

# Оптическое распознавание шрифтов на изображениях, полученных со смартфонов, и его использование для определения подлинности документов, удостоверяющих личность\*

Ю.С. ЧЕРНЫШОВА<sup>1,II</sup>, М.А. АЛИЕВ<sup>1,II</sup>, А.В. ШЕШКУС<sup>I</sup>

<sup>I</sup> ООО «Смарт Энджинс Сервис», Москва, Россия

<sup>II</sup> Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, Москва, Россия

**Аннотация.** В работе рассматривается способ оптического распознавания шрифтов при помощи сверточной нейронной сети на изображениях, полученных с малоразмерных цифровых камер. Исследуется возможность применения подобных сетей для определения подлинности документов, удостоверяющих личность, со шрифтами, заданными стандартами. Приведены результаты экспериментов по обучению сетей для определения верности шрифта, использованного для печати машиночитаемых зон и номера паспорта гражданина РФ. Результаты показывают, что предложенный метод может быть применен для определения стандартных шрифтов и, при дальнейшем совмещении с другими подходами, для определения подлинности документов.

**Ключевые слова:** OCR, OFR, сверточные нейронные сети, синтез обучающей выборки.

**DOI:** 10.14357/20790279180521

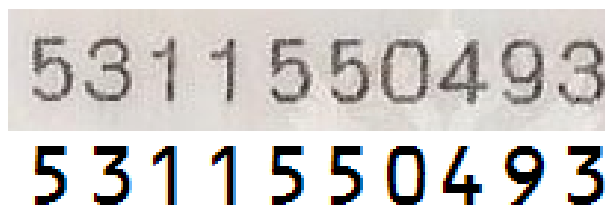
## Введение

Документы, удостоверяющие личность и данные из них, требуются для получения множества услуг. Самой очевидной и опасной в случае пропуска подделки является пограничный контроль, где в условиях постоянно растущего пассажиропотока ручная проверка документов становится практически невозможной [1]. При предоставлении финансовых услуг [2] от поддельных документов страдают банки, страховые и телекоммуникационные компании. По оценке Бюро кредитных историй ежегодные потери только в кредитном сегменте превышают 150 млрд рублей [3]. Несмотря на все защитные меры, количество поддельных документов растет год от года [4].

При этом, все больше компаний позволяют оформлять услуги, требующие идентификационных документов, с помощью смартфонов и мобильных устройств [5]. В таких условиях возникает задача автоматической проверки подлинности документов по фотографиям или видеопотоку видимого спектра.

Целью нашей работы является разработка метода определения подлинности документов

по изображениям, полученным со смартфонов, с помощью искусственных нейронных сетей, определяющих верность использованного шрифта. К примеру, на рис. 1 представлена строка из напечатанной неправильным шрифтом машиночитаемой зоны и символы, соответствующие стандарту [6].



**Рис. 1.** Примеры машиночитаемой строки, набранной неверным шрифтом (вверху), и ее верного аналога (внизу)

В данной статье в разделе 1 рассматриваются известные методы определения фальшивых документов по изображениям и оптического распознавания шрифтов. В разделе 2 ставится задача и изложена ее особенности. В разделе 3 описаны обучающая и тестовая выборки, а также приведены результаты экспериментов.

\* Работа выполнена при частичной поддержке РФФИ, гранты 17-29-03236 и 17-29-07093.

## 1. Обзор существующих методов

### 1.1 Методы определения подлинности документов по изображениям

Одним из основных и наиболее частых способов подделки документов (как идентификационных, так и прочих), особенно при использовании цифровых способов передачи данных, является подделка текстовых данных документа [7].

