



Math-Net.Ru

Общероссийский математический портал

I. V. Vinokurov, Распознавание цифровых последовательностей с использованием свёрточных нейронных сетей, *Программные системы: теория и приложения*, 2023, том 14, выпуск 3, 3–36

DOI: 10.25209/2079-3316-2023-14-3-3-36

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 18.119.253.198

26 декабря 2024 г., 05:30:06



УДК 004.932.75'1+004.89

doi 10.25209/2079-3316-2023-14-3-3-36



Распознавание цифровых последовательностей с использованием свёрточных нейронных сетей

Игорь Викторович **Винокуров**

Финансовый Университет при Правительстве Российской Федерации, Москва, Россия

✉ igvvinokurov@fa.ru

Аннотация. В статье показана актуальность задачи преобразования в текстовый формат элементов изображений, содержащих последовательности машинописных цифр. На примере распознавания табличной информации из отсканированных документов ШПК «Роскадастр» предложено возможное решение этой задачи с использованием свёрточных нейронных сетей (CNN). Описаны принципы формирования наборов данных и моделей CNN для распознавания последовательностей из двух, трёх и четырёх цифр. Приведены результаты экспериментального исследования этих моделей и показана их эффективность. Описана интеграция моделей CNN в разрабатываемую в настоящее время информационную систему (ИС), предназначенную для автоматизированного перевода отсканированных документов в их текстовые аналоги.

Ключевые слова и фразы: Распознавание цифр, свёрточные нейронные сети, CNN, Keras, Python

Для цитирования: Винокуров И.В. *Распознавание цифровых последовательностей с использованием свёрточных нейронных сетей* // Программные системы: теория и приложения. 2023. Т. 14. № 3(58). С. 3–36. https://psta.psisras.ru/read/psta2023_3_3-36.pdf

Введение

Процесс перехода на электронный документооборот, реализуемый в настоящее время многим организациями, предполагает перевод электронных копий документов в их тестовые аналоги. В ППК «Роскадастр» основными документами, подлежащими распознаванию и переводу в текстовые форматы (txt, json и другие), являются изображения таблиц с кадастровыми координатами объектов. Детектирование таблиц на изображениях документов достаточно эффективно реализуется с использованием сетей глубокого обучения YOLO [1]. Не менее эффективно с использованием CNN, могут быть распознаны и одиночные символы машинописного текста [2]. Однако при плохом качестве отсканированного документа его сегментация с использованием и стандартных (например, *findContours()* библиотеки OpenCV [3]) и собственных методов их обнаружения может привести к выявлению регионов (ROI, *Region Of Interest, англ.*), содержащих более одного символа. В этих случаях предложенный в [2] подход неприменим: требуется распознать на фрагментах изображений непрерывные последовательности из нескольких цифр.

Пример сегментов (фрагментов) изображения, содержащих от одной до трёх цифр, приведён на рисунке 1. Выявление регионов, содержащих более одной цифры, как это показано на рисунке 1б, объясняется наличием градаций серого цвета, которые пороговая подсистема трансформации цветов в ИС [2] воспринимает как элементы изображения и, как следствие, относит их к текущему ROI.



(а) Координаты объекта

(б) Формирование ROI

РИСУНОК 1. Сегментация координат объекта

Решение задачи распознавания цифровых последовательностей, при массовом автоматизированном преобразовании отсканированных изображений в их текстовые аналоги, является в настоящее время актуальной и значимой для ППК «Роскадастр».

Ниже, в разделе 1 приведены существующие в настоящее время подходы к распознаванию последовательностей цифр. Раздел 2 посвящён формированию и исследованию моделей CNN для распознавания последовательностей из двух, трёх и четырёх цифр. В разделе 3 описаны особенности выбора сформированных моделей CNN для распознавания документов с кадастровыми координатами объектов.

1. Анализ возможных подходов к решению задачи

В настоящее время существует достаточно большое количество работ, посвящённых задаче распознавания буквенно-цифровых последовательностей. Интерес к этой задаче объясняется тем, что результаты её решения могут быть использованы в нетривиальном компьютерном зрении с целью распознавания содержимого электронных копий документов, обработки форм с результатами опросов, идентификации почтовых индексов и адресов и т.п.

Определённый интерес с точки зрения поиска возможного подхода к решению задачи представляют методы исследовательских групп из разных стран по распознаванию рукописных буквенно-цифровых последовательностей, представленные на Международной конференции по распознаванию рукописного текста (ICFHR) 2014 [4]. Все эти методы предполагают наличие значимого этапа начальной предобработки изображения. Например, в пекинском методе из [4] для удаления шумов из простых изображений применяется бинаризация Otsu [5]. Для сложных изображений предлагается использовать два многослойных перцептрона [6] с последующей бинаризацией Сауволя [7]. После предобработки изображения генерируются шаблоны символов кандидатов, которые затем классифицируются с помощью полиномиального классификатора.

В шанхайском методе из [4] цифры сегментируются на основе адаптивной бинаризации и анализа связанных компонентов. Собственно классификация реализуется с использованием машины опорных векторов (SVM) [8] и самоорганизующихся карт Кохонена (SOM) [9]. Результаты распознавания для каждой цифры последовательности упорядочиваются по их достоверности.

Сингапурский метод [4] реализует предварительную обработку изображения с целью извлечения векторов последовательных признаков. Для этой цели используется гистограмма ориентированных градиентов (HOG) [10]. Распознавание цифровых последовательностей реализуется рекуррентной нейронной сетью (RNN) [11].

Бразильский метод [4] предварительно обрабатывает изображения с целью обнаружения цифровых последовательностей, для распознавания которых используется классификатор, состоящий из многомерной рекуррентной нейронной сети (MDRNN) [12] и машины опорных векторов (SVM) [13].

И, наконец, алжирский метод Tébéssa I [4] использует стратегию прямой сегментации, основанную на ориентированном скользящем окне [14]. Распознавание осуществляется с помощью преобразования Радона

[15] в сочетании с SVM [8]. Другой аналогичный метод, Tébéssa П [4], реализует те же действия, что и выше, но с несколькими углами ориентации скользящего окна.

В [16] предложен алгоритм, который обучает k-разрывные автокодеры и использует их в качестве предварительно обученных скрытых слоёв CNN. Для достижения высоких скоростей распознавания применяется техника раздвижных окон.

Комбинация модели ResNet [17] с двунаправленными слоями краткосрочной памяти предложена в [18]. Высокая производительность модели ResNet для извлечения функций из изображений привела к значительному улучшению распознавания цифровых последовательностей, особенно при работе с маркировкой длинных строк.

В [19] представлена CNN для маркировки рукописных строк, в которой для обработки последовательностей произвольной длины строки используется пул пространственных пирамид. Сегментация цифр в строковых последовательностях в этой работе не используется.

Интересная модель для задач маркировки буквенно-цифровых последовательностей приведена в [20]. В основе модели лежит архитектура CRNN–CNN без недостатков RNN. Такая модель немного улучшила результаты из [19], в которой использовались двунаправленные рекуррентные блоки в сочетании с моделью ResNet.

В [21] разработана система, в которой для распознавания цифр на изображениях с низким разрешением используется детектор цели на уровне символов. В основе детектора цели лежат координаты контура символа.

В большинстве приведенных выше работах использовалась комбинация нетривиальных методов предварительной обработки изображения с последующим применением нейросетевых классификаторов. Однако в настоящее время существенные успехи в области нейронных сетей привели к появлению ряда работ, в которых для распознавания в том числе и буквенно-цифровых последовательностей используются только нейронные сети того или иного типа. Этап предварительной обработки или отсутствует или крайне незначителен. Например, в [22] для обнаружения и распознавания элементов буквенно-цифровых последовательностей предложено использовать модель на основе нейронной сети глубокого обучения (DNN) YOLO [23]. В этой работе полностью исключена предварительная обработка и сегментация изображений, кроме того, YOLO хорошо справляется с распознаванием длинных последовательностей.

Помимо YOLO, для распознавания последовательностей машинописных и рукописных цифр, изолированных или контактных, могут быть использованы и CNN, например так, как это сделано в [24]. В этой работе показано, что *распознавание цифровых последовательностей может быть эффективно реализовано без существенной предварительной обработки изображений*. Критически важным для такого подхода является незашумленность (чёткость) изображения.

Ещё один пример эффективного использования CNN для распознавания строк рукописных цифр с низким разрешением приведён в [25]. Повышение точности работы CNN реализуется за счёт добавления двух выходных ветвей для классификации и регрессии по ограничивающим контурам символов. С целью дальнейшего повышения точности используются некатегоризированное подавление и тонкая минипакетная настройка.

Исходя из существующей в настоящее время тенденции распознавания буквенно-цифровых последовательностей, использование CNN с незначительной предварительной обработкой изображений представляется достаточно эффективным. В этой работе был выбран именно такой подход к распознаванию цифровых последовательностей из документов ППК «Роскадастр».

2. Формирование и исследование моделей CNN

Как показал анализ результатов сегментирования произвольной выборки из 2000 документов ППК «Роскадастр», последовательности из 5-и и более цифр встречаются крайне редко, рисунок 2. Исключения

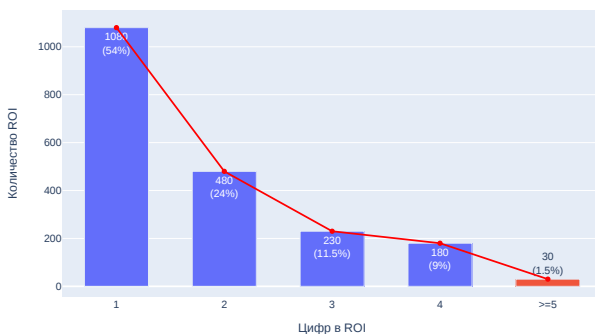


Рисунок 2. Количество ROI в выборке из 2000 отсканированных документов составляют сильно зашумленные документы или документы с неприемлемым качеством сканирования. Для таких документов, как правило,

используется ручной разбор содержимого. Как следствие, было принято решение о распознавании последовательностей, состоящих не более чем из четырёх цифр.

