



Распознавание табличной информации с использованием свёрточных нейронных сетей

Игорь Викторович **Винокуров**

Финансовый Университет при Правительстве Российской Федерации, Москва, Россия

 igvvinokurov@fa.ru

Аннотация. Показана актуальность выявления табличной информации и распознавания её содержимого для обработки отсканированных документов. Описано формирование набора данных для обучения, валидации и тестирования нейронной сети глубокого обучения (DNN) YOLOv5s для обнаружения простых таблиц. Отмечена эффективность использования этой DNN при работе с отсканированными документами. С использованием Keras Functional API сформирована свёрточная нейронная сеть (CNN) для распознавания основных элементов табличной информации – цифр, основных знаков препинания и букв кириллицы. Приведены результаты исследования работы этой CNN. Описана реализация выявления и распознавания табличной информации на отсканированных документах в разработанной ИС актуализации информации в базах данных системы ЕГРН Росреестра.

Ключевые слова и фразы: Свёрточные нейронные сети, нейронные сети глубокого обучения, CNN, DNN, YOLOv5s, Keras, Python

Для цитирования: Винокуров И. В. *Распознавание табличной информации с использованием свёрточных нейронных сетей* // Программные системы: теория и приложения. 2023. Т. 14. № 1(56). С. 3–30. https://psta.psiras.ru/read/psta2023_1_3-30.pdf

Введение

Реализуемый в настоящее время процесс цифровой трансформации бизнес-моделей различного вида организаций предполагает в том числе и их переход на электронные носители информации. Следствием этого является необходимость распознавания содержимого электронных документов с целью последующей обработки их содержимого. В случае эффективного распознавания содержимого электронного документа его автоматическая (в ряде случаев автоматизированная) обработка существенно повышает скорость, и в ряде случаев, качество этого процесса. Отсюда решение задачи распознавания информации на отсканированных копиях документов, являющейся первым этапом преобразования бумажного документа в его электронный аналог, представляет определённый практический и научный интерес.

Одним из возможных подходов к решению этой задачи с приемлемой точностью, наряду с традиционными методами, является использование CNN. В [1] описаны общие принципы распознавания отсканированного изображения с использованием CNN, обученной на характерном для содержимого текстовых документов наборе данных. Однако достаточно часто наиболее важная информация в документе представляется в табличном виде. Выявление только таблицы и распознавание её содержимого позволит для подобного вида документов повысить скорость распознавания необходимой для последующей обработки информации и, как следствие, эффективность этого процесса.

Целью данной работы является разработка ИС, реализующей локализацию таблицы на отсканированном документе и распознавание её содержимого. В этой работе осуществляется обнаружение в отсканированных документах простых типов таблиц, характерных для большинства документов и содержащих только текстовую или числовую информацию. Ограничение типа таблицы и особенности содержащейся в ней информации даёт основание сформировать свой собственный набор данных для обучения и выбрать любую из известных моделей DNN или CNN, демонстрирующих приемлемые результаты детектирования различного типа объектов. После выявления расположения таблицы в документе, для распознавания её содержимого предлагается использовать небольшую модификацию изложенного в [1] метода.

Полученные в работе результаты последовательного применения DNN для детектирования таблиц и разработанной CNN нашли практическое применение для актуализации базы данных системы ЕГРН Росреестра, переходящего в настоящее время на электронные носители информации.

1. Локализация таблицы на изображении

Существующие на данный момент времени методы локализации таблицы на изображении можно разделить на традиционные методы, реализующие обработку изображений и современные методы, в основе которых лежит использование нейронных сетей с архитектурами различных типов.

Традиционные методы локализации таблиц основаны на специфичном структурировании информации и присутствия, в большинстве случаев, горизонтальных и вертикальных линий, их признаков и наличия различных критериев близости и подобия. Так в работах [2] и [3] были предложены методы нахождения табличной информации между ключевыми словами или их комбинациями в начале и конце таблицы и в областях пересечения её горизонтальных и вертикальных линий соответственно. Недостатки подходов очевидны – отсутствие ключевых слов и линий (границ) таблицы приводит к невозможности или крайней неэффективности использования этих методов.

В основе другого подхода [4] лежит метод определения длины строк и столбцов таблицы. После выявления на изображении документа всех горизонтальных и вертикальных линий по их длине, для каждой из них формируется набор из признаков низкого уровня, которые передаются в машину опорных векторов (SVM), реализующей детектирование таблицы. В случае отсутствия линий метод не работает.

Еще один метод [5] предполагает реализовать локализацию и извлечение областей таблицы из изображения документа исходя локальных порогов для межсловного интервала и высоты строки. Основным ограничением этого метода является то, что он обнаруживает области таблицы вместе с окружающими текстовыми областями. Следовательно, его нельзя использовать только для локализации табличной информации.

К традиционным методам можно отнести и метод детектирования в документе табличных структур [6] на основе контура случайно выбранного слова. Метод предполагает, что аналогичные по форме контуры, выявленные по горизонтали и по вертикали и определяют таблицу. Наличие пустых строк в таблице может привести к неэффективности и даже невозможности обнаружения всей табличной информации.

Большую эффективность локализации таблиц на электронном документе реализуют методы, в основе которых лежат CNN или DNN. В таких методах нет явной привязки к структуре таблицы и отсутствует необходимость в предварительной подготовке документа. Однако обучение таких моделей требует большого количества изображений документов, содержащих таблицы – в этом и заключается основная сложность этих методов. Существует несколько наборов данных, которые могут быть использованы для обучения моделей детектирования таблиц – TableBlank

[7], Marmot [8], ICDAR [9], UNLV[10], ICDAR 2019 cTDaR [11] и других. Однако достаточно часто возникает необходимость формирования и собственных наборов данных, что является совсем не тривиальной задачей.

Наиболее эффективными моделями глубокого обучения в настоящее время считаются Fast RCNN и Faster RCNN [12], YOLO [13]. Особенностями работы Fast RCNN и Faster RCNN являются нахождение потенциальных объектов на изображении и разбиение их на регионы с помощью метода селективного поиска, извлечение признаков каждого полученного региона с помощью CNN, последующая классификация методом опорных векторов и уточнение границ регионов с помощью линейной регрессии.

Первой для обнаружения таблиц в документе [14] была использована модель Faster RCNN. Однако в настоящее время лидером по детектированию объектов является модель YOLO. Эта модель реализует меньшее количество действий по сравнению с Faster RCNN и реализует поиск ограничивающих объекты рамок и вероятность их принадлежности тому или иному классу изображений.

Существует несколько версий этой DNN; для решения практических задач в настоящее время используются версии v5, v6 и v7 [15], предварительно обученные на наборе данных *COCO (Common Objects in COntext)*^{[URL](#)}. Время обучения, работы и такие параметры модели как Precision, Recall и AP [16] повышаются от v5 к v7, следовательно, формально версию v7 можно считать более эффективной. Однако на практике эффективность той или иной версии модели определяется особенностями её использования и обучения. Поскольку особых требований к скорости и точности детектирования таблиц нет, а версия v5 [17] показывает вполне приемлемый результат в сочетании с низким потреблением ресурсов и при этом имеет развитые средства сбора информации о процессе обучения и метриках на каждой эпохе, для реализации задачи локализации таблицы на электронном документе была выбрана именно эта версия. Примеры эффективного использования DNN YOLO, обученных на наборе данных ICDAR [11] для детектирования таблиц в электронных копиях документах приведены в [18, 19].

При известной организации этой DNN, единственное, что необходимо сделать — это сформировать наборы данных для её дообучения, валидации и тестирования. Формирование дополнительного обучающего набора данных осуществлялась на нескольких сотнях (порядка 400) отсканированных документов Росреестра с простыми таблицами, содержащими небольшой поясняющий текст в заголовках таблицы и числовую информацию в виде совокупности координат объектов капитального строительства или земельных участков. Для каждого из этих документов с использованием ПО с открытым исходным кодом *labelImg*^{[URL](#)} были сформированы текстовые файлы с метками таблиц — их нормализованными координатами внутри

документа. Обучение сети осуществлялось на 70% от общего количества изображений, валидация на 20% и тестирование на 10%. Как показали результаты проведённых экспериментальных исследований, 30 эпох обучения позволили получить вполне приемлемый результат по основным параметрам работы DNN и её обобщающих способностей, рисунок 1.

