Метод верификации изображений удостоверений личности по ограниченному фрагменту изображения

А.В. Самарин[I], В.А. Малых[II], П.С. Калайдин[I]

[I] ООО «В Контакте», г. Санкт-Петербург, Россия
[II] Казанский (Приволжский) федеральный университет, г. Казань, Россия

Аннотация. В статье рассмотрена задача распознавания фотографий документов, удостоверяющих личность человека. Описан подход к решению, основанный на использовании модифицированных вариантов архитектуры VGG для получения наиболее выразительных в контексте данной задачи дескрипторов изображений, и приведена комбинированная схема нейросетевой архитектуры для проверки изображения документа. Предлагаемое решение представляет собой составную модель, включающую в себя автокодировщик для извлечения наиболее выразительных характеристик изображения удостоверения личности. Рассмотрены преимущества подхода по сравнению с альтернативными современными решениями. Приведены результаты работы предложенного метода и аналогов на собственных наборах данных отдела восстановления доступа социальной сети «ВКонтакте».
Ключевые слова: *оптическое распознавание текстов, классификация изображений, устойчивость к шуму, нейронные сети.*
DOI: 10.14357/20790279200102

Введение

В настоящее время во многих интернет-сервисах, от платежных систем до схем восстановления доступа к учетным записям в социальных сетях, активно используются инструменты распознавания документов. Специфика построения механизма распознавания и проверки документа зависит от контекста использования. Например, в таких системах, как портал «Госуслуги»[1] и платежная система «Webmoney»[2], целью обработки документа является получение всей персональной информации о клиенте, например, в случае паспорта РФ это фамилия, имя, серия и номер паспорта, место и дата выдачи, код подразделения, машиночитаемая зона, и установление ее подлинности. Так как подобные системы зачастую основаны на регистрации конкретных фрагментов персональных данных, водяных знаков и их взаимного расположения на странице документа, размещенной в кадре строго определенном образом (для упрощения задачи автоматического распознавания персональной информации), к фотографиям документов предъявляются строгие требования (например, на едином портале государственных услуг[3]), которые регламентируют ракурс, условия освещения и другие параметры, обуславливающие качество механизмов распознавания надписей (таких как «Tesseract»[4]). В социальных сетях, таких как «ВКонтакте»[5], для восстановления доступа к учетной записи пользователя, не требуется получать всю персональную информацию, указанную на странице документа, удостоверяющего личность пользователя. В большинстве случаев достаточно проверить изображение определенной части документа, содержащей некоторую окрестность фотографии пользователя. Данная окрестность должна полностью содержать фотографию пользователя, имя, фамилию. Помимо наличия перечисленных данных, к исследуемой области предъявляются такие требования, как отсутствие искусственных дополнений, внесенных пользователем. Также не допускаются артефакты, являющиеся следствием условий съемки и влекущие за собой очевидные трудности при распознавании перечисленной ранее информации человеческим глазом. Разрешение фотографии должно обеспечивать должный уровень детализации для однозначной идентификации типа документа по

[1] https://www.gosuslugi.ru
[2] https://www.webmoney.ru
[3] https://51.мвд.рф/folder/2411299
[4] https://github.com/tesseract-ocr/tesseract
[5] https://vk.com/

характеру фона и наличию других специфических признаков. В настоящее время, множество различных подходов к распознаванию документов (таких как [Error: Reference source not found, Error: Reference source not found, Error: Reference source not found, Error: Reference source not found, Error: Reference source not found]) основаны на детектировании и разборе семантической информации из значимых областей (фотографий граждан, ключевых надписей, водяных знаков) и их взаимного расположения на фотографиях изображений удостоверений личности. Однако, данные методы неприменимы, когда мы хотим распознавать изображения окрестностей документов, содержащих фото пользователей, допуская различные условия освещения и искусственные изменения в фотографии, деформирующие часть персональных данных и предоставляющих лишь информацию (рис. 1), необходимую для подтверждения наличия у пользователя подлинного документа удостоверения личности.