В [1] авторы анализируют соответствие данных в основной части документа и в машиночитаемой зоне, а также проводят кросс-проверку изображений одних и тех же символов между собой. В [8 и 9] предлагаются методы обнаружения подделок вида “copy-paste” – авторы анализируют сходство символов, которые были скопированы и вставлены в пределах одного документа исходя из их пиксельного анализа. В [10] авторы предлагают использовать индивидуальные признаки шрифтов, а в [7 и 11] в продолжение [10] используют случайные марковские поля для получения оценки шрифта по индивидуальным типографическим признакам и соответствия его остальному окружению в документе. Однако представленные до настоящего момента решения не могут быть полноценно применены для анализа и проверки изображений документов, полученных с малоразмерных цифровых камер в неконтролируемых условиях съемки [12], так как в них рассматриваются изображения с простым фоном, полученные при хорошем равномерном освещении [1], и, как правило, разрешением не ниже 300 dpi [5-9]. В работе [10] рассматривается возможность сжатия изображения (с использованием JPEG с коэффициентом сжатия от 1 до 1.25), что ведет к снижению качества входных данных, однако его недостаточно для имитации условий съемки смартфоном [12].

### 1.2 Методы оптического распознавания шрифтов

Шрифты обычно распознают по последовательностям символов, однако без распознавания символов как таковых. В [13] предлагается метод распознавания шрифта из более чем 2000 классов по изображению слова с использованием классификатора по ближайшему среднему на основе дескрипторов особых точек. В работе [14] предлагается использование декомпозиции многослойной сверточной нейронной сети с использованием глубокого сверточного автоэнкодера. В работах [13 и 14] обучающие данные для классификатора генерируются по известным общедоступным шрифтам. Во многих работах по распознаванию шрифтов авторы рассматривают только изображение высокого разрешения, например, в [15] минимальное разрешение – 400 dpi.

## 2. Постановка задачи и ее особенности

Задано  $F$  – множество всех существующих шрифтов. Также заданы шрифты  $F'$  – истинный шрифт, который следует использовать по стандарту, и  $F''$  – множество шрифтов, доступных для создания отрицательной выборки.

Дано множество изображений символов  $X = \{I\}$ , для которого определены функция шрифта  $f: X \rightarrow F$  и функция соответствия  $v: X \rightarrow \{0, 1\}$ , задающая множество  $G$ , такая что:

$$v(I) = \begin{cases} 0, & f(I) \in F' \\ 1, & f(I) \in F \setminus F'. \end{cases}$$

Задан алфавит распознавания  $A = \{T_i\}_{i=1}^M$ ,  $M$  – размер алфавита.

Требуется обучить классификаторы, реализующие функции  $C': X \rightarrow A \times G$ , способные разделять как по алфавитным, так и по шрифтовым признакам, и  $C'': X \rightarrow G$ , разделяющую только по шрифтовым признакам. При этом, в случае выявления подделки нам не требуется точно знать, какой шрифт использовался.

В качестве объектов распознавания использовались изображения паспорта гражданина РФ [16]. В этом документе встречается подделка либо номера, либо машиночитаемых строк. При этом заменено может быть как все поле, так и отдельные символы. Бланк паспорта относится к защищенной полиграфической продукции уровня «А» [17], поэтому все необходимое для его производства, в том числе и шрифт, используемый для печати номера, является конфиденциальной информацией [17,18].

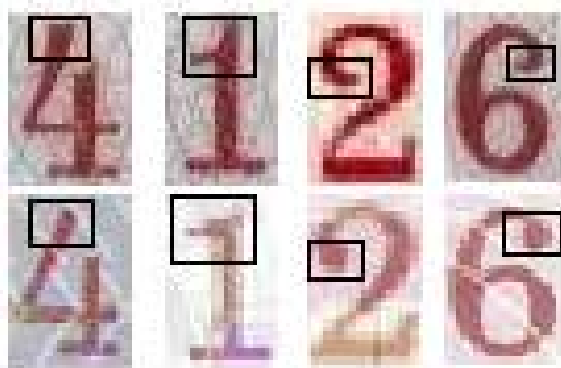
Особенностями рассматриваемых изображений можно считать:

- Низкое качество, так как они получены с камер смартфонов в неконтролируемых условиях съемки, наличие артефактов сжатия, а также низкое разрешение – 96 dpi.
- Наличие сложного фона, а также различное качество бумаги и печати.

На рис. 2 в первой строке представлены примеры цифр, напечатанных верным шрифтом для номера паспорта, а во второй – их аналогов из поддельных документов.

## 3. Эксперименты

Мы обучили две сети, реализующие функцию  $C'$ : для распознавания символов и шрифта машиночитаемой зоны (МЧЗ) и номера паспорта гражданина РФ. В качестве алфавита для классификаторов было решено использовать только цифры, то есть  $M = 10$ , а общее количество классов – 20.