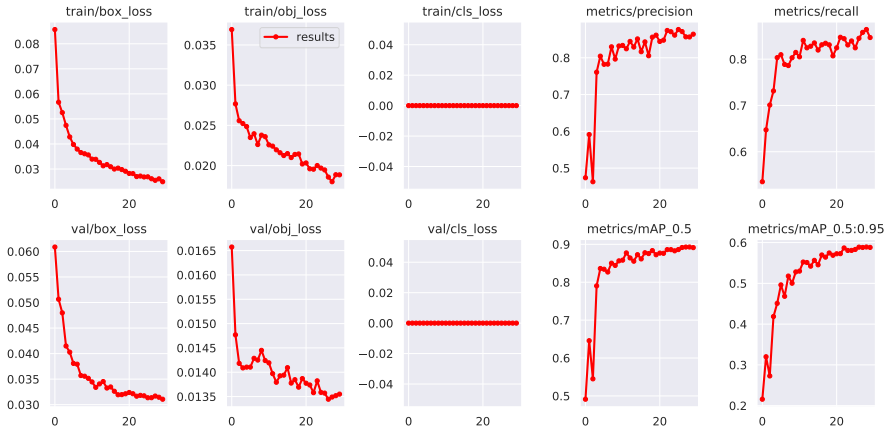


Рисунок 1. Графики функций потерь, precision, recall и mAP_0.5:0.95 при 30 эпохах обучения DNN YOLOv5s

На валидационной выборке при 30 эпохах обучения были получены следующие результаты: $Precision = 0.871$, $Recall = 0.828$, $mAP_{0.5} = 0.592$. Зависимость Precision-Recall приведена на рисунке 2.

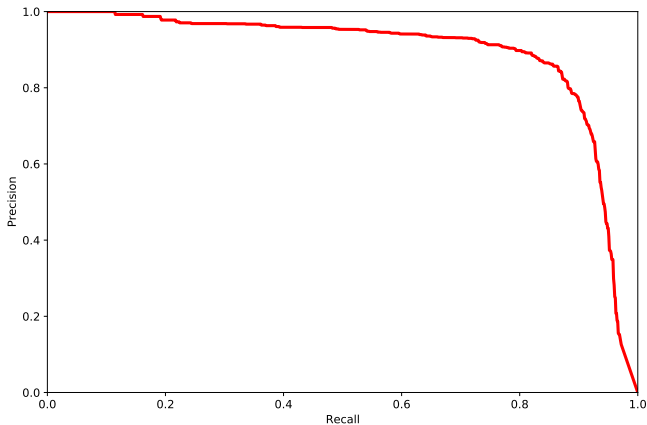


Рисунок 2. Precision-Recall при 30 эпохах обучения DNN YOLOv5s

Дообученная на своём наборе данных DNN YOLOv5s продемонстрировала хороший результат обнаружения таблиц на электронных копиях документах хорошего качества – порядка 98-99%. Пример результата работы сети приведен на рисунке 3.

Table 0.78

Описание земельных участков. Раздел "Описание границ"					
Кадастровый квартал XX:XX:XX:XX			Изменение №		
Сведения о вновь образованных и прекращающих существование узловых и поворотных точках границ					
Условное обознач. точки	Координаты		f, м	Описание закрепления точки	Кадастровая запись
	X	Y			
1н	5976781.18	2226562.32	7.5	По естественным рубежам	
2н	5976907.56	2226513.13	7.5	По естественным рубежам	
3н	5976997.66	2226537.35	7.5	По естественным рубежам	
4н	5877213.76	2226611.14	7.5	По естественным рубежам	
5н	5977406.32	2226628.65	7.5	По естественным рубежам	
6н	5977596.63	2226652.87	7.5	По естественным рубежам	
7н	5977823.31	2226759.99	7.5	По естественным рубежам	
8н	5977829.71	2226853.41	7.5	По естественным рубежам	
9н	5977903.99	2226976.25	7.5	По естественным рубежам	
10н	5977956.51	2227056.59	7.5	По естественным рубежам	

Рисунок 3. Результат обнаружения таблицы с помощью DNN YOLOv5s

2. Распознавание табличной информации

В [1] приведены общие принципы формирования CNN для реализации многоклассовой классификации элементов текста на отсканированных изображениях плохого качества. Продолжающиеся в настоящее исследование этой CNN и набора данных для её обучения позволили сделать следующий вывод – улучшение результатов классификации может быть достигнуто в большей степени за счёт параллельного выявления устойчивых признаков и их последующего суммирования. Иными словами, усложнение последовательной модели CNN из [1] даёт худшие результаты по сравнению с её функциональной моделью, реализуемой *Keras Functional API*^{[URL](#)}.

Для распознавания табличной информации предлагается структура CNN, имеющая два начальных контура. Первый контур состоит из 2-х свёрточных слоев (*Conv2D*^{[URL](#)}) и одного слоя максимизирующего пуллинга (*MaxPooling2D*^{[URL](#)}). Второй – из одного свёрточного слоя (*Conv2D*) и одного слоя максимизирующего пуллинга (*MaxPooling2D*). Все остальные слои CNN повторяют слои из [1] – один свёрточный слой (*Conv2D*) для выявления признаков из результатов суммирования работы 2-х входных контуров, один линейаризирующий слой (*Flatten*^{[URL](#)}) и два полносвязных слоя (*Dense*^{[URL](#)}). Структура CNN приведена на рисунке 4.

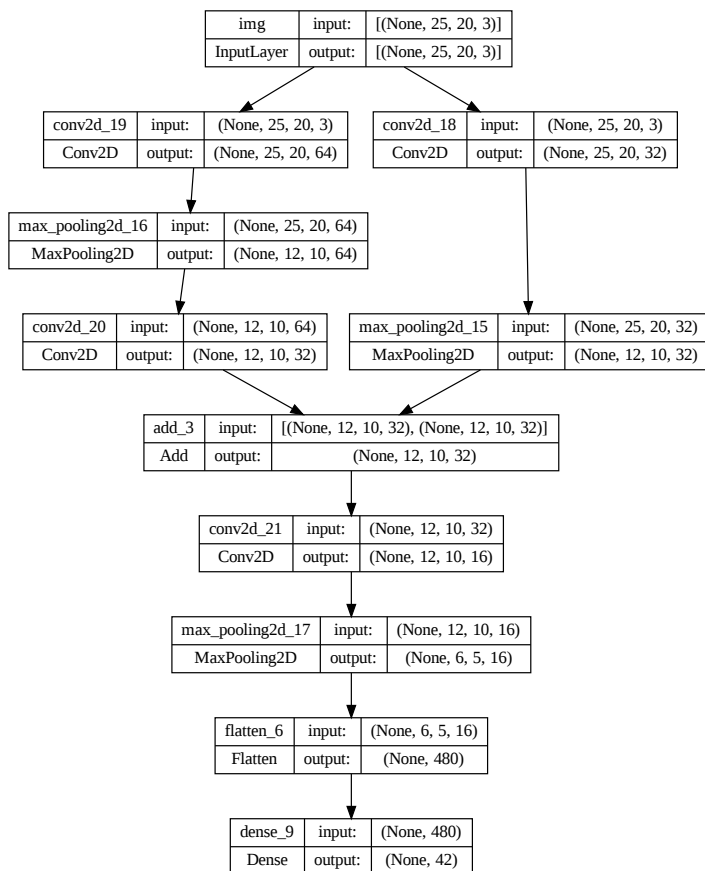


Рисунок 4. Структура CNN для распознавания табличной информации

Все изображения для обучения и валидации приводятся к размеру 20×25 [1]. Функции активации всех нейронов CNN – «*sigmoid*»^(URL). Параметры компиляции и обучения выбраны стандартными для многоклассовой классификации: оптимизатор – «*adam*»^(URL), метрика – «*acc*»^(URL), функция потерь – «*categorical-crossentropy*»^(URL).