2.1. Формирование наборов данных для обучения и тестирования моделей

На начальном этапе формирования наборов данных для обучения и тестирования моделей CNN был использован предложенный в [2] подход формирования изображений цифр от 0 до 9 основными шрифтами кадастровых документов. Далее, в зависимости от количества цифр в последовательности, реализуется их случайный выбор, аугментация, варьирующая необходимым образом угол наклона и масштабирование и слияние изображений в одно, рисунок 3.

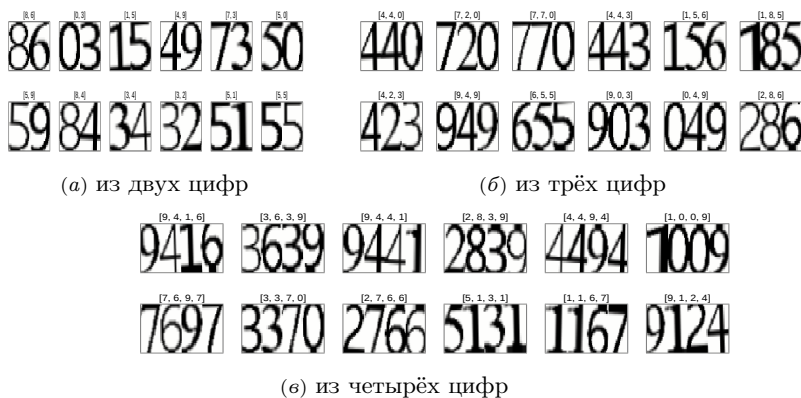


Рисунок 3. Примеры сформированных изображений последовательностей, сверху указаны метки классов

При проведении экспериментальных исследований, для последовательностей из двух, трёх и четырёх цифр, были сформированы 1500, 2000 и 3000 изображений соответственно, гарантирующих перебор всех возможных цифровых комбинаций. Для обучения CNN использовались 70% этих изображений и 30% – для валидации и тестирования. Количество изображений цифр (классов) в каждой из цифровых последовательностей одинаково, как следствие, наборы данных для обучения и тестирования являются сбалансированными.

2.2. Формирование моделей CNN

Экспериментальные исследования с использованием языка Python и библиотеки Keras [26] показали, что модели CNN, обеспечивающие приемлемые точности распознавания цифровых последовательностей, должны быть последовательными и иметь в двух свёрточных слоях Conv2D¹ по 32 фильтра с размером ядра 3×3 , при этом размер выборки в двух дискретизирующих слоях с типом MaxPooling2D⁵ должен быть 2×2 .

Компиляция всех моделей осуществлялась со следующими параметрами: функция потерь *categorical_crossentropy*², оптимизатор *adam*³ и основная метрика *accuracy*⁴, дающая приемлемую оценку точности для сбалансированных наборов данных.

Все модели CNN имеют одинаковые свёрточные и дискретизирующие слои и отличаются только выходными полносвязными слоями Dense⁵, количество которых определяется числом распознаваемых моделью CNN цифр, рисунок 4.

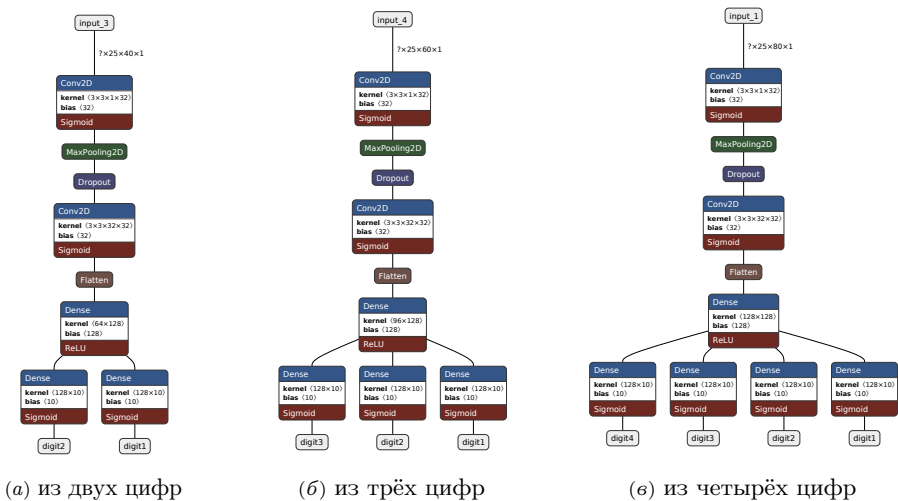


Рисунок 4. Модели CNN для распознавания последовательностей

Каждый из этих слоёв возвращает вероятность принадлежности цифры к одному из 10-ти классов.

¹см. описание *Keras layers API*^{URL}

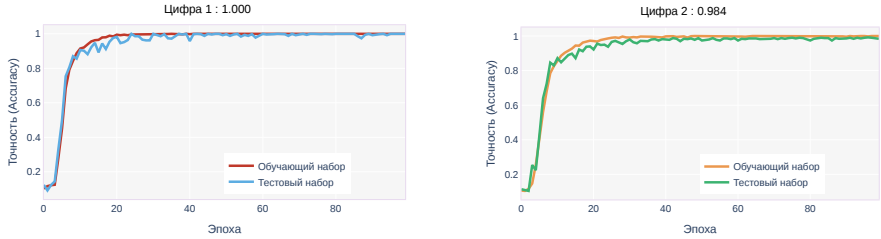
²см. описание *Keras.Losses*^{URL}

³см. описание *Keras.Optimizers*^{URL}

⁴см. описание *Keras.Accuracy metrics*^{URL}

2.2.1. Исследование моделей CNN

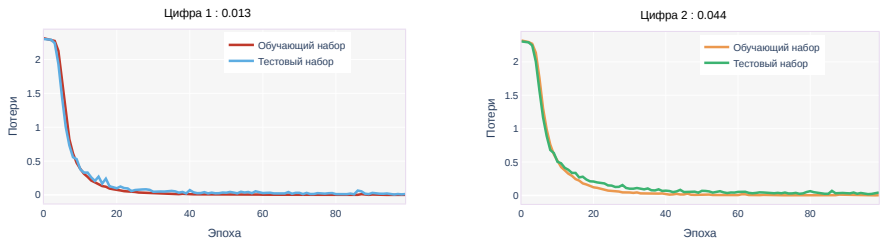
Результаты исследований моделей CNN из рисунка 4 при 100 (оптимальное значение) эпохах их обучения в виде графиков точности, функции потерь и ROC-кривых для распознавания последовательностей из двух цифр приведены на рисунке 5, 6, 7.



(а) для первой цифры

(б) для второй цифры

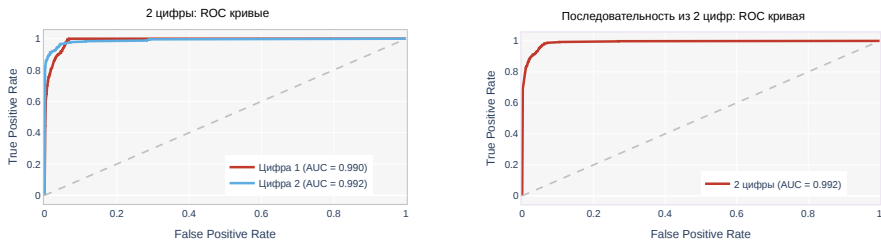
Рисунок 5. Точность распознавания пары цифр



(а) для первой цифры

(б) для второй цифры

Рисунок 6. Потери распознавания пары цифр

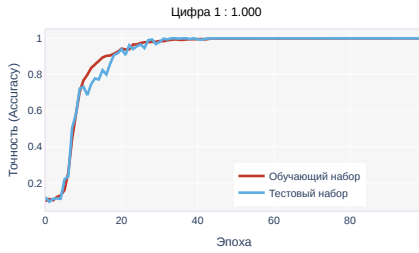


(а) каждой цифры

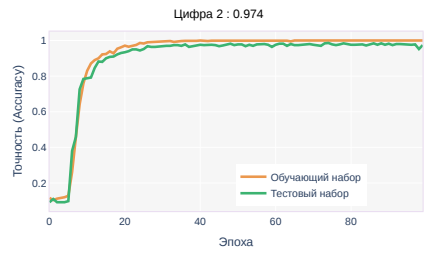
(б) последовательности цифр

Рисунок 7. ROC кривые при распознавании пары цифр

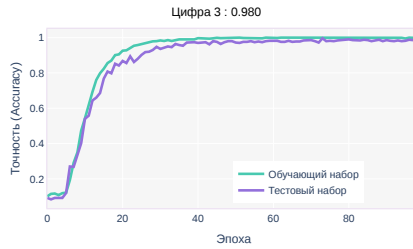
Для распознавания последовательностей из трёх цифр аналогичные результаты приведены на рисунках 8, 9, 10.



