Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

УДК 004.93’1; 004.932

КУЗЬМИЦКИЙ
Николай Николаевич

ДЕТЕКТИРОВАНИЕ И РАСПОЗНАВАНИЕ РАЗНОТИПНЫХ ТЕКСТОВЫХ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ ПРОИЗВОЛЬНЫХ СЦЕН СРЕДСТВАМИ СВЕРТОЧНЫХ НЕЙРОСЕТЕЙ

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

по специальности 05.13.17 – Теоретические основы информатики

Минск 2016

Работа выполнена в учреждении образования «Брестский государственный технический университет».

Научный руководитель	**Дереченник Станислав Станиславович**, кандидат технических наук, доцент, заведующий кафедрой электронных вычислительных машин и систем учреждения образования «Брестский государственный технический университет»
Официальные оппоненты:	**Татур Михаил Михайлович**, доктор технических наук, профессор, профессор кафедры электронных вычислительных машин учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»
	Богуш Рихард Петрович, кандидат технических наук, доцент, заведующий кафедрой вычислительных систем и сетей учреждения образования «Полоцкий государственный университет»
Оппонирующая организация	Государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси»

Защита состоится «_20_» октября 2016 г. в 14.00 на заседании совета по защите диссертаций Д 02.15.04 при учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, г. Минск, ул. П. Бровки, 6, корп. 1, ауд. 232, тел. 293-89-89, e-mail: dissovet@bsuir.by.

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

Автореферат разослан «_19_» сентября 2016 г.

Ученый секретарь совета
по защите диссертаций, кандидат
технических наук, доцент П.Ю. Бранцевич

КРАТКОЕ ВВЕДЕНИЕ

Текущий период развития человечества называют «информационным веком», т. к. окружающий нас мир как никогда прежде наполнен информацией, наиболее распространенной искусственной формой которой является текстовая. Повышение доступности вычислительных и оптических устройств сделало возможным ее автоматический анализ. Благодаря исследованиям в области машинного зрения созданы системы обработки текста, удовлетворяющие в той или иной мере практическим потребностям. При этом достигнутые в этой области отечественные результаты все еще недостаточны, и главным образом базируются на зарубежных прототипах, имеющих высокую коммерческую стоимость наряду со сложностью получения их пакетных тестовых версий.

Кроме того, основным условием эффективности внедрений зачастую является ориентация разработчиков на получение частных решений с учетом предполагаемых ограничений характеристик входных данных: шрифтовые и рукопечатные образы; изображения, создаваемые специальным оборудованием либо в специальных условиях, и др. Но данный подход не позволяет достичь приемлемых результатов в целом ряде областей, начиная от обработки документов с различными способами синтеза текстовых образов, заканчивая робототехникой. При этом методы классической парадигмы *OCR* (*optical character recognition*, оптическое распознавание образов) в своей базовой реализации не могут использоваться для обработки изображений реальных сцен ввиду большого разнообразия композиции и средств форматирования текста.

Данные факторы обосновывают актуальность создания новых методик и алгоритмов, повышения адаптивности существующих методов, построения эффективных программных средств анализа текстовых данных. Результатом диссертационного исследования являются авторские разработки, применимые в различных практических приложениях: детектировании текстовых объектов на изображениях реальных сцен (автоматический учет транспортных средств) автоматизации документооборота (оцифровка документов и потоковый ввод данных), системах дополненной реальности (навигация в городе) и др.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Связь работы с крупными научными программами и темами

Тема диссертации соответствует приоритетным направлениям научных исследований в Республике Беларусь на 2011–2015 гг., утвержденным Постановлением Совета Министров Республики Беларусь от 12 августа 2010 г. № 1196, п. 5.4 «Математические и интеллектуальные методы, информационные

технологии и системы распознавания и обработки образов, сигналов, речи и мультимедийной информации». Диссертация выполнялась в рамках научно-исследовательских работ кафедры электронных вычислительных машин и систем учреждения образования «Брестский государственный технический университет» (БрГТУ), а также гранта исследовательского подразделения компании Microsoft (Microsoft Research) и Московского государственного университета имени М.В. Ломоносова на участие в Microsoft Computer Vision School 2011 (г. Москва, Российская Федерация, 28.07.2011 – 03.08.2011).

Цель и задачи исследования

Объект исследования – растровые изображения с текстовыми объектами.

Предмет исследования – методы обработки растровых изображений, анализа текстовых данных, машинного обучения (искусственные нейросети).

Целью диссертационной работы является разработка нейросетевых методик и адаптивных алгоритмов детектирования и распознавания текстовых объектов растровых изображений с произвольной композицией сцены, различными яркостными свойствами и способами синтеза текстовых образов.

Для достижения поставленной цели необходимо решить следующие задачи:

1) локализовать текстовые области на изображении произвольной сцены;

2) структурировать текстовые данные по основным уровням группировки: блок, строка, слово и символ;

3) классифицировать разнотипные текстовые образы, включая шрифтовые, рукописные и искаженные.

Научная новизна

1. С учетом структурных и яркостных свойств текста модифицированы известные методы анализа изображений, что позволило для контуризации Canny увеличить целостность формируемых контуров, для бинаризации Niblack повысить качество разделения классов (текста и фона) при их яркостной неоднородности.

2. Разработан способ локализации и сегментации текстовых блоков на изображениях произвольных сцен с помощью сверточных нейросетей, при этом ресурсоемкость их применения снижена более чем на два порядка.

3. Впервые исследована проблема «хрупкости» (низкой перекрестной точности) типовых сверточных нейросетей LeNet-5, для преодоления которой разработана оригинальная технология создания текстового классификатора, существенно повышающая (более чем на 6 %) перекрестную точность одиночных нейросетей без применения высокопроизводительных вычислений.

4. Синтезированы комитеты сверточных нейросетей, обладающие точностью распознавания рукописных образов цифр MNIST (99,65 %) и заглавных букв

английского алфавита NIST (98,17 %) на уровне лучших мировых результатов (комитетов нейросетей в 1,6 и 2,7 раза более громоздких, чем использованные).

5. Созданы классификаторы образов цифр, букв английского и русского алфавитов, перекрестная точность которых на разнотипных образах (более 95 %) превышает уровень ведущих коммерческих аналогов, что позволяет применять классификаторы в решении сложных прикладных задач.

6. Впервые собрана представительная база изображений маркированных текстовых объектов русского языка номинальным объемом 106500 образов, являющаяся эффективным ресурсом для сравнения методов детектирования, сегментации и распознавания образов.

Положения, выносимые на защиту

1. Модель текстового детектора в виде неглубокой сверточной нейросети и способа ее применения модификацией мультимасштабного фрагментирования изображения с последовательным уточнением положения текстовых объектов по откликам нейросети в близких масштабах соседних строк изображения.

2. Методика синтеза компактного комитета сверточных нейросетей путем их обучения на выборках текстовых образов различного (фиксированного для каждой нейросети) масштаба и селекции по модифицированному критерию эффективности, обеспечивающая необходимое разнообразие членов комитета.

3. Технология создания классификатора разнотипных текстовых образов различного алфавита в виде нейросетевого комитета с тремя уровнями: отнесения входного образа к группе схожих по начертанию; выбора класса в группе; коррекции решения с учетом оценки уверенности распознавания и типовых ошибок, выявляемых в ходе практического применения классификатора.

