

УДК 004.932.75'1, 004.89

doi 10.25209/2079-3316-2024-15-1-3-30



# Распознавание кадастровых координат с использованием свёрточно-рекуррентных нейронных сетей

Игорь Викторович **Винокуров**

Финансовый Университет при Правительстве Российской Федерации, Москва, Россия

[igvvinokurov@fa.ru](mailto:igvvinokurov@fa.ru)

**Аннотация.** В статье исследуется применение свёрточно-рекуррентных нейронных сетей (CRNN) для распознавания изображений кадастровых координат объектов на отсканированных документах ППК «Роскадастр». Комбинированная архитектура CRNN, объединяющая свёрточные нейронные сети (CNN) и рекуррентные нейронные сети (RNN), позволяет использовать преимущества каждой из них для обработки изображений и распознавания содержащихся в них непрерывных цифровых последовательностей. При проведении экспериментальных исследований были сформированы изображения, состоящие из заданного количества цифр, построена и исследована CRNN модель. Формирование изображений цифровых последовательностей заключалось в предобработке и конкатенации изображений образующих их цифр из собственного набора данных. Анализ значений функции потерь и метрик Accuracy, Character Error Rate (CER) и Word Error Rate (WER) показал, что использование предложенной CRNN модели позволяет достичь высокой точности распознавания кадастровых координат на их отсканированных изображениях.

(Связанные тексты статьи на русском и на английском языках)

**Ключевые слова и фразы:** свёрточно-рекуррентная нейронная сеть, CRNN, распознавание изображений, цифровые последовательности, глубокое обучение, Keras, Python

Для цитирования: Винокуров И. В. *Распознавание кадастровых координат с использованием свёрточно-рекуррентных нейронных сетей* // Программные системы: теория и приложения. 2024. Т. 15. № 1(60). С. 3–30. (Русс.+англ.). [https://psta.psir.ru/read/psta2024\\_1\\_3-30.pdf](https://psta.psir.ru/read/psta2024_1_3-30.pdf)

## Введение

Распознавание буквенно-цифровых последовательностей на изображениях является значимой задачей в области компьютерного зрения и обработки изображений. Эффективное решение этой задачи имеет большое значение для автоматизации и оптимизации различных процессов, связанных с идентификацией и классификацией объектов.

В последние годы с использованием нейронных сетей глубокого обучения, особенно CNN, были получены впечатляющие результаты в распознавании и классификации изображений. Однако классические CNN ориентированы на выявление признаков входных данных, что ограничивает их применимость для распознавания последовательностей переменной длины. С целью преодоления этого ограничения была разработана архитектура CRNN, которая объединяет преимущества свёрточных и рекуррентных нейронных сетей. Основной особенностью архитектуры CRNN является совмещение свёрточных слоев CNN для извлечения локальных и пространственных признаков из изображений и рекуррентных слоёв RNN для учёта контекста и последовательности информации. Свёрточные слои позволяют обнаружить важные особенности изображений на разных уровнях абстракции, в то время как рекуррентные слои моделируют зависимости и последовательность входных данных [1].

В наиболее общем случае архитектура CRNN состоит из трёх основных компонентов – свёрточного, рекуррентного и компонента классификации. Свёрточный компонент содержит несколько свёрточных слоев, выполняющих обнаружение и извлечение признаков из изображений. Рекуррентный компонент включает рекуррентные слои долгой краткосрочной памяти (LSTM) [2] или управляемые рекуррентные блоки (GRU) [3], позволяющие учитывать контекст и последовательность входных данных. Компонент классификации выполняет распознавание и классификацию последовательностей на основе предсказаний рекуррентного компонента.

В данной статье проводится исследование эффективности использования CRNN для распознавания изображений цифровых последовательностей переменной длины. В разделе 1 осуществляется обоснование необходимости исследований и постановка задачи. Раздел 2 посвящён обзору и анализу работ по использованию сетей CRNN для распознавания текста и цифровых последовательностей. Создание набора данных для обучения модели описано в разделе 3. Формирование и исследование CRNN модели для распознавания цифровых последовательностей на изображениях приведено в разделе 4. В заключении приведены выводы по результатам проведённых исследований.

## 1. Постановка цели и задач исследования

Целью исследования является разработка CRNN модели, способной реализовать приемлемую точность распознавания кадастровых координат на отсканированных документах ППК «Роскадастр». В [4], [5] и [6] описаны реализованные в информационной системе (ИС) этой организации подходы к решению задачи преобразования изображений кадастровых координат в их текстовые аналоги с использованием моделей CNN. В [6] показывается эффективность использования CNN для последовательностей, состоящих не более чем из 4-х цифр. По результатам проведённых в [6] исследований, можно сделать вывод, что при распознавании большего количества цифр, структура CNN усложняется, качество распознавания при этом либо остаётся прежним, либо увеличивается незначительно. Учитывая, что максимальное количество символьно-цифровых элементов в кадастровых координатах может превышать 10, одним из целесообразных и эффективных средств решения задачи их распознавания может являться использование CRNN моделей. Поставленная в работе цель может быть достигнута за счёт решения следующих основных задач:

- (1) Формирование набора данных, заключающееся в подготовке набора изображений элементов кадастровых номеров и их аннотирование (сопоставление с кадастровым номером); формирование изображений кадастровых координат на заданное количество их цифровых элементов; разделении набора данных на выборки для обучения и валидации.
- (2) Формирование CRNN модели – выбор свёрточных и рекуррентных слоёв для извлечения признаков и распознавания цифровых последовательностей соответственно.
- (3) Обучение модели и анализ значений функций потерь и метрик точности, CER и WER.

## 2. Анализ основных работ по распознаванию буквенно-цифровых последовательностей

Впервые CRNN модель, объединяющая свёрточные и рекуррентные слои для обработки изображений с текстовыми последовательностями описана в [1]. Достоинством модели является сочетание свёрточных и рекуррентных слоёв, позволяющее выявлять как локальные, так и глобальные зависимости в изображениях, содержащих последовательности, и достаточно эффективно реализовывать их распознавание. К недостаткам предложенной модели могут быть отнесены достаточно большие объёмы данных для обучения и значительное время распознавания длинных последовательностей.

В [7] описана CRNN модель, которая может распознавать неизвестные ей слова, используя значимую контекстную информацию. Модель является устойчивой к различным искажениям изображения, не зависит от заранее определённого словаря и может обрабатывать произвольные предложения. Недостаток модели – плохое распознавание текста с низкой контрастностью, нечёткими границами и искажениями, что требует дополнительных методов предварительной обработки изображения.

Модель из [8] предназначена для распознавания текста на изображениях с искажением перспективы, изогнутым расположением символов и т. д. Предложенная CRNN модель способна обеспечить приемлемую читаемость и распознаваемость искажённого текста и превосходит аналогичные модели при распознавании текста с разными значениями входного шума и наклона. Кроме этого, модель демонстрирует высокую точность и производительность при обучении на больших наборах данных. К недостатку предложенной модели может быть отнесено плохое распознавание текста, если шум или деформация сильно изменяют форму символов, что особенно сильно заметно для текста с малым шрифтом или низким разрешением.

В статье [9] авторы применяют CRNN для распознавания текста на изображениях, сосредотачиваясь на сложных случаях, таких как сцены с плохим освещением или низким разрешением. Они предлагают модель, которая использует свёрточные слои для извлечения признаков из изображений и рекуррентные слои для моделирования последовательностей из этих символов. В экспериментах демонстрируется высокая точность распознавания текста на различных изображениях текстовых последовательностей.