(a) для первой цифры

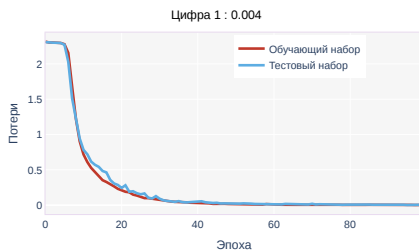


(б) для второй цифры

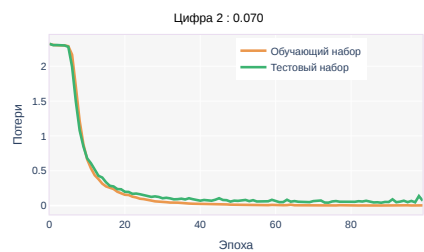


(в) для третьей цифры

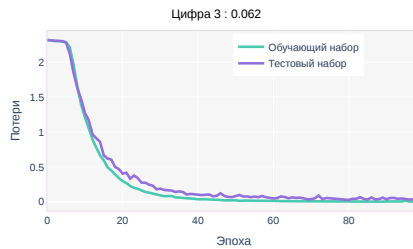
РИСУНОК 8. Точность распознавания тройки цифр



(a) для первой цифры

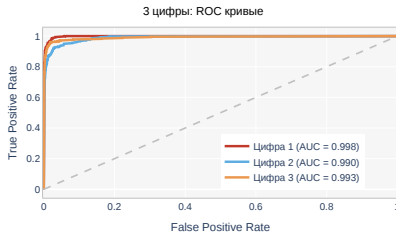


(б) для второй цифры

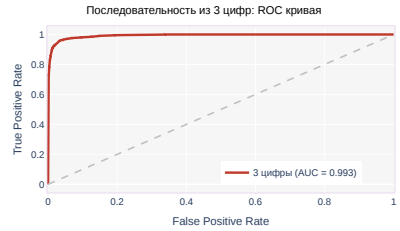


(в) для третьей цифры

РИСУНОК 9. Потери распознавания тройки цифр



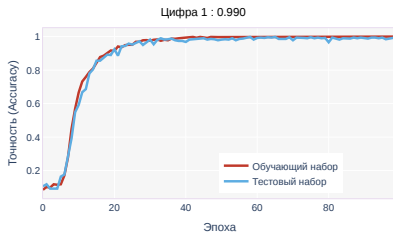
(а) каждой цифры



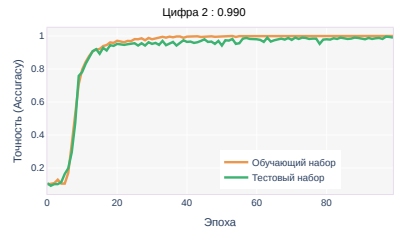
(б) последовательности цифр

Рисунок 10. ROC кривые при распознавании тройки цифр

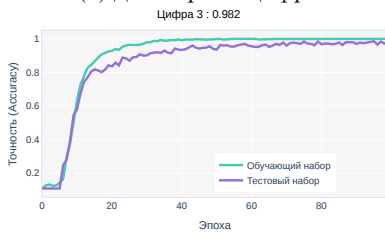
Результаты распознавания последовательностей из четырёх цифр представлены на рисунках 11, 12 и 13.



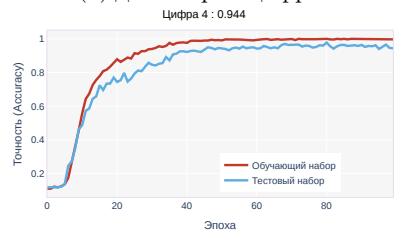
(а) для первой цифры



(б) для второй цифры



(в) для третьей цифры



(г) для четвёртой цифры

Рисунок 11. Точность распознавания четвёрки цифр

Здесь и ниже по тексту над графиками приведено значение исследуемого параметра на последней эпохе обучения CNN.

Количество эпох обучения определено экспериментально и позволяет произвести точную настройку моделей CNN для распознавания всех возможных комбинаций цифр, сгенерированных на этапе их аугментации. С целью устранения возможного переобучения, в каждую из моделей был добавлен слой Dropout⁵.

⁵см. описание *Keras layers API*^{URL}

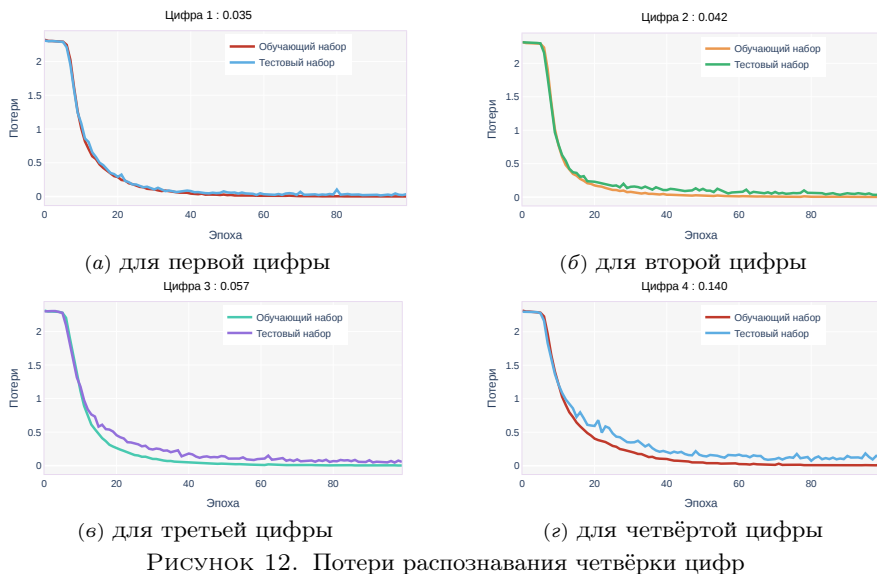


Рисунок 12. Потери распознавания четвёрки цифр

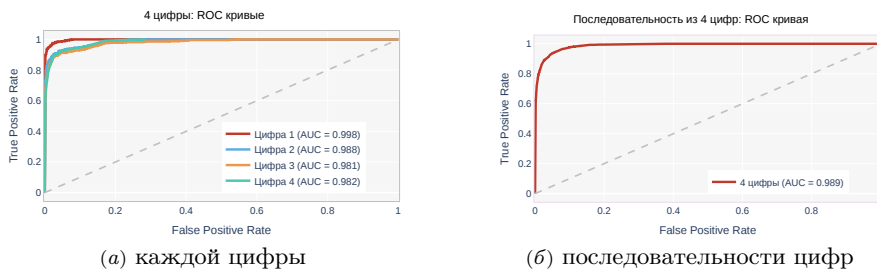


Рисунок 13. ROC кривые при распознавании четвёрки цифр

Наилучшие значения метрик *accuracy*, *recall* и *f1*⁶, полученных в результате проведения экспериментальных исследований моделей, приведены в таблице 1.

ТАБЛИЦА 1. Значения основных метрик точности моделей

Цифр в последовательности	Отдельная цифра (<i>accuracy</i> , <i>recall</i> , <i>f1</i>)	Последовательность цифр (<i>accuracy</i> , <i>recall</i> , <i>f1</i>)
2	(0.996, 0.984, 0.908)	(0.988, 0.987, 0.897)
3	(0.989, 0.980, 0.890)	(0.980, 0.984, 0.876)
4	(0.977, 0.970, 0.824)	(0.912, 0.945, 0.822)

⁶см. описание *Keras.Accuracy metrics*^[9]

Исследование моделей CNN показало вполне ожидаемый результат – с увеличением количества цифр в последовательности снижается точность их распознавания. Например, ошибка распознавания отдельной цифры в последовательностях, приведённых на рисунке 3в, встречается чаще, чем в последовательностях из рисунка 3а и 3б. Пример ошибочного распознавания последовательности из четырёх цифр показан на рисунке 14.

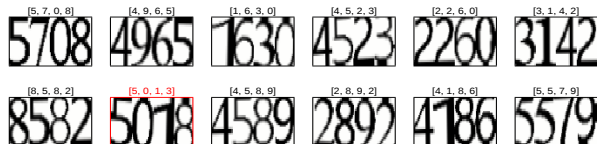


Рисунок 14. Ошибка при распознавании последовательностей из четырёх цифр

Ошибки распознавания не снижает значимости полученных результатов, поскольку они встречаются крайне редко – 1 раз на несколько десятков документов. В текущей реализации ИС ППК «Роскадастр» ошибки распознавания цифровых последовательностей не обрабатываются.

3. Реализация распознавания последовательностей цифр в ИС

Описанные выше модели CNN были реализованы в ИС ППК «Роскадастр», предназначенной для автоматизированного распознавания содержимого отсканированных документов, в том числе, и с координатами объектов. Расположение таблиц внутри документа детектируются с использованием сети глубокого обучения YOLOv5s [1]. Далее внутри таблиц выявляются контуры ROI и их размеры – для этой цели используется стандартная функция *findContours()* библиотеки OpenCV [3]. Например, для координат, приведённых на рисунке 1, диапазоны размеров ROI приведены на рисунке 15.

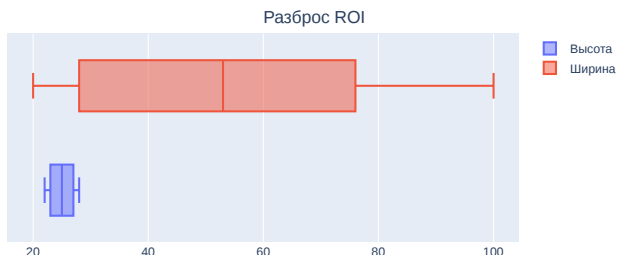


Рисунок 15. Диапазоны значений высоты и ширины ROI

Размеры ROI (в большинстве случаев это только ширина) используются для оценки количества цифр в последовательности с использованием метода кластеризации k-средних [27]. Результаты кластеризации ROI этим методом для координат из рисунка 1 показаны на рисунке 16.

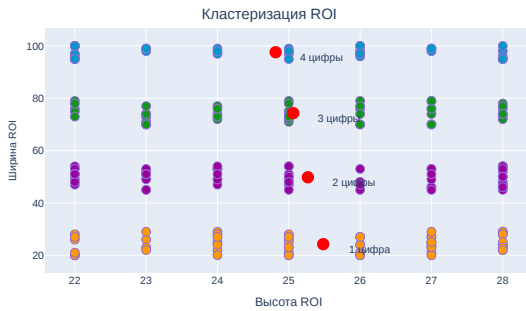


Рисунок 16. Кластеризация ROI. Красным показаны центры кластеров

Собственно процесс распознавания заключается в переборе всех ROI, соотношения их размеров с центроидами четырёх кластеров, и выборе соответствующей модели CNN, рисунок 17.