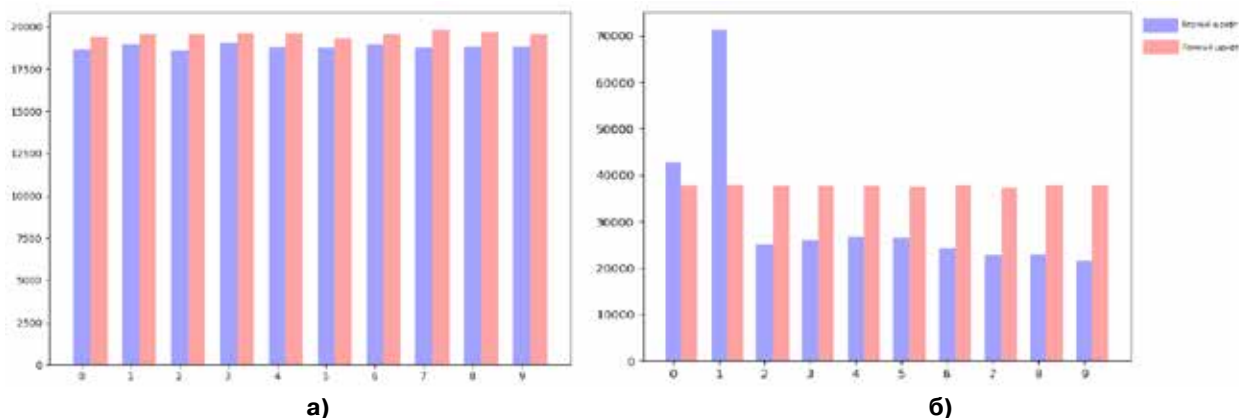


**Рис. 2.** Примеры символов из настоящих (вверху) и поддельных (внизу) номеров паспортов РФ

Также мы обучили сеть, реализующую функцию  $C''$ , для номера паспорта РФ.

### 3.1. Архитектура сети

Для оптического распознавания шрифтов предлагается использовать сверточную нейронную сеть с архитектурой, сверточная часть которой представлена в табл. 1. Данная архитектура обладает достаточно малым количеством весов, а именно примерно 7500 весов для сети вида  $C'$  и  $M=10$  и порядка 3500 для сети вида  $C''$ , что делает ее пригодной для дальнейшего использования в приложениях для смартфонов.



**Рис. 3.** Распределение символов в обучающих выборках для машиночитаемой зоны (а); б – номера паспорта РФ

### 3.2. Обучающие выборки

В рамках данной работы проведены эксперименты для цифр из двух разных зон паспорта. Способы создания выборок для них различаются.

Для обучения классификатора цифр машиночитаемой зоны все данные были синтезированы при помощи подхода, описанного в [19]. Для создания выборок использовались шрифты, доступные на Google Fonts [20]. В качестве истинного шрифта использован указанный в постановлениях о машиночитаемых зонах OCR-B [21, 22]. Для создания отрицательной выборки использовалось 17 шрифтов, похожих на OCR-B по стилю.

Однако шрифт, используемый в номере паспорта, общедоступным не является, и сгенерировать искусственные данные с ним невозможно. Соответственно, при обучении сети в качестве выборки изображений истинного шрифта использовались аугментированные натуральные данные, а для создания отрицательной выборки – 46 общедоступных шрифтов.

Распределение символов представлено на рис. 3.

### 3.3. Тестовая база изображений

Для проведения экспериментов используется стенд изображений паспортов граждан РФ, состоящий из 3238 различных изображений па-

**Табл. 1**

Архитектура сверточной части используемых искусственных нейронных сетей

Слой	Кол-во фильтров	Размер фильтра	Шаг	Отступ
Сверточный	4	3 × 3	1	0 × 0
Сверточный	8	5 × 5	2	2 × 2
Сверточный	8	3 × 3	1	1 × 1
Сверточный	8	5 × 5	2	2 × 2
Сверточный	8	3 × 3	1	1 × 1
Сверточный	8	3 × 3	1	1 × 1

спортов граждан РФ, полученных с помощью малоформатных цифровых камер. Также было собрано 173 изображения поддельных паспортов, на 45 из которых фальсифицирован номер. Для всех изображений была вручную сделана посимвольная разметка.

