Department of Information Technology, Financial University under the Government of the Russian Federation, Moscow, the Russian Federation.

Статья поступила в редакцию 20.03.2025; одобрена после рецензирования 14.04.2025; принята к публикации 21.04.2025.

The article was submitted 20.03.2025; approved after reviewing 14.04.2025; accepted for publication 21.04.2025.