Личный вклад соискателя ученой степени

Результаты диссертационного исследования и положения, выносимые на защиту, получены автором самостоятельно. Вклад научного руководителя кандидата технических наук, доцента С.С. Дереченника связан с постановкой целей и задач исследования, определением возможных путей их решения, обсуждением полученных результатов.

Апробация результатов диссертации и информация об использовании ее результатов

Основные результаты диссертационного исследования докладывались и обсуждались на следующих научных мероприятиях: Международные научные конференции «Информационные технологии и системы» – Минск, 2011, 2012; VII международная конференция «Neural Networks and Artificial Intelligence» –

Минск, 2012; International Conference on Computer Graphics and Vision GraphiCon'2013 – Владивосток, 2013; XVIII научно-практическая конференция «Комплексная защита информации» – Брест, 2013; Республиканская научная конференция молодых ученых и студентов «Современные проблемы математики и вычислительной техники» – Брест, 2009, 2011.

Использование результатов диссертации подтверждено актами внедрения в разрабатываемые в ООО «ГЕРСИС СОФТВЕР» системы автоматической обработки изображений документов и в учебный процесс БрГТУ.

Опубликование результатов диссертации

По теме диссертационного исследования опубликовано 15 работ, включая 5 статей в рецензируемых научных журналах, 3 статьи в сборниках научных работ, 7 докладов в трудах международных и республиканских конференций. Общий объем публикаций по теме диссертации составляет 5,67 авторских листа, из них 2,99 в научных журналах.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, библиографического списка и приложения. Общий объем диссертации составляет 136 страниц, из них 103 страницы текста, 63 рисунка на 17 страницах, 11 таблиц на 3 страницах, 1 приложение на 3 страницах, библиографический список из 122 наименований на 10 страницах.

ОСНОВНАЯ ЧАСТЬ

Во ***введении*** обоснована актуальность темы диссертационной работы, сформулированы цель и задачи исследования.

В ***первой главе*** рассмотрены основные аспекты анализа растровых изображений, содержащих текстовые данные. *Главную цель* их автоматической обработки можно сформулировать следующим образом: имеется изображение произвольного размера с распределенными текстовыми объектами различного способа синтеза и формата, необходимо получить их структурное описание и выполнить перевод в цифровую форму. Актуальность систем обработки текста существенно возросла ввиду постоянного расширения сферы их практического применения (например, потоковый ввод данных, оперативное информирование водителей и т. п.), а также в связи с повышением доступности оптических цифровых устройств (в частности, камер смартфонов).

Учитывая сложность композиции и стилистическое разнообразие изображений, можно выделить следующие классы: *изображения документов*,

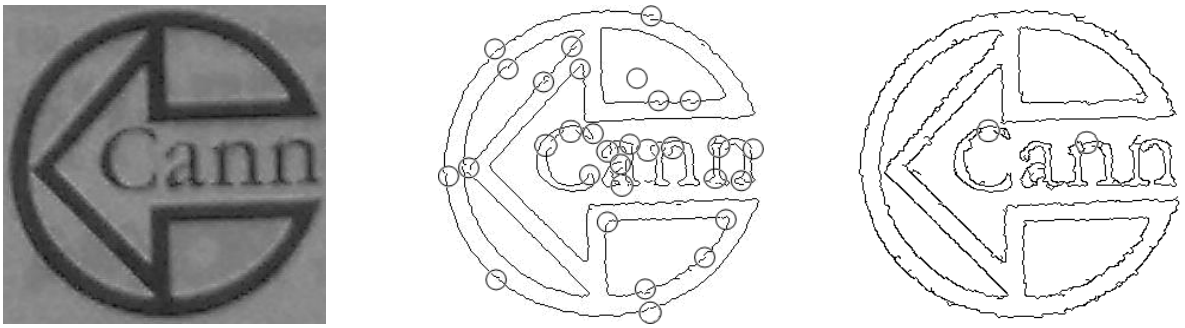
печатных изданий, *реальных сцен*. Их обработка предполагает решение общих для уровней иерархии текста (блок, строка, слово и символ) задач, сложность которых для указанных классов различна, включая локализацию текстового объекта, высокоуровневое описание, классификацию и добавление в иерархию. Эффективность решения зависит от универсальности контекстной информации, адаптивности и ресурсоемкости алгоритмов.

Анализ существующих подходов к обнаружению и структуризации текстов на изображениях выявил, что применимость методов, разработанных в рамках традиционной парадигмы OCR, ограничена изображениями документов с условием оптимальной настройки соответствующих параметров методов. Установлена необходимость применения для обработки изображений реальных сцен комплексных подходов, комбинирующих различные поисковые стратегии, а также использующих текстовые классификаторы. Обзор методик формирования признаков показал, что большинство основано на простых зрительных моделях и интуитивно подобранных характеристиках (чаще яркостных и структурных), эффективность которых ограничена машинопечатным текстом, имеющим низкую вариативность начертания. При этом адаптивная пороговая обработка признаков рукописных, а в широкой постановке задачи – произвольно синтезированных образов, существенно затруднена, в частности, из-за наличия принципиально разных вариантов начертания символов одного класса (например, 'Д'/'𝔇').

Анализ методов классификации текстовых образов позволил определить необходимость учета, при выборе модели классификатора, внутриклассовой изменчивости образов, что зачастую не позволяет использовать методы, требующие точных предположений о статистических свойствах классов. Более эффективными представляются классификаторы в виде искусственных нейросетей, формирующих разделяющие функции непосредственно в ходе своего обучения, являющиеся также экстракторами высокоуровневых признаков. В частности, Р.Х. Садыхов и М.Е. Ваткин на основе неокогнитрона создали классификатор рукопечатных образов заглавных букв русского алфавита с точностью на уровне 91,5 %, однако предложенное ими решение не обладает достаточной универсальностью из-за проблемы «хрупкости» методов машинного обучения.

Во ***второй главе*** рассмотрены реализации двух подходов к локализации текстовых объектов на изображениях: *нейросетевого* и *OCR-подхода*. Их общая проблема – разное представление объектов при изменении условий съемки, что обусловливает применение признаков, отражающих универсальные свойства. В рамках OCR-подхода таковыми являются контуры, исследование которых выявило ограничения метода Canny (наиболее применимого в анализе текстов): связывание ориентации градиента в точке только с одним из четырех направлений, кратных 45º; отбрасывание точек, соединяющих обособленные сегменты контура объектов; отсутствие учета близости характеристик смежных контурных точек.

Разработана *модифицированная версия метода Canny,* обеспечивающая бóльшую целостность контуров (рисунок 1) за счет назначения точке до трех направлений прослеживания, замены условия локальной экстремальности модуля градиента в точке его достаточной контрастностью по направлению, что позволяет выделять слабосвязные сегменты контура. Фильтрация по монотонности изменения ориентации вдоль контуров, их утончение (аналогичное полутоновому) и удаление «петель» широких перепадов яркости, приводят контуры к единичной толщине. Недостатки модификации – повышение ресурсоемкости и числа параметров. На основе метода разработан *алгоритм контурной сегментации изображений* путем кластеризации контурных сегментов на базе критерия их пространственной близости и фильтрации кластеров по геометрическим и статистическим признакам яркости, с учетом при пороговом ограничении особенностей структуры текстов. Сложность адаптивного подбора порогов ограничивает применимость алгоритма изображениями документов и печатных изданий, имеющих общие принципы распределения текстовых данных и схожие стилистические характеристики.



