

ОТЗЫВ

члена диссертационного совета Николая Владимировича Кузнецов на диссертационную работу Алексея Владимировича Самарина на тему «Комбинированные нейросетевые модели для классификации специфичных изображений», представленную на соискание учёной степени кандидата физико-математических наук по специальности 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей

На сегодняшний день область компьютерного зрения представляет собой широкое поле для исследований и оказывает существенное воздействие на развитие анализа данных. Следует признать, что эта область находится на пике развития, и в то же время её научный потенциал по-прежнему остается значительным. Крупнейшим разделом компьютерного зрения является область классификации изображений. Современные методы достигли здесь впечатляющих успехов, однако, имеются многочисленные задачи, где общие методы не способны обеспечить необходимую точность и достоверность результатов. К таким задачам относится анализ изображений с текстом, которые получены в нестандартных/трудных условиях, например, при низкой освещенности, что снижает качество и точность распознавания. Следует отметить, что отдельные классы таких задач могут иметь разные требования по качеству и точности распознавания.

В своей диссертационной работе А. В. Самарин предлагает решения для двух классов задач классификации изображений с текстом. Первый класс предполагает, что специальная обработка текста изображения невозможна или не нужна, в рамках второго класса задач требуется анализ текстовой информации, так как последняя оказывается определяющей при классификации. Во втором случае сложность задачи связана с тем, что полное семантическое распознавание текста является дорогостоящей операцией и, кроме того, часто затруднительно в виду нестандартных условий получения изображения, а также в виду широкого использования различных визуальных атрибутов текста – размера шрифтов, цветовой гаммы текстовой области, наличия художественных элементов в текстовой области изображения и т.д. Находкой диссертанта явилось предположение о том, что именно эти значимые визуальные эффекты текста можно эффективно использовать для классификации, опустив семантическое распознавание текста.

Для выделенных выше классов задач А. В. Самарин предложил комбинированные нейросетевые архитектуры, использующих глубокие сверточные нейронные сети. Эти архитектуры, а также созданные в их рамках методы анализа изображений, а именно, предварительный анализ изображения, использование автокодировщиков, построение дескрипторов текстовой области на основе параллельного обхода несколькими агентами, являющиеся **теоретической ценностью** данной диссертационной работы и обладают неоспоримой **научной новизной**, что подтверждено многочисленными публикациями в российских и зарубежных журналах.

В ходе работы были получены следующие результаты.

1. Комбинированная нейросетевая архитектура для классификации класса изображений с текстом, не нуждающимся в явном распознавании; изображения этого класса содержат явно выраженные контуры, могут иметь дефекты (блики, расфокусированные области и пр.), а также допустимые пользовательские изменения.
2. Комбинированная нейросетевая архитектура для классификации класса изображений с тестом, нуждающимся в явном распознавании; изображения получены в разных условиях съемки (с камер различных видов, с использованием различных углов съемки, с применением разного выравнивания объектов в кадре, с использованием разной цветовой схемы и т.д.), содержат дефекты (блики, засветы, шумы).
3. Предложены алгоритмы построения дескрипторов для архитектуры предыдущего класса, для которых доказано, что они имеют линейную сложность от размера изображения и гиперпараметров.
4. Экспериментальное исследование качества классификации разработанных архитектур по сравнению с современными методами классификации изображений общего плана в контексте рассматриваемых задач.

Сделаем краткий обзор диссертационной работы по главам.

Во введении обосновывается актуальность данных исследований, приведена постановка задачи и результаты, выносимые на защиту.

В первой главе дается обзор базовых нейросетевых архитектур, глубоких нейросетевых классификаторов изображений общего плана, задачи детектирование текста и методов оптического распознавания текста, а также комбинированных нейросетевых архитектур.