Для проверки работоспособности предложенного подхода были созданы три тестовые выборки, состоящие из изображений отдельных символов следующих видов:

- Цифры из номера паспорта;
- Цифры из машиночитаемой зоны;
- Цифры из кода подразделений и дат.

Распределение символов в собранных выборках представлено на рис. 4.

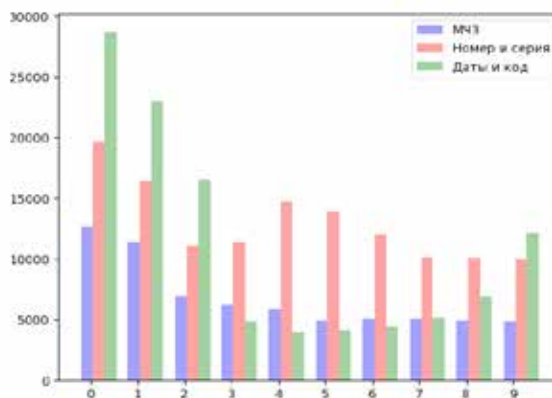


Рис. 4. Распределение символов тестовых выборок по классам

### 3.4. Результаты экспериментов

Работоспособность итоговых сетей была проверена на вышеописанных тестовых выборках. В рамках проведенных экспериментов мы считаем ответ верным, если были правильно определены и символ из алфавита, и класс шрифта. В табл. 2 представлено качество распознавания, демонстрируемое сетью вида  $C'$  для определения шрифта OCR-B, обученной только на синтезированных данных. Среднее качество классификации, продемонстрированное сетью, на выборке верного шрифта 78.32%, а на выборках шрифтов номера паспорта и дат – 79.29 и 88.41% соответственно. В табл. 3 представлены результаты работы сети вида  $C'$  для классификации цифр номера паспорта гражданина РФ. Среднее качество классификации изображений, выполненных правильным шрифтом, 87.25%, а выполненных шрифтами из машиночитаемой зоны и из дат и кодов подразделений – 84.55 и 94.44%.

По полученным таблицам видно, что качество распознавания некоторых символов шрифтов, например, '0' и '8' для OCR-B, сильно хуже среднего. У сети, обученной для определения подлинности шрифта номера паспорта, сильное падение относительно среднего значения наблюдается только у символа '0' верного шрифта. Из этого можно сделать вывод, что данные цифры в этих шрифтах не имеют достаточно специфических признаков и похожи на символы из отрицательной выборки. При дальнейшем использовании предложенных сетей для определения подделок документов можно учитывать данные символы с меньшими весовыми коэффициентами или же всегда считать требующими проверки. Такое поведение считаем оправданным, так как при проверке подлинности требуется выделять максимально возможное количество подделок.

В табл. 4 представлено качество классификации, показанное сетью вида  $C''$  для номера паспорта.

Табл. 2

Качество нейронной сети вида  $C'$  для машиночитаемой зоны

Символ	Качество (%) на тестовой выборке		
	Номер	МЧЗ	Даты и код
0	32.35	68.49	80.92
1	98.54	79.36	92.14
2	87.83	90.29	86.36
3	57.08	86.49	93.20
4	66.25	83.17	98.08
5	98.01	77.86	93.43
6	90.41	86.31	91.09
7	96.96	88.24	80.65
8	98.64	39.73	95.31
9	97.65	88.65	93.36

Табл. 3

Качество нейронной сети вида  $C'$  для номера паспорта РФ

Символ	Качество (%) на тестовой выборке		
	Номер	МЧЗ	Даты и код
0	70.64	58.30	94.07
1	85.86	91.94	88.87
2	91.94	93.94	95.79
3	86.39	94.80	98.15
4	94.53	80.98	99.13
5	91.57	92.13	98.01
6	94.48	90.88	96.32
7	94.35	92.63	89.56
8	80.26	81.60	91.68
9	92.35	93.36	93.64