а б в

а – исходное; б – результат метода Canny; в – модификации метода (○ – разрыв контура)

Рисунок 1. – Формирование контурного представления изображения

Поиск текста на изображениях реальных сцен – более сложная задача ввиду низкого качества изображений, разнообразия свойств текстовых объектов, сложной композиции сцен и т. д., что обусловливает необходимость применения подходов, основанных на методах машинного обучения. Анализ показал, что для обработки изображений одними из наиболее эффективных являются *сверточные нейросети (CHC, CNN)*, основанные на трех архитектурных идеях: разделяемые веса, локальные рецептивные поля и пространственные подвыборки. Для решения задачи обнаружения текста разработана *модель детектора в виде неглубокой сверточной нейросети* (рисунок 2), первые четыре слоя которой являются экстрактором высокоуровневых признаков, а последние два – классификатором. Выход нейросети интерпретируется как решение о наличии (отсутствии) на входном изображении текстового образа. Нейросеть содержит 409912 связей и 2272 настраиваемых параметра, а ее обучение выполняется на основе метода Левенберга–Марквардта, увеличивающего скорость сходимости процесса.

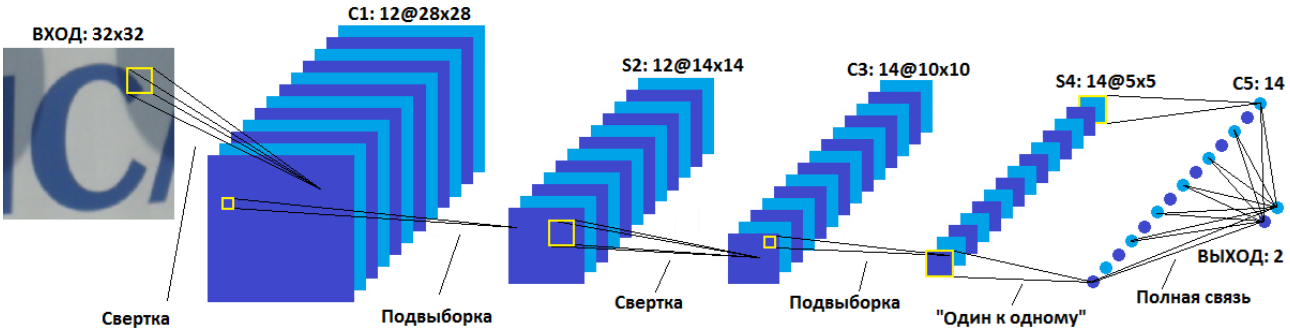


Рисунок 2. – Архитектура нейросетевого детектора текстовых образов

Преимущество модели перед методом опорных векторов (SVM) – интеграция синтеза признаков (в режиме «черного ящика») и классификации. В сравнении с аналогичными (Delakis и Garcia, 2008) модель более универсальна: позволяет выполнять как посимвольное, так и построчное детектирование, при этом ее применимость не ограничена сценами с однотипными текстовыми объектами (в отличие от предложенной А.А. Друки, 2014 для локализации регистрационных номеров автотранспорта). Кроме того, она имеет компактную архитектуру (в 40 раз меньшую, чем у созданной Wang и др., 2012), снижающую ресурсоемкость детектирования и не требующую бесконтрольного обучения, что повышает эффективность применения модели как универсального бинарного классификатора текстовых образов (так, в исследовании она использовалась для распознавания образов с низким уровнем межгруппового отличия, в частности букв 'Е'/'Ё').

Основной разработанный способ применения модели – модификация *алгоритма мультимасштабного фрагментирования изображения* скользящим окном с последующей локализацией и сегментацией текстовых блоков по соответствующим алгоритмам. Модификация заключается в расчете откликов детектора одновременно для всех позиций скользящего окна в одном масштабе изображения путем вычисления сверток с шагами h_i откликов карт $(i-1)$-го слоя с фильтрами текущего i-го слоя и группировкой разреженных результатов сверток, учитывая связность и тип слоев, что позволяет на два порядка сократить ресурсоемкость применения модели детектора, а также аналогичных ей:

$$M_i^j(x,y) = f_i\left(s_i^j + \sum_{l\in V_i^j}\sum_{p=-a_i,h_i}^{a_i}\sum_{t=-b_i,h_i}^{b_i} w_i^l(p+a_i,t+b_i)\cdot M_{i-1}^l(x+p,y+t)\right), \quad (1)$$

где M_i^j – отклики j-й карты i-го слоя;

f_i – функция активации нейронов i-го слоя;

s_i^j – величина смещения нейронов j-й карты i-го слоя;

V_i^j – номера карт $(i-1)$-го слоя, связанных с j-й картой i-го слоя;

w_i^l – фильтр размера $(2a_i+1)\times(2b_i+1)$ j-й карты i-го слоя, связывающий каждый из ее нейронов с подмножеством нейронов l-й карты $(i-1)$-го слоя.

Этапами созданного *алгоритма локализации текстовых блоков* являются: вычисление матриц уверенности детектора для выбранных масштабов изображения; кластеризация «уверенных откликов» отдельных масштабов в первичные блоки; объединение первичных и выбор итоговых текстовых блоков. За счет совместной обработки откликов детектора в близких строках и соседних масштабах алгоритм, в отличие от аналогов, снижает требования к точности детектора и позволяет локализовать блоки произвольной ориентации. Оценка качества локализации на изображениях базы ICDAR 2013, дорожных сцен и страниц паспорта показала работоспособность модели и способа ее применения при сложной композиции объектов и высокой стилистической вариативности текстовых образов (рисунок 3).

а б в

а – базы ICDAR 2013; б – дорожной сцены; в – страницы паспорта

Рисунок 3. – Локализация текста с помощью нейросетевого детектора на изображении

В OCR-подходе часто применяются бинарные изображения, синтез которых может быть осложнен составным фоном, зашумленностью и др. Учитывая это, создана *методика адаптивной бинаризации изображения текстового блока*, в отличие от аналогов (Chen, 2004) объединяющая два вида порогового ограничения: глобальный порог, рассчитанный по методу Otsu, применяется для коррекции локальных порогов метода Niblack. Исключение фона, с использованием оценки толщины образов, повышает качество бинаризации изображений с неравномерным освещением, а оптимизация расчетов с помощью «интегральных» изображений существенно снижает ресурсоемкость обработки. Тестирование показало бóльшую эффективность методики по сравнению с методами Otsu и Niblack (типовыми для OCR) при яркостной неоднородности разделяемых классов (текста и фона).

Традиционными способами группировки текста являются строка и слово, обособляемые пробельными интервалами. Разработанные *алгоритмы сегментации изображения текстового блока на строки и слова* основаны на объединении связных бинарных сегментов блока с учетом их близости (в OCR-подходе) либо на анализе матрицы уверенности текстового детектора для него (в нейросетевом). Первые применимы для обработки изображений документов, вторые, учитывая низкую уверенность решений детектора в пробельных интервалах, позволяют

локализовать на изображениях реальных сцен строки и слова с ориентацией, отличающейся от строго горизонтальной, путем выделения их «главных осей» (проходящих через локальные максимумы матрицы) аналогично методу Хафа.