Модель для многоуровневого распознавание рукописного текста на изображениях приведена в [10]. На первом уровне для распознавания слов, которые часто встречаются в тексте, используется CNN. Если слово не распознается этой моделью, оно переходит на второй уровень, на котором используется полностью свёрточной сети (FCN). Экспериментальное исследование модели проводилось с использованием NIST19 в качестве набора данных для обучения и рукописного текста в качестве набора тестовых данных и показало вполне приемлемый результат распознавания.

В [11] представлены две модели для распознавания последовательности цифр. В первой кодером и декодером последовательностей являются CNN и LSTM. Во второй – гистограмма ориентированного градиента (HOG) и параллельные полносвязные (Dense) слои соответственно. Обучение и тестирование осуществлялось на наборе данных Street View Number House

(SVHN). В результате проведённых исследований показано преимущество CNN в отношении кодирования изображений и преимущество LSTM в предсказании последовательностей.

Сеть глубокого обучения DIGI-Net, которая способна изучать общие характеристики трех различных форматов цифр (рукописные, естественные изображения, печатный шрифт) и распознавать их описана в [12]. Эксперименты, проведённые на наборах данных MNIST, CVL и Chars74K, продемонстрировали высокую точность распознавания непрерывных цифровых последовательностей.

Предложенная в [13] CRNN модель имеет свёрточный слой, слой слияния признаков, рекуррентный слой и слой транскрипции. Свёрточный слой, используемый для извлечения признаков, формирует два результата для входного текстового изображения. Слой объединения объектов объединяет результаты работы свёрточного слоя в один, из которого рекуррентный слой извлекает последовательности. Окончательный результат выводит слой транскрипции. Предложенная модель за счёт слияния признаков реализует лучшую точность распознавания текста на наборах текстовых данных Street View Text (SVT), IIT-5K, ICDAR2003 и ICDAR2013.

В [14] для распознавания рукописных цифр предлагается использовать гибридную архитектуру – визуальные трансформеры (ViTs) и многослойные перцептроны (MLP). Проведённые исследования на наборах данных EMNIST и DIDA показали хорошую точность распознавания машинописных цифровых, в том числе, и на зашумлённых изображениях.

В [15] предлагается симметричная многомасштабная архитектура под названием Circular Dilated Convolutional Neural Network (CDIL-CNN), где каждый элемент текущего уровня имеет равные шансы получить информацию от других элементов с предыдущих уровней. Предлагаемая CRNN модель позволяет сформировать логиты (logits) классификации для всех элементов, в результате чего становится возможным применение простого ансамблевого обучения для принятия лучшего решения. По результатам тестирования CDIL-CNN на длинных последовательных наборах данных показано, что CDIL-CNN позволяет получить приемлемый по точности результат распознавания.

Лучший подход к формированию моделей для распознавания последовательностей, с точки зрения автора этой работы, приведён в [16] и [17]. Для распознавания текстовых последовательностей предлагается использовать кодер в виде CNN и декодер в виде двунаправленной

долгосрочной краткосрочной памяти (BSTM) с использованием коннекционистской временной классификации (СТС). СТС является алгоритмом, используемым для обучения RNN на последовательностях с переменной длиной и сопоставления их с соответствующими метками. В задачах распознавания текста и цифровых последовательностей, СТС может справиться с проблемой их переменной длины. Он позволяет модели прогнозировать переменное количество букв или цифр в последовательности без их предварительного разделения или выравнивания. Алгоритм СТС вычисляет вероятность выходной последовательности и рекуррентно обновляет веса модели на основе разницы между прогнозами и метками последовательностей. Эффективность предложенного подхода показана на собственном наборе данных в [16] и на наборах данных арабских букв MADCAT, ANTID/MW и IFN/ENIT в [17].

### 3. Формирование набора данных

Для обучения CRNN модели и исследования её работы был сформирован собственный набор данных.

На первом этапе, по аналогии с [4], формировались чёрно-белые изображения элементов цифровых последовательностей с использованием основных шрифтов документов ППК «Роскадастр». Количество классов изображений выбрано равным 12 – 10 классов для цифр от 0 до 9, 1 класс для символов-разделителей «.» и «,» и ещё один на отсутствие символа в последовательности. Значения 2-х последних классов выбраны равными 10 и 11 соответственно. Для каждого класса изображений было сформировано по 10 и 5 изображений размером 20 × 25 пикселей для обучения и валидации модели соответственно.

На втором этапе формировались изображения кадастровых координат, состоящие из 2-х цифр в дробной части и от 4-х до 7-и в целой. Одновременно с формированием изображений формировались и их СТС-метки. Все изображения в наборе данных приводились к одному размеру 200 × 32 пикселей. Пример сформированных изображений кадастровых координат и соответствующих им СТС-меток приведён на рисунке 1.

Сформированный таким образом набор данных приведён в таблице 1 и состоит из 24240 и 12240 изображений кадастровых координат и их меток для обучения и валидации соответственно.

### 4. Формирование и исследование CRNN модели

Формирование модели осуществлялось с использованием библиотеки Keras. Все слои этой модели, помимо основных слоёв CRNN – свёрточных

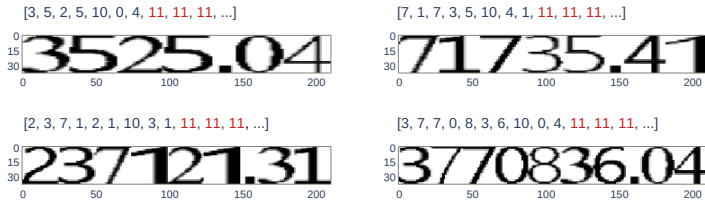


Рисунок 1. Примеры изображений кадастровых координат. Сверху указаны их CTC-метки

Таблица 1. Количество цифровых последовательностей для обучения и валидации

Цифр в последовательности	Последовательностей для обучения	Последовательностей для валидации
6	4848	2448
7	5656	2856
8	6464	3264
9	7272	3672

и рекуррентных (Conv2D и Bidirectional соответственно), приведены в таблице 2.

Субдискретизирующий слой MaxPooling2D уменьшает размерность пространства признаков, выделяя самые значимые из них.

Слой BatchNormalization нормализует данные по мини-пакетам, что позволяет ускорить сходимость обучения и уменьшить вероятность переобучения. Также он помогает стабилизировать распределение активаций между слоями.

Слой Dropout является регуляризатором, его цель – снижение переобучения за счёт предотвращения активации случайно выбранных нейронов. Это заставляет модель обучаться более устойчивыми признаками и уменьшает вклад каждого отдельного нейрона.

Слой Activation (активационная функция) применяет функцию активации к выводу предыдущего слоя. В данной модели это ReLU, которая активирует нейроны только при положительных значениях, Tanh – при положительных и отрицательных значениях и Softmax, генерирующая вероятности для разных классов.