В данной работе приведено описание нейросетевого метода, основанного на извлечении устойчивых признаков для проверки изображения, удостоверяющего личность человека, устойчивого к изменению ракурса и условий освещения, а также допускающего искусственные изменения с целью сокрытия полной персональной информации. Данный подход успешно зарекомендовал себя собственных наборах данных социальной сети «ВКонтакте» и успешно интегрирован в систему подтверждения личности пользователя при восстановлении доступа к учетным записям, что позволило существенно уменьшить нагрузку и

Рис. 1. Примеры изображений удостоверения личности, пригодных для идентификации пользователя: а) изображение, не содержащее никаких видимых искажений; б) изображение, содержащее допустимые пользовательские дополнения, скрывающие часть персональных данных; в) изображение, содержащее проективные искажения, возникшие вследствие варьирования ракурсов съемки; г) изображение, содержащее артефакты возникшие при варьировании условий освещенности (блики). (Изображения получены путем фотосъемки паспорта одного из авторов статьи, вся персональная информация скрыта с помощью размытия)

автоматизировать работу отдела восстановления доступа[6].

1. Постановка задачи

Задача распознавания типа удостоверения личности на изображениях, рассмотренная в данной работе, может быть сформулирована следующим образом. Изображение удостоверение личности Q должно быть отнесено к одному из классов $C = C_i$, где $i \in [0, N]$.

C_0 - класс прочих изображений, соответствующий значению «не является фотографией удостоверения личности, пригодной для идентификации пользователя».

Отметим, что помимо формальной постановки задачи классификации, к решению предъявляются требования по устойчивости к шумам и различным условиям съемки, так как данный метод должен обрабатывать все заявки на восстановление доступа к учетным записям для социальной сети «ВКонтакте». В частности, допускаются засветы, затенеуия, проективные искажения, пользовательские дополнения за пределами областей с фотографией и ФИО (рис. 1).

Воспользуемся следующими предположениями. Изображения пользователей получены с помощью камеры мобильного устройства неподготовленным пользователем при съемке в помещении, следовательно:

а) может содержать визуальные дефекты – блики, расфокусированные области, шум;
б) ракурс, кадрирование, освещение и цветовой баланс неизвестны;
в) изображения могут содержать области, содержащие искусственные изменения, добавленные пользователем для сокрытия части персональных данных.

На изображения, полученные от пользователя наложен ряд ограничений, несоблюдение которых влечет однозначную классификацию изображения документа, как C_0 (непригодных для идентификации пользователя).

Область удостоверения личности, содержащая фотографии пользователей, а также имя и фамилию не содержит искусственных изменений. Блики и шумы, перекрывающие область с фотографией, а также имя и фамилию не мешают визуальной идентификации при распознавании изображения человеком (сотрудником отдела восстановления доступа социальной сети «ВКонтакте»);

К пригодным, для процесса подтверждения личности, для граждан Российской Федерации, относятся фотографии документов из перечня[7]. Для граждан других государств, для которых поддерживается процедура автоматизированной верификации, необходима фотография паспорта.

2. Обзор существующих архитектур

Исходя из заявленной постановки, рассмотренный этап процедуры подтверждения личности при восстановлении доступа к учетным записям пользователей, сводится к задаче классификации фотографий документов, отправляемых пользователями в процессе восстановления доступа к учетной записи. Так как наиболее прогрессивные и эффективные методы классификации изображений на данный момент основаны на сверточных нейросетях, которые с большим отрывом превзошли все остальные подходы на открытых наборах данных для классификации изображений [Error: Reference source not found, Error: Reference source not found, Error: Reference source not found, Error: Reference source not found], модель для валидации изображений документов изначально разрабатывалась на основе использования нейросетевых моделей.

Каждая из перечисленных нейросетей была построена на основе инновационных для своего времени идей, эффективность применения которых исследовалась при классификации набора фотографий документов (раздел Error: Reference source not found).

Модель VGG [Error: Reference source not found] представляет собой последовательность сверточных, полносвязных слоев и слоев субдискретизации [Error: Reference source not found]. В данной работе рассмотрены варианты VGG16 и VGG19, отличающиеся друг от друга количеством слоев - 16 и 19 соответственно (рис. 6D и 6E).

Архитектура Inception [Error: Reference source not found] предложила использование специальных блоков, комбинирующих сверточные и полносвязные слои, для наиболее эффективного извлечения ключевых признаков из исходной информации и ее промежуточных представлений.