Предшествующим распознаванию является этап выделения образов, для выполнения которого в рамках OCR-подхода предложен *алгоритм сегментации изображения слова*, рассматривающий локальные экстремумы суммарной яркости в столбцах как вероятные границы символов с дальнейшим выбором оптимальных. Алгоритм включает методику адаптивной бинаризации (при этом бинарное изображение является маской полутонового в отличие от А.В. Брухтий, 2014) и оценки характеристик как отдельных секущих (локальной контрастности), так и их серий (близости расстояний между секущими). Универсальность алгоритма повышена за счет наклонных секущих, соединяющих локальные экстремумы кривизны внешнего контура слова (позволяют отделять образы вариативного начертания), выбора лучшей серии секущих с учетом оценки классификаторов. Сравнение с ABBYY FineReader 12 на выборке искаженных изображений дат показало бóльшую адаптивность алгоритма (качество сегментации выше на 9 %).

В ***третьей главе*** исследованы технологии распознавания, основанные на сверточных нейросетях архитектуры LeNet-5 (рисунок 4). Для оценки свойств классификаторов проведена *систематизация множества текстовых образов*, в ходе которой выделены их значимые типы: *машинопечатный, рукописный* и *синтезированный*. Для последнего разработана *процедура генерации*, основанная на «волновых» искажениях. Сгенерированы (KNI1, KNI2*)* и отобраны (FONT, ETL6, USPS, NIST, ETL1, MNIST) базы, содержащие сотни тысяч образов.

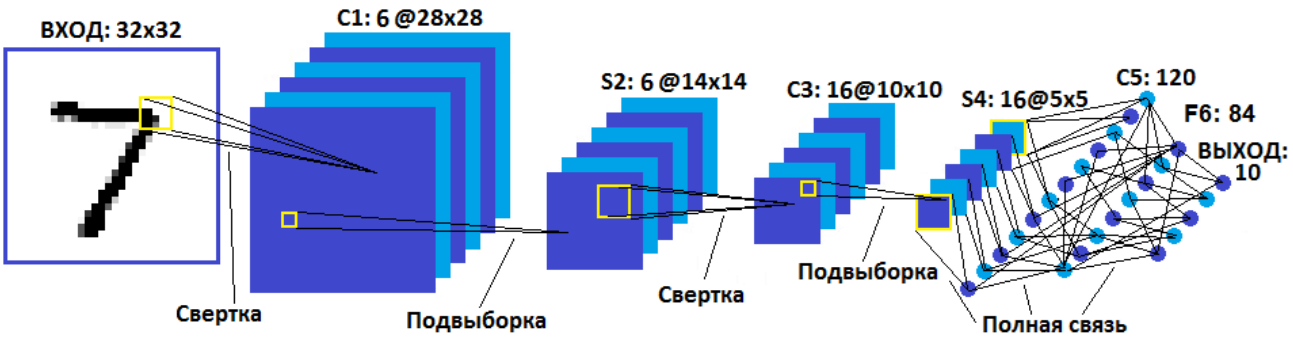


Рисунок 4. – Архитектура LeNet-5 классификатора текстовых образов (Y. LeCun, 1998)

Обучение нейросети выполняется с искажением тренировочных образов и дообучением, улучшающими ее обобщающие свойства. Тестирование показало эффективность LeNet-5 (CNN_h20, таблица 1) лишь в распознавании образов, подобных тренировочным. Проблема *«хрупкости»* (Seewald, 2010) присуща и другим моделям машинного обучения: создание классификатора, который после обучения на одном подмножестве N-классового множества образов для других подмножеств обеспечивал точность распознавания выше заданного уровня T.

Таблица 1. – Точность распознавания (в %) одиночной нейросетью (CNN) и нейросетевыми комитетами (COM) образов рукописных и шрифтовых цифр

Классификатор	MNIST_test	NIST	USPS	FONT	Среднее
CNN_h20	99,39	99,13	98,77	95,23	98,13
MAX_COM	99,52	99,16	99,12	96,08	98,47
AVER_COM	**99,58**	**99,21**	99,03	**96,85**	**98,66**
MAJOR_COM	99,57	99,18	**99,05**	96,71	98,62

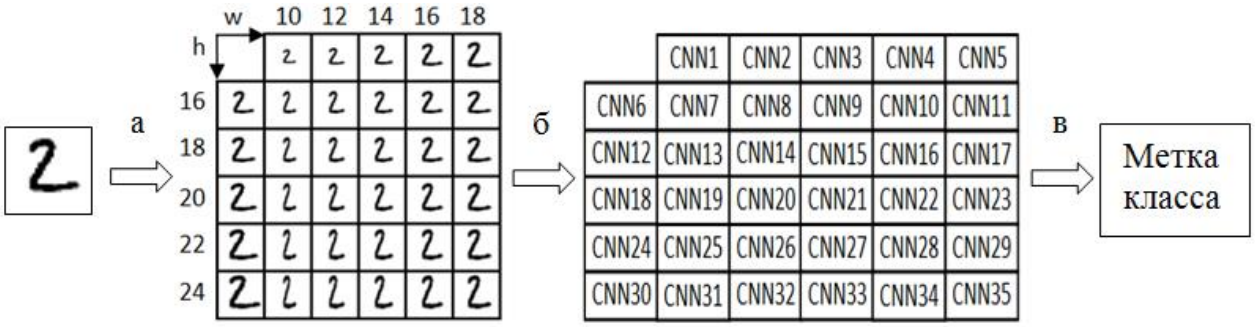
Известно, что принятие решения на основе объединения классификаторов зачастую более эффективно, чем с помощью лучшего из них (Dietterich, 2000), поэтому для преодоления проблемы хрупкости разработана *методика синтеза комитета сверточных нейросетей* (схема применения показана на рисунке 5):

1) выберем диапазоны вариации высоты h и ширины w образов (учитывая размер в 20×20 пикселей, применены $h\in[16,24]$ и $w\in[10,18]$ с шагом, равным 2);

2) масштабируем образы к выбранным размерам: к 25 с фиксацией обеих размерностей, к 10 – с фиксацией одной из них и сохранением отношения w/h;

3) создадим из обучающего множества 35 выборок той же мощности, но с различными размерами образов, с помощью которых обучим 35 нейросетей;

4) сформируем нейросетевой комитет и выполним селекцию его членов.



а – масштабирование образа; б – распознавание; в – объединение решений членов

Рисунок 5. – Схема применения комитета сверточных нейросетей

Преимущество методики – применимость к неглубоким нейросетям, обучение которых проводится на образах различного (фиксированного для каждой) масштаба, за счет чего нейросети (с одинаковыми размерами фильтров) выделяют уникальные признаки образа, изменяя степень приближения к нему (наподобие «увеличительного стекла»). При этом из-за высокой ресурсоемкости полной реализации методики в рамках исследования применялись не все масштабы. В частности, на базе MNIST обучены 10 нейросетей при следующих размерах образов: $\{h=16, 18, 20, 22, 24;\ w/h$ сохранено$\}$ и $\{h=20;\ w=10, 12, 14, 16, 18\}$.

Для формирования нейросетевых комитетов (COM) применялись схемы голосования, решение которых соответствует: *AVER* – наибольшему среднему, *MAX* – максимальному, *MAJOR* – преобладающему отклику членов комитета.

Тестирование (см. таблицу 1) показало, что схема *AVER* ввиду высокой корреляции откликов нейросетей (обученных на образах одной базы) наиболее эффективна.