Полносвязный слой Dense объединяет все выходы предыдущего слоя и применяет линейные преобразования для получения финального вывода модели. Он связывает выходные сигналы всех нейронов предыдущего

ТАБЛИЦА 2. Слои CRNN модели

Тип слоя	Функция активации	Количество фильтров	Входной вектор
InputLayer	–	–	[(None, 200, 32, 1)]
Conv2D	–	32	(None, 200, 32, 32)
BatchNormalization	–	–	(None, 200, 32, 32)
Activation	ReLu	–	(None, 200, 32, 32)
MaxPooling2D	–	–	(None, 100, 16, 32)
Conv2D	–	64	(None, 100, 16, 64)
BatchNormalization	–	–	(None, 100, 16, 64)
Activation	ReLu	–	(None, 100, 16, 64)
MaxPooling2D	–	–	(None, 50, 8, 64)
Dropout	–	–	(None, 50, 8, 64)
Conv2D	–	128	(None, 50, 8, 128)
BatchNormalization	–	–	(None, 50, 8, 128)
Activation	ReLu	–	(None, 50, 8, 128)
MaxPooling2D	–	–	(None, 50, 4, 128)
Dropout	–	–	(None, 50, 4, 128)
Conv2D	–	256	(None, 50, 4, 256)
BatchNormalization	–	–	(None, 50, 4, 256)
Activation	ReLu	–	(None, 50, 4, 256)
MaxPooling2D	–	–	(None, 50, 2, 256)
Dropout	–	–	(None, 50, 2, 256)
Reshape	–	–	(None, 32, 800)
Dense	–	–	(None, 32, 25)
Bidirectional	Tanh	–	(None, 32, 320)
Bidirectional	Tanh	–	(None, 32, 320)
Dense	–	–	(None, 32, 12)
Activation	Softmax	–	(None, 32, 12)

слоя с каждым нейроном в текущем слое и является основным слоем классификации в нейронных сетях.

Слой Lambda позволяет задать собственную лямбда-функцию для нестандартного преобразования данных, например, изменения их размерности. В сформированной модели он используется совместно с ground-truth метками, используемыми для задач распознавания непрерывных последовательностей. На этапе обучения эти метки подвергаются согласованию и сравниваются с истинными метками, в результате чего возникает ошибка, используемая для оптимизации модели.

Структура CRNN модели приведена на рисунке 2. Для её обучения был использован метод обратного распространения ошибки с оптимизатором Adam.



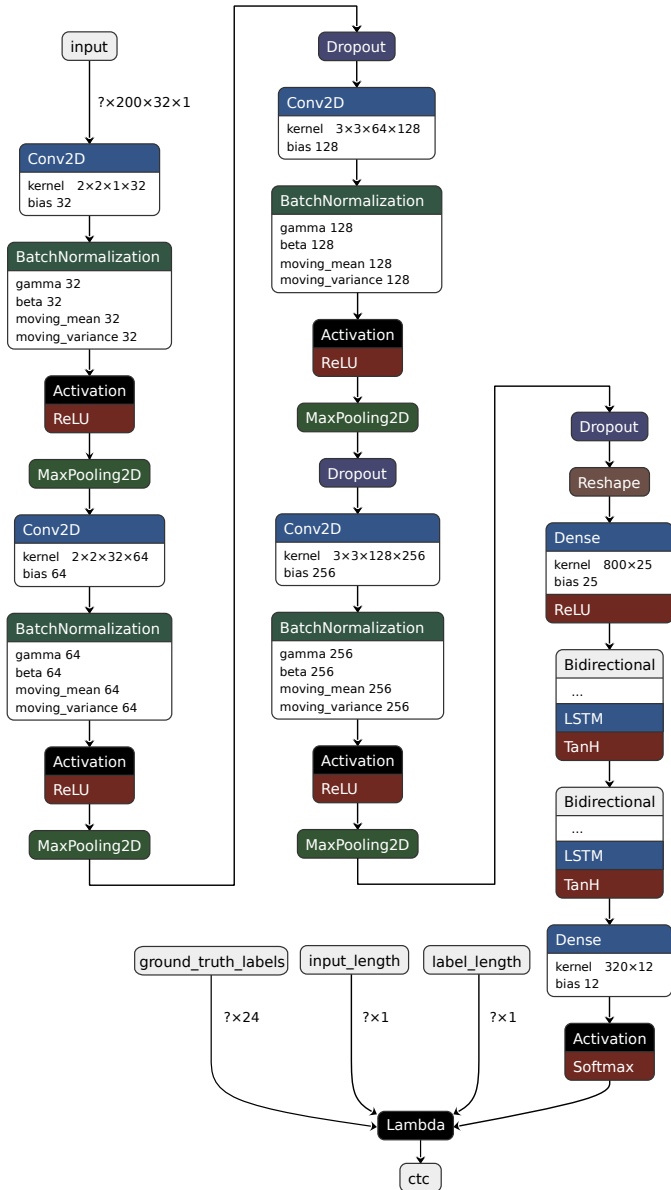


РИСУНОК 2. CRNN модель

Для оценки качества работы модели и её способности решать задачу осуществлялось вычисление значений функции потерь и метрик точности, CER и WER. Функция потерь (Loss) измеряет разницу между фактическими (истинными) значениями меток и значениями, предсказанными моделью. На рисунке 3 приведены потери модели для 9 эпох обучения. Такое количество эпох обучения найдено экспериментальным путём и является оптимальным.

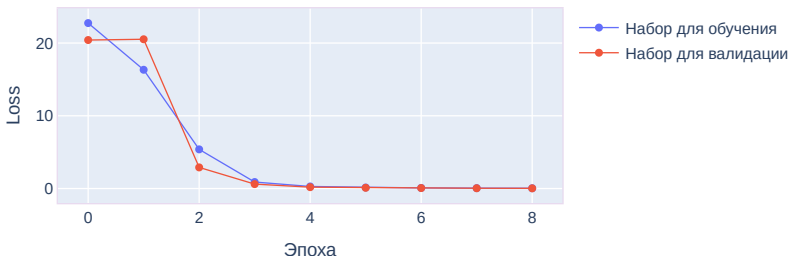


Рисунок 3. Потери (Loss) модели

Метрика точности (Accuracy) определяет долю правильных предсказаний, сделанных моделью, по отношению к общему количеству примеров. Она позволяет оценить, насколько хорошо модель может классифицировать или предсказывать правильный класс или значение для данного набора данных, рисунок 4.

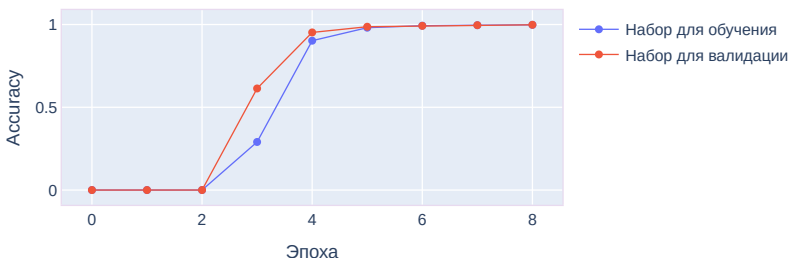


Рисунок 4. Точность (Accuracy) модели

Числовые значения функций потерь и метрик точности для наборов обучения и валидации на каждой из эпох обучения приведены в таблице 3.

Вычисление значений функции потерь и метрик точности является обычным подходом к оценке качества нейросетевых моделей. Для моделей, ориентированных на распознавание последовательностей, вычисляются ещё две – Character Error Rate (CER) и Word Error Rate (WER). Обе метрики используются для сравнительной оценки различных систем распознавания и анализа их точности.