Структура ResNet [Error: Reference source not found] предполагает использование остаточных связей, что позволяет строить сети большей глубины.

MobileNet [Error: Reference source not found] предлагает использование специальную структуру сверток базовых блоков, которая позволяет сделать нейросетевую модель пригодной для использования без значительных потерь в эффективности.

[6] https://vk.com/blog/detecting-documents

[7] http://www.consultant.ru/cons/cgi/online.cgi?req=doc;base=LAW;n=149244\#008807190564687595

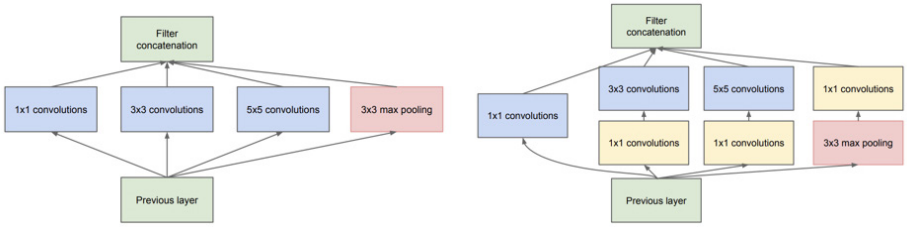

Рис. 2. Структура блока Inception [Error: Reference source not found].Error: Reference source not found

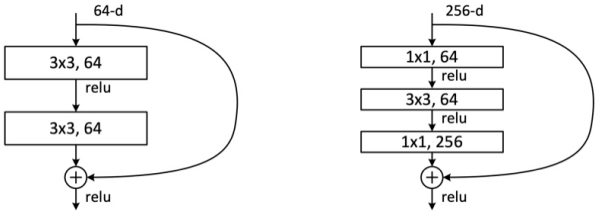

Рис. 3. Примеры остаточных связей ResNet [Error: Reference source not found].Error: Reference source not found

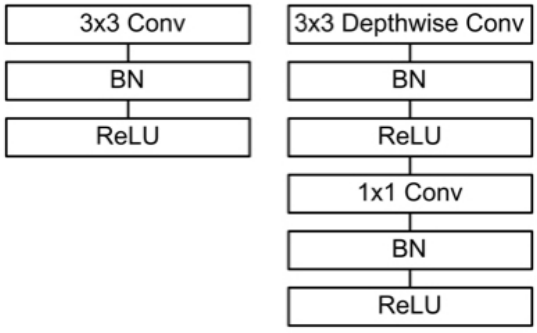

Рис. 4. Структура блоков MobileNet [Error: Reference source not found].Error: Reference source not found

3. Предлагаемая архитектура

Для решения рассматриваемой задачи классификации фотографий документов, перечисленные выше нейросетевые архитектуры были обучены и протестированы на тренировочных и тестовых наборах данных, соответственно (см. раздел Error: Reference source not found). При рассмотрении данных архитектур и заявленных результатов их работы на наборах данных общего характера (таблица 1), таких как [Error: Reference source not found] и на наборе фотографий документов, было выявлено, что лучшие результаты на наборах общих снимков не гарантируют более высокой точности классификации снимков документов, а именно, была установлена склонность моделей [Error: Reference source not found, Error: Reference source not found, Error: Reference source not found, Error: Reference source not found] к ложноположительным срабатываниям. Это можно проинтерпретировать, как следствие специфичности исследуемых данных.

Удостоверения личности характеризуются наличием ярко выраженных контуров, что отличает их от изображений общего плана, но следует также отметить важность и не столь явно выраженной контурной информации, такой, например, как орнамент фона. Помимо перечисленных признаков, большое значение имеет и взаимное расположение различных фрагментов информации, что в значительной степени обуславливает отличие фотографий одного типа документов от другого. Для фокусировки модели на перечисленных особенностях удостоверений личности, была рассмотрена комбинация специального автокодировщика (см. раздел 3.1) с последующей классификацией генерируемых изображений с помощью моделей [Error: Reference source not found, Error: Reference source not found, Error: Reference source not found, Error: Reference source not found]. Однако и данный подход не принес желаемого результата. Составленные таким образом классификаторы проявили склонность к ложноотрицательным срабатываниям.