Снижение размерности комитетов обусловило необходимость *селекции членов* – выбора компактного подкомитета с точностью не меньшей, чем у всего комитета. Вклад *I* каждого члена в эффективность комитета оценен модификацией алгоритма *EPIC* (Lu, 2010), бинарное множество значений коэффициентов α_{ij}, β_{ij}, λ_{ij} которого заменено на непрерывный диапазон $[0,1]$ откликов нейросети:

$$I_i = \sum_{j=1}^{N} \alpha_{ij} \left(2v^j_{y^1_j} - v^j_{y^*_j} \right) + \beta_{ij} v^j_{y^2_j} + \lambda_{ij} \left(v^j_{y^*_j} - v^j_{c_i(\mathbf{x}_j)} - v^j_{y^1_j} \right), \qquad (2)$$

где $c_i(\mathbf{x}_j)$ – метка класса для образа $\mathbf{x}_j$ ($j = \overline{1, N}$) по решению *i*-го члена;
v^j_m – количество членов, относящих *j*-й образ к классу с меткой *m*;
y^1_j, y^2_j и y^*_j – первая, вторая преобладающие в ответах членов и верная метки.

Селекция позволила создать *комитет четырех нейросетей с точностью распознавания рукописных образов цифр MNIST 99,65 %,* соответствующей уровню лучших мировых результатов. Выбор нейросетей (10, 1, 9 и 6-й по точности), обученных образам размеров $\{h = 16, 18, 24;\ w/h$ сохранено$\}$ и $\{h = 20;\ w = 10, 18\}$, обусловлен балансом их точности и разнообразия – условием эффективности комитета. Достигнутая точность на MNIST превышает все ненейросетевые модели (SVM, k-NN и т. д.) и является наивысшей для архитектуры LeNet-5. Лучшие результаты (0,23 % Ciresan и др., 2012) и (0,21 % Wan и др., 2013) получены комитетами из 35 и 5 нейросетей, имеющих в 1,6 и 2,7 раза больше настраиваемых параметров, чем LeNet-5. Кроме того, по методике синтеза сформирован комитет *с точностью распознавания рукописных образов заглавных букв английского алфавита NIST 98,17 %,* соответствующей лучшему результату для данной базы.

Таким образом, установлено, что методика синтеза повышает точность распознавания, однако результат для базы FONT (см. таблицу 1) показал, что ее эффективность ограничена выборками одного типа, поэтому исследованы образы цифр и заглавных букв английского алфавита трех выделенных типов. Первыми были созданы «частные» цифровые / заглавные нейросети, обученные на образах одного из типов баз FONT_dig, MNIST, KNI1_dig / FONT_big, NIST_big, KNI1_big (_dig, _big и _test – цифровая, заглавная и тестовая части баз). Тестирование выявило низкую универсальность частных нейросетей, т.к. их средняя перекрестная точность составила 92,58 % / 91,90 %, поэтому они объединены в комитет с MAX-голосованием, что увеличило точность на 5,89 % / 5,80 %. Далее по методике синтеза созданы частные комитеты (COM) и их объединения (COM_COM). Точность первых превысила частные нейросети на 2,27 % / 2,75 %, а точность вторых выросла по сравнению с MAX-комитетами частных нейросетей на 0,62 % / 0,18 % (таблица 2).

Таблица 2. – Точность распознавания (в %) сверточными нейросетями (CNN), их комитетами (COM) и коммерческими классификаторами (NicomsoftOCR, SmartZoneICR) образов цифр (dig) и заглавных букв английского алфавита (big)

Комитет	**MNIST_test**	**FONT_dig_test**	**KNI1_dig_test**	**Среднее**
MAX_COM_dig	98,23	98,62	98,56	98,47
COM_MNIST_dig	**99,58**	96,85	89,77	95,40
COM_COM_dig	98,95	99,27	99,06	99,09
CNN_MKF	98,82	99,01	99,55	99,12
COM_MKF	99,15	**99,56**	**99,86**	**99,52**
SmartZoneICR	97,65	95,07	89,04	93,92
NicomsoftOCR	97,50	97,44	85,84	93,59
Комитет	**NIST_big_test**	**FONT_big_test**	**KNI1_big_test**	**Среднее**
MAX_COM_big	96,62	98,27	98,21	97,70
COM_NIST_big	**97,94**	89,28	86,95	91,39
COM_COM_big	96,61	98,43	98,60	97,88
CNN_NKF	97,17	98,10	98,07	97,78
COM_NKF	97,68	**98,80**	**98,81**	**98,43**
SmartZoneICR	92,07	93,85	86,46	90,79
NicomsoftOCR	85,55	97,76	85,26	89,52

Так как выявлена низкая универсальность частных нейросетей, исследованы общие нейросети, которые обучались на смешанных по типу образов выборках, собранных из равных частей баз, примененных для обучения частных. Тестирование показало, что точность общих нейросетей и их комитетов (_MKF и _NKF) превысила как точность частных комитетов (MAX_COM_dig – на 0,65 %, MAX_COM_big – на 0,73 %, COM_COM_dig – на 0,03 %, COM_COM_big – на 0,55 %), так и коммерческих аналогов, что подтвердило эффективность разнотипных выборок.

На следующем этапе исследования проблема хрупкости рассмотрена применительно к распознаванию образов букв произвольного типа русского алфавита, для которого по сравнению с английским более значимыми факторами являются как высокое отличие образов одного класса, так и весьма низкое отличие образов разных классов. Учитывая это, при перегруппировке множества образов сформированы 33 заглавных и 26 строчных классов букв с различным начертанием, что имеет существенное значение для обучаемости нейросетей: его продолжительности и возможности достижения приемлемого локального минимума функции ошибки из-за ограниченной емкости архитектуры LeNet-5.

Ввиду отсутствия общедоступных баз текстовых объектов русского языка с помощью специально разработанных форм получено 250 образцов почерка студентов БрГТУ. Номинальный объем собранной базы составил: цифры – по 750, заглавные буквы – по 2000, строчные – по 1000 образов каждого класса, итого – 106500. В тренировочных и тестовых выборках также применялись шрифтовые (ОС Windows) и синтезированные образы, обеспечивалось равное

представительство классов образов и вариантов их различного начертания. Отдельные нейросетевые комитеты для заглавных и строчных букв строились с помощью предложенной методики синтеза. При этом, т. к. в некоторых классах объединены несколько исходных классов, комитеты дополнены «составными» нейросетями, базовой моделью для которых являлась архитектура текстового детектора (количество выходных нейронов соответствовало числу разделяемых классов). В результате были построены 13 нейросетей для следующих классов: 'б/д', 'г'/'ч', 'ґ/з' 'е'/'ё', 'ш'/'щ', 'ъ'/'ь', 'Г'/'Т', 'Е'/'Ё', 'З'/'Э', 'И'/'Й', 'Л'/'П', 'Ш'/'Щ', 'Ь'/'Ъ'.

В таблице 3 приведена точность созданных комитетов (COM1_RUS_big и COM1_RUS_lit), которые содержат навигационный подкомитет многоклассовых нейросетей и множество составных нейросетей, используемых при выборе подкомитетом метки составного класса. Тестирование комитетов показало, что большинство ошибок допущено при распознавании близких по начертанию классов: 'А'/'Д', 'У'/'Ч'/'Ц', 'И'/'Н'/'М', 'в'/'ь'/'ъ' и 'т'/'ж'/'ш'. Их разделение можно выполнить с помощью *нейросетей-корректоров*, аналогичных составным, но определяемых на этапе эксплуатации комитета, что повышает его адаптивность. В итоге точность классификаторов COM2_RUS_big и COM2_RUS_lit существенно превысила уровень ведущего коммерческого аналога – ABBYY FlexiCapture 10.