Наборы данных для обучения и валидации этой CNN остался таким же, как и в [1], – 42 класса различных изображений цифр, кириллических букв и 2-х знаков пунктуации «.» и «,». Поскольку эти наборы данных являются небольшими – по 10 и 5 изображений на каждый из символов соответственно, при обучении и последующем исследовании работы

CNN осуществлялась их аугментация с использованием генератора пакетов трансформированных данных *ImageDataGenerator*^{URL}. Учитывая особенности предметной области, были выбраны следующие параметры трансформации и их значения: $zoom_range = 0.125$, $shear_range = 0.15$, $rotation_range = 0.15$, $width_shift_range = 1.2$, $height_shift_range = 1.2$, $horizontal_flip = \text{False}$ и $fill_mode = \text{«nearest»}$.

На рисунках 5 и 6 приведены результаты экспериментальных

Результирующая точность модели: 0.975

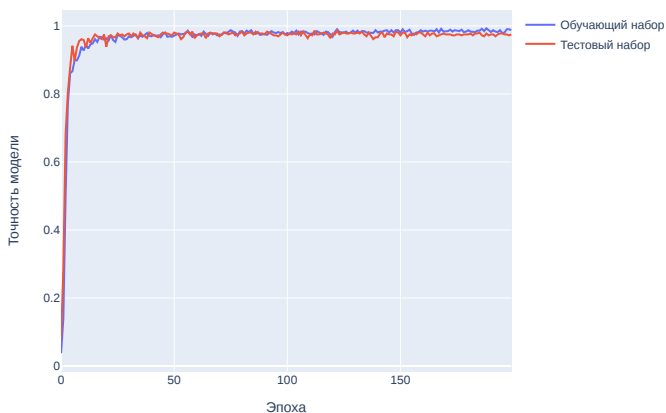


Рисунок 5. Точность модели

Результирующие потери модели: 0.056

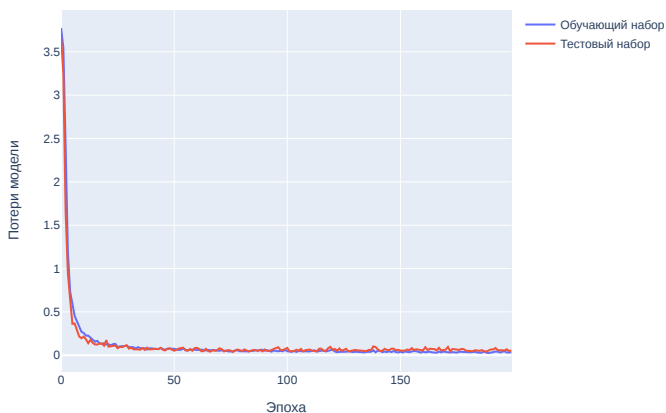


Рисунок 6. Потери модели

исследований соответственно точности и потерь модели при 200 эпохах её обучения, являющимся оптимальным количеством по результатам экспериментальных исследований.

Значения параметров gescall и F1 для модели составили 0.98055 и 0.98029 соответственно, что является достаточно хорошим результатом. Качество работы сформированной CNN при классификации элементов текста показывает следующая матрица ошибок CNN рисунок 7. Как

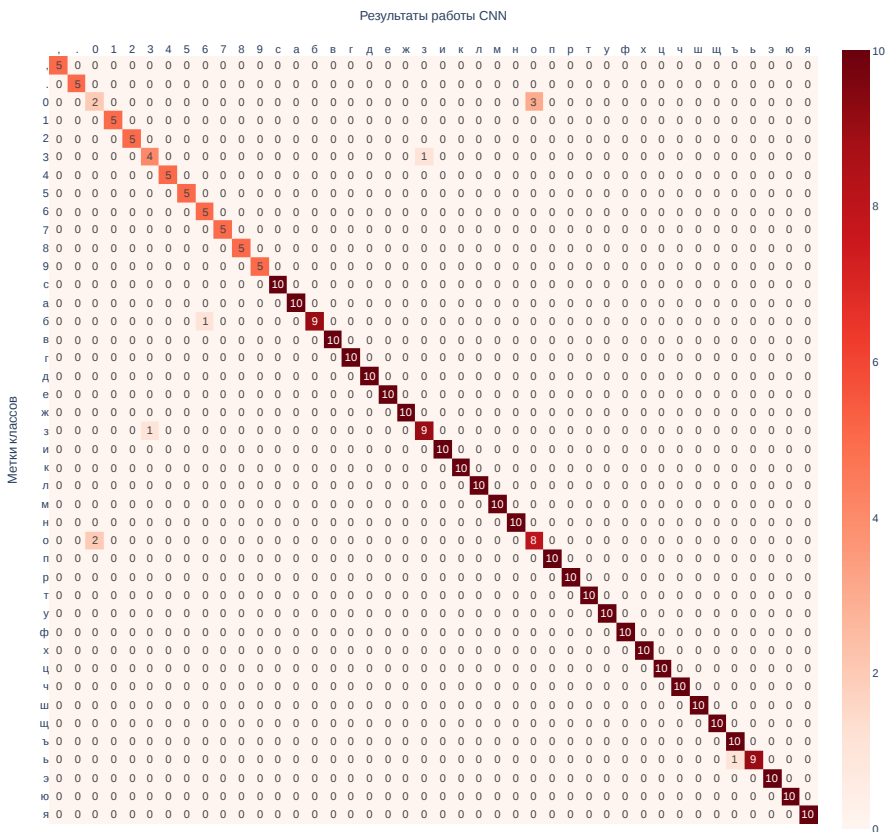


Рисунок 7. Матрица ошибок классификации

видно из этого рисунка, ошибки при классификации элементов текста заключались в схожести цифр 0, 3 и букв «о» и «з».

Пример ошибочной классификации этих цифр и букв приведён на рисунках 8а, и 8б.



(а) Изображения цифр



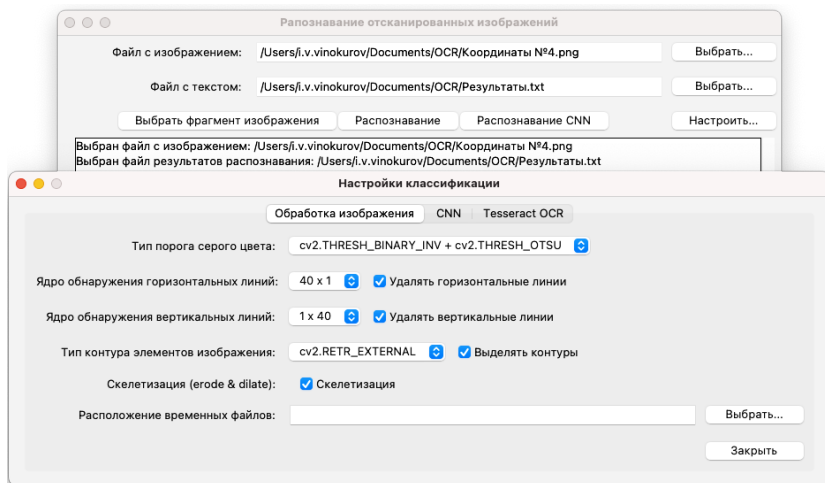
(б) Результаты классификации цифр

Рисунок 8. Ошибки классификации цифр

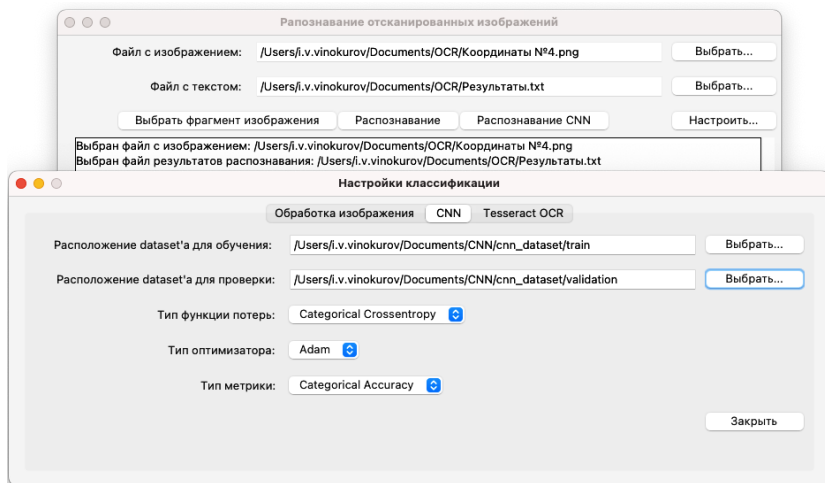
Ошибок распознавания кириллических букв и знаков пунктуации «.» и «,» нет, что является существенным улучшением результатов, полученных в [1].