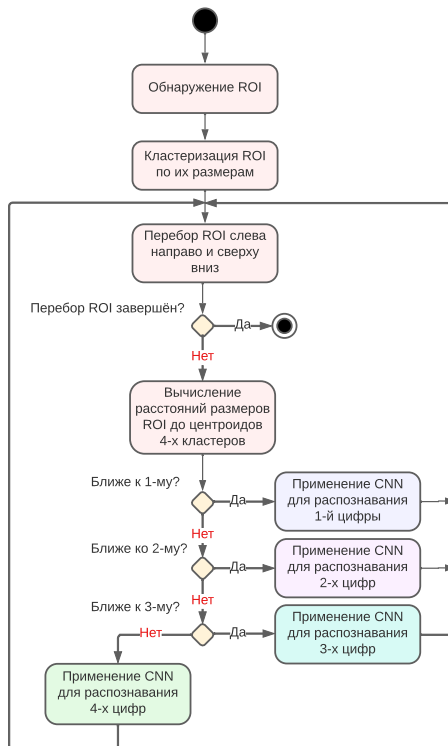


Рисунок 17. Диаграмма основных этапов работы ИС по выбору модели CNN

Пример распознавания координат объекта в ИС ППК «Роскадастр» приведён на рисунке 18.

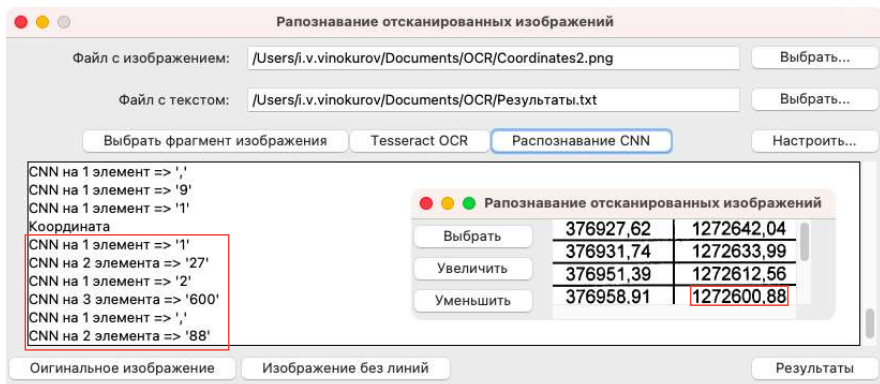


Рисунок 18. Реализация распознавания координат объекта с использованием свёрточных нейронных сетей

4. Достоинства и недостатки предлагаемого подхода

К очевидным достоинствам распознавания цифровых последовательностей с использованием CNN, применительно к документам ППК «Роскадастр», относятся отсутствие необходимости в предварительной обработке изображений, как это было отмечено в [24], и достаточно высокая точность распознавания. Незначительное снижение точности распознавания второй, третьей и четвёртой цифры (см. рисунок 5б, 8в, 11г) объясняется структурой CNN, приводящей к накоплению погрешности распознавания последующих цифр.

К недостатку можно отнести неверную в ряде случаев кластеризацию ROI. Причиной этого является использование в документах ППК «Роскадастр» немоноширинных шрифтов. Неверная кластеризация приводит к тому, что в ИС выбирается модель CNN, не соответствующая цифровой последовательности.

Например, цифровая последовательность из трёх единиц оказывается ближе к центроиду кластера двух цифровых последовательностей, цифровая последовательность из четырёх единиц ближе к центроиду кластера трёх цифровых последовательностей и т.д. Подобные ситуации встречаются крайне редко, тем не менее, их появление является вполне возможным.

Заключение

В результате проведённых исследований были разработаны модели CNN, позволившие достаточно эффективно решить задачу распознавания последовательностей из двух, трёх и четырёх цифр. Математические и экспериментальные исследования этих моделей показали приемлемую точность распознавания как отдельных цифр, так и их непрерывных последовательностей.

Разработанные модели были использованы в ИС ППК «Роскадастр» для решения достаточно значимой на данный момент задачи преобразования отсканированных документов в их текстовые аналоги. По предварительным оценкам, использование предложенного в работе подхода распознавания цифровых последовательностей с использованием CNN, позволило повысить эффективность (скорость, качество) распознавания кадастровых координат на отсканированных документах невысокого качества на 20%–40% в зависимости от качества последних.

Список литературы

- [1] Винокуров И. В. *Распознавание табличной информации с использованием свёрточных нейронных сетей* // Программные системы: теория и приложения.– 2023.– Т. 14.– № 1(56).– С. 3–30. [doi](#) [*](#) [URL](#) ↑4, 14
- [2] 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(54).– С. 29–43. [doi](#) [*](#) [URL](#) ↑4, 8
- [3] Мюллер А., Гвидо С. *Введение в машинное обучение с помощью Python. Руководство для специалистов по обработке данных.*– СПб.: Изд-во «ООО Альфа-книга».– 2017.– ISBN 978-5-907515-03-1.– 480 С. ↑4, 14
- [4] M. Diem, S. Fiel, F. Kleber, R. Sablatnig, J. M. Saavedra, D. Contreras, J. M. Barrios, L. S. Oliveira *ICFHR 2014 competition on handwritten digit string recognition in challenging datasets (HDSRC 2014)* // *2014 14th International Conference on Frontiers in Handwriting Recognition* (Hersonissos, Greece, 01–04 September 2014).– pp. 779–784. [doi](#) ↑5, 6
- [5] N. Otsu *A threshold selection method from gray-level histograms*, IEEE Trans. Syst. Man Cybern.– 1979.– Vol. 9.– No. 1.– Pp. 62–66. [doi](#) ↑5
- [6] J. C. B. Rabelo, C. Zanchettin, C. A. B. Mello and B. L. D. Bezerra *A multi-layer perceptron approach to threshold documents with complex background*, 2011 IEEE International Conference on Systems, Man, and Cybernetics (Anchorage, AK, USA, 09–12 October 2011).– Pp. 2523–2530. [doi](#) ↑5
- [7] J. Sauvola, M. Pietikäinen *Adaptive document image binarization* // *Pattern Recognition*.– 2000.– Vol. 33.– No. 2.– Pp. 225–236. [doi](#) ↑5
- [8] C. Cortes, V. Vapnik *Support-vector networks* // *Mach. Learn.*– 1995.– Vol. 20.– No. 3.– Pp. 273–297. [doi](#) ↑5, 6

- [9] T. Kohonen *Self-organized formation of topologically correct feature maps* // Biol. Cybern.– 1982.– Vol. **43**.– No. 1.– Pp. 59–69. doi ↑5
- [10] N. Dalal, B. Triggs *Histograms of oriented gradients for human detection* // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'05 (San Diego, CA, USA, 20–25 June 2005).– ISBN 0-7695-2372-2.– pp. 886–893. doi ↑5
- [11] J. J. Hopfield *Neural networks and physical systems with emergent collective computational abilities*, Proc. Nat. Acad. Sci. USA.– 1982.– Vol. **79**.– No. 8.– Pp. 2554–2558. doi ↑5
- [12] A. Graves, J. Schmidhuber *Offline handwriting recognition with multidimensional recurrent neural networks* // *Guide to OCR for Arabic Scripts*, eds. Märgner V., El Abed H., London: Springer.– ISBN 978-1-4471-4072-6.– pp. 297–313. doi ↑5
- [13] B. L. D. Bezerra, C. Zanchettin, V. B. de Andrade *A MDRNN-SVM hybrid model for cursive offline handwriting recognition*, Artificial Neural Networks and Machine Learning — ICANN 2012, Lecture Notes in Computer Science.– vol. **7553**, Berlin: Springer.– 2012.– ISBN 978-3-642-33265-4.– pp. 246–254. doi ↑5
- [14] A. Gattal, Y. Chibani *Segmentation strategy of handwritten connected digits (SSHCD)*, Image Analysis and Processing (ICIAP).– 2011.– Pp. 248–254. doi ↑5
- [15] M. Hejazi, G. Shevlyakov and Y.-S. Ho *Modified discrete radon transforms and their application to rotation-invariant image analysis*, 2006 IEEE Workshop on Multimedia Signal Processing, Lecture Notes in Computer Science.– vol. **6979**, Berlin–Heidelberg: Springer.– 2006.– ISBN 978-3-642-24088-1.– pp. 429–434. doi ↑6
- [16] R. Saabni *Recognizing handwritten single digits and digit strings using deep architecture of neural networks* // 2016 Third International Conference on Artificial Intelligence and Pattern Recognition, AIPR (Lodz, Poland, 19–21 September 2016).– 2016.– pp. 1–6. doi ↑6
- [17] K. He, X. Zhang, S. Ren, J. Sun *Deep residual learning for image recognition* // 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR (Las Vegas, NV, USA, 27–30 June 2016).– pp. 770–778. doi ↑6
- [18] H. Zhan, Q. Wang, Y. Lu *Handwritten digit string recognition by combination of residual network and RNN-CTC*, ICONIP 2017: Neural Information Processing, Lecture Notes in Computer Science.– vol. **10639**, Cham: Springer.– 2017.– ISBN 978-3-319-70136-3.– pp. 583–591. doi ↑6
- [19] Q. Wang, Y. Lu *A sequence labeling convolutional network and its application to handwritten string recognition*, Twenty-Sixth International Joint Conference on Artificial Intelligence (Melbourne, Australia, 19–25 August 2017).– 2017.– Pp. 2950–2956. doi ↑6
- [20] H. Zhan, S. Lyu, X. Tu, Y. Lu *Residual CRNN and its application to handwritten digit string recognition*, ICONIP 2019: Neural Information Processing, Communications in Computer and Information Science.– vol. **1143**, Cham: Springer.– 2019.– ISBN 978-3-030-36802-9.– pp. 49–56. doi ↑6
- [21] Y. Ma, J. Guo, W. Wei *An exceedingly fast model for low resolution handwritten digit string recognition* // 2019 IEEE 7th International Conference on Computer Science and Network Technology, ICCSNT (Dalian, China, 9–20 October 2019).– pp. 282–288. doi ↑6