ТАБЛИЦА 3. Числовые значения Loss и Accuracy

Номер эпохи	Loss		Accuracy	
	Набор для обучения	Набор для валидации	Набор для обучения	Набор для валидации
Epoch	Train set	Validation set	Train set	Validation set
0	22.7430438	20.4115753	0.0	0.0
1	16.3242683	20.5168876	0.0	0.0
2	5.3856644	2.9053363	0.0	0.0000817
3	0.8966413	0.6119849	0.2904703	0.6136437
4	0.2865071	0.1913659	0.9022276	0.9521241
5	0.1421806	0.1023578	0.9807755	0.9866012
6	0.0855829	0.0680020	0.9928630	0.9919934
7	0.0573494	0.0432046	0.9958745	0.9959967
8	0.0411685	0.0314824	0.9980198	0.9982843

CER позволяет оценить точность распознавания моделью отдельных символов. Зависимости значений этой метрики для набора данных и набора для валидации от номера эпохи обучения показано на рисунке 5.

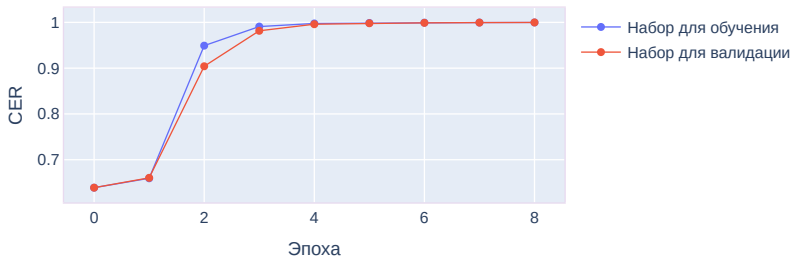


Рисунок 5. Точность распознавания символов (CER)

Точность распознавания моделью целых слов позволяет оценить метрика WER. Зависимости значений метрики для набора данных и набора для валидации от номера эпохи обучения показано на рисунке 6.

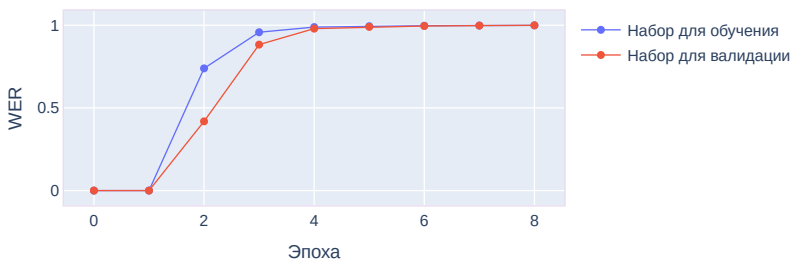


Рисунок 6. Точность распознавания слов (WER)

Числовые значения метрик CER и WER для наборов обучения и

валидации на каждой из эпох обучения приведены в таблице 4.

Таблица 4. Значения метрик CER и WER

Номер эпохи	CER		WER	
	Набор для обучения	Набор для валидации	Набор для обучения	Набор для валидации
0	0.6388888	0.6388888	0.0	0.0
1	0.6594093	0.6603758	0.0	0.0
2	0.9491560	20.9041156	0.7390676	0.4186274
3	0.9908536	0.9816210	0.9576320	0.8828431
4	0.9976399	0.9960001	0.9887788	0.9794117
5	0.9987451	0.9981072	0.9940182	0.9900327
6	0.9993949	0.9989787	0.9970297	0.9953431
7	0.9996046	0.9995132	0.9981435	0.9977942
8	0.999804	0.9998400	0.9992161	0.9994281

Результаты распознавания моделью нескольких кадастровых координат на 6,7,8 и 9 цифр соответственно из тестового набора данных приведены на рисунке 7.



Рисунок 7. Результаты распознавания кадастровых координат. Одна из координат распознана с ошибкой























Предложенная CRNN модель реализована в ИС ППК «Роскадастр» и, как показали результаты её экспериментального исследования, точность распознавания отдельных символов и кадастровых координат в целом составила 99.98% и 99.94% соответственно. Изображения кадастровых координат выделялись из отсканированного документа по координатам, формируемым подсистемой контуризации этой ИС. Общие принципы работы этой подсистемы описаны в [4].

## Заключение

В статье проведены исследования применения архитектуры CRNN для задачи распознавания изображений кадастровых координат. По результа-

там обучения модели были построены графики функции потерь (Loss) и точности (Accuracy). Графики показали, что модель успешно сходится и способна достичь высокой точности распознавания кадастровых координат. Для оценки качества работы модели были использованы метрики CER (Character Error Rate) и WER (Word Error Rate), которые позволили измерить процент ошибок на уровне символов и слов соответственно. По результатам экспериментальных исследований можно сделать вывод, что модель способна распознать кадастровые координаты с высоким уровнем точности и минимальным количеством ошибок. Как следствие, применение CRNN модели позволит значительно улучшить эффективность и достоверность геопространственных анализов и принятия решений. Дальнейшие исследования могут быть направлены на расширение набора данных, включение разных типов шрифтов и стилей, обучение модели на более разнообразных данных и исследование эффективности CRNN для других задач распознавания и классификации. Также возможно применение дополнительных методов предобработки данных или аугментации для улучшения точности модели.

### Список использованных источников

- [1] Shi B., Bai X., Yao C. *An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence.– 2017.– Vol. **39**.– No. 11.– Pp. 2298–2304. arXiv:1507.05717 [cs.CV]     4, 5
- [2] Hochreiter S., Schmidhuber J. *Long short-term memory* // Neural Computation.– 1997.– Vol. **9**.– No. 8.– Pp. 1735–1780.  
- [3] Chung J., Gulcehre C., Cho K., Bengio Y. *Gated feedback recurrent neural networks* // Proceedings of Machine Learning Research.– 2015.– Vol. **37**.– Pp. 2067–2075. arXiv:1502.02367 [cs.NE]  
- [4] 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.     5, 8, 14
- [5] Винокуров И. В. *Распознавание табличной информации с использованием свёрточных нейронных сетей* // Программные системы: теория и приложения.– 2023.– Т. **14**.– № 1(56).– С. 3–30.     5
- [6] Винокуров И. В. *Распознавание цифровых последовательностей с использованием свёрточных нейронных сетей* // Программные системы: теория и приложения.– 2023.– Т. **14**.– № 3(58).– С. 3–36.     5
- [7] He P., Huang W., Qiao Y., Change Loy C., Tang X. *Reading scene text in deep convolutional sequences*, AAAI'16: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (Phoenix, Arizona, USA, February 12–17, 2016) // Proceedings of the AAAI Conference on Artificial Intelligence.– 2016.– Vol. **30** **1**.– Pp. 3501–3508.   6
- [8] Shi B., Wang X., Lv P., Yao C., Bai X. *Robust scene text recognition with automatic rectification*, 2016 IEEE Conference on Computer Vision and Pattern Recognition