3. Реализация ИС

По результатам проведённых исследований была доработана ИС, общие принципы организации и функционирования которой приведены в [1]. В доработанной ИС, с использованием DNN YOLOv5s, реализовано распознавание расположения таблиц. Координаты таблицы определяют область документа с табличной информацией. Собственно распознавание табличной информации обученной CNN осуществляется по описанным в [1] принципам – преобразование её в градации серого цвета, удаление горизонтальных и вертикальных линий таблицы на основе выбора значений ядер их детектирования и определение расположения границ элементов текста с возможным применением механизма скелетизации последних, рисунок 9а. Параметры обучения CNN остались такими же, как и в [1] – указание расположения наборов для обучения и валидации, выбор типа оптимизатора, функций оценки точности модели и её потерь, рисунок 9б.



(а) Параметры распознавания элементов текста



(б) Параметры обучения CNN

Рисунок 9. Выбор параметров распознавания текста и обучения CNN

В случае невозможности обнаружения таблицы с помощью DNN YOLOv5s на отсканированных документах плохого качества, её расположение в документе может быть выделено вручную (мышью), рисунок 10

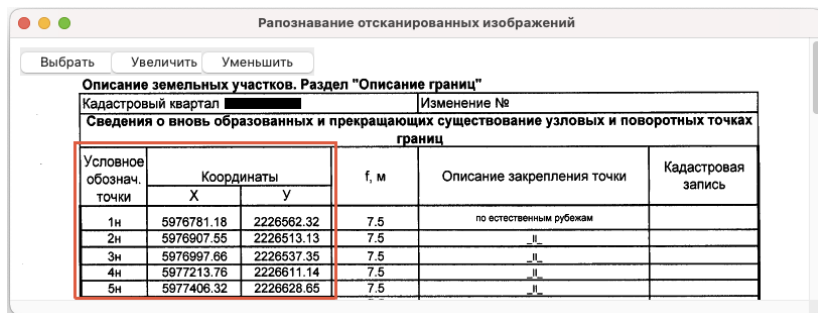


Рисунок 10. Выделение таблицы вручную

Используемое в текущей версии ИС автоматическое распознавание расположение таблицы с использованием DNN YOLOv5s и ограничение распознаванием только табличной информации как наиболее значимой в документах Росреестра позволило повысить скорость её распознавания на 20-30% по сравнению с начальной версией этой ИС.

Заключение

Проведены экспериментальные исследования DNN YOLOv5s для обнаружения простых таблиц на отсканированных документах Росреестра. Для дообучения YOLOv5s был использован собственный набор данных из предметной области актуализации баз данных системы ЕНРН, состоящий из 400 изображений и соответствующих им текстовых файлов с нормализованными координатами таблиц. Дообученная DNN YOLOv5s показала вполне приемлемые результаты их детектирования на отсканированных документах.



Для распознавания элементов текста внутри найденной области с табличной информацией предложена CNN с двумя контурами выявления признаков. Результаты исследования этой CNN показали лучший по сравнению с [1] результат классификации цифр, кириллических букв и основных знаков пунктуации.

По полученным в результате работы результатам была повышена эффективность работы ИС [1], за счёт распознавания только табличной информации как наиболее значимой информации в документах Росреестра.

Список литературы

- [1] Винокуров И. В. *Использование свёрточной нейронной сети для распознавания элементов текста на отсканированных изображениях плохого качества // Программные системы: теория и приложения.* – 2022. – Т. 13. – № 3. – С. 29–43.

- [2] Harit G., Bansal A. *Table detection in document images using header and trailer patterns* // *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP'12* (December 16–19, 2012, Mumbai, India), New York: ACM.– 2012.– ISBN 978-1-4503-1660-6.– id. 62.– 8 pp. doi ↑5
- [3] Gatos B., Danatsas D., Pratikakis I., Perantonis S. *Automatic table detection in document images*, ICAPR 2005: Pattern Recognition and Data Mining, Lecture Notes in Computer Science.– vol. **3686**, Berlin–Heidelberg: Springer.– 2005.– ISBN 978-3-540-28757-5.– Pp. 609–618. doi ↑5
- [4] Kasar T., Barlas P., Adam S., Chatelain C., Paquet T. *Learning to detect tables in scanned document images using line information*, 2013 12th International Conference on Document Analysis and Recognition (25–28 August 2013, Washington, DC, USA).– 2013.– Pp. 1185–1189. doi ↑5
- [5] Jahan M. A., Ragel R. G. *Locating tables in scanned documents for reconstructing and republishing*, 7th International Conference on Information and Automation for Sustainability (22–24 December 2014, Colombo, Sri Lanka).– 2014.– Pp. 1–6. doi ↑5
- [6] Kieninger T. G. *Table structure recognition based on robust block segmentation* // *Document Recognition V*, Photonics West'98 Electronic Imaging (1998, San Jose, CA, United States), Proc. SPIE.– vol. **3305**.– 1998.– Pp. 22–32. doi ↑5
- [7] Minghao Li, Lei Cui, Shaohan Huang, Furu Wei, Ming Zhou, Zhoujun Li *TableBank: A benchmark dataset for table detection and recognition*.– 2020.– 9 pp. doi ↑6
- [8] Fang J., Tao X., Tang Z., Qiu R., Liu Y. *Dataset, ground-truth and performance metrics for table detection evaluation*, 2012 10th IAPR International Workshop on Document Analysis Systems (27–29 March 2012, Gold Coast, QLD, Australia).– 2012.– Pp. 445–449. doi ↑6
- [9] Gobel M., Hassan T., Oro E., Orsi G. *Icdar 2013 table competition*, 2013 12th International Conference on Document Analysis and Recognition (15 October 2013, Washington, DC, USA).– 2013.– Pp. 1449–1453. doi ↑6
- [10] Shahab A., Shafait F., Kieninger T., Dengel A. *An open approach towards the benchmarking of table structure recognition systems* // *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, DAS'10* (June 9–11, 2010, Boston, Massachusetts, USA).– 2010.– ISBN 978-1-60558-773-8.– Pp. 113–120. doi ↑6
- [11] Gao L., Huang Y., Dejean H., Meunier J.-L., Yan Q., Fang Y., Kleber F., Lang E. *ICDAR 2019 competition on table detection and recognition (cTDaR)*, 2019 International Conference on Document Analysis and Recognition (ICDAR) (20–25 September 2019, Sydney, NSW, Australia).– 2019.– Pp. 1510–1515. doi ↑6
- [12] Ren S., He K., Girshick R., Sun J. *Faster R-CNN: towards real-time object detection with region proposal networks* // *IEEE transactions on pattern analysis and machine intelligence*.– 2016.– Vol. **39**.– No. 6.– Pp. 1137–1149. doi ↑6
- [13] Redmon J., Divvala S., Girshick R., Farhadi A. *You only look once: Unified, real-time object detection* // *Proceedings of the IEEE conference on computer vision and pattern recognition*.– 2016.– Pp. 779–788. doi ↑6
- [14] Gilani A., Qasim S. R., Malik I., Shafait F. *Table detection using deep learning* // *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*.– V. 1 (09–15 November 2017, Kyoto, Japan).– 2017.– Pp. 771–776. doi ↑6

- [15] Banerjee A. *YOLOv5 vs YOLOv6 vs YOLOv7*, Learn With A Robot, <https://www.learnwitharobot.com/p/yolov5-vs-yolov6-vs-yolov7>.– 2022–2023. ↑₆
- [16] Lebedzinski P. *A single number metric for evaluating object detection models*, Towards Data Science, <https://towardsdatascience.com/a-single-number-metric-for-evaluating-object-detection-models-c97f4a98616d>.– 2021. ↑₆
- [17] Surya Gutta *Object Detection Algorithm — YOLO v5 Architecture*, <https://medium.com/analytics-vidhya/object-detection-algorithm-yolo-v5-architecture-89e0a35472ef>.– Analytics Vidhya.– 2021. ↑₆
- [18] Zixin Ning, Xinjiao Wu, Jing Yang, Yanqin Yang *MT-YOLOv5: Mobile terminal table detection model based on YOLOv5*, The Fourth International Conference on Physics, Mathematics and Statistics (ICPMS) 2021 (19–21 May 2021, Kunming, China) // Journal of Physics: Conference Series.– 2021.– Vol. **1978**.– id. 012010.  ↑₆
- [19] Yilun Huang, Qinqin Yan, Yibo Li, Yifan Chen, Zhi Tang *A YOLO-based table detection method*, 2019 International Conference on Document Analysis and Recognition (ICDAR) (20–25 September 2019, Sydney, NSW, Australia).– 2019.  ↑₆

Поступила в редакцию 23.11.2022;
одобрена после рецензирования 28.11.2022;
принята к публикации 12.12.2022;
опубликована онлайн 13.02.2023.