Для того, чтобы чтобы избавиться от негативных сторон рассмотренных выше решений, в окончательную версию модели была заложена нейросеть, содержащая два основных потока вычислений, называемая в дальнейшем «VGG combined with Autoencoders» (VCA, рис. 5). Данная модель включила в себя две основные параллельные ветви – кодировщик на основе классификатора общего назначения и комбинированный классификатор с предварительным использованием автокодировщика.

Табл. 1

Результаты работы моделей на тестовом наборе фотографий общего плана ImageNet

Model	Top-1 Accuracy	Top-5 Accuracy
ResNet50	0.749	0.921
MobileNet	0.704	0.895
VGG16	0.713	0.901
VGG19	0.713	0.900
InceptionV3	0.779	0.937

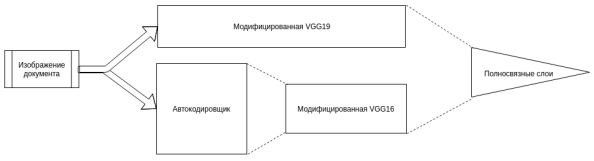

Рис. 5. Архитектура VCA.

Первая ветвь состоит из модифицированной VGG19 [Error: Reference source not found], сокращенной до предпоследнего слоя субдискретизации (рис. 6E). Данная часть архитектуры представляет собой своеобразный слабый классификатор, который позволяет устойчиво относительно шумов, ракурса и условий освещения определить наличие удостоверения личности на изображении. Однако, как показали проведенные эксперименты, при использовании подобного классификатора, возникает проблема в дифференцировании различных типов удостоверений личности между собой и определении документов, которые подходят для идентификации пользователя (например, удостоверение члена ассоциации рыболовов не считается удостоверением личности, несмотря на наличие фотографии в этом типе документов). Используя описанный механизм, возможно классифицировать изображения общего плана и изображения, содержащие удостоверения личности. Однако, данная модель не показывает должной эффективности при отделении фотографий одного типа удостоверений личности от другого.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

- Часть оригинальной архитектуры VGG, использованная в качестве слабого классификатора

- Часть оригинальной архитектуры VGG, использованная для обработки результата применения автокодировщика к исходному изображению

Рис. 6. Оригинальная архитектура VGG и ее части, использованные при построении предлагаемого классификатора VCA.

Для повышения способности VCA устойчиво различать типы документов между собой, в данную архитектуру добавлена дополнительную ветвь, которая содержит видоизмененную VGG16 (рис. 6D), которая обрабатывает результат применения автокодировщика [Error: Reference source not found] к анализируемому изображению. Автокодировщик интегрирован в модель для выделения информации о контурах в условиях присутствия шумов и вариативности факторов освещения. VGG16 здесь сокращена до предпоследнего слоя субдискретизации (аналогично VGG19 в другой ветви, рис. 6E). Следует отметить, что комбинация автокодировщика и VGG не позволяет хорошо распознавать подходящие изображения документов при сильном варьировании условий съемки. Однако, данная ветвь позволяет устойчиво разделять удостоверения личности разных типов (паспорта граждан различных государств). В итоге, при соединении вышеописанных потоков в один добавлением верхнего полносвязного слоя с сигмоидальной активацией, получается сеть, которая позволяет хорошо верифицировать и дифференцировать между собой удостоверения личности разного типа.

3.1. Автокодировщик для выделения характерных особенностей контуров

Рассмотрим подробнее ветвь предложенной архитектуры, содержащую в себе автокодировщик. Ключевой способностью автокодировщика является получение компактного представления информации, содержащейся во входных данных. Обычно автокодировщики обучаются восстанавливать исходный тензор из своего скрытого представления, однако, в VCA данная модель использована для устойчивого выделения дополнительной информации о контурах на исходном изображении, что позволяет извлекать информацию о взаимном расположении элементов (фотографии пользователя и областей с надписями и знаками, содержащими персональную информацию). Для достижения данного эффекта автокодировщик обучается восстанавливать из исходного изображения представление, полученное из исходного с помощью набора фильтров, задачей которого является выделение из исходного изображения максимально детализированного и избавленного от шумов контурного рисунка. Для достижения данного эффекта были использованы фильтры на основе фильтров Собеля [Error: Reference source not found] и Гаусса [Error: Reference source not found]. Фильтр Гаусса позволяет избавиться от высокочастотных шумов, путем сглаживания исходного изображения.