Табл. 4

Качество нейронной сети вида  $C''$  для номера паспорта РФ

Качество (%) на тестовой выборке		
Номер	МЧЗ	Даты и код
85.83	83.45	90.31

Табл. 5

Применение сети типа  $C'$  для определения нестандартного шрифта номера паспорта способом 1

Порог	Качество (%)		Чувствительность (%)		Специфичность (%)	
	Полный алфавит	Урезанный алфавит	Полный алфавит	Урезанный алфавит	Полный алфавит	Урезанный алфавит
0.50	95.38	96.53	86.67	88.89	98.44	99.22
0.60	95.38	96.53	88.89	91.11	97.66	98.44
0.70	93.06	95.95	88.89	91.11	94.53	97.66
0.80	90.75	93.64	91.11	91.11	90.63	94.53
0.90	82.08	83.24	93.33	91.11	78.13	80.47
1.00	67.05	83.24	95.56	91.11	57.03	80.47

Табл. 6

Применение сети типа  $C'$  для определения нестандартного шрифта номера паспорта способом 2

Порог	Качество (%)		Чувствительность (%)		Специфичность (%)	
	Полный алфавит	Урезанный алфавит	Полный алфавит	Урезанный алфавит	Полный алфавит	Урезанный алфавит
0.50	91.12	92.90	69.39	75.51	100.00	100.00
0.60	93.49	94.08	77.55	79.59	100.00	100.00
0.70	92.31	94.67	77.55	81.63	98.33	100.00
0.80	89.35	92.90	79.59	81.63	93.33	97.50
0.90	82.84	86.39	83.67	81.63	82.50	88.33
1.00	68.05	86.39	85.71	81.63	60.83	88.33

Табл. 7

Применение сети типа  $C''$  для определения нестандартного шрифта номера паспорта способом 3

Порог	Качество (%)		Чувствительность (%)		Специфичность (%)	
	Полный алфавит	Урезанный алфавит	Полный алфавит	Урезанный алфавит	Полный алфавит	Урезанный алфавит
0.50	95.98	97.13	84.44	88.89	100.00	100.00
0.60	96.55	97.13	88.89	91.11	99.22	99.22
0.70	94.83	97.13	88.89	91.11	96.70	99.22
0.80	92.53	93.68	91.11	91.11	93.02	94.57
0.90	82.76	82.76	93.33	91.11	79.07	79.84
1.00	64.94	82.76	100.00	91.11	52.71	79.84

Сети для номера паспорта были встроены в систему распознавания паспорта РФ с камеры смартфона тремя разными способами:

1. Сначала применялась распознающая сеть, а потом сеть вида  $C'$  вызывалась только на цифрах, распознанных с уверенностью больше 0.5;
2. Сеть вида  $C'$  применялась и как распознающая и для классификации шрифта.
3. Сначала применялась распознающая сеть, а потом сеть вида  $C''$  вызывалась только на цифрах, распознанных с уверенностью больше 0.5;

Отбраковка паспортов производилась в зависимости от количества символов, признанных поддельными.

В табл. 5-7 представлены результаты поиска паспортов с фальшивым номером при разных пороговых значениях доли символов, напечатанных неверным шрифтом. Классификатор применялся с полным и сокращенным, то есть без учета символов, распознанных как '0' и '8', алфавитами.

### Заключение

В работе предложен метод детекции поддельных документов, удостоверяющих личность, на основе оптического распознавания шрифтов. В качестве классификаторов использованы сверточные нейронные сети, определяющие либо только верность использованного шрифта, либо и символ, изображенный на картинке, и верным ли шрифтом он напечатан. Предложенный подход проверен на примере паспорта гражданина РФ. Рассмотрены случаи, когда все шрифты, необходимые для создания обучающей выборки, находятся в общем доступе, и когда выборку верного шрифта приходится получать из изображений настоящих документов.

Предложенные классификаторы были встроены в систему распознавания паспорта РФ. Решение о подлинности паспорта производилось в зависи-

мости от количества символов, распознанных как напечатанные фальшивым шрифтом. При использовании полного алфавита (10 цифр) при пороге 0.8 была достигнута чувствительность 91.11% и специфичность 93.02%, при пороге 0.95 чувствительность возросла до 100.00%, но специфичность снизилась до 52.71%. Если же учитывать не весь алфавит, исключив '0' и '8', при пороге 0.8 достигается чувствительность 91.11% и специфичность 94.57%.