- [22] A. G. Hochuli, A. S. Britto, J. P. Barddal, R. Sabourin, L. E. S. Oliveira *An End-to-End approach for recognition of modern and historical handwritten numeral strings // 2020 International Joint Conference on Neural Networks, IJCNN (Glasgow, UK, 19–24 July 2020).*– 2020.– pp. 1–8. doi ↑6
- [23] J. Redmon, S. Divvala, R. Girshick, A. Farhadi *You only look once: Unified real-time object detection // 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR (Las Vegas, NV, USA, 27–30 June 2016).*– pp. 779–788. doi ↑6
- [24] A. G. Hochuli, L. S. Oliveira, A. S. Britto, R. Sabourin *Handwritten digit segmentation: Is it still necessary?*Pattern Recognition.– 2018.– Vol. **78**.– Pp. 1–11. doi ↑7, 16
- [25] J. Guo, Y. Xu *Low Resolution Handwritten Digit String Recognition based on Object Detection Network // 2020 International Joint Conference on Neural Networks, IJCNN (Glasgow, UK, 19–24 July 2020).*– pp. 1–7. doi ↑7
- [26] J. Moolayil *Learn Keras for Deep Neural Networks. A Fast-Track Approach to Modern Deep Learning with Python.*– Berkeley: Apress.– 2019.– ISBN 978-1-4842-4240-7.– xv+182 Pp. doi ↑9
- [27] S. Datta *Learning OpenCV 3 Application Development.*– Packt Publishing.– 2016.– ISBN 9781784391454.– 294 Pp. ↑14

Поступила в редакцию 14.04.2023;
одобрена после рецензирования 04.07.2023;
принята к публикации 04.07.2023;
опубликована онлайн 13.08.2023.

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.

UDC 004.932.75'1+004.89

 10.25209/2079-3316-2023-14-3-3-36

Recognition of digital sequences using convolutional neural networks

Igor Victorovich **Vinokurov**[✉]

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

[✉]igvvinokurov@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: I. V. Vinokurov. *Recognition of digital sequences using convolutional neural networks*. Program Systems: Theory and Applications, 2023, 14:3(58), pp. 3–36. https://psta.psir.ru/read/psta2023_3_3-36.pdf

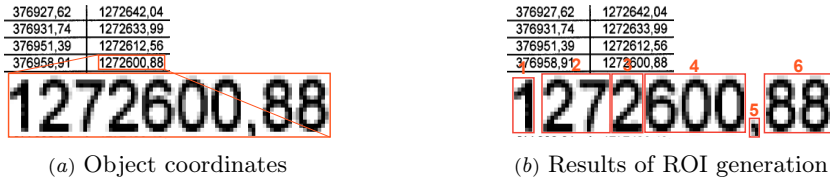


FIGURE 1. Segmentation of object coordinates

Introduction

The process of transition to electronic document management, currently implemented by many organizations, involves the conversion of electronic copies of documents into their text analogues. One of such organizations is the «Roskadastr» PLC. In this organization, the main documents subject to recognition and translation into text formats (txt, json and others) are images of tables with cadastral coordinates of objects. Table detection on document images, as it was shown in [1], is realized quite effectively using YOLO deep learning networks. Single characters of typewritten text [2] can be recognized equally effectively using CNNs. However, if the quality of the scanned document is poor, its segmentation using both standard (e.g., *findContours()* of the OpenCV [3] library) and proprietary detection methods can lead to the detection of regions (ROI, *Region Of Interest*) containing more than 1 character. In these cases, the approach proposed in [2] cannot be used, so there is a need to develop a new one that allows to recognize continuous sequences of several digits in image fragments.

An example of image segments (fragments) containing from 1 to 3 digits is given in Figure 1. Detection of regions containing more than 1 digit, as shown in Figure 1b, is explained by the presence of grayscale colors, which the threshold subsystem of color transformation in the IS [2] perceives as image elements and, as a consequence, relates them to the current ROI.

The solution of this problem, with mass automated conversion of scanned images into their text analogs, is currently urgent and significant for the «Roskadastr» PLC.

Below, in section 1 the existing approaches to the recognition of sequences of digits are given. The section 2 is devoted to the formation and investigation of CNN models for recognizing sequences of 2, 3, and 4 digits. Section 3 describes the peculiarities of selecting the generated CNN models for recognizing documents with cadastral coordinates of objects.

1. Analysis of possible approaches to solving the problem

At present, there are quite a large number of works devoted to the task of recognising alphanumeric sequences. The interest in this problem is explained by the fact that the results of its solution can be used in non-trivial computer vision to recognise the contents of electronic copies of documents, to process forms with survey results, to identify postal codes and addresses, etc.

Of definite interest from the point of view of finding a possible approach to solving the problem are the methods of research groups from different countries for recognising handwritten alphanumeric sequences presented at the International Conference on Handwriting Recognition (ICFHR) 2014 [4]. All these methods involve a meaningful initial image preprocessing step. For example, the Beijing method from [4] applies Otsu [5] binarisation to remove noise from simple images. For complex images, two multilayer [6] perceptrons followed by Sauvola [7] binarisation is proposed. After image preprocessing, candidate character templates are generated and then classified using a polynomial classifier.

In the Shanghai method from [4] digits are segmented based on adaptive binarisation and connected component analysis. The actual classification is implemented using a support vector machine (SVM) [8] and self-organising Kohonen maps (SOM) [9]. The recognition results for each digit of the sequence are ordered according to their confidence.

The Singapore [4] method implements image preprocessing to extract sequence feature vectors. The histogram of oriented gradients (HOG) [10] is used for this purpose. The recognition of digital sequences is realised by a recurrent neural network (RNN) [11].

The Brazilian [4] method pre-processes the images to detect digital sequences, for the recognition of which a classifier consisting of a multidimensional recurrent neural network (MDRNN) [12] and a support vector machine (SVM) [13] is used.

Finally, the Algerian Tébessa I [4] method uses a direct segmentation strategy based on an oriented sliding window [14]. Recognition is

performed using Radon transform [15] combined with SVM [8]. Another similar method, Tébéssa II [4], implements the same steps as above, but with multiple sliding window orientation angles.

In [16], an algorithm is proposed that trains k-breaking autoencoders and uses them as pre-trained hidden layer CNNs. A sliding window technique is applied to achieve high recognition rates.

A combination of the ResNet model [17] with bidirectional short-term memory layers is proposed in [18]. The high performance of the ResNet model for feature extraction from images led to a significant improvement in digital sequence recognition, especially when dealing with long string labelling.

A CNN for handwritten string labelling is presented in [19], which uses a pool of spatial pyramids to handle sequences of arbitrary string length. Segmentation of digits in string sequences is not used in this work.

An interesting model for the task of labelling alphanumeric sequences is given in [20]. The model is based on the CRNN architecture — CNN without the drawbacks of RNN. This model slightly improved the results from [19], which used bidirectional recurrent blocks in combination with the ResNet model.

In [21], a system is developed in which a character-level target detector is used to recognise digits in low-resolution images. The target detector is based on the symbol contour coordinates.

Most of the above works used a combination of non-trivial methods of image preprocessing followed by the application of neural network classifiers. However, at present, significant progress in the field of neural networks has led to a number of works in which only neural networks of one or another type are used to recognise alphanumeric sequences. The preprocessing stage is either absent or extremely insignificant. For example, in [22], a model based on the deep learning neural network (DNN) YOLO [23] is proposed to detect and recognise alphanumeric sequence elements. In this work, image preprocessing and segmentation are completely eliminated, and in addition, YOLO performs well in recognising long sequences.

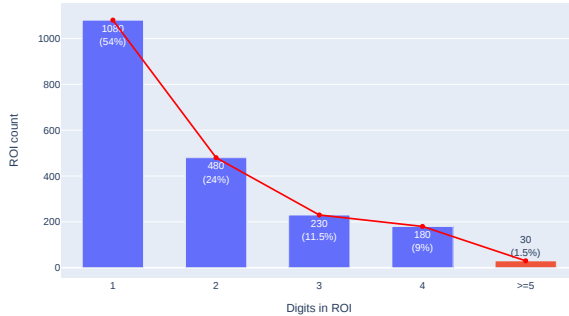


FIGURE 2. Number of ROIs in a sample of 2000 scanned documents

Besides YOLO, CNNs can also be used to recognise sequences of typewritten and handwritten digits, isolated or contacted, for example as done in [24]. In this work, it is shown that *digit sequence recognition can be efficiently implemented without significant image preprocessing*. Critical to this approach is the unnoised (sharpness) of the image.

Another example of effective use of CNNs for low-resolution handwritten digit string recognition is given in [25]. The accuracy improvement of the CNN is realised by adding 2 output branches for classification and regression on bounding character contours. In order to further improve the accuracy, unclassified suppression and fine minipacket tuning are used.

Based on the current trend of alphanumeric sequence recognition, the use of CNNs with minor image preprocessing appears to be quite effective. In this work, exactly such an approach was chosen for recognising numeric sequences from «Roskadastr» PLC documents.