Рекомендовал к публикации

д.ф.-м.н. А. М. Елизаров

Информация об авторе:



Игорь Викторович Винокуров

Кандидат технических наук (PhD), ассоциированный профессор в Финансовом Университете при Правительстве Российской Федерации. Область научных интересов: информационные системы, информационные технологии, технологии обработки данных.



0000-0001-8697-1032

e-mail: igvvinokurov@fa.ru

Автор заявляет об отсутствии конфликта интересов.



Tabular information recognition using convolutional neural networks

Igor Victorovich **Vinokurov**[✉]

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

[✉]iguvinokurov@fa.ru

Abstract. The relevance of identifying tabular information and recognizing its contents for processing scanned documents is shown. The formation of a data set for training, validation and testing of a deep learning neural network (DNN) YOLOv5s for the detection of simple tables is described. The effectiveness of using this DNN when working with scanned documents is shown. Using the Keras Functional API, a convolutional neural network (CNN) was formed to recognize the main elements of tabular information — numbers, basic punctuation marks and Cyrillic letters. The results of a study of the work of this CNN are given. The implementation of the identification and recognition of tabular information on scanned documents in the developed IS updating information in databases for the Unified State Register of Real Estate system is described.

Key words and phrases: Convolutional Neural Networks, Deep Learning Neural Networks, CNN, DNN, YOLOv5s, Keras, Python

2020 *Mathematics Subject Classification:* 68T20; 68T07, 68T45

For citation: Igor V. Vinokurov. *Tabular information recognition using convolutional neural networks*. Program Systems: Theory and Applications, 2023, 14:1(56), pp. 3–30. https://psta.psiras.ru/read/psta2023_1_3-30.pdf

Introduction

The process of digital transformation of business models of various types of organizations currently being implemented involves, among other things, their transition to electronic media. This, in turn, leads to the need to recognize the contents of electronic documents for the purpose of subsequent processing of their contents. In the case of effective recognition of the contents of an electronic document, its automatic (in some cases automated) processing significantly increases the speed, and in some cases, the quality of this process. Hence, the solution of the problem of information recognition on scanned copies of documents, which is the first stage of converting a paper document into its electronic counterpart, is of particular practical and scientific interest.

One of the possible approaches to solve this problem with acceptable accuracy, along with traditional methods, is the use of CNN. In [1] describes the general principles of scanned image recognition using a CNN trained on a dataset specific to the content of text documents. Quite often, however, the most important information in a document is presented in tabular form. Revealing only the table and recognizing its contents will allow for this type of documents to increase the speed of recognition of the information necessary for subsequent processing and, as a result, the efficiency of this process.

The purpose of this work is to develop an IS that implements the localization of a table on a scanned document and recognition of its contents. In this work, simple types of tables are detected in scanned documents, which are typical for most documents and contain only textual or numeric information. The restriction of the type of table and the features of the information contained in it gives a reason to form your own data set for training and choose any of the known DNN or CNN models that demonstrate acceptable results for detecting various types of objects. After identifying the location of the table in the document, to recognize its contents, it is proposed to use a slight modification of the method described in [1].

The results obtained in the work of the consistent application of DNN for table detection and the developed CNN have found practical application for updating the database of the USRN system of Rosreestr, which is currently being transferred to electronic media.

1. Table localization on the image

The currently existing methods for localizing a table on an image can be divided into traditional methods that implement image processing and modern methods based on the use of neural networks with architectures of various types. Traditional methods of table localization are based on the specific structuring of information and the presence, in most cases, of horizontal and vertical lines, their signs and the presence of various proximity and similarity criteria. So in the works [2] and [3] methods for finding tabular information between keywords or their combinations at the beginning and end of the table and in the areas of intersection of its horizontal and vertical lines, respectively, were proposed. The disadvantages of the approaches are obvious—the absence of keywords and lines (borders) of the table leads to the impossibility or extreme inefficiency of using these methods.

Another approach [4] is based on the method of determining the length of rows and columns of a table. After identifying all horizontal and vertical lines along their length on the document image, for each of them a set of low-level features is formed, which are transferred to the support vector machine (SVM) that implements table detection. If there are no lines, the method does not work.

In [5] involves implementing the localization and extraction of table regions from a document image based on local thresholds for word spacing and line height. The main limitation of this method is that it detects table areas along with surrounding text areas. Therefore, it cannot be used only for tabular information localization.

Traditional methods include the method of detecting [6] tabular structures in a document based on the outline of a randomly selected word. The method assumes that contours similar in shape, identified horizontally and vertically, determine the table. Having empty rows in a table can lead to inefficiencies and even failure to discover all of the table information.

Greater efficiency of table localization on an electronic document is implemented by methods based on CNN or DNN. In such methods, there is no explicit binding to the structure of the table and there is no need for preliminary preparation of the document. However, the training of such models requires a large number of images of documents containing tables - this is the main difficulty of these methods. There are several datasets that can be used to train table detection models—TableBlank [7], Marmot

[8], ICDAR [9], UNLV[10], ICDAR 2019 cTDAr [11] and others. However, quite often there is a need to form your own data sets, which is not at all a trivial task.

Fast RCNN and Faster RCNN [12], YOLO [13] are currently considered the most efficient deep learning models. The features of Fast RCNN and Faster RCNN are finding potential objects in the image and dividing them into regions using the selective search method, extracting features of each obtained region using CNN, subsequent classification using the support vector method and refining the boundaries of regions using linear regression.

The Faster RCNN model was used first to detect tables in the [14] document. However, the YOLO model is currently the leader in object detection. This model implements a smaller number of actions compared to Faster RCNN and implements the search for bounding boxes and the probability of their belonging to one or another image class.

There are several versions of this DNN; to solve practical problems, versions v5, v6 and v7 [15] are currently used, pre-trained on the data set *COCO (Common Objects in COntext)*^(URL). Training time, work time and model parameters such as Precision, Recall and AP [16] increase from v5 to v7, therefore, formally v7 version can be considered more efficient. However, in practice, the effectiveness of a particular version of the model is determined by the peculiarities of its use and training. Since there are no special requirements for the speed and accuracy of table detection, and version v5 [17] shows quite acceptable results in combination with low resource consumption and, at the same time, has developed tools for collecting information about the learning process and metrics at each epoch to implement the localization problem table on the electronic document, this version was chosen. Examples of efficient use of YOLO DNNs trained on the ICDAR [11] dataset to detect tables in electronic copies of documents are given in [18, 19].