В рамках дальнейшего исследования планируется провести замеры времени работы предложенных классификаторов на смартфоне. Также планируется провести эксперименты с полями, где возможно использование нескольких стандартных шрифтов и протестировать классификаторы с разделением шрифтов на 3 и более класса, например создать единый классификатор шрифтов для паспорта гражданина РФ.

### Литература

1. *Young-bin Kwon, Jeong-hoon Kim.* Recognition based Verification for the Machine Readable Travel Documents // International Workshop on Graphics Recognition (GREC 2007), Curitiba, Brazil, 2007.
2. *De Koker L.* Money laundering compliance – the challenges of technology // Financial Crimes: Psychological, Technological, and Ethical Issues, Springer, 2016, pp. 329-347.
3. *Герасимов А.* Цифровое мошенничество: риски и ущерб. – URL: <https://bosfera.ru/bo/cifrovomoshennichestvo-riski-i-ushcherb>
4. *Equifax and EFX.* The New Reality of Synthetic ID Fraud. // Atlanta, Georgia, 2015. – URL: [https://www.equifax.com/assets/IFS/syntheticID-fraud\\_wp.pdf](https://www.equifax.com/assets/IFS/syntheticID-fraud_wp.pdf)
5. *Gai K., Qiu M. and Sun X.* “A survey on fintech,” Journal of Network and Computer Applications, 2017.

6. *ICAO Doc 9303*. 2015. Machine Readable Travel Documents. Seventh Edition. In twelve volumes. International Civil Aviation Organization.
7. *Bertrand R., Terrades O.R., Gomez-Krämer P., Franco P., Ogier J.-M.* A Conditional Random Field model for font forgery detection // 13<sup>th</sup> International Conference on Document Analysis and Recognition (ICDAR), pp. 576-580, 2015.
8. *Bertrand R., Terrades O.R., Gomez-Krämer P., Franco P., Ogier J.-M.* A system based on intrinsic features for fraudulent document detection // 12<sup>th</sup> International Conference on Document Analysis and Recognition, pp. 106-110, 2013.
9. *Abramova S., Böhme R.* Detecting copy-move forgeries in scanned text documents // *Electronic Imaging 2016.8*, 2016.
10. *Zramdini A., Ingold R.* Optical Font Recognition Using Typographical Features // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Volume 20. pp. 877-882, 1998.
11. *Chen G., Yang J., Jin H., Brandt J., Shechtman, Agarwala A., Han T.X.* Large-Scale Visual Font Recognition // *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
12. *Bulatov K., Arlazarov V., Chernov T. et al.* «SmartIDReader: Document recognition in video stream», in 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 6, IEEE, 2017, pp. 39–44. DOI: 10.1109/ICDAR.2017.347
13. *Chen G., Yang J., Jin H., Brandt J., Shechtman, Agarwala A., Han T.X.* Large-Scale Visual Font Recognition // *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
14. *Wang Z., Yang J., Jin H., Shechtman E., Agarwala A., Brandt J., Huang T.S.* DeepFont: Identify Your Font from An Image // *Proc. of ACM Multimedia*, 2015.
15. *Berenguel A., Terrades O.R., Lladós J., Cañero C.* e-Counterfeit: a mobile-server platform for document counterfeit detection // arXiv: 1708.06126, 2017.
16. *Постановление от 8 июля 1997 г. № 828* «Об утверждении положения о паспорте гражданина Российской Федерации, образца бланка и описания паспорта гражданина Российской Федерации» // Правительство Российской Федерации.
17. *Приказ Минфина РФ от 7 февраля 2003 г. № 14н* «О реализации постановления Правительства Российской Федерации от 11 ноября 2002 г. № 817.
18. *Постановление от 29 июля 2016 г. № 727* «О лицензировании деятельности по производству и реализации защищенной от подделок полиграфической продукции» // Правительство Российской Федерации.
19. *Chernyshova Y., Gayer A., Sheshkus A.* Generation method of synthetic training data for mobile OCR system // *Proc. of 10<sup>th</sup> International Conference on Machine Vision*, 2017. DOI: 10.1117/12.2310119
20. *Google Fonts*. – URL: <https://fonts.google.com>
21. *Приказ от 13 ноября 2017 г. № 851* «Об утверждении административного регламента министерства внутренних дел Российской Федерации по предоставлению государственной услуги по выдаче, замене паспортов гражданина Российской Федерации, удостоверяющих личность гражданина Российской Федерации на территории Российской Федерации» // Министерство внутренних дел Российской Федерации.
22. *Постановление от 27 мая 2011 г. № 424* «О машиночитаемой записи в паспорте гражданина Российской Федерации» // Правительство Российской Федерации.