- (CVPR) (Las Vegas, NV, USA, June 27–30, 2016).– 2016.– Pp. 4168–4176. [doi](#) [arXiv:1603.03915 \[cs.CV\]](#) ↑<sub>6</sub>
- [9] Yin F., Wu Y.-C., Zhang X.-Y., Liu C.-L. *Scene text recognition with sliding convolutional character models.*– 2017.– 10 pp. [arXiv:1709.01727 \[cs.CV\]](#) ↑<sub>6</sub>
- [10] Nirmalasari D. A., Suciati N., Navastara D. A. *Handwritten text recognition using fully convolutional network* // IOP Conference Series: Materials Science and Engineering.– 2021.– Vol. **1077**.– No. 1.– id. 012030.– 9 pp. [doi](#) ↑<sub>6</sub>
- [11] Liu X., Deng Y., Sun Y., Zhou Y. *Multi-digit recognition with convolutional neural network and long short-term memory* // 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (Huangshan, China, July 28–30, 2018).– IEEE.– 2018.– Pp. 1187–1192. [doi](#) ↑<sub>6</sub>
- [12] Madakannu A., Selvaraj A. *DIGI-Net: a deep convolutional neural network for multi-format digit recognition* // Neural Computing and Applications.– 2020.– Vol. **32**.– Pp. 11373–11383. [doi](#) ↑<sub>7</sub>
- [13] Zou L., He Z., Wang K., Wu Z., Wang Y., Zhang G., Wang X. *Text recognition model based on multi-scale fusion CRNN* // Sensors.– 2023.– Vol. **32**.– No. 16.– id. 7034.– 18 pp. [doi](#) ↑<sub>7</sub>
- [14] Agrawal V., Jagtap J. *Convolutional vision transformer for handwritten digit recognition.*– Research Square.– 2022.– 11 pp. [doi](#) ↑<sub>7</sub>
- [15] Cheng L., Khalitov R., Yu T., Yang Z. *Classification of long sequential data using circular dilated convolutional neural networks.*– 2022.– 16 pp. [arXiv:2201.02143 \[cs.LG\]](#) ↑<sub>7</sub>
- [16] Bhat R. S. *Text recognition with CRNN-CTC network.*– W&B Fully Connected.– 2022. [URL](#) ↑<sub>7, 8</sub>
- [17] Khamekhem S., Sourour A., Kessentini Y. *Domain and writer adaptation of offline Arabic handwriting recognition using deep neural networks* // Neural Computing and Applications.– 2022.– Vol. **34**.– Pp. 2055–2071. [doi](#) ↑<sub>7, 8</sub>

Поступила в редакцию 29.09.2023;  
 одобрена после рецензирования 27.11.2023;  
 принята к публикации 27.11.2023;  
 опубликована онлайн 11.03.2024.

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.

[ID](#) 0000-0001-8697-1032  
 e-mail: [igvvinokurov@fa.ru](mailto:igvvinokurov@fa.ru)

Декларация об отсутствии личной заинтересованности: *благополучие автора не зависит от результатов исследования.*



## Recognition of cadastral coordinates using convolutional recurrent neural networks

Igor Victorovich **Vinokurov**

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

 [iguvinokurov@fa.ru](mailto:iguvinokurov@fa.ru)

**Abstract.** The article examines the use of convolutional recurrent neural networks (CRNN) for recognizing images of cadastral coordinates of objects on scanned documents of the «Roskadastr» PLC. The combined CRNN architecture, combining convolutional neural networks (CNN) and recurrent neural networks (RNN), allows you to take advantage of each of them for image processing and recognition of continuous digital sequences contained in them. During experimental studies, images consisting of a given number of digits were generated, and a CRNN model was built and studied. The formation of images of digital sequences consisted of preprocessing and concatenation of images of the digits forming them from one's own data set. Analysis of the values of the loss function and Accuracy, Character Error Rate (CER), and Word Error Rate (WER) metrics showed that the use of the proposed CRNN model makes it possible to achieve high accuracy in recognizing cadastral coordinates in their scanned images. (*Linked article texts in Russian and in English*).

**Key words and phrases:** convolutional recurrent neural network, CRNN, image recognition, digital sequences, deep learning, Keras, Python

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

**For citation:** Igor V. Vinokurov. *Recognition of cadastral coordinates using convolutional recurrent neural networks*. Program Systems: Theory and Applications, 2024, **15**:1(60), pp. 3–30. (*In Russian, in English*).

[https://psta.psir.ru/read/psta2024\\_1\\_3-30.pdf](https://psta.psir.ru/read/psta2024_1_3-30.pdf)

## Introduction

Recognition of alphanumeric sequences in images is a significant problem in the field of computer vision and image processing. An effective solution to this problem is of great importance for the automation and optimization of various processes associated with the identification and classification of objects.

In recent years, impressive results in image recognition and classification have been achieved using deep learning neural networks, especially CNNs. However, classical CNNs are focused on identifying features of the input data, which limits their applicability for recognizing sequences of variable length. To overcome this limitation, the CRNN architecture was developed, which combines the advantages of convolutional and recurrent neural networks. The main feature of the CRNN architecture is the combination of CNN convolutional layers to extract local and spatial features from images and RNN recurrent layers to take into account context and sequence information. Convolutional layers allow you to discover important features of images at different levels of abstraction, while recurrent layers model the dependencies and sequence of input data [1].

In the most general case, the CRNN architecture consists of three main components: convolutional, recurrent, and classification components. The convolutional component contains several convolutional layers that perform detection and feature extraction from images. The recurrent component includes recurrent layers of long short-term memory (LSTM) [2] or managed recurrent units (GRU) [3] to take into account the context and sequence of the input data. The classification component performs sequence recognition and classification based on the predictions of the recurrent component.

This article studies the effectiveness of using CRNN for image recognition of digital sequences of variable length. In section 1 the need for research and the formulation of the problem are substantiated. Section 2 is devoted to the review and analysis of works on the use of CRNN networks for text recognition and digital sequences. Creating a dataset for training the model is described in section 3. The formation and study of a CRNN model for recognizing digital sequences in images is given in section 4. In conclusion, conclusions are presented based on the results of the research.



## 1. Setting the purpose and objectives of the study

The purpose of the study is to develop a CRNN model capable of achieving acceptable recognition accuracy of cadastral coordinates on scanned documents of the «Roskadastr» PLC. [4], [5] and [6] describe approaches implemented in the information system (IS) of this organization to solving the problem of converting images of cadastral coordinates into their text counterparts using CNN models. [6] shows the effectiveness of using CNN for sequences consisting of no more than 4 digits. Based on the results of studies carried out in [6], we can conclude that when recognizing a larger number of digits, the structure of the CNN becomes more complex, while the quality of recognition either remains the same or increases slightly. Considering that the maximum number of symbolic-digital elements in cadastral coordinates can exceed 10, one of the expedient and effective means of solving the problem of their recognition may be the use of CRNN models.

The goal set in the work can be achieved by solving the following main tasks:

- (1) Formation of a dataset, which consists in preparing a set of images of elements of cadastral numbers and their annotation (comparison with the cadastral number); generation of images of cadastral coordinates for a given number of their digital elements; dividing the dataset into samples for training and validation.
- (2) Formation of a CRNN model – selection of convolutional and recurrent layers for feature extraction and digital sequence recognition, respectively.
- (3) Model training and analysis of loss function values and accuracy metrics, CER, and WER.

## 2. Analysis of the main works on alphanumeric sequence recognition

For the first time, a CRNN model combining convolutional and recurrent layers for processing images with text sequences is described in [1]. The advantage of the model is the combination of convolutional and recurrent layers, which makes it possible to identify both local and global dependencies in images containing sequences and to implement their

recognition quite effectively. The disadvantages of the proposed model include fairly large amounts of data for training and significant time for recognizing long sequences.

The paper [7] describes a CRNN model that can recognize words unknown to it using meaningful contextual information. The model is resistant to various image distortions, does not depend on a predefined dictionary, and can process arbitrary sentences. The disadvantage of the model is poor text recognition with low contrast, unclear boundaries, and distortions, which requires additional image pre-processing methods.