Последующее применение операторов Собеля с разным размером ядра, позволяет выделять информацию о контурах объектов, представленных на анализируемом изображении. После применения операторов Собеля, к результату фильтрации применяется адаптивное пороговое преобразование, учитывающее локальное значение средней яркости изображения (для обеспечения инвариантности при изменении освещения). Следует отметить, что в общем случае, оптимальных параметров (размеров ядер для операторов Гаусса и Собеля, а также размеров окрестности и порогу активации при пороговом преобразовании) не существует. Поэтому при создании обучающего набора данных для автокодировщика использовано сразу несколько вариантов выделения характерных особенностей контуров с различными параметрами, то есть каждому входному изображению сопоставляется множество P тензоров следующего вида:

$$P = \{t_i \square h_j \square v_k \square g_l(I) \mid t_i \in T, h_j \in H, v_k \in V, g_l \in G\}$$

где I - исходное изображение удостоверения личности, G - набор гауссовских операторов сглаживания с различными размерами ядер, H - набор горизонтальных операторов Собеля с различными размерами ядер, V - набор вертикальных операторов Собеля с различными размерами ядер, T - набор набор бинарных пороговых преобразований с различными значениями порогов. Далее, для каждого исходного изображения отбирается наилучший вариант из соответствующего ему множества тензоров, согласно контурной характеристике, в роли которой выступает дисперсия L_2 - норм областей из центральной части фотографии (для избежания влияния контурных характеристик краев страницы документа). Таким образом, каждому входному изображению сопоставляется выходной тензор b, удовлетворяющий условию:

$$b = \underset{p \in P(I)}{argminVar}\left[n\left(s\left(c(p)\right)\right)\right],$$

где I - исходное изображение удостоверения личности, P - набор результатов обработки исходного изображения с помощью фильтров и порогового преобразования (как показано выше), c - отображение, сопоставляющее изображению p его версию без рамки (обрезается 1/5 по высоте и 1/5 по ширине от исходных размеров изображения), s - отображение, производящее разбиение исходного на набор непересекающихся окон (ширина и высота окна - 1/10 от размеров аргумента - всего 100 окон, соответственно), n - отображение, принимающее на вход набор тензоров и возвращающее набор чисел, каждое из которых соответствует L_2 - норме одного из элементов входного множества, то есть $n(S) = \{||s_i|| \mid s_i \in S\}$.

4. Эксперименты

При разработке описанной выше архитектуры были исследованы методы на основе таких нейросетей, как VGG16, VGG19, Inception [Error: Reference source not found], ResNet [Error: Reference source not found], MobileNet [Error: Reference source not found], показавших высокую эффективность при решении задачи классификации изображений общего плана. Также были исследованы методы, представляющие собой комбинацию автокодировщика (см. раздел 3.1), результат применения которого обрабатывается с помощью перечисленных выше нейросетевых надстроек. Общая схема данных комбинированных моделей представлена на рис. 7.

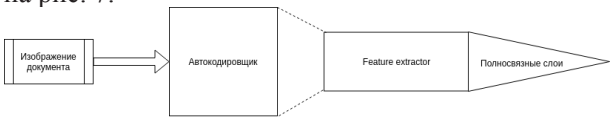

Рис. 7. Общая архитектура комбинированных моделей, содержащих автокодировщик. Feature extractor блок представляет собой сверточную нейронную сеть (VGG, Inception, ResNet или MobileNet) для выделения вектора признаков, который обрабатывается с помощью полносвязной сети.

Результаты работы данных архитектур (автокодировщик (далее AE) + VGG16, AE + VGG19, AE + Inception, AE + ResNet, AE + MobileNet) представлены в таблице 2.