**Чернышова Юлия Сергеевна.** ООО «Смарт Энджинс Сервис», г. Москва, Россия. Научный сотрудник-программист. Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, г. Москва, Россия. Аспирантка. Количество печатных работ: 3. Область научных интересов: синтез искусственных данных для обучения, аугментация данных. E-mail: [chernyshovayulia07@gmail.com](mailto:chernyshovayulia07@gmail.com)

**Алиев Михаил Александрович.** ООО «Смарт Энджинс Сервис», г. Москва, Россия. Научный сотрудник-программист. Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, г. Москва, Россия. Математик 1-й категории. Количество печатных работ: 2. Область научных интересов: обработка изображений. E-mail: [aliev.michael@gmail.com](mailto:aliev.michael@gmail.com)

**Шешкус Александр Владимирович.** ООО «Смарт Энджинс Сервис», г. Москва, Россия. Старший научный сотрудник-программист. Количество печатных работ: 18. Область научных интересов: глубокое обучение, компьютерное зрение, проективно-инвариантная сегментация изображений. E-mail: [astdcall@gmail.com](mailto:astdcall@gmail.com)

## Optical font recognition of images captured with mobile devices and its application for detecting identity documents forgery

Y.S. Chernyshova<sup>1,II</sup>, M.A. Aliev<sup>1,II</sup>, A.V. Sheshkus<sup>1</sup>

<sup>1</sup> LLC «Smart Engines Service», Moscow, Russia

<sup>II</sup> Institute for Systems Analysis, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia

**Abstract.** This paper addresses a method of optical font recognition via convolutional neural networks on images, acquired by small-scale digital cameras. We research applicability of such networks for detecting forgery of identity documents with fonts, specified in the standards. We provide experimental results of training networks for authenticity estimation of fonts, used in machine readable zones and numbers in Russian national passport. Results indicate that the proposed method can be used for specified fonts and for effective forgery detection if combined with other methods.