With a known organization of this DNN, the only thing that needs to be done is to generate data sets for its additional training, validation and testing. The formation of an additional training data set was carried out on several hundred (about 400) scanned documents of Rosreestr with simple tables containing a small explanatory text in the table headers and numerical information in the form of a set of coordinates of capital construction objects or land plots. For each of these documents, using open source software *labelImg*^(URL), text files were generated with table labels—their normalized coordinates within the document. Network training was

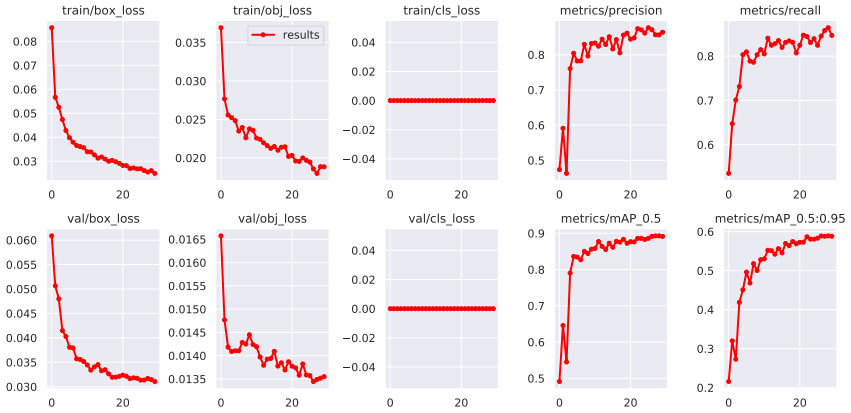


FIGURE 1. Plots of loss functions, precision, recall and $mAP_{0.5:0.95}$ at 30 training epochs DNN YOLOv5s

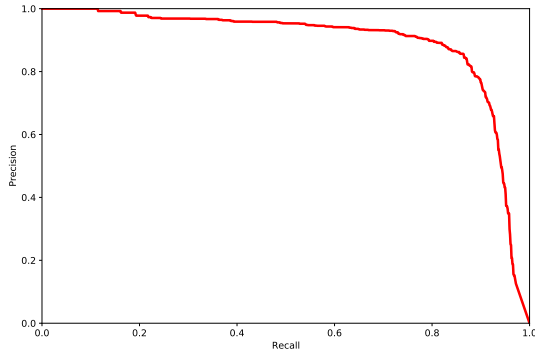


FIGURE 2. Precision-Recall at 30 training epochs DNN YOLOv5s

carried out on 70% of the total number of images, validation on 20% and testing on 10%. As the results of the conducted experimental studies showed, 30 training epochs made it possible to obtain a quite acceptable result in terms of the main parameters of the DNN operation and its generalizing abilities, Figure 1.

On the validation set with 30 training epochs, the following results were obtained: $Precision = 0.871$, $Recall = 0.828$, $mAP_{0.5} = 0.592$. The dependence Precision-Recall is shown in Figure 2.

Table 0.78

Описание земельных участков. Раздел "Описание границ"					
Кадастровый квартал XX:XX:XX:XX			Изменение №		
Сведения о вновь образованных и прекращающих существование узловых и поворотных точек границ					
Условное обознач. точки	Координаты		f, м	Описание закрепления точки	Кадастровая запись
	X	Y			
1н	5976781.18	2226562.32	7.5	По естественным рубежам	
2н	5976907.56	2226513.13	7.5	По естественным рубежам	
3н	5976997.66	2226537.35	7.5	По естественным рубежам	
4н	5877213.76	2226611.14	7.5	По естественным рубежам	
5н	5977406.32	2226628.65	7.5	По естественным рубежам	
6н	5977596.63	2226652.87	7.5	По естественным рубежам	
7н	5977823.31	2226759.99	7.5	По естественным рубежам	
8н	5977829.71	2226853.41	7.5	По естественным рубежам	
9н	5977903.99	2226976.25	7.5	По естественным рубежам	
10н	5977956.51	2227056.59	7.5	По естественным рубежам	

FIGURE 3. Table discovery result with DNN YOLOv5s

DNN YOLOv5s, retrained on its own data set, showed a good result of detecting tables on electronic copies of good quality documents—about 98-99%. An example of the result of the network operation is shown in Figure 3.

2. Recognition of tabular information

In [1] provides the general principles of CNN formation for the implementation of multiclass classification of text elements on scanned images of poor quality. The ongoing research of this CNN and the dataset for its training led to the following conclusion – the improvement of classification results can be achieved to a greater extent due to the parallel detection of stable features and their subsequent summation. In other words, the complication of the sequential CNN model from [1] gives worse results compared to its functional model implemented by *Keras Functional API*^{[URL](#)}.

To recognize tabular information, a CNN structure is proposed that has two initial contours. The first loop consists of 2 convolution layers (*Conv2D*^{[URL](#)}) and one maximizing pooling layer (*MaxPooling2D*^{[URL](#)}). The second one consists of one convolutional layer (*Conv2D*) and one maximizing pooling layer (*MaxPooling2D*). All other CNN layers repeat layers from [1]—one convolutional layer (*Conv2D*) for feature detection from summation results of 2 input circuits, one linearizing layer (*Flatten*^{[URL](#)}) and

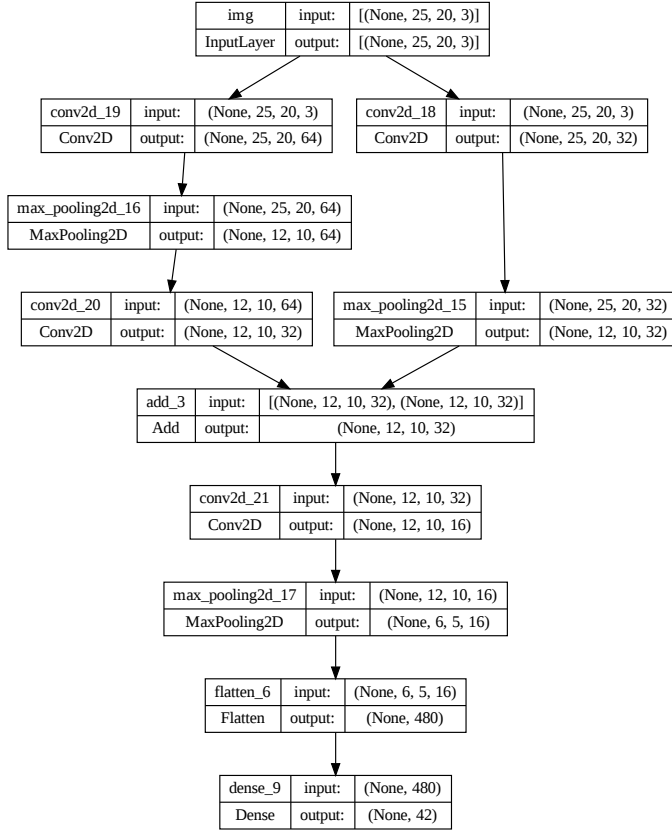


FIGURE 4. CNN structure for tabular information recognition

two fully connected layers ($Dense^{\text{URL}}$). The structure of the CNN is shown in Figure 4.

All training and validation images are resized to 20×25 [1]. The activation functions of all CNN neurons are $\ll sigmoid \gg^{\text{URL}}$. Compilation and training parameters are chosen as standard for multiclass classification: optimizer — $\ll adam \gg^{\text{URL}}$, metric — $\ll acc \gg^{\text{URL}}$, the loss function is $\ll ategorical-crossentropy \gg^{\text{URL}}$.

The data sets for training and validation of this CNN remained the same as in [1], — 42 classes of different images of numbers, Cyrillic letters and 2 punctuation marks «.» and «,». Since these datasets are small — 10 and 5

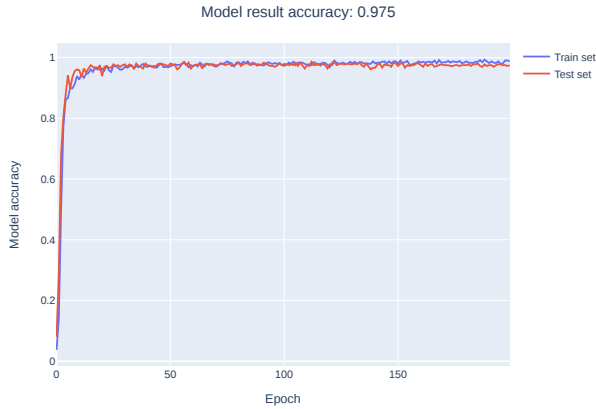


FIGURE 5. Model accuracy

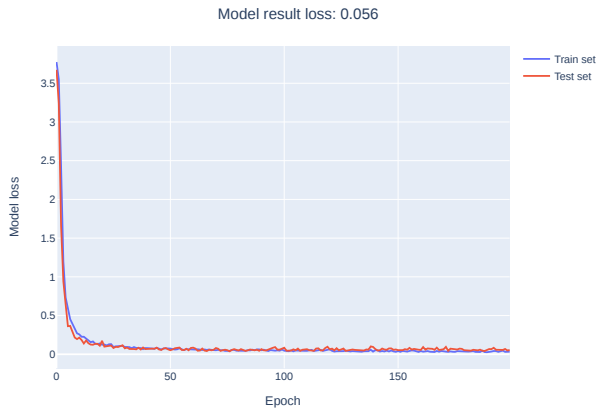


FIGURE 6. Model loss

images for each of the symbols, respectively, during training and subsequent study of the CNN operation, they were augmented using the transformed data batch generator *ImageDataGenerator*^{[URL](#)}. Taking into account the specifics of the subject area, the following transformation parameters and their values were chosen: *zoom_range* = 0.125, *shear_range* = 0.15, *rotation_range* = 0.15, *width_shift_range* = 1.2, *height_shift_range* = 1.2, *horizontal_flip* = False and *fill_mode* = «nearest».