2. Formation and exploration of CNN models

As the analysis of the results of segmentation of a random sample of 2000 documents of the «Roskadastr» PLC has shown, sequences of 5 and more digits are extremely rare, Figure 2. Exceptions are highly noisy documents or documents with unacceptable scanning quality. For such

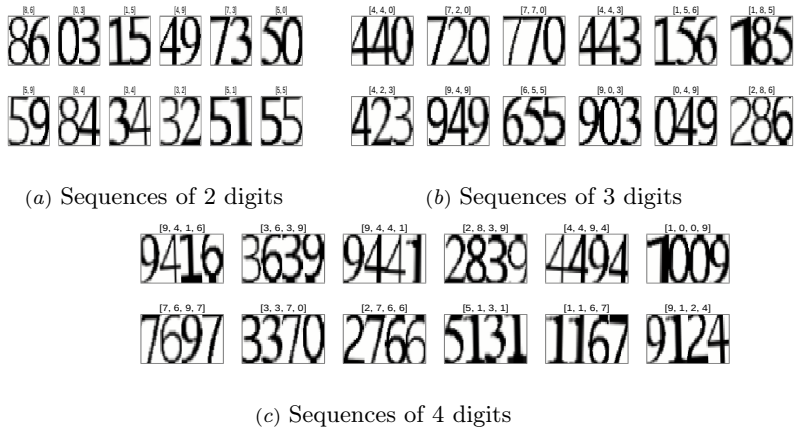


FIGURE 3. Examples of generated image sequences of 2, 3 and 4 digits. The class labels are shown at the top

documents, manual content parsing is usually used. As a consequence, it was decided to recognise sequences consisting of no more than 4 digits.

2.1. Generating datasets for training and testing models

At the initial stage of datasets generation for training and testing of CNN models, the approach proposed in [2] to form images of digits from 0 to 9 by the main fonts of cadastral documents was used. Then, depending on the number of digits in the sequence, their random selection, augmentation, varying the slope angle as necessary and scaling and merging the images into one, Figure 3.

In experimental studies, for sequences of 2, 3, and 4 digits, 1500, 2000 and 3000 images were generated, respectively, guaranteeing that all possible digit combinations were enumerated. About 70% of these images were used for training the CNN and 30% were used for validation and testing. The number of digit images (classes) in each of the digit sequences is the same; as a consequence, the training and testing datasets are balanced.

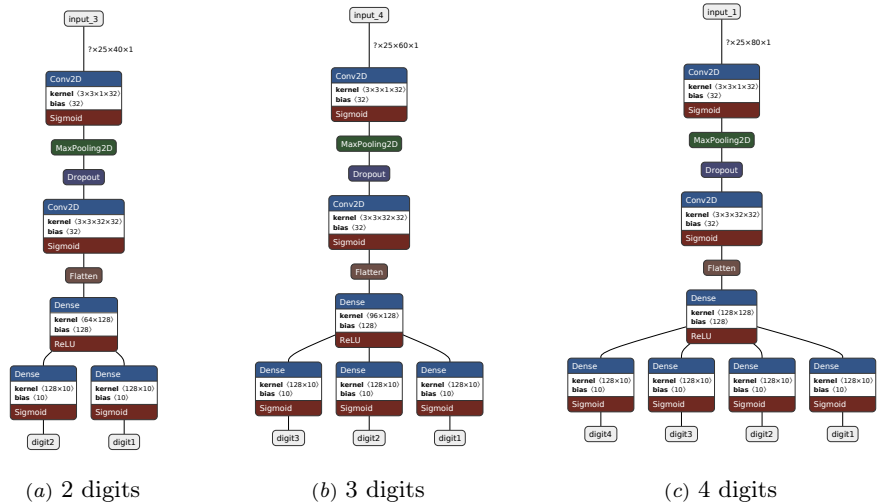


FIGURE 4. CNN models for recognising sequences of 2, 3 and 4 digits

2.2. Shaping the CNN models

Experimental researches using the Python language and the Keras [26] library have shown that CNN models that provide acceptable digital sequence recognition accuracies should be sequential and have 32 filters in 2 convolutional Conv2D¹ layers with a kernel size of 3×3 , while the sample size in 2 sampling layers with the MaxPooling2D¹ type should be 2×2 .

All models were compiled with the following parameters: loss function — *categorical_crossentropy*², optimiser — *adam*³, and basic metric — *accuracy*⁴, which gives an acceptable accuracy estimate for balanced datasets.

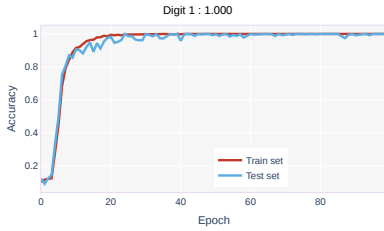
All CNN models have the same convolution and discretisation layers and differ only in the output full-link Dense¹ layers, the number of which is determined by the number of digits recognised by the CNN model, Figure 4. Each of these layers returns the probability that a digit belongs to one of 10 classes.

¹see the description *Keras.layers.API*^{URL}

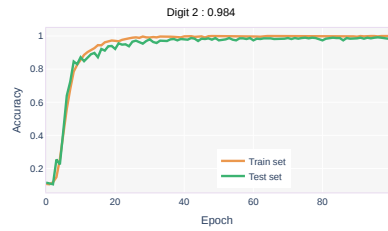
²see the description *Keras.Losses*^{URL}

³see the description *Keras.Optimizers*^{URL}

⁴see the description https://keras.io/api/metrics/accuracy_metrics/^{URL}

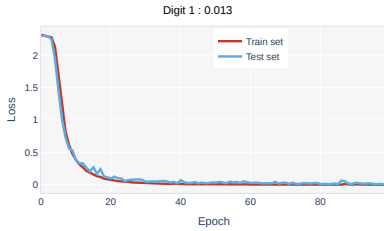


(a) Recognition accuracy of the 1st digit

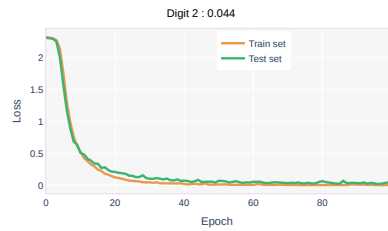


(b) Recognition accuracy of the 2nd digit

FIGURE 5. Recognition accuracy of 2 digits

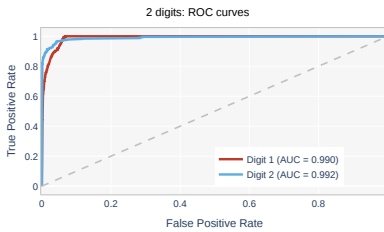


(a) Recognition loss of the 1st digit

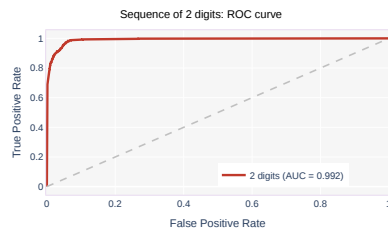


(b) Recognition loss of the 2nd digit

FIGURE 6. Recognition loss of 2 digits



(a) ROC when each of the 2 digits is recognised

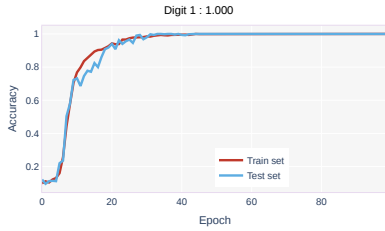


(b) ROC in recognition of sequences of 2 digits

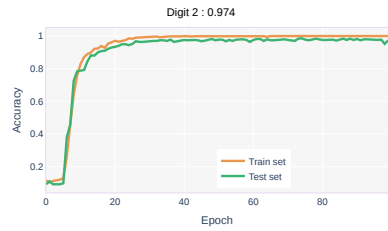
FIGURE 7. ROC curves for 2-digit recognition

2.2.1. CNN models research

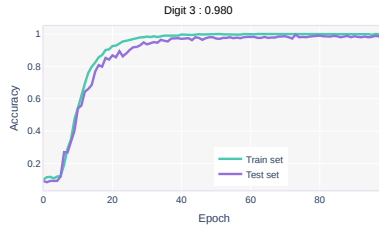
The results of the CNN models from Figure 4 at 100 (optimal) training epochs in the form of accuracy plots, loss functions and ROC curves for recognising sequences of 2 digits are shown in Figures 5, 6, 7.



(a) Recognition accuracy of the 1st digit

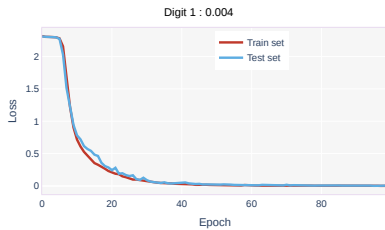


(b) Recognition accuracy of the 2nd digit

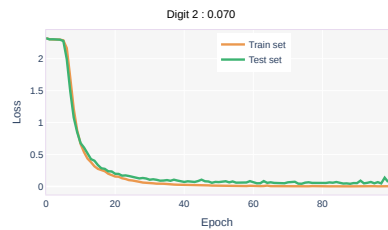


(c) Recognition accuracy of the 3rd digit

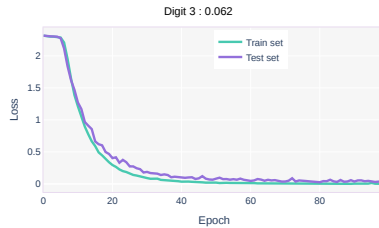
FIGURE 8. Recognition accuracy of 3 digits



(a) Recognition loss of the 1st digit



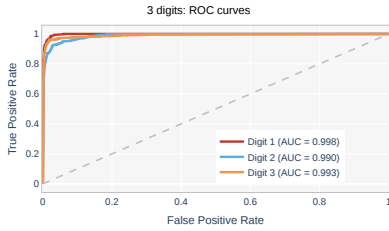
(b) Recognition loss of the 2nd digit



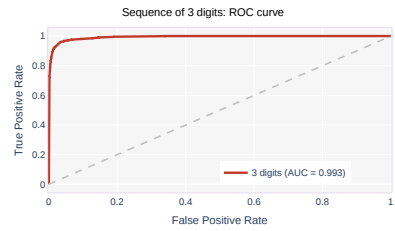
(c) Recognition loss of the 3rd digit

FIGURE 9. Recognition loss of 3 digits

For the recognition of sequences of 3 digits the results are shown in Figures 8, 9, 10.