Все перечисленные выше модели были обучены на собственных данных социальной сети «ВКонтакте». В качестве набора данных для обучения моделей были использованы фотографии документов сотрудников (18000 образцов, принадлежащие к 12 классам), аугментированные путем добавления искусственных элементов, имитирующих артефакты, возникающие при съемке в различных условиях освещения, а также различные вариации пользовательских дополнений, скрывающих часть персональной информации. В последствии, к набору данных применялись афинные преобразования для расширения количества поддерживаемых ракурсов съемки. Количество образцов распределено по классам равномерно. Следует отметить что расширение набора данных путем добавления дисторсии и других дополнительных артефактов помимо вышеперечисленных (см. раздел 1) не применялось, так как образцы снимков, содержащие сильные искажения данного типа не подходят под изначальные требования, а наличие

Табл. 2

Результаты работы моделей на наборах данных социальной сети «<ВКонтакте»>, где P – точность (precision), R – полнота (recall), F_1 – метрика F_1.

	Результаты на наборах данных								
Модели	**Смешанные**			**С дополнениями**			**Сложные ракурсы**		
	P	R		P	R		P	R	
VGG16	0.863	0.914	0.888	0.626	0.587	0.606	0.674	0.622	0.647
VGG19	0.871	0.912	0.891	0.631	0.576	0.602	0.672	0.638	0.655
MobileNet	0.897	0.931	0.914	0.678	0.593	0.633	0.691	0.683	0.687
ResNet	0.901	0.933	0.917	0.676	0.624	0.649	0.712	0.698	0.705
Inception	0.913	0.944	0.928	0.699	0.641	0.669	0.722	0.702	0.712
AE+VGG16	0.928	0.810	0.865	0.814	0.733	0.771	0.839	0.708	0.768
AE+VGG19	0.930	0.808	0.865	0.801	0.731	0.764	0.843	0.711	0.771
AE+MobileNet	0.951	0.896	0.922	0.927	0.812	0.866	0.867	0.787	0.825
AE+ResNet	0.962	0.883	0.920	0.920	0.818	0.866	0.865	0.772	0.816
AE+Inception	0.963	0.923	0.943	0.924	0.822	0.870	0.894	0.815	0.853
VCA	0.996	0.971	0.983	0.951	0.894	0.922	0.941	0.871	0.905

слабых искажений не оказало значительного влияния на результаты работы моделей. Тестовый же набор состоял из дополнительных 6000 изображений. Результаты тестирования работы перечисленных архитектур на тестовом наборе данных представлены в таблице 2.

Для оценки качества классификации была использована мера F_1 [Error: Reference source not found], вычисленная следующим образом:

$$Precision(P) = \frac{TP}{TP+FP}, \quad (1)$$

$$Recall(R) = \frac{TP}{TP+FN}, \quad (2)$$

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision+Recall} \quad (3)$$

где TP - количество образцов, верно распознанных в качестве образцов исследуемого класса; FP - количество образцов, неверно распознанных в качестве образцов исследуемого класса; и FN - количество образцов, неверно распознанных в качестве образцов, не принадлежащих к исследуемому классу. Таким образом, F_1 формулируется в терминах бинарной классификации. Для оценки качества классификации для нескольких классов F_1 усредняется по результатам, полученным для каждого исследуемого класса.

Также, для исследования устойчивости подобных архитектур при классификации изображений, содержащих приемлемые пользовательские дополнения, перечисленные выше модели были протестированы на наборах из 1000 изображений документов, содержащих допустимые дополнения и на 1000 фотографиях при сильно варьированных условиях освещения и ракурсах (углах съемки). Результаты тестирования представлены в таблице 2. Исходя из полученных результатов, предложенный метод, оказался значительно эффективней рассмотренных базовых методов и продемонстрировал качество, позволившее использовать его в качестве ядра механизма автоматизированного восстановления страниц в социальной сети «ВКонтакте», что в свою очередь позволило существенно снизить нагрузку на отдел восстановления доступа (20% заявок отклоняются на стадии проверки документов[8]) и убирать большую часть изображений с документами из поиска в социальной сети «<ВКонтакте»> (каждый день порядка 8000 изображений удостоверений личности убирается из поискового индекса).