Figure 5 and Figure 6 show the results of experimental studies, respec-

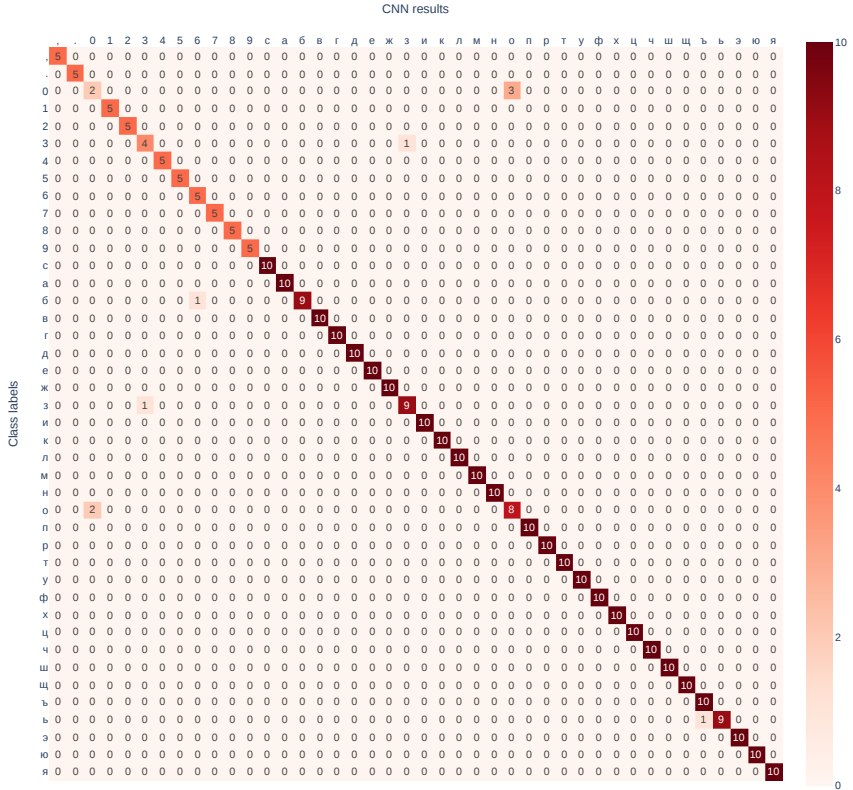


FIGURE 7. Confusion matrix

tively, of the accuracy and loss of the model at 200 epochs of its training, which is the optimal number according to the results of experimental studies.

The values of the recall and F1 parameters for the model were 0.98055 and 0.98029, respectively, which is a very good result. The quality of the generated CNN when classifying text elements is shown by the following CNN confusion matrix in Figure 7.

As can be seen from this figure, errors in the classification of text elements consisted in the similarity of the numbers 0, 3 and the letters «о» and «3». An examples of the erroneous classification of these numbers and letters are shown in Figure 8a, Figure 8b. No error appears in recognizing

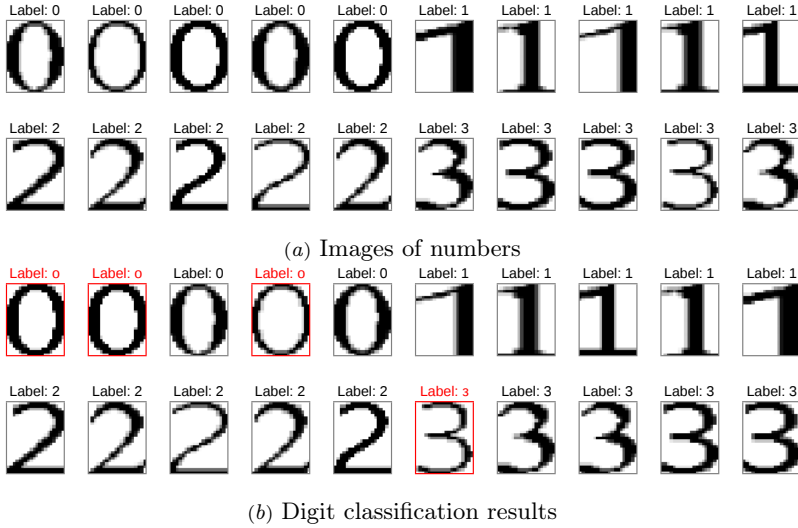


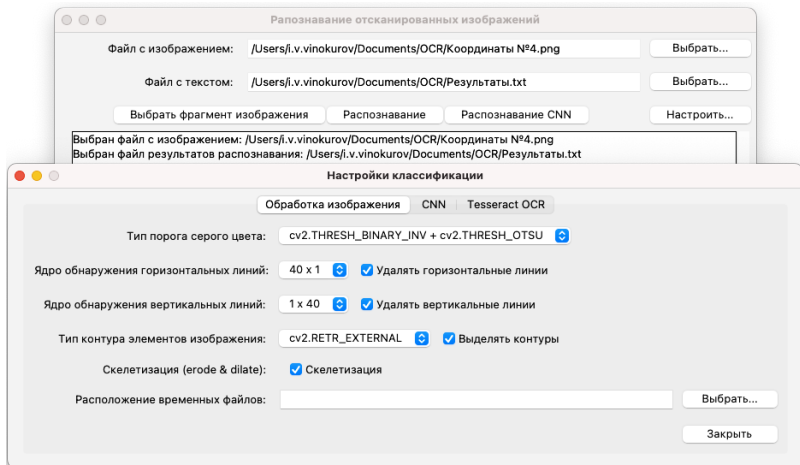
FIGURE 8. Digits classification errors

Cyrillic letters and punctuation marks «.» and «,», that is a significant improvement on the results obtained in [1].

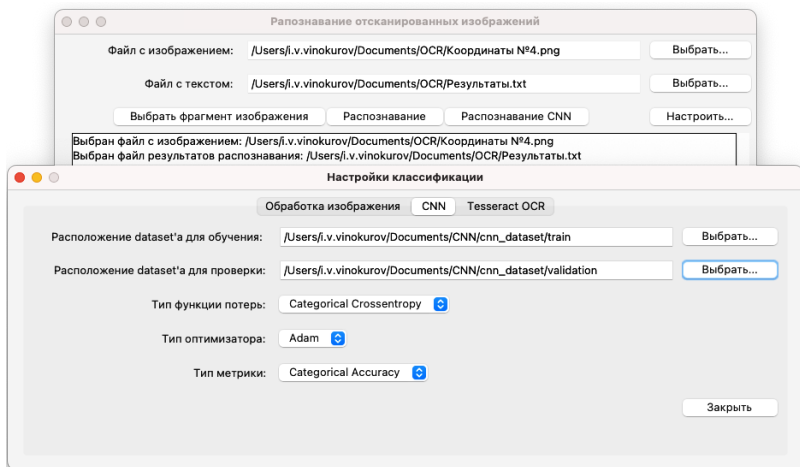
3. IS implementation

Based on the results of the research, the IS was finalized, the general principles of organization and functioning of which are given in [1]. In the modified IS, using DNN YOLOv5s, recognition of the location of tables is implemented. Table coordinates define the area of the document with tabular information.

The actual recognition of the tabular information of the trained CNN is carried out according to the principles described in [1]—converting it to grayscale, deleting horizontal and vertical lines of the table based on the choice of the values of their detection kernels and determining the location of the boundaries of text elements with the possible use of the skeletonization mechanism of the latter, Figure 9a. The training parameters of the CNN remained the same as in [1]—specifying the location of sets for training and validation, choosing the type of optimizer, functions for estimating the accuracy of the model and its losses, Figure 9b.