The model from [8] is designed for text recognition in images with perspective distortion, curved character placement, etc. The proposed CRNN model is able to provide acceptable readability and recognition of distorted text and outperforms similar models in recognizing text with different values of input noise and slope. In addition, the model demonstrates high accuracy and performance when trained on large datasets. A disadvantage of the proposed model may be poor text recognition if noise or deformation greatly changes the shape of the characters, which is especially noticeable for text with a small font or low resolution.

In the paper [9] the authors apply CRNN to recognize text in images, focusing on difficult cases such as scenes with poor lighting or low resolution. They propose a model that uses convolutional layers to extract features from images and recurrent layers to model sequences from these characters. The experiments demonstrate high accuracy of text recognition on various images of text sequences.

A model for multi-level recognition of handwritten text in images is given in [10]. At the first level, CNN is used to recognize words that appear frequently in text. If a word is not recognized by this model, it moves to the second layer, which uses a fully convolutional network (FCN). An experimental study of the model was conducted using NIST19 as the training dataset and handwriting as the test dataset and showed quite acceptable recognition result.

In [11] presented two models for recognizing sequences of digits. In the first, the encoder and decoder of the sequences are CNN and LSTM. In the second – histogram of oriented gradient (HOG) and parallel fully connected (Dense) layers, respectively. Training and testing were carried

out on the Street View Number House dataset (SVHN). As a result of the conducted research, the advantage of CNN in terms of image encoding and the advantage of LSTM in sequence prediction were shown.

The DIGI-Net deep learning network, which is capable of learning the common characteristics of three different digit formats (handwritten, natural images, printed font) and recognizing them, is described in [12]. Experiments conducted on the MNIST, CVL and Chars74K, demonstrated high accuracy of recognition of continuous digital sequences.

The CRNN model proposed in [13] has a convolutional layer, a feature fusion layer, a recurrent layer, and a transcription layer. The convolutional layer used for feature extraction produces two outputs for the input text image. The feature pooling layer combines the results of the convolutional layer into one, from which the recurrent layer extracts sequences. The final result is output by the transcription layer. The proposed model, due to the fusion of features, realizes better text recognition accuracy on text datasets Street View Text (SVT), IIT-5K, ICDAR2003 and ICDAR2013.

In [14] it is proposed to use a hybrid architecture of visual transformers (ViTs) and multilayer perceptrons (MLP) to recognize handwritten digits. Conducted research on the EMNIST and DIDA showed good accuracy in recognizing typewritten digital data, including on noisy images.

Paper [15] proposes a symmetrical multi-scale architecture called Circular Dilated Convolutional Neural Network (CDIL-CNN), where each element in the current layer has an equal chance of receiving information from other elements in previous layers. The proposed CRNN model allows the generation of classification logits for all elements, as a result of which it becomes possible to use simple ensemble learning to make the best decision. Based on the results of testing CDIL-CNN on long sequential datasets, it is shown that CDIL-CNN allows one to obtain recognition results that are acceptable in terms of accuracy.

The best approach to forming models for sequence recognition, from the point of view of the author of this work, is given in [16] and [17]. To recognize text sequences, it is proposed to use an encoder in the form of a CNN and a decoder in the form of a bidirectional long-term short-term memory (BSTM) using connectionist temporal classification (CTC). CTC is an algorithm used to train an RNN on variable length sequences and

match them with corresponding labels. In text and digit recognition tasks, CTC can cope with the problem of their variable length. It allows the model to predict a variable number of letters or numbers in a sequence without first separating or aligning them. The CTC algorithm calculates the probability of the output sequence and recursively updates the model weights based on the difference between the predictions and the sequence labels. The effectiveness of the proposed approach is shown on its own dataset in [16] and on the Arabic letters datasets MADCAT, AHTID/MW and IFN/ENIT in [17].

### 3. Creating the dataset

To train the CRNN model and study its operation, our own dataset was generated.

At the first stage, by analogy with [4], black and white images of elements of digital sequences were formed using the main fonts of «Roskadastr» PLC documents. The number of image classes is chosen to be 12 – 10 classes for numbers from 0 to 9, 1 class for delimiter characters «.» and «,», and one more for the absence of a character in the sequence. The values of the last 2 classes are chosen to be 10 and 11, respectively. For each class of images, 10 and 5 images of  $20 \times 25$  pixels were generated for training and validating the model, respectively.

At the second stage, images of cadastral coordinates were formed, consisting of 2 digits in the fractional part and from 4 to 7 in the whole. Simultaneously with the formation of images, their CTC labels were also formed. All images in the dataset were reduced to the same size of  $200 \times 32$  pixels. An example of the generated images of cadastral coordinates and the corresponding CTC labels is shown in Figure 1.

The dataset thus generated is shown in Table 1 and consists of 24240 and 12240 images of cadastral coordinates and their labels for training and validation, respectively.

### 4. Formation and research of a CRNN model

The model was formed using the `Keras` library. All layers of this model, in addition to the main CRNN layers – convolutional and recurrent

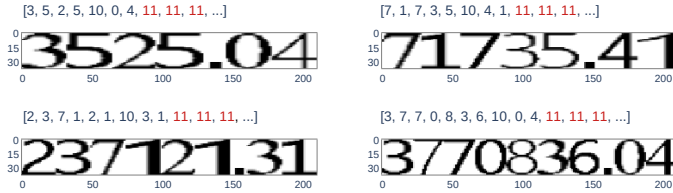


FIGURE 1. Examples of images of cadastral coordinates. Their CTC labels are indicated at the top

TABLE 1. Number of digital sequences for train and validation

Digits in sequence	Train sequences	Validation sequences
6	4848	2448
7	5656	2856
8	6464	3264
9	7272	3672

(Conv2D and Bidirectional respectively), are given in Table 2.

The downsampling layer MaxPooling2D reduces the dimension of the feature space, highlighting the most significant ones.

Layer BatchNormalization normalizes data in mini-batches, which speeds up training convergence and reduces the likelihood of overfitting. It also helps stabilize the distribution of activations between layers.

The Dropout layer is a regularizer, its goal is to reduce overfitting by preventing the activation of randomly selected neurons. This forces the model to learn more stable features and reduces the contribution of each individual neuron.

Layer Activation (activation function) applies an activation function to the output of the previous layer. In this model, this is ReLU, which activates neurons only for positive values, Tahn – for positive and negative values and Softmax, generating probabilities for different classes.

The fully connected layer Dense combines all outputs from the previous layer and applies linear transformations to produce the final model output. It connects the outputs of all neurons in the previous layer to each neuron in the current layer and is the main classification layer in neural networks.