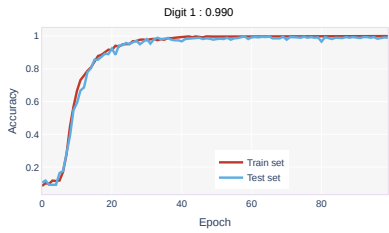


(a) ROC when each of the 3 digits is recognised

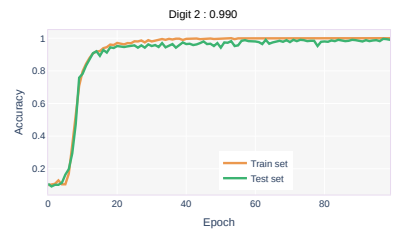


(b) ROC in recognition of sequences of 3 digits

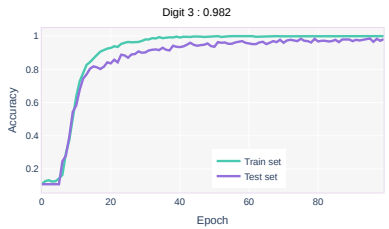
FIGURE 10. ROC curves for 3-digit recognition



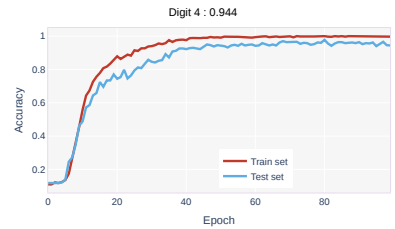
(a) Recognition accuracy of the 1st digit



(b) Recognition accuracy of the 3rd digit



(c) Recognition accuracy of the 3rd digit



(d) Recognition accuracy of the 4th digit

FIGURE 11. Recognition accuracy of 4 digits

The results of recognising sequences of 4 digits are shown in Figures 11, 12 and 13. Here and below the text above the graphs, the value of the investigated parameter at the last CNN training epoch is given.

The number of training epochs is determined experimentally and allows us to fine-tune the CNN models to recognise all possible combinations of digits generated at the stage of their augmentation. In order to eliminate possible overtraining, a Dropout⁵ layer was added to each of the models.

⁵see the description *Keras layers API*^{URL}

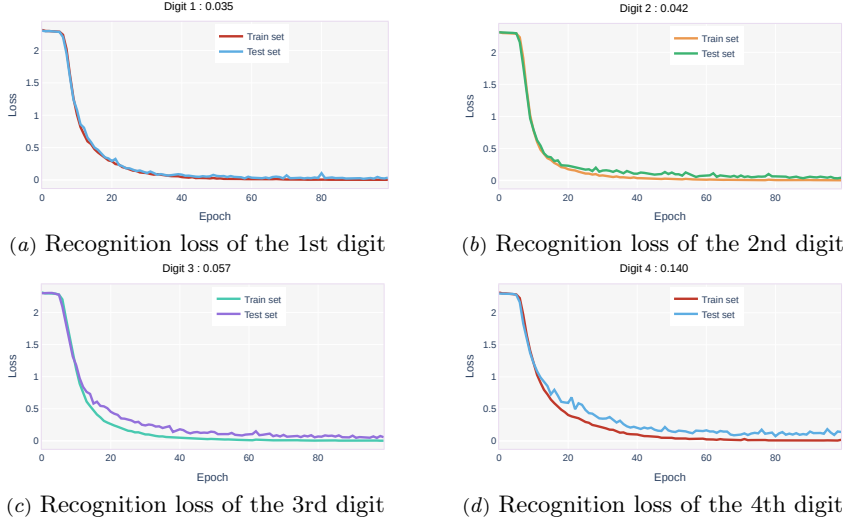


FIGURE 12. Recognition loss of 4 digits

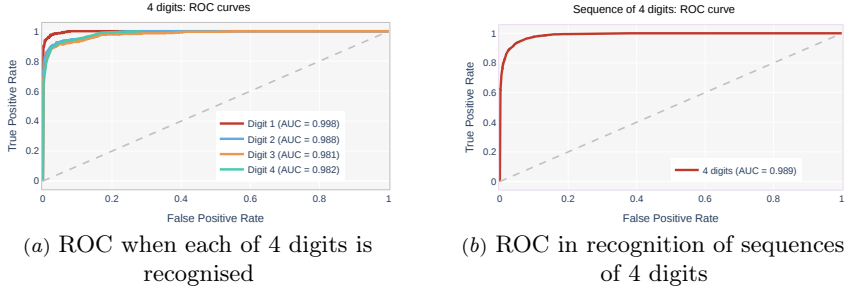


FIGURE 13. ROC curves for 4-digit recognition

TABLE 1. Values of the main metrics of model accuracy

Digits in sequences	An individual digit (accuracy, recall, f1)	A sequence of digits (accuracy, recall, f1)
2	(0.996, 0.984, 0.908)	(0.988, 0.987, 0.897)
3	(0.989, 0.980, 0.890)	(0.980, 0.984, 0.876)
4	(0.977, 0.970, 0.824)	(0.912, 0.945, 0.822)

The best values of the accuracy, recall and $f1^6$ metrics obtained from the experimental studies of the models are summarised in Table 1.

⁶see the description https://keras.io/api/metrics/accuracy_metrics/^(URL)

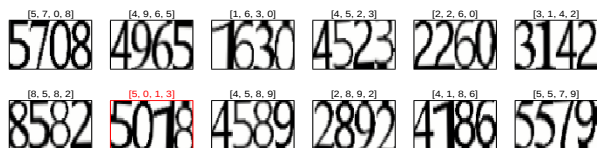


FIGURE 14. Error in recognising 4-digit sequences

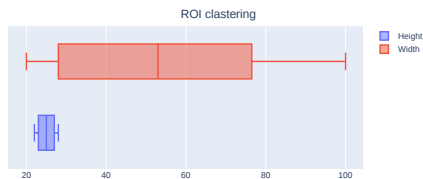


FIGURE 15. ROI height and width value ranges

The study of CNN models has shown the expected result — as the number of digits in the sequence increases, the accuracy of their recognition decreases. For example, the recognition error of a single digit in the sequences shown in Figure 3c is more common than the sequences in Figure 3a and autoreffig:inputs3. An example of erroneous recognition of a sequence of 4 digits is shown in Figure 14. Recognition errors do not reduce the significance of the obtained results, as they occur extremely rarely — 1 time per several dozens of documents. In the current implementation of the «Roskadastr» PLC IS digital sequence recognition errors are not processed.

3. Realisation of digit sequence recognition in IS

The CNN models described above have been implemented in the «Roskadastr» PLC IS designed for automated recognition of the contents of scanned documents, including those with object coordinates. The location of tables inside the document are detected using YOLOv5s deep learning network [1]. Next, ROI contours and their sizes are detected within the tables – the standard function *findContours()* of the OpenCV [3] library is used for this purpose. For example, for the coordinates shown in Figure 1, the ranges of ROI sizes are given in Figure 15.

The dimensions of the ROI (in most cases only the width) are used to estimate the number of digits in the sequence using the K-means clustering method [27]. The results of ROI clustering by this method for the coordinates from Figure 1 are shown in Figure 16.

The actual recognition process consists of enumerating all ROIs,

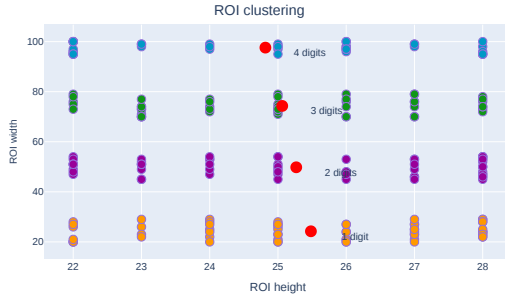


FIGURE 16. ROI clustering. Centroids of clusters are shown in red

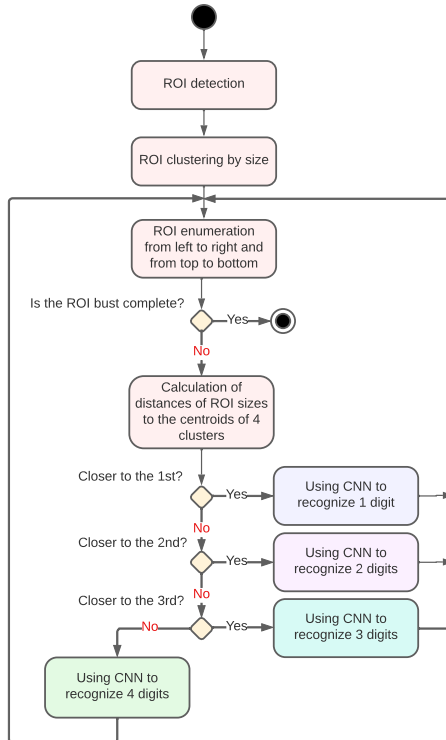


FIGURE 17. Diagram of the main stages of the IS's work on CNN model selection

correlating their sizes with the centroids of the 4 clusters, and selecting the corresponding CNN model, Figure 17.