(a) Text recognition options



(b) CNN training parameters

FIGURE 9. Selecting text recognition and CNN training parameters

If it is impossible to detect a table using DNN YOLOv5s on scanned documents of poor quality, its location in the document can be selected manually (with the mouse), Figure 10.

The automatic recognition of the table location used in the current

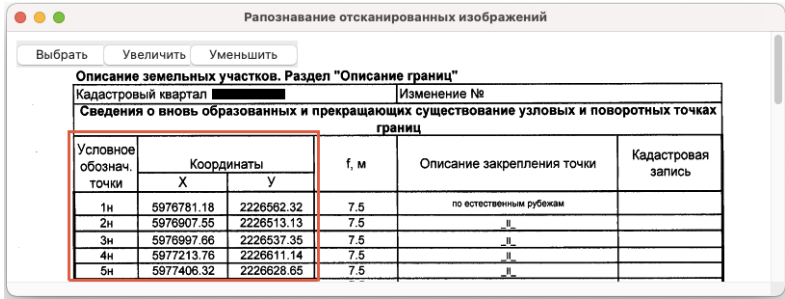


FIGURE 10. Selecting a table manually

version of the IS using DNN YOLOv5s and the limitation of recognizing only tabular information as the most significant in Rosreestr documents made it possible to increase the speed of its recognition by 20-30 compared to the initial version of this IS.



Conclusion


Experimental studies of DNN YOLOv5s were carried out to detect simple tables on scanned documents of Rosreestr. For additional training of YOLOv5s, we used our own data set from the subject area of updating the databases of the ENPH system, consisting of 400 images and corresponding text files with normalized table coordinates. The retrained DNN YOLOv5s showed quite acceptable results of their detection on scanned documents.












To recognize text elements within the found area with tabular information, a CNN with two features detection contours is proposed. The results of the study of this CNN showed a better result in the classification of numbers, Cyrillic letters and basic punctuation compared to [1].




According to the results obtained as a result of the work, the efficiency of the IS [1] was increased, due to the recognition of only tabular information as the most significant information in the Rosreestr documents.

References

[1] I. V. Vinokurov. “Using a convolutional neural network to recognize text elements in poor quality scanned images”, *Programmnyye sistemy: teoriya i prilozheniya*, 13:3 (2022), pp. 29–43 (in Russian).  [URL](#)  ↑18, 22, 23, 26, 28

[2] G. Harit, A. Bansal. “Table detection in document images using header and trailer patterns”, *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP’12* (December 16–19, 2012, Mumbai, India), ACM, New York, 2012, ISBN 978-1-4503-1660-6, id. 62, 8 pp.  ↑19

- [3] B. Gatos, D. Danatsas, I. Pratikakis, S. Perantonis. “Automatic table detection in document images”, ICAPR 2005: Pattern Recognition and Data Mining, Lecture Notes in Computer Science, vol. **3686**, Springer, Berlin–Heidelberg, 2005, ISBN 978-3-540-28757-5, pp. 609–618.  ↑¹⁹
- [4] T. Kasar, P. Barlas, S. Adam, C. Chatelain, T. Paquet. “Learning to detect tables in scanned document images using line information”, 2013 12th International Conference on Document Analysis and Recognition (25–28 August 2013, Washington, DC, USA), 2013, pp. 1185–1189.  ↑¹⁹
- [5] M. A. Jahan, R. G. Ragel. “Locating tables in scanned documents for reconstructing and republishing”, 7th International Conference on Information and Automation for Sustainability (22–24 December 2014, Colombo, Sri Lanka), 2014, pp. 1–6.  ↑¹⁹
- [6] T. G. Kieninger. “Table structure recognition based on robust block segmentation”, *Document Recognition V*, Photonics West’98 Electronic Imaging (1998, San Jose, CA, United States), Proc. SPIE, vol. **3305**, 1998, pp. 22–32.  ↑¹⁹
- [7] Li Minghao, Cui Lei, Huang Shaohan, Wei Furu, Zhou Ming, Li Zhoujun. *TableBank: A benchmark dataset for table detection and recognition*, 2020, 9 pp.  ↑¹⁹
- [8] J. Fang, X. Tao, Z. Tang, R. Qiu, Y. Liu. “Dataset, ground-truth and performance metrics for table detection evaluation”, 2012 10th IAPR International Workshop on Document Analysis Systems (27–29 March 2012, Gold Coast, QLD, Australia), 2012, pp. 445–449.  ↑²⁰
- [9] M. Gobel, T. Hassan, E. Oro, G. Orsi. “Icdar 2013 table competition”, 2013 12th International Conference on Document Analysis and Recognition (15 October 2013, Washington, DC, USA), 2013, pp. 1449–1453.  ↑²⁰
- [10] A. Shahab, F. Shafait, T. Kieninger, A. Dengel. “An open approach towards the benchmarking of table structure recognition systems”, *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, DAS’10 (June 9–11, 2010, Boston, Massachusetts, USA), 2010, ISBN 978-1-60558-773-8, pp. 113–120.  ↑²⁰
- [11] L. Gao, Y. Huang, H. Dejean, J. -L. Meunier, Q. Yan, Y. Fang, F. Kleber, E. Lang. “ICDAR 2019 competition on table detection and recognition (cTDaR)”, 2019 International Conference on Document Analysis and Recognition (ICDAR) (20–25 September 2019, Sydney, NSW, Australia), 2019, pp. 1510–1515.  ↑²⁰
- [12] S. Ren, K. He, R. Girshick, J. Sun. “Faster R-CNN: towards real-time object detection with region proposal networks”, *IEEE transactions on pattern analysis and machine intelligence*, **39**:6 (2016), pp. 1137–1149.  ↑²⁰
- [13] J. Redmon, S. Divvala, R. Girshick, A. Farhadi. “You only look once: Unified, real-time object detection”, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.  ↑²⁰

- [14] A. Gilani, S. R. Qasim, I. Malik, F. Shafait. “Table detection using deep learning”, *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. V. 1 (09–15 November 2017, Kyoto, Japan), 2017, pp. 771–776.  [↑₂₀](#)
- [15] A. Banerjee. *YOLOv5 vs YOLOv6 vs YOLOv7*, Learn With A Robot, <https://www.learnwitharobot.com/p/yolov5-vs-yolov6-vs-yolov7>, 2022–2023. [↑₂₀](#)
- [16] P. Lebedzinski. *A single number metric for evaluating object detection models*, Towards Data Science, <https://towardsdatascience.com/a-single-number-metric-for-evaluating-object-detection-models-c97f4a98616d>, 2021. [↑₂₀](#)
- [17] Surya Gutta. *Object Detection Algorithm — YOLO v5 Architecture*, <https://medium.com/analytics-vidhya/object-detection-algorithm-yolo-v5-architecture-89e0a35472ef>, Analytics Vidhya, 2021. [↑₂₀](#)
- [18] Zixin Ning, Xinjiao Wu, Jing Yang, Yanqin Yang. “MT-YOLOv5: Mobile terminal table detection model based on YOLOv5”, The Fourth International Conference on Physics, Mathematics and Statistics (ICPMS) 2021 (19–21 May 2021, Kunming, China), *Journal of Physics: Conference Series*, **1978** (2021), id. 012010.  [↑₂₀](#)
- [19] Yilun Huang, Qinqin Yan, Yibo Li, Yifan Chen, Zhi Tang. “A YOLO-based table detection method”, 2019 International Conference on Document Analysis and Recognition (ICDAR) (20–25 September 2019, Sydney, NSW, Australia), 2019.  [↑₂₀](#)

Received
approved after reviewing
accepted for publication
published online

23.11.2022;
28.11.2022;
12.12.2022;
13.02.2023.

Recommended by

prof. A. M. Elizarov

Information about the author:



Igor Victorovich Vinokurov

Candidate of Technical Sciences (PhD), Associate Professor at the Financial University under the Government of the Russian Federation. Research interests: information systems, information technologies, data processing technologies.



0000-0001-8697-1032

e-mail: igvvinokurov@fa.ru

The author declare no conflicts of interests.