TABLE 2. CRNN model layers

Layer	Activation	Filters	Input
InputLayer	–	–	[(None, 200, 32, 1)]
Conv2D	–	32	(None, 200, 32, 32)
BatchNormalization	–	–	(None, 200, 32, 32)
Activation	ReLu	–	(None, 200, 32, 32)
MaxPooling2D	–	–	(None, 100, 16, 32)
Conv2D	–	64	(None, 100, 16, 64)
BatchNormalization	–	–	(None, 100, 16, 64)
Activation	ReLu	–	(None, 100, 16, 64)
MaxPooling2D	–	–	(None, 50, 8, 64)
Dropout	–	–	(None, 50, 8, 64)
Conv2D	–	128	(None, 50, 8, 128)
BatchNormalization	–	–	(None, 50, 8, 128)
Activation	ReLu	–	(None, 50, 8, 128)
MaxPooling2D	–	–	(None, 50, 4, 128)
Dropout	–	–	(None, 50, 4, 128)
Conv2D	–	256	(None, 50, 4, 256)
BatchNormalization	–	–	(None, 50, 4, 256)
Activation	ReLu	–	(None, 50, 4, 256)
MaxPooling2D	–	–	(None, 50, 2, 256)
Dropout	–	–	(None, 50, 2, 256)
Reshape	–	–	(None, 32, 800)
Dense	–	–	(None, 32, 25)
Bidirectional	Tanh	–	(None, 32, 320)
Bidirectional	Tanh	–	(None, 32, 320)
Dense	–	–	(None, 32, 12)
Activation	Softmax	–	(None, 32, 12)

The Lambda layer allows you to define your own lambda function for non-standard data transformation, for example, changing its dimension. In the generated model, it is used in conjunction with ground-truth labels used for recognition tasks of continuous sequences. During the training phase, these labels are matched and compared with the ground truth labels, resulting in an error that is used to optimize the model.

The structure of the CRNN model is shown in Figure 2. To train it, the backpropagation method with the optimizer Adam was used.

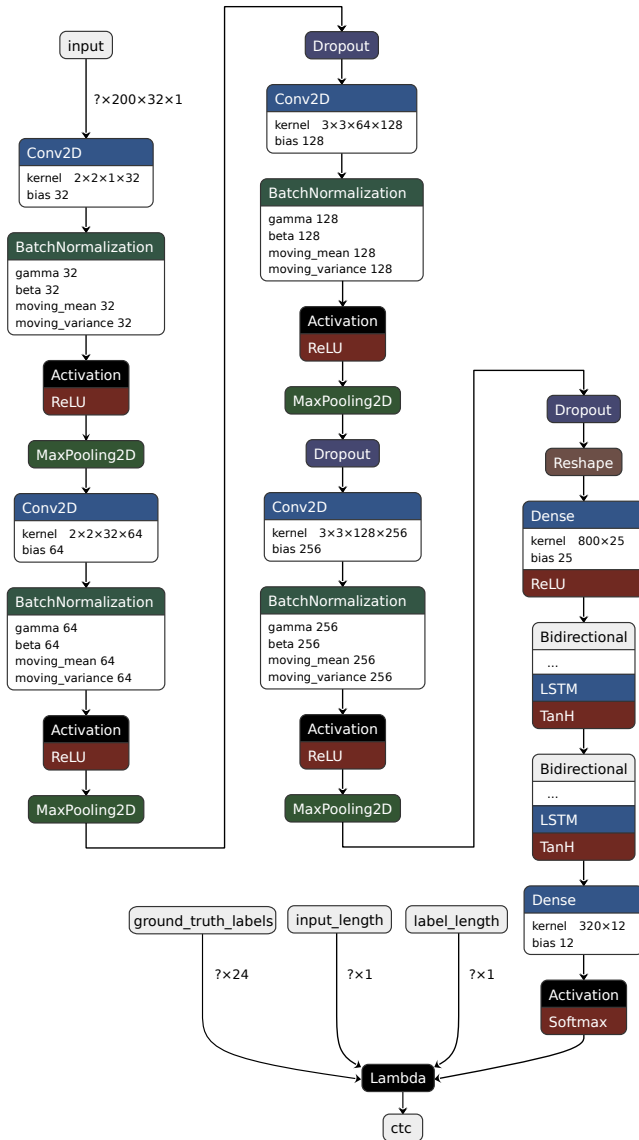


FIGURE 2. CRNN model

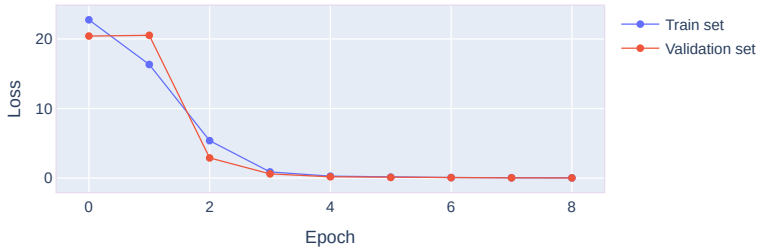


FIGURE 3. Model loss

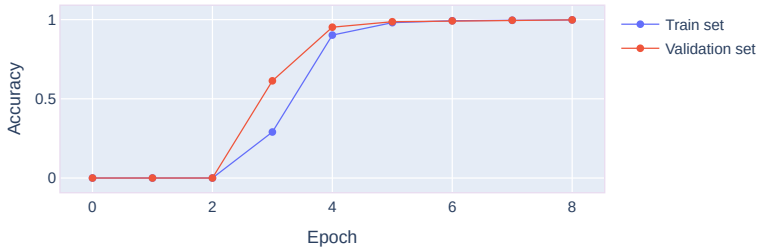


FIGURE 4. Model accuracy

To assess the quality of the model and its ability to solve the problem, the values of the loss function and accuracy metrics, CER, and WER were calculated. The loss function (**Loss**) measures the difference between the actual (true) label values and the values predicted by the model. In Figure 3 shows the model losses for 9 training epochs. This number of training epochs was found experimentally and is optimal.

The accuracy metric (**Accuracy**) measures the proportion of correct predictions made by the model relative to the total number of examples. It evaluates how well a model can classify or predict the correct class or value for a given set of data, Figure 4.

The numerical values of the loss functions and accuracy metrics for the train and validation sets at each of the training epochs are given in Table 3.

Calculating loss function values and accuracy metrics is a common approach to assessing the quality of neural network models. For models focused on sequence recognition, two more are calculated – Character Error Rate (**CER**) and Word Error Rate (**WER**). Both metrics are used to comparatively evaluate different recognition systems and analyze their accuracy.



TABLE 3. Numerical values of Loss and Accuracy

Epoch	Loss		Accuracy	
	Train set	Validation set	Train set	Validation set
0	22.7430438	20.4115753	0.0	0.0
1	16.3242683	20.5168876	0.0	0.0
2	5.3856644	2.9053363	0.0	0.0000817
3	0.8966413	0.6119849	0.2904703	0.6136437
4	0.2865071	0.1913659	0.9022276	0.9521241
5	0.1421806	0.1023578	0.9807755	0.9866012
6	0.0855829	0.0680020	0.9928630	0.9919934
7	0.0573494	0.0432046	0.9958745	0.9959967
8	0.0411685	0.0314824	0.9980198	0.9982843

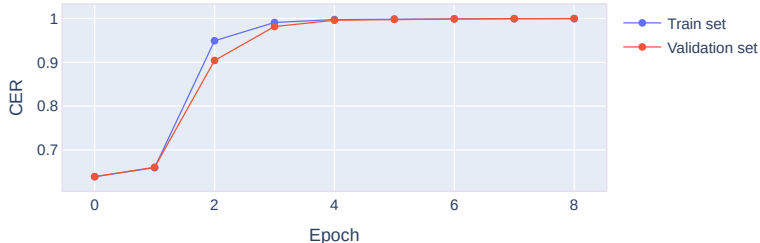


FIGURE 5. Character recognition accuracy

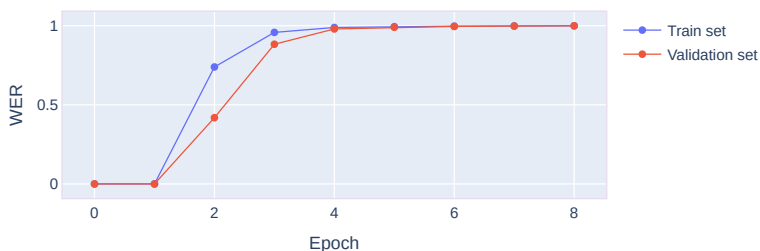


FIGURE 6. Word recognition accuracy

CER allows you to evaluate the accuracy of a model's recognition of individual characters. The dependences of the values of this metric for the training dataset and the validation dataset on the training epoch number are shown in Figure 5.