Таблица 3. – Точность (в %) созданных (COM) и коммерческого (ABBY) классификаторов разнотипных образов букв полного алфавита русского языка

Классификатор	**Rus_big_font**	**Rus_big_cell**	**Rus_big_gen**	**Rus_big_word**
COM1_RUS_big	99,07	97,56	96,51	95,85
COM2_RUS_big	**99,26**	**98,11**	**97,64**	**96,37**
ABBYY FlexiCapture 10	94,81	89,08	88,75	75,21
Классификатор	**Rus_lit_font**	**Rus_lit_cell**	**Rus_lit_gen**	**Rus_lit_word**
COM1_RUS_lit	99,10	92,29	93,72	91,03
COM2_RUS_lit	**99,30**	**95,25**	**94,95**	**92,75**
ABBYY FlexiCapture 10	87,62	82,05	79,38	77,34

Обобщение полученных результатов позволило определить *технологию создания классификатора разнотипных текстовых образов*:

1) *Систематизация множества образов*:

- выделение существенных для рассматриваемой задачи типов образов;
- объединение классов образов с низким межгрупповым отличием;
- увеличение количества классов до числа их различных начертаний.

2) *Создание тренировочных и тестовых выборок*:

- обеспечение равного представительства классов и типов образов;
- контроль минимального уровня отличия образов в каждом классе;
- расширение выборок до требуемого объема с помощью процедуры генерации образов на базе пространственных и волновых искажений.

3) *Формирование трехуровневого комитета сверточных нейросетей*:

- построение с помощью методики синтеза навигационного подкомитета многоклассовых нейросетей, определяющего начальный класс входного образа $\mathbf{x}$: $COM\_CNN = (CNN_i, S_i, A, vote)$, где CNN_i – нейросеть, A – алфавит, $S_i = h_i \times w_i$ – масштаб $\mathbf{x}$, *vote* – схема голосования, $y = COM\_CNN(\mathbf{x})$ – метка класса $\mathbf{x}$;

- создание множества составных нейросетей для классов, объединяющих несколько исходных классов алфавита A: $CNN\_set1 = \{(CNN_i, A_i)\}$, где CNN_i – составная нейросеть, A_i – ее алфавит, $z = CNN_m(\mathbf{x})$ – метка класса $\mathbf{x}$ по решению m-й нейросети, при этом $(y \in A_m) \wedge (y \notin A_k \,|\, \forall k \neq m)$;

- определение типичных перекрестных ошибок и создание множества корректирующих нейросетей: $CNN\_set2 = \{(CNN_i, A_i)\}$, где CNN_i – нейросеть-корректор $t_n = CNN_p(\mathbf{x})$ – метка класса $\mathbf{x}$ по решению p-й нейросети, при этом $t_0 = z$, $(t_{n-1} \in A_p) \wedge (t_n \notin A_m \,|\, \forall m \neq p)$, где n – номер этапа коррекции.

Далее исследовалась задача создания классификатора образов букв полного алфавита русского языка, который содержал 39 классов: образы, одинаковые по начертанию в обоих регистрах (например, 'ж'/'Ж') объединены в один класс, также созданы составные классы схожих букв: 'П'/'п'/'Л'/'л', 'ш'/'щ'/'Ш'/'Щ', 'в'/'В'. Средняя точность сформированного по разработанной технологии классификатора на заглавных и строчных выборках составила 97,31 % и 94,95 %. Это *превысило уровень ведущего коммерческого аналога* (ABBYY FlexiCapture 10) на 8,5 % и 15,1 % соответственно, подтвердив бóльшую универсальность классификатора, что наряду с его адаптивностью (возможностью добавления членов на каждом уровне) подтверждает перспективность практического применения технологии.

В ***четвертой главе*** диссертации описаны прикладные разработки, основанные на представленных в предыдущих главах методиках и алгоритмах.

Создан *модуль детектирования текста* на изображениях с произвольной композицией. Применение модуля для автоматической фиксации автотранспорта показало, что его эффективность (полнота локализации номеров – 0,96, точность – 0,65) выше уровня аналога на основе каскад Хаара (0,84 и 0,46 соответственно). Показаны варианты снижения ресурсоемкости детектирования, основанные на контекстной информации. Собрана представительная база текстовых образов, применимая для создания детекторов, специализированных к различным сценам.

Разработана *система структуризации документов*, позволяющая решать задачи обработки изображений с однородным расположением текста, включая локализацию, сегментацию на различных уровнях иерархии и распознавание. Описаны реализация системы и пути повышения ее производительности для платформы Microsoft .NET, C# 5.0. Тестирование на изображениях белорусских паспортов показало, что качество локализации и сегментации текста созданной системой выше, чем известным аналогом (ABBYY PassportReader SDK 1.5), при этом точность распознавания образов одинаково высокого уровня ($\approx 99\,\%$).

Разработана *модель системы распознавания форм*, предназначенной для автоматизации ввода и анализа текстовых данных форм, заполняемых вручную респондентами. Универсальные контекстные классы и настраиваемые шаблоны форм позволяют применять систему для обработки информации различной предметной области. Тестирование системы на выборке форм «Индивидуальная карта студента» показало высокую точность распознавания образов цифр и букв русского алфавита на уровне 99,60 % и 97,98 % соответственно.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. С учетом структурных и яркостных свойств текста модифицирован метод контуризации Canny: упрощены требования, предъявляемые к свойствам отдельных контурных точек (в частности, локальная экстремальность заменена достаточной контрастностью модуля градиента) и расширены требования к свойствам последовательностей (прослеживаемость контура из любой точки), что позволило увеличить целостность формируемых контуров объектов. Создана методика адаптивной бинаризации изображения текстового блока, в отличие от аналогичных объединяющая глобальный и локальный подходы к пороговому ограничению (методы Otsu и Niblack). Процедура исключения фона повысила качество разделения классов (текста и фона) при их яркостной неоднородности, а «интегральные» изображения существенно снизили ресурсоемкость [7, 8, 11].

2. Создана модель текстового детектора в виде неглубокой сверточной нейросети, а также способа ее применения, основанного на мультимасштабном фрагментировании изображения с последующим анализом откликов нейросети. Модель в отличие от аналогичных менее требовательна к точности детектора, снижает ресурсоемкость применения детектора (как и аналогичных) более чем на два порядка, позволяет обнаруживать на изображениях реальных сцен текстовые блоки с ориентацией, отличающейся от горизонтальной, за счет совместной обработки нейросетевых откликов в соседних строках и близких масштабах изображения, а также сегментировать блоки на строки и слова [4, 9].

3. Разработан алгоритм сегментации изображения слова, основанный на его специализированной предобработке и анализе характеристик как отдельных секущих, отделяющих текстовый образ, так и их серий. Адаптивность алгоритма повышена за счет использования, помимо вертикальных, наклонных секущих, проходящих через участки существенной кривизны внешнего контура слова, при этом корректность отсечения оценивается текстовыми классификаторами, способствующими выбору оптимального варианта сегментации. Тестирование показало, что алгоритм применим для обработки изображений слов с частично

разрушенными и слитыми символами, при этом оценка качества их сегментации алгоритмом превысила уровень ведущего коммерческого аналога на 9 % [6, 15].

4. Создана методика синтеза комитета сверточных нейросетей, обучаемых на множествах образов различного (фиксированного для каждой нейросети) масштаба с последующей селекцией членов комитета, которая в отличие от аналогичных не требует высокопроизводительных вычислений для синтеза и практического применения комитета. Эффективность методики подтверждена созданием комитетов с точностью распознавания баз рукописных образов цифр MNIST (99,65 %) и заглавных букв английского алфавита NIST (98,17 %) соответствующей уровню лучших мировых результатов [1, 2, 3, 10, 13].