Во второй главе рассматривается класс задач по классификации изображений с текстом, являющихся фотографиями документов, которые имеют различные области, характеристики и порядок следования которых оказываются значимыми для классификации. Изображения могут содержать дефекты, допустимы пользовательские модификации, а также заранее неизвестны и могут сильно варьироваться ракурс съемки, освещенность и цветовой баланс изображения и некоторые другие параметры. Наконец, в рамках данного класса задач не требуется глубокий анализ изображений. Следует отметить, что данный класс задач определен автором весьма отчетливо. Далее в этой главе излагаются принципы построения и функционирования комбинированной нейросетевой архитектуры, предназначенной для решения данного класса задач, в том числе кодировщики для извлечения характеристических признаков и автокодировщики для извлечения информации о контурах. Также приводится пример задачи данного класса, использующий предложенную архитектуру. Наконец, для этой решенной задачи приводятся результаты экспериментов, в рамках которых разные методы классификации документов сравниваются с предложенной архитектурой. Эксперименты позволяют сделать вывод об эффективности предложенной нейросетевой архитектуры.

В третьей главе рассматривается класс задач по классификации изображений с текстом, характеризующихся вариативностью параметров камер съемки и условий съемки,

наличием визуальных дефектов, а также разнообразных визуальных средств оформления текста, и, наконец, отсутствием выпуклых элементов. Следует отметить тщательность, с которой автор выполнил описание класса данного класса задач. Существенными компонентами нейросетевой архитектуры, предлагаемой для данного класса задач, является метод предобработки изображений, использование OCR-модуля, а также алгоритмы построения комбинированных дескрипторов области с текстом. Для этих алгоритмов доказываются свойства, характеризующие их сложность. Представлено решение конкретной задачи из описанного класса, реализованное на основе предложенной архитектуры – классификация фотографий фасадов коммерческих строений с рекламными вывесками. Наконец, представлены результаты экспериментов, в которых предложенная архитектура сравнивается с существующими методами классификации изображений общего плана. Данные результаты свидетельствуют об эффективности предложенной архитектуры.

В заключении диссертации кратко суммированы результаты исследования, и приведены направления дальнейших исследовательских работ.

Следует отметить легкий и весьма доступный для понимания стиль изложения диссертации. Очень интересна обзорная глава, освещающая широкий спектр вопросов, связанных с классификацией изображений. Также является важным, что автор весьма строго определяет классы изображений, на которые ориентированы его методы, при этом он не впадает в излишний формализм, что является общепринятой тенденцией в современной науке о данных.

К представленной диссертации имеется ряд замечаний.

1. В названии п. 2.1 предлагаемая нейросетевая архитектура названа моделью, хотя на стр. 12 даны точные определения термина «модель» и «архитектура», из которых следует, что они не являются синонимами.
2. В тексте диссертации не объясняется мотивация выбора существующих методов для сравнения с архитектурой VCA (см. раздел 2.6.).
3. К сожалению, в тексте работы не приводятся примеры задач для каждого из рассмотренных классов в количестве большем, чем одна задача на класс. Формально говоря, возникает вопрос о невырожденности этих классов.
4. Ряд важных моментов исследования следовало описать более детально, например, создание обучающей выборки для автокодировщика архитектуры VCA.
5. В тексте работы содержится ряд опечаток, например, на стр. 55 следует писать $[0, \dots, N]$ вместо $[0..N]$ (как и в ряде аналогичных случаях на стр. 58); на стр. 63 слово “Длина” в середине предложения почему-то написано с заглавной буквы и т.д.

Тем не менее, указанные замечания не являются существенными и не влияют на корректность и научную новизну представленных результатов.

Итак, можно сделать вывод, что представленная к защите диссертационная работа заслуживает положительной оценки, соответствует основным требованиям, установленным Приказом от 19.11.2021 № 11181/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», а сам соискатель Самарин Алексей Владимирович заслуживает присуждения 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.

Член диссертационного совета,
доктор физико-математических наук,
член-корреспондент РАН, профессор,
заведующий кафедрой прикладной
кибернетики Санкт-Петербургского
государственного университета

Николай Владимирович Кузнецов

15.02.2024