The accuracy of recognition of whole words by the model can be assessed by the WER metric. The dependences of the values of this metric for the training dataset and the validation dataset on the training epoch number are shown in Figure 6.

The numerical values of the CER and WER metrics for the train and validation sets at each training epoch are given in Table 4.

The results of recognition by the model of several cadastral coordinates

TABLE 4. Values of the CER and WER metrics

Epoch	CER		WER	
	Train set	Validation set	Train set	Validation set
0	0.6388888	0.6388888	0.0	0.0
1	0.6594093	0.6603758	0.0	0.0
2	0.9491560	20.9041156	0.7390676	0.4186274
3	0.9908536	0.9816210	0.9576320	0.8828431
4	0.9976399	0.9960001	0.9887788	0.9794117
5	0.9987451	0.9981072	0.9940182	0.9900327
6	0.9993949	0.9989787	0.9970297	0.9953431
7	0.9996046	0.9995132	0.9981435	0.9977942
8	0.999804	0.9998400	0.9992161	0.9994281



FIGURE 7. Results of recognition of cadastral coordinates. One of the coordinates was recognized with an error

of 6, 7, 8, and 9 digits, respectively, from the test dataset are shown in Figure 7.

The model proposed by CRNN was implemented in the IS «Roskadastr» PLC and, as the results of its experimental study showed, the accuracy of recognition of individual symbols and cadastral coordinates as a whole was 99.98% and 99.94%, respectively. Images of cadastral coordinates were extracted from the scanned document according to the coordinates generated by the contourization subsystem of this IS. The general principles of operation of this subsystem are described in [4].

## Conclusion

The article studies the use of CRNN architecture for the task of recognizing images of cadastral coordinates. Based on the results of training the model, graphs of the loss function (Loss) and accuracy (Accuracy) were constructed. The graphs showed that the model converges successfully and

is capable of achieving high accuracy in recognizing cadastral coordinates. To assess the quality of the model, the CER (Character Error Rate) and WER (Word Error Rate) metrics were used, which made it possible to measure the percentage of errors at the character and word levels, respectively. Based on the results of experimental studies, we can conclude that the model is capable of recognizing cadastral coordinates with a high level of accuracy and a minimum number of errors. As a result, the use of the CRNN model will significantly improve the efficiency and reliability of geospatial analyzes and decision making. Future research could focus on expanding the dataset, including different font types and styles, training the model on more diverse data, and investigating the effectiveness of CRNN for other recognition and classification tasks. It is also possible to use additional data preprocessing or augmentation methods to improve the accuracy of the model.

## References

- [1] B. Shi, X. Bai, C. Yao. “An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**:11 (2017), pp. 2298–2304. [arXiv:1507.05717 \[cs.CV\]](#) [doi](#) ↑18, 19
- [2] S. Hochreiter, J. Schmidhuber. “Long short-term memory”, *Neural Computation*, **9**:8 (1997), pp. 1735–1780. [doi](#) [URL](#) ↑18
- [3] J. Chung, C. Gulcehre, K. Cho, Y. Bengio. “Gated feedback recurrent neural networks”, *Proceedings of Machine Learning Research*, **37** (2015), pp. 2067–2075. [arXiv:1502.02367 \[cs.NE\]](#) [URL](#) ↑18
- [4] 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. [doi](#) [URL](#) ↑19, 22, 28
- [5] I. V. Vinokurov. “Tabular information recognition using convolutional neural networks”, *Program Systems: Theory and Applications*, **14**:1(56) (2023), pp. 3–30. [doi](#) [URL](#) ↑19
- [6] I. V. Vinokurov. “Recognition of digital sequences using convolutional neural networks”, *Program Systems: Theory and Applications*, **14**:3(58) (2023), pp. 3–36. [doi](#) [URL](#) ↑19
- [7] P. He, W. Huang, Y. Qiao, Change Loy C., X. Tang. “Reading scene text in deep convolutional sequences”, *AAAI’16: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (Phoenix, Arizona, USA, February 12–17, 2016), *Proceedings of the AAAI Conference on Artificial Intelligence*, **30**:1 (2016), pp. 3501–3508. [doi](#) ↑20
- [8] B. Shi, X. Wang, P. Lv, C. Yao, X. Bai. “Robust scene text recognition with automatic rectification”, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV, USA, June 27–30, 2016), 2016, pp. 4168–4176. [doi](#) [arXiv:1603.03915 \[cs.CV\]](#) ↑20

- [9] F. Yin, Y. -C. Wu, X. -Y. Zhang, C. -L. Liu. *Scene text recognition with sliding convolutional character models*, 2017, 10 pp. arXiv:1709.01727 [cs.CV] [↑20](#)
- [10] D. A. Nirmalasari, N. Suciati, D. A. Navastara. “Handwritten text recognition using fully convolutional network”, *IOP Conference Series: Materials Science and Engineering*, **1077**:1 (2021), id. 012030, 9 pp. [doi](#) [↑20](#)
- [11] X. Liu, Y. Deng, Y. Sun, Y. Zhou. “Multi-digit recognition with convolutional neural network and long short-term memory”, *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (Huangshan, China, July 28–30, 2018), IEEE, 2018, pp. 1187–1192. [doi](#) [↑20](#)
- [12] A. Madakannu, A. Selvaraj. “DIGI-Net: a deep convolutional neural network for multi-format digit recognition”, *Neural Computing and Applications*, **32** (2020), pp. 11373–11383. [doi](#) [↑21](#)
- [13] L. Zou, Z. He, K. Wang, Z. Wu, Y. Wang, G. Zhang, X. Wang. “Text recognition model based on multi-scale fusion CRNN”, *Sensors*, **32**:16 (2023), id. 7034, 18 pp. [doi](#) [↑21](#)
- [14] V. Agrawal, J. Jagtap. *Convolutional vision transformer for handwritten digit recognition*, Research Square, 2022, 11 pp. [doi](#) [↑21](#)
- [15] L. Cheng, R. Khalitov, T. Yu, Z. Yang. *Classification of long sequential data using circular dilated convolutional neural networks*, 2022, 16 pp. arXiv:2201.02143 [cs.LG] [↑21](#)
- [16] R. S. Bhat. *Text recognition with CRNN-CTC network*, W&B Fully Connected, 2022. [URL](#) [↑21](#), 22
- [17] S. Khamekhem, A. Sourour, Y. Kessentini. “Domain and writer adaptation of offline Arabic handwriting recognition using deep neural networks”, *Neural Computing and Applications*, **34** (2022), pp. 2055–2071. [doi](#) [↑21](#), 22

Received 29.09.2023;  
 approved after reviewing 27.11.2023;  
 accepted for publication 27.11.2023;  
 published online 11.03.2024.

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

[ID](#) 0000-0001-8697-1032  
 e-mail: [igvvinokurov@fa.ru](mailto:igvvinokurov@fa.ru)

*The author declare no conflicts of interests.*