5. Разработана технология создания классификатора разнотипных текстовых образов различного алфавита в виде нейросетевого комитета с тремя уровнями: отнесения входного образа к группе схожих по начертанию; выбора класса в группе; коррекции итогового решения с учетом оценки уверенности распознавания и типовых ошибок, выявляемых при практическом применении классификатора. С помощью технологии созданы классификаторы образов цифр, букв английского и русского алфавитов, универсальность (перекрестная точность более 95 %) которых превышает уровень ведущих коммерческих аналогов [5, 12, 14].

6. Разработаны шаблон текстовой формы и алгоритм ее сегментации, которые позволили собрать представительную базу изображений маркированных текстовых объектов русского языка номинальным объемом 106500 образов, являющуюся эффективным ресурсом для сравнения методов обработки текста [5].

Рекомендации по практическому использованию результатов

Результаты диссертационного исследования могут использоваться для создания систем структуризации текстовых данных изображений документов, детектирования текста на изображениях с произвольной композицией объектов, в интеллектуальных комплексах, основанных на универсальном распознавании текстовых образов, системах электронного документооборота и потокового ввода данных. Их положительный эффект достигается за счет автоматизации труда операторов, повышения точности и скорости обработки данных. Кроме того, разработки могут применяться в приложениях «дополненной реальности»: повышение восприятия текстовой информации при низком уровне зрения, недостаточном знании иностранных языков и др., а также в исследованиях по машинному зрению, не связанных с анализом текстовых объектов.

Внедрение результатов диссертации в разрабатываемые в ООО «ГЕРСИС СОФТВЕР» программные системы показало повышение качества обработки изображений документов (точности выделения и распознавания текстовых образов). Результаты диссертации также внедрены в учебный процесс БрГТУ. Внедрения подтверждены актами, представленными в приложении диссертации.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ УЧЕНОЙ СТЕПЕНИ

Статьи в рецензируемых научных журналах

1. Кузьмицкий, Н.Н. Сверточная нейросетевая модель в задаче классификации изображений изолированных цифр / Н.Н. Кузьмицкий // Доклады БГУИР. – 2012. – № 7 (69). – С. 64–70.

2. Кузьмицкий, Н.Н. Актуальные вопросы использования сверточных нейронных сетей и их комитетов в распознавании образов цифр / Н.Н. Кузьмицкий // Вестник БрГТУ. – 2012. – № 5 : Физика, математика, информатика. – С. 6–10.

3. Кузьмицкий, Н.Н. К вопросу оценки эффективности применения технологии CAPTCHA для защиты сетевых ресурсов / Н.Н. Кузьмицкий, С.С. Дереченник, Д.А. Костюк // ЭЛЕКТРОНИКА инфо. – 2013. – № 6. – С. 151–153.

4. Кузьмицкий, Н.Н. Обнаружение фрагментов текста на изображениях реальных сцен на базе сверточной нейросетевой модели / Н.Н. Кузьмицкий // ИНФОРМАТИКА. – 2015. – № 2 (46). – С. 12–21.

5. Кузьмицкий, Н.Н. Построение универсальных классификаторов текстовых образов русского языка на базе сверточных нейросетей / Н.Н. Кузьмицкий // Доклады БГУИР. – 2015. – № 4 (90). – С. 33–39.

Статьи в научных сборниках

6. Кузьмицкий, Н.Н. Сегментация искаженного растрового изображения алфавитно-цифровой последовательности / Н.Н. Кузьмицкий // Сборник конкурсных научных работ студентов и магистрантов : в 2 ч. / БрГТУ. – Брест, 2009. – Ч. 1. – С. 134–137.

7. Кузьмицкий, Н.Н. Контурная сегментация цифровых изображений документов / Н.Н. Кузьмицкий // Сборник конкурсных научных работ студентов и магистрантов : в 2 ч. / БрГТУ. – Брест, 2010. – Ч. 1. – С. 278–282.

8. Кузьмицкий, Н.Н. Адаптивная бинаризация цифровых изображений текстовых блоков / Н.Н. Кузьмицкий // Сборник конкурсных научных работ студентов и магистрантов : в 2 ч. / БрГТУ. – Брест, 2010. – Ч. 1. – С. 283–286.

Статьи в сборниках материалов научных конференций

9. Кузьмицкий, Н.Н. Организация эффективной обработки цифровых изображений средствами .NET-платформы / Н.Н. Кузьмицкий // Современные проблемы математики и вычислительной техники : материалы VI Республ. науч. конф. молодых ученых и студентов : в 2 ч., Брест, 26–28 ноября 2009 г. / БрГТУ. – Брест, 2009. – Ч. 1. – С. 48–49.

10. Кузьмицкий, Н.Н. Выбор признаков и модели классификатора в задаче распознавания изображений реальных сцен / Н.Н. Кузьмицкий // Современные проблемы математики и вычислительной техники : материалы VII Респ. науч. конф. молодых ученых и студентов : в 2 ч., Брест, 24–26 ноября 2011 г. / БрГТУ. – Брест, 2011. – Ч. 1. – С. 14–17.

11. Кузьмицкий, Н.Н. Построение целостных контуров объектов на полутоновых изображениях / Н.Н. Кузьмицкий, С.С. Дереченник // Информационные технологии и системы : материалы междунар. науч. конф., Минск, 26 октября 2011 г. / БГУИР. – Минск, 2011. – С. 175–176.

12. Kuzmitsky, N. The problem of "brittleness" of convolutional neural networks in recognition of digit patterns with different writing styles / N. Kuzmitsky, S. Derechennik // Neural Networks and Artificial Intelligence : proc. of the 7th Int. Conf., Minsk, 10–12 October 2012 / BSUIR. – Minsk, 2012. – P. 235–238.

13. Кузьмицкий, Н.Н. Об одном значимом результате в распознавании образов рукописных цифр / Н.Н. Кузьмицкий // Информационные технологии и системы : материалы междунар. науч. конф., Минск, 24 октября 2012 г. / БГУИР. – Минск, 2012. – С. 228–229.

14. Кузьмицкий, Н.Н. Создание универсальных классификаторов текстовых образов на основе сверточных нейросетевых технологий / Н.Н. Кузьмицкий // ГрафиКон’2013 : материалы междунар. конф. по компьютерной графике и зрению, Владивосток, 16–20 сентября 2013 г. / ДФУ. – Владивосток, 2013. – С. 234–237.

Тезисы докладов на научных конференциях и семинарах

15. Кузьмицкий, Н.Н. Обнаружение и сегментация текстовой информации растровых изображений / Н.Н. Кузьмицкий // Телекоммуникации: сети и технологии, алгебраическое кодирование и безопасность данных : тез. докл. междунар. науч.-техн. семинара, Браслав, 20–24 сентября 2010 г. / БГУИР. – Минск, 2010. – С. 32.

РЭЗЮМЭ

Кузьміцкі Мікалай Мікалаевіч

Дэтэктыраванне і распазнаванне рознатыповых тэкставых аб'ектаў на відарысах адвольных сцэн сродкамі згортваючых нейрасетак

Ключавыя словы: відарыс, рэальная сцэна, згортваючая нейрасетка, тэкставы аб'ект, дэтэктыраванне, распазнаванне, класіфікацыя, сегментацыя, бінарызацыя, контур, камітэт, селекцыя, універсальнасць.

