

ОТЗЫВ

члена диссертационного совета Гавриловой Татьяны Альбертовны на диссертацию

Алексея Владимировича Самарина

на тему:

«Комбинированные нейросетевые модели для классификации специфичных изображений»,

представленную на соискание учёной степени кандидата физико-математических наук по научной специальности 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей

На сегодняшний день компьютерное зрение является одной из ключевых областей анализа данных. Совместно с глубокими нейросетями и новейшими методами вычислений компьютерное зрение расширяет научные и практические горизонты в сферах здравоохранения, машиностроения, логистики, маркетинга, безопасности и пр., оказываясь одним из основных составляющих современных технологий и внося значительный вклад в развитие общества.

Задача классификации изображений, полученных фотосъемкой реальных объектов физической реальности оказывается одной из важнейших задач компьютерного зрения. Ею занимались значительное время и на настоящий момент задача классификация изображений общего плана, фактически, решена.

Вместе с тем, классификация специфичных изображений – в составе различных сервисов и бизнес-процессов, для изображений плохого качества, содержащих текст и т.д. – продолжает оставаться открытой: универсальные решения показывают здесь неудовлетворительные результаты. Соответственно, востребованы новые решения, ориентирующиеся на определённые параметры классифицируемых изображений, а также на тот сервис/бизнес-процесс, который будет использовать результаты классификации.

Диссертация А.В. Самарина нацелена на классификацию двух классов специальных изображений. Первый класс является фотографиями документов, содержит контурную информацию и разбит на фрагменты, наличие которых и их взаимное расположение оказываются важными при классификации. Кроме того, изображения этого класса могут содержать различные дефекты, а также допустимые изменения пользователя. Наконец, изображения этого вида хоть и содержат текст, но его распознавание не требуется – классификация изображений используется в облегчённом режиме с тем, чтобы облегчить работу людям, которые принимают окончательное решение. Второй класс изображений не имеет

контурной информации, также содержит текст, который требуется распознавать, и также содержит дефекты съемки, но, дополнительно к предыдущему случаю, имеет особенности съемки – углы съемки, выравнивание объектов в кадре и т.д. Таким образом, в отличии от первого класса изображения данного класса создаются в более сложной обстановке. В итоге в диссертации получены следующие новые результаты.

1. Комбинированная нейросетевая архитектура для классификации изображений с текстом, которые не нуждаются в явном распознавании текстовой информации.
2. Комбинированная нейросетевая архитектура для классификации класса изображений с текстом, нуждающихся в явном распознавании текста. Предложена идея не выполнять полное распознавание текста, а использовать графические характеристики для классификационных признаков. В рамках данной архитектуры предложены специальные дескрипторы текстовых фрагментов изображения, для которых доказано, что они имеют линейную сложность от размера изображения и гиперпараметров.
3. Экспериментальное исследование, продемонстрировавшее эффективность предложенных нейросетевых архитектур.

Диссертация состоит из введения, трёх глав и заключения, имеет объем 92 страницы, список литературы содержит 108 наименований. Имеется достаточное количество публикаций по теме диссертации, которые выполнены в авторитетных зарубежных изданиях по данной тематике, а также в российских журналах из перечня ВАК.

Сделаем краткий обзор диссертационной работы по главам.

Во **введении** описывается актуальность и степень исследованности задачи классификации изображений, дана постановка задачи исследования, перечислены основные результаты диссертации и объяснено их соответствие паспорту специальности 2.3.5, кратко изложена их научная новизна, а также теоретическая и практическая ценность диссертации.

В **первой главе** представлен обзор основных идей и понятий нейросетевых моделей и архитектур, рассмотрена ключевая для диссертации область – классификация изображений и детектирование текста.