Заключение

В работе показана эффективность метода верификации изображения удостоверения личности с использованием составной нейросетевой архитектуры. Модель используется отделом восстановления доступа социальной сети «ВКонтакте» и продемонстрировала точность выше 99,5%. Следует отметить, что при решении данной задачи была создана модель, позволяющая извлекать устойчивый относительно заявленных условий получения фотографий дескриптор изображения документа, что позволило успешно решить текущую задачу. В дальнейшем планируются исследования возможности использования спроектированной архитектуры для классификации изображений документов разного рода, а также рукописных текстов, в частности удостоверений личности, заполненных от руки. Исходя из полученных результатов, даль-

[8] https://vk.com/blog/detecting-documents

нейшее развитие работы может быть ориентировано как на исследование свойств метода на других классах распознаваемых документов, так и на улучшение его характеристик за счёт использования более выразительных дескрипторов изображений и оптимизации производительности модели.

Литература

1 *Javidi B., Horner J.L.* Optical pattern recognition for validation and security verification //Optical engineering. – 1994. – Т. 33. – №. 6. – С. 1752-1757.

2 *Visilter Y.V., Zheltov S.Y., Lukin A.A.* Development of OCR system for portable passport and visa reader //Document Recognition and Retrieval VI. – International Society for Optics and Photonics, 1999. – Т. 3651. – С. 194-199.

3 *Limonova E. et al.* Slant rectification in Russian passport OCR system using fast Hough transform // Ninth International Conference on Machine Vision (ICMV 2016). – International Society for Optics and Photonics, 2017. – Т. 10341. – С. 103410P.

4 *Kim K.B., Oh A., Woo Y.W.* PCA-based face verification and passport code recognition using improved FKCN algorithm //2008 Eighth International Conference on Intelligent Systems Design and Applications. – IEEE, 2008. – Т. 2. – С. 51-57.

5 *Kim K.B., Kim S.* A passport recognition and face verification using enhanced fuzzy ART based RBF network and PCA algorithm //Neurocomputing. – 2008. – Т. 71. – №. 16-18. – С. 3202-3210.

6 *Simonyan K., Zisserman A.* Very deep convolutional networks for large-scale image recognition //arXiv preprint arXiv:1409.1556. – 2014.

7 *Szegedy C. et al.* Going deeper with convolutions // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2015. – С. 1-9.

8 *He K. et al.* Deep residual learning for image recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 770-778.

9 *Howard A.G. et al.* Mobilenets: Efficient convolutional neural networks for mobile vision applications //arXiv preprint arXiv:1704.04861. – 2017.

10 *Tianyi Liu, Shuangsang Fang, Yuehui Zhao, Peng Wang, and Jun Zhang.* Implementation of training convolutional neural networks //*CoRR*, abs/1506.01195. - 2015.

11 *Deng J. et al.* Imagenet: A large-scale hierarchical image database //2009 IEEE conference on computer vision and pattern recognition. – Ieee, 2009. – С. 248-255.

12 *Hinton G.E., Salakhutdinov R.R.* Reducing the dimensionality of data with neural networks // Science (New York, NY). – 2006. – Т. 313. – №. 5786. – С. 504-507.

13 *Gupta S., Mazumdar S.G.* Sobel edge detection algorithm //International journal of computer science and management Research. – 2013. – Т. 2. – №. 2. – С. 1578-1583.

14 *Elboher E., Werman M.* Efficient and accurate Gaussian image filtering using running sums //2012 12th International Conference on Intelligent Systems Design and Applications (ISDA). – IEEE, 2012. – С. 897-902.

15 *Metz J. et al.* Comparing published multi-label classifier performance measures to the ones obtained by a simple multi-label baseline classifier //arXiv preprint arXiv:1503.06952. – 2015.

Самарин Алексей Владимирович. ООО «ВКонтакте», г. Санкт-Петербург, Россия. Старший разработчик отдела прикладных исследований. Количество печатных работ: 5. Область научных изображений: обработка изображений, нейронные сети. Email: alexey.samarin@vk.com

Малых Валентин Андреевич. Казанский (Приволжский) федеральный университет, г. Казань, Россия. Научный сотрудник, кандидат технических наук. Количество печатных работ: 23. Область научных интересов: обработка естественного языка, устойчивость к шуму. E-mail: valem.mipt@gmail.com