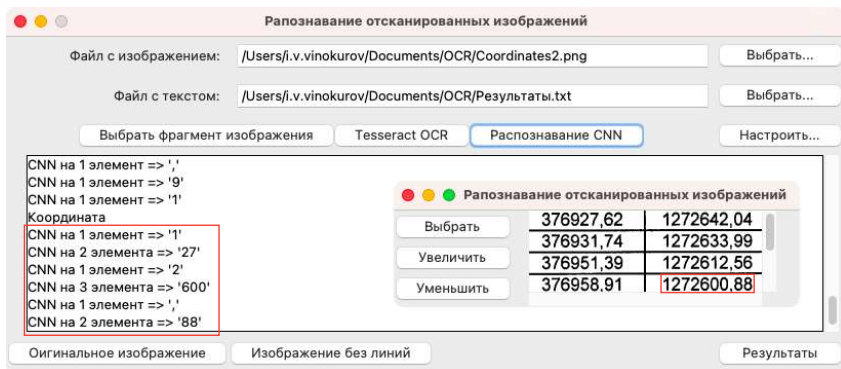


FIGURE 18. Implementation of object coordinates recognition using CNN models

An example of recognizing the coordinates of an object in the «Roskadastr» PLC IS is shown in Figure 18.

4. Advantages and disadvantages of the proposed approach

The obvious advantages of recognition of digital sequences using CNN, as applied to the documents of the «Roskadastr» PLC, include the absence of the need for preliminary image processing, as it was noted in [24], and rather high recognition accuracy. A slight decrease in the recognition accuracy of the 2nd, 3rd and 4th digits (see Figures 5*b*, 8*c*, 11*d*) can be explained by the structure of the CNN, which leads to the accumulation of the recognition error of subsequent digits.

A drawback is that ROI clustering is incorrect in some cases. The reason for this is the use of non-monospaced fonts in «Roskadastr» PLC documents. Incorrect clustering leads to the fact that the IS selects a CNN model that does not correspond to the numeric sequence.











For example, a numerical sequence of 3 units turns out to be closer to the centroid of a cluster of 2 numerical sequences, a numerical sequence of 4 units is closer to the centroid of a cluster of 3 numerical sequences, etc. Such situations are extremely rare, nevertheless, their occurrence is quite possible.

Conclusion






As a result of this research, CNN models were developed that allowed us to solve the problem of recognising sequences of 2, 3 and 4 digits quite effectively. Mathematical and experimental studies of these models have shown acceptable accuracy of recognition of both individual digits and their continuous sequences.

The developed models were used in the «Roskadastr» PLC IS to solve a rather significant at the moment task of converting scanned documents into their text analogues. According to preliminary estimates, the use of the proposed approach of digital sequence recognition using CNN allowed to increase the efficiency (speed, quality) of cadastral coordinates recognition on scanned documents of low quality by 20%–40% depending on the quality of the latter.

References

- [1] I.V. Vinokurov. “Tabular information recognition using convolutional neural networks”, *Program Systems: Theory and Applications*, **14**:1(56) (2023), pp. 3–30 (in Russian).   ↑²¹, 31
- [2] I.V. Vinokurov. “Using a convolutional neural network to recognize text elements in poor quality scanned images”, *Program Systems: Theory and Applications*, **13**:3(54) (2022), pp. 45–59.   ↑²¹, 25
- [3] A. C. Müller, S. Guido. *Introduction to Machine Learning with Python*, O’Reilly Media, Inc., 2016, ISBN 9781449369897, 338 pp. ↑²¹, 31
- [4] M. Diem, S. Fiel, F. Kleber, R. Sablatnig, J. M. Saavedra, D. Contreras, J. M. Barrios, L. S. Oliveira. “ICFHR 2014 competition on handwritten digit string recognition in challenging datasets (HDSRC 2014)”, *2014 14th International Conference on Frontiers in Handwriting Recognition* (Hersonissos, Greece, 01–04 September 2014), pp. 779–784.  ↑²², 23
- [5] N. Otsu. “A threshold selection method from gray-level histograms”, *IEEE Trans. Syst. Man Cybern.*, **9**:1 (1979), pp. 62–66.  ↑²²
- [6] J. C. B. Rabelo, C. Zanchettin, C. A. B. Mello and B. L. D. Bezerra. “A multi-layer perceptron approach to threshold documents with complex background”, 2011 IEEE International Conference on Systems, Man, and Cybernetics (Anchorage, AK, USA, 09–12 October 2011), pp. 2523–2530.  ↑²²
- [7] J. Sauvola, M. Pietikäinen. “Adaptive document image binarization”, *Pattern Recognition*, **33**:2 (2000), pp. 225–236.  ↑²²
- [8] C. Cortes, V. Vapnik. “Support-vector networks”, *Mach. Learn.*, **20**:3 (1995), pp. 273–297.  ↑²², 23
- [9] T. Kohonen. “Self-organized formation of topologically correct feature maps”, *Biol. Cybern.*, **43**:1 (1982), pp. 59–69.  ↑²²

- [10] N. Dalal, B. Triggs. “Histograms of oriented gradients for human detection”, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, CVPR’05 (San Diego, CA, USA, 20–25 June 2005), ISBN 0-7695-2372-2, pp. 886–893. [doi](#) ↑²²
- [11] J. J. Hopfield. “Neural networks and physical systems with emergent collective computational abilities”, *Proc. Nat. Acad. Sci. USA*, **79**:8 (1982), pp. 2554–2558. [doi](#) ↑²²
- [12] A. Graves, J. Schmidhuber. “Offline handwriting recognition with multidimensional recurrent neural networks”, *Guide to OCR for Arabic Scripts*, eds. Märgner V., El Abed H., Springer, London, ISBN 978-1-4471-4072-6, pp. 297–313. [doi](#) ↑²²
- [13] B. L. D. Bezerra, C. Zanchettin, V. B. de Andrade. “A MDRNN-SVM hybrid model for cursive offline handwriting recognition”, *Artificial Neural Networks and Machine Learning — ICANN 2012, Lecture Notes in Computer Science*, vol. **7553**, Springer, Berlin, 2012, ISBN 978-3-642-33265-4, pp. 246–254. [doi](#) ↑²²
- [14] A. Gattal, Y. Chibani. “Segmentation strategy of handwritten connected digits (SSHCD)”, *Image Analysis and Processing (ICIAP)*, 2011, pp. 248–254. [doi](#) ↑²²
- [15] M. Hejazi, G. Shevlyakov and Y.-S. Ho. “Modified discrete radon transforms and their application to rotation-invariant image analysis”, *2006 IEEE Workshop on Multimedia Signal Processing, Lecture Notes in Computer Science*, vol. **6979**, Springer, Berlin–Heidelberg, 2006, ISBN 978-3-642-24088-1, pp. 429–434. [doi](#) ↑²³
- [16] R. Saabni. “Recognizing handwritten single digits and digit strings using deep architecture of neural networks”, *2016 Third International Conference on Artificial Intelligence and Pattern Recognition*, AIPR (Lodz, Poland, 19–21 September 2016), 2016, pp. 1–6. [doi](#) ↑²³
- [17] K. He, X. Zhang, S. Ren, J. Sun. “Deep residual learning for image recognition”, *2016 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR (Las Vegas, NV, USA, 27–30 June 2016), pp. 770–778. [doi](#) ↑²³
- [18] H. Zhan, Q. Wang, Y. Lu. “Handwritten digit string recognition by combination of residual network and RNN-CTC”, *ICONIP 2017: Neural Information Processing, Lecture Notes in Computer Science*, vol. **10639**, Springer, Cham, 2017, ISBN 978-3-319-70136-3, pp. 583–591. [doi](#) ↑²³
- [19] Q. Wang, Y. Lu. “A sequence labeling convolutional network and its application to handwritten string recognition”, *Twenty-Sixth International Joint Conference on Artificial Intelligence (Melbourne, Australia, 19–25 August 2017)*, 2017, pp. 2950–2956. [doi](#) ↑²³
- [20] H. Zhan, S. Lyu, X. Tu, Y. Lu. “Residual CRNN and its application to handwritten digit string recognition”, *ICONIP 2019: Neural Information Processing, Communications in Computer and Information Science*, vol. **1143**, Springer, Cham, 2019, ISBN 978-3-030-36802-9, pp. 49–56. [doi](#) ↑²³
- [21] Y. Ma, J. Guo, W. Wei. “An exceedingly fast model for low resolution handwritten digit string recognition”, *2019 IEEE 7th International Conference on Computer Science and Network Technology*, ICCSNT (Dalian, China, 9–20 October 2019), pp. 282–288. [doi](#) ↑²³

- [22] A. G. Hochuli, A. S. Britto, J. P. Barddal, R. Sabourin, L. E. S. Oliveira. “An End-to-End approach for recognition of modern and historical handwritten numeral strings”, *2020 International Joint Conference on Neural Networks, IJCNN* (Glasgow, UK, 19–24 July 2020), 2020, pp. 1–8.  ↑²³
- [23] J. Redmon, S. Divvala, R. Girshick, A. Farhadi. “You only look once: Unified real-time object detection”, *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR* (Las Vegas, NV, USA, 27–30 June 2016), pp. 779–788.  ↑²³
- [24] A. G. Hochuli, L. S. Oliveira, A. S. Britto, R. Sabourin. Handwritten digit segmentation: Is it still necessary? *Pattern Recognition*, **78** (2018), pp. 1–11.  ↑²⁴, 33
- [25] J. Guo, Y. Xu. “Low Resolution Handwritten Digit String Recognition based on Object Detection Network”, *2020 International Joint Conference on Neural Networks, IJCNN* (Glasgow, UK, 19–24 July 2020), pp. 1–7.  ↑²⁴
- [26] J. Moolayil. *Learn Keras for Deep Neural Networks. A Fast-Track Approach to Modern Deep Learning with Python*, Apress, Berkeley, 2019, ISBN 978-1-4842-4240-7, xv+182 pp.  ↑²⁶
- [27] S. Datta. *Learning OpenCV 3 Application Development*, Packt Publishing, 2016, ISBN 9781784391454, 294 pp. ↑³¹

Received	14.04.2023;
approved after reviewing	04.07.2023;
accepted for publication	04.07.2023;
published online	13.08.2023.

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.