Разработанная автором нейросетевая архитектура для классификации специфических фотографий документов представлена во **второй главе**. При этом автор делает акцент на «легковесной» классификации, которая не требует извлечения детальной информации из документов и предназначенная для

уменьшения «ручной» работы операторов. Такая классификация в большинстве случаев оказывается частью некоторого бизнес-процесса компании, например, сервиса по восстановлению учётной записи в системе «ВКонтакте». При этом рассматриваемые изображения могут содержать дефекты, пользовательские изменения и др. неизвестные заранее параметры. Наконец, в главе представлены эксперименты, показывающие эффективность предложенной архитектуры в контексте одной прикладной задачи.

В **третьей главе** автор рассматривает задачу классификации изображений с текстом, в которой важно распознавание текста. Особенности данной задачи являются вариативностью параметров камер съемки и условий съемки при создании изображений, отсутствие выпуклых элементов, разнообразие визуальных средств оформления текста, а также наличие визуальных дефектов. Предложенная архитектура состоит из двух частей: сверточной нейронной сети и модуля детектирования и распознавания текста (Optical Character Recognition). Интересной идеей автора является не полное, а частичное распознавание текста – использование его визуальных атрибутов (цвет фона, размер шрифтов и т.д.), на основе которых создаются признаки классификации. Такой подход оказывается эффективным и значительно сокращает трудозатратность обработки текста при классификации изображения. В главе представлены также результаты экспериментов для данной архитектуры в контексте практической задачи по классификации вывесок коммерческой недвижимости.

В **заключении** текста диссертации представлены основные результаты, а также кратко описаны направления дальнейших работ по этой теме.

Таким образом, **теоретическая ценность** работы заключается в создании двух нейросетевых архитектур, которые могут применяться к различным задачам классификации специфических изображений. При этом в диссертации детально описаны характеристики соответствующих классов изображений. Также важным теоретическим результатом является доказательство линейной зависимости от величины изображения времени работы алгоритмов построения дескрипторов для фрагментов изображений с текстом. Реализация предложенных архитектур для конкретных индустриальных задач является несомненной **практической ценностью** данной диссертации.

Между тем, к работе имеются следующие замечания.

1. Не смотря на обилие в тексте диссертации рисунков и схем, что в целом повышает понимаемость текста работы, диссертация существенно выиграла бы от более развитой визуальной культуры. Так, количество рисунков в главе 1 выглядит избыточным, они заимствованы из других источников и далеко не всегда действительно что-то поясняют. Напротив, в

- главах 2 и 3 недостаёт рисунков и диаграмм самого автора, которые бы наглядно поясняли те или иные детали предлагаемых архитектур.
2. В раздела 2.6 (стр. 47) не ясно, что это за 12 классов изображений, по которым классифицируются обучающий и тестовый наборы.
 3. В тексте работы присутствует некоторое число опечаток, например, «архитекутры» (стр. 2, 13), «распознавания» (стр. 7), «дальнейшей» (стр. 11), «архитекткры» (стр. 12), «предназначенной» (стр. 190, «вслествии» (стр. 30), «улучшения» (стр. 75).

В заключении можно сделать вывод о том, что диссертация А.В.Самарина на тему «Комбинированные нейросетевые модели для классификации специфичных изображений» соответствует требованиям, установленным Приказом от 19.11.2021 № 11181/1 «О порядке присуждения научных степеней в Санкт-Петербургском государственном университете», а сам соискатель Алексей Владимирович Самарин заслуживает присуждения ему учёной степени кандидата физико-математических наук по специальности 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей. Нарушения пунктов 9 и 11 указанного Порядка в диссертации не обнаружены.

19.02.2024

Член диссертационного совета,



Гаврилова Татьяна Альбертовна

Доктор технических наук,
профессор кафедры информационных технологий в менеджменте

Высшая школа менеджмента
Санкт-Петербургский государственный Университет
Волховский пер.3, 199004, СПб,

gavrilova@gsom.spbu.ru

<https://gsom.spbu.ru/about-gsom/faculty/gavrilova/>