**Keywords:** OCR, OFR, convolutional neural networks, training data synthesis.

**DOI:** 10.14357/20790279180521

### References

1. *Young-bin Kwon, Jeong-hoon Kim.* Recognition based Verification for the Machine Readable Travel Documents // International Workshop on Graphics Recognition (GREC 2007), Curitiba, Brazil, 2007.
2. *De Koker L.* Money laundering compliance – the challenges of technology // Financial Crimes: Psychological, Technological, and Ethical Issues, Springer, 2016, pp. 329-347.
3. *Gerasimov A.* Tsifrovoe moshennichestvo: riski i ushcherb. – URL: <https://bosfera.ru/bo/cifrovoe-moshennichestvo-riski-i-ushcherb>
4. *Equifax and EFX.* The New Reality of Synthetic ID Fraud. // Atlanta, Georgia, 2015. – URL: [https://www.equifax.com/assets/IFS/syntheticID-fraud\\_wp.pdf](https://www.equifax.com/assets/IFS/syntheticID-fraud_wp.pdf)
5. *K. Gai, M. Qiu, and X. Sun.* “A survey on fintech,” Journal of Network and Computer Applications, 2017.
6. *ICAO Doc 9303.* 2015. Machine Readable Travel Documents. Seventh Edition. In twelve volumes. International Civil Aviation Organization.
7. *Bertrand R., Terrades O.R., Gomez-Krämer P., Franco P., Ogier J.-M.* A Conditional Random Field model for font forgery detection // 13<sup>th</sup> International Conference on Document Analysis and Recognition (ICDAR), pp. 576-580, 2015.
8. *Bertrand R., Terrades O.R., Gomez-Krämer P., Franco P., Ogier J.-M.* A system based on intrinsic features for fraudulent document detection // 12<sup>th</sup> International Conference on Document Analysis and Recognition, pp. 106-110, 2013.
9. *Abramova S., Böhme R.* Detecting copy-move forgeries in scanned text documents // Electronic Imaging 2016.8, 2016.
10. *Zramdini A., Ingold R.* Optical Font Recognition Using Typographical Features // IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume 20. pp. 877-882, 1998.
11. *Chen G., Yang J., Jin H., Brandt J., Shechtman, Agarwala A, Han T.X.* Large-Scale Visual Font Recognition // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
12. *Bulatov K., Arlazarov V., Chernov T. et al.* «SmartIDReader: Document recognition in video stream», in 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 6, IEEE, 2017, pp. 39–44. DOI: 10.1109/ICDAR.2017.347
13. *Chen G., Yang J., Jin H., Brandt J., Shechtman, Agarwala A, Han T.X.* Large-Scale Visual Font Recognition // Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
14. *Wang Z., Yang J., Jin H., Shechtman E., Agarwala A., Brandt J., Huang T.S.* DeepFont: Identify Your Font from An Image // Proc. of ACM Multimedia, 2015.
15. *Berenguel A., Terrades O.R., Lladós J., Cañero C.* e-Counterfeit: a mobile-server platform for document counterfeit detection // arXiv: 1708.06126, 2017.
16. *Postanovlenie ot 8 iyulya 1997 g.* N 828 «Ob utverzhdenii polozheniya o pasporte grazhdanina Rossijskoj Federatsii, obraztsa blanka i opisaniya pasporta grazhdanina Rossijskoj Federatsii» // Pravitel'stvo Rossijskoj Federatsii
17. *Prikaz Minfina RF ot 7 fevralya 2003 g.* N 14n «O realizatsii postanovleniya Pravitel'stva Rossijskoj Federatsii ot 11 noyabrya 2002 g. N 817»
18. *Postanovlenie ot 29 iyulya 2016 g.* N 727 «O litsenzirovanii deyatel'nosti po proizvodstvu i realizatsii zashchishchennoj ot poddelok



- poligraficheskoy produktsii» // Pravitel'stvo Rossijskoj Federatsii
19. *Chernyshova Y., Gayer A., Sheshkus A.* Generation method of synthetic training data for mobile OCR system // Proc. of 10<sup>th</sup> International Conference on Machine Vision, 2017. DOI: 10.1117/12.2310119
20. *Google Fonts.* – URL: <https://fonts.google.com>
21. *Prikaz ot 13 noyabrya 2017 g.* N 851 «Ob utverzhdenii administrativnogo reglamenta ministerstva vnutrennikh del Rossijskoj Federatsii
- po predostavleniyu gosudarstvennoj uslugi po vydache, zamene pasportov grazhdanina Rossijskoj Federatsii, udostoverayushchikh lichnost' grazhdanina Rossijskoj Federatsii na territorii Rossijskoj Federatsii» // Ministerstvo vnutrennikh del Rossijskoj Federatsii.
22. *Postanovlenie ot 27 maya 2011 g.* N 424 «O mashinochitaemoj zapisi v pasporte grazhdanina Rossijskoj Federatsii» // Pravitel'stvo Rossijskoj Federatsii

**Y.S. Chernyshova.** LLC “Smart Engines Service”, Prospekt 60-Letiya Oktyabrya, 9, Moscow, 117312, Russia.  
E-mail: [chernyshovayulia07@gmail.com](mailto:chernyshovayulia07@gmail.com)

**M.A. Aliev.** LLC “Smart Engines Service”, Prospekt 60-Letiya Oktyabrya, 9, Moscow, 117312, Russia.  
E-mail: [aliev.michael@gmail.com](mailto:aliev.michael@gmail.com)

**A.V. Sheshkus.** LLC “Smart Engines Service”, Prospekt 60-Letiya Oktyabrya, 9, Moscow, 117312, Russia.  
E-mail: [astdcall@gmail.com](mailto:astdcall@gmail.com)