Калайдин Павел Станиславович. ООО «ВКонтакте», г. Санкт-Петербург, Россия. Руководитель отдела прикладных исследований. Количество печатных работ: 3. Область научных изображений: перспективные методы обработки данных, нейронные сети. E-mail: pavel.kalaidin@vk.com

ID verification method using limited image area

A.V. Samarin[I], V.A. Malykh[II], P.S. Kalaidin[I]

[I] JSC "VK.com", Saint-Petersburg, Russia

[II] Kazan (Volga Region) Federal University, Kazan, Russia

Abstract. In this work we study a ID document recognition task. We propose an approach based on VGG network variants, which allows to get the most expressive docmument descriptors for this task. We also propose a combined neural network architecture effective for document verification. The proposed architecture includes auto-encoder generating an expressive image descriptor. This apphiach we compare to the existing analogues, and provide results of evaluation on specific dataset of ID documents owned by VK.com.

Keywords: *optical character recognition, image classification, noise robustness, neural networks*.
DOI: 10.14357/20790279200102

References

1 *Javidi B., Horner J.L.* Optical pattern recognition for validation and security verification //Optical engineering. – 1994. – Т. 33. – №. 6. – С. 1752-1757.

2 *Visilter Y.V., Zheltov S.Y., Lukin A.A.* Development of OCR system for portable passport and visa reader //Document Recognition and Retrieval VI. – International Society for Optics and Photonics, 1999. – Т. 3651. – С. 194-199.

3 *Limonova E. et al.* Slant rectification in Russian passport OCR system using fast Hough transform // Ninth International Conference on Machine Vision (ICMV 2016). – International Society for Optics and Photonics, 2017. – Т. 10341. – С. 103410P.

4 *Kim K.B., Oh A., Woo Y.W.* PCA-based face verification and passport code recognition using improved FKCN algorithm //2008 Eighth International Conference on Intelligent Systems Design and Applications. – IEEE, 2008. – Т. 2. – С. 51-57.

5 *Kim K.B., Kim S.* A passport recognition and face verification using enhanced fuzzy ART based RBF network and PCA algorithm //Neurocomputing. – 2008. – Т. 71. – №. 16-18. – С. 3202-3210.

6 *Simonyan K., Zisserman A.* Very deep convolutional networks for large-scale image recognition //arXiv preprint arXiv:1409.1556. – 2014.

7 *Szegedy C. et al.* Going deeper with convolutions // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2015. – С. 1-9.

8 *He K. et al.* Deep residual learning for image recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 770-778.

9 *Howard A.G. et al.* Mobilenets: Efficient convolutional neural networks for mobile vision applications //arXiv preprint arXiv:1704.04861. – 2017.

10 *Tianyi Liu, Shuangsang Fang, Yuehui Zhao, Peng Wang, and Jun Zhang.* Implementation of training convolutional neural networks //*CoRR*, abs/1506.01195. - 2015.

11 *Deng J. et al.* Imagenet: A large-scale hierarchical image database //2009 IEEE conference on computer vision and pattern recognition. – Ieee, 2009. – С. 248-255.

12 *Hinton G.E., Salakhutdinov R.R.* Reducing the dimensionality of data with neural networks // Science (New York, NY). – 2006. – Т. 313. – №. 5786. – С. 504-507.

13 *Gupta S., Mazumdar S.G.* Sobel edge detection algorithm //International journal of computer science and management Research. – 2013. – Т. 2. – №. 2. – С. 1578-1583.

14 *Elboher E., Werman M.* Efficient and accurate Gaussian image filtering using running sums //2012 12th International Conference on Intelligent Systems Design and Applications (ISDA). – IEEE, 2012. – С. 897-902.

15 *Metz J. et al.* Comparing published multi-label classifier performance measures to the ones obtained by a simple multi-label baseline classifier //arXiv preprint arXiv:1503.06952. – 2015.

A.V. Samatin. JSC «VK.com», Saint-Petersburg, Russia. Senior Software Engineer of Applied Research Department. Email: alexey.samarin@vk.com

V.A. Malykh. PhD. Kazan (Volga Region) Federal University, Kazan, Russia. E-mail: valem.mipt@gmail.com

P.S. Kalaidin. JSC «VK.com», Saint-Petersburg, Russia. Head of Applied Research Department. E-mail: pavel.kalaidin@vk.com