Мэта працы: распрацоўка нейрасеткавых методык і адаптыўных алгарытмаў дэтэктыравання і распазнавання тэкставых аб'ектаў растравых відарысаў з адвольнай кампазіцыяй сцэны, рознымі яркаснымі ўласцівасцямі і спосабамі сінтэзу тэкставых выяў.

Метады даследвання: метады апрацоўкі растравых відарысаў, аналізу тэкставых дадзеных, машыннага навучання (штучныя нейрасеткі).

Атрыманыя вынікі і іх навізна: распрацавана мадыфікаваная версія метаду Canny, якая забяспечвае большы ўзровень цэласнасці контураў за кошт спрашчэння крытэрыяў контурнай кропкі. Створана методыка адаптыўнай бінарызацыі відарыса тэкставага блока шляхам аб'яднання метадаў Otsu і Niblack, выключэння фону, аптымізацыі вылічэнняў на аснове «інтэгральных» відарысаў. Распрацавана мадэль тэкставага дэтэктара ў выглядзе згортваючай нейрасеткі неглыбокай архітэктуры, з дапамогай якой можна лакалізаваць тэкставыя блокі на відарысах рэальных сцэн, а таксама сегментаваць іх на радкі і словы. Створана методыка сінтэзу камітэта згортваючых нейрасетак, абучаных на выявах розных памераў з наступнай селекцыяй членаў, без прымянення высокапрадукцыйных вылічэнняў, эфектыўнасць якой пацверджана стварэннем камітэтаў, якія валодаюць адпаведнымі лепшым вынікамі распазнання рукапісных выяў MNIST (дакладнасць – 99,65 %) і NIST (дакладнасць – 98,17 %). Распрацавана тэхналогія стварэння класіфікатара рознатыповых тэкставых выяў адвольнага алфавіта ў выглядзе трохузроўневага камітэта згортваючых нейрасетак, абучаных на выбарках, сфарміраваных з улікам розных графічных уяўленняў класаў, міжгрупавога падабенства і тыпавых памылак распазнавання. Тэхналогія дазволіла стварыць класіфікатары выяў лічбаў, літар англійскага і рускага алфавітаў, універсальнасць (перакрыжаваная дакладнасць) якіх перавышае ўзровень вядучых аналагаў.

Рэкамендацыі па выкарыстанні: прымяненне распрацовак у складзе інтэлектуальных комплексаў аўтаматычнай апрацоўкі тэкставых дадзеных.

Вобласць ужывання: аблічбоўка тэксту растравых відарысаў дакументаў, струменевы ўвод дадзеных і аналіз відарысаў рэальных сцэн.

РЕЗЮМЕ

Кузьмицкий Николай Николаевич

Детектирование и распознавание разнотипных текстовых объектов на изображениях произвольных сцен средствами сверточных нейросетей

Ключевые слова: изображение, реальная сцена, сверточная нейросеть, текстовый объект, детектирование, распознавание, классификация, сегментация, бинаризация, контур, комитет, селекция, универсальность.

Цель работы: разработка нейросетевых методик и адаптивных алгоритмов детектирования и распознавания текстовых объектов растровых изображений с произвольной композицией сцены, различными яркостными свойствами и способами синтеза текстовых образов.

Методы исследования: методы обработки растровых изображений, анализа текстовых данных, машинного обучения (искусственные нейросети).

Полученные результаты и их новизна: разработана модифицированная версия метода Canny, обеспечивающая бóльший уровень целостности контуров, за счет упрощения критериев контурной точки. Создана методика адаптивной бинаризации изображения текстового блока путем объединения методов Otsu и Niblack, исключения фона, оптимизации вычислений на основе «интегральных» изображений. Разработана модель текстового детектора в виде сверточной нейросети неглубокой архитектуры, с помощью которой можно локализовать текстовые блоки на изображениях реальных сцен, а также сегментировать их на строки и слова. Создана методика синтеза комитета сверточных нейросетей, обученных на образах различных размеров с последующей селекцией членов без применения высокопроизводительных вычислений, эффективность которой подтверждена созданием комитетов, обладающих соответствующими лучшим результатами распознавания рукописных образов MNIST (точность – 99,65 %) и NIST (точность – 98,17 %). Разработана технология создания классификатора разнотипных текстовых образов различного алфавита в виде трехуровневого комитета сверточных нейросетей, обученных на выборках, сформированных с учетом разных графических представлений классов, межгруппового подобия и типовых ошибок распознавания. Технология позволила создать классификаторы образов цифр, букв английского и русского алфавитов, универсальность (перекрестная точность) которых превышает уровень ведущих аналогов.

Рекомендации по использованию: применение разработок в составе интеллектуальных комплексов автоматической обработки текстовых данных.

Область применения: оцифровка текста растровых изображений документов, потоковый ввод данных и анализ изображений реальных сцен.

SUMMARY

Kuzmitsky Nikolay Nikolaevich

Detection and recognition of different types text objects on images of arbitrary scenes by means of convolutional neural networks

Keywords: image, real scene, convolutional neural network, text object, detection, recognition, classification, segmentation, binarization, contour, committee, selection, universality.

Aim of the work: development of neural-network techniques and adaptive algorithms for text objects detection and recognition on raster images with arbitrary scene composition, different brightness properties and methods for text patterns synthesizing.

Research methods: raster image processing methods, text data analysis, machine learning (artificial neural networks).

Obtained results and their novelty: The modified version of Canny method is developed, providing a higher level of contours integrity by simplifying contour point criterion. The technique for adaptive binarization of text block image is developed by combining the methods of Otsu and Niblack, background exception with calculations optimization based on «integral» images. The model of text detector in form of convolution neural network with non-deep architecture is created, which can be used to localize text blocks in real scene images and to segment them into words and lines. The technique is created for synthesis of committee of convolution neural networks trained on images of different sizes, followed by members selection, without use of high performance computing, efficiency of which is confirmed by creation of the committees with relevant best results of MNIST (accuracy – 99,65 %) and NIST (accuracy – 98,17 %) handwriting patterns recognition. The technology for creating classifier of different type text patterns of various alphabet is developed in form a three-level committee of convolution neural networks, trained on sets of patterns which are formed related to different graphic representations of classes, intergroup similarity and recognition errors. Technology has allowed to create the classifiers of digit and letter patterns of the English and Russian alphabets with universality (cross accuracy) exceeding level of leading analogues.

Recommendations for use: using the results as a part of intelligent systems for automatic text data processing.

Application area: digitizing text of raster document images, data capture and analysis of real scene images.

Научное издание

Кузьмицкий Николай Николаевич

ДЕТЕКТИРОВАНИЕ И РАСПОЗНАВАНИЕ РАЗНОТИПНЫХ ТЕКСТОВЫХ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ ПРОИЗВОЛЬНЫХ СЦЕН СРЕДСТВАМИ СВЕРТОЧНЫХ НЕЙРОСЕТЕЙ

АВТОРЕФЕРАТ
диссертации на соискание ученой степени кандидата технических наук

по специальности 05.13.17 – Теоретические основы информатики

Подписано в печать 12.09.2016. Формат 60x84 1/16. Бумага офсетная. Гарнитура «Таймс».
Отпечатано на ризографе. Усл. печ. л. 1,63. Уч. -изд. л. 1,4. Тираж 60 экз. Заказ 269.

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 1/238 от 24.03.2014,
№ 2/113 от 07.04.2014, № 3/615 от 07.04.2014.
ЛП № 02330/264 от 14.04.2014.
220013, Минск, П. Бровки, 6.