

## ОТЗЫВ

председателя диссертационного совета

Парилиной Елены Михайловны

на диссертационную работу Самарина Алексея Владимировича 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

Компьютерное зрение на настоящий момент является одним из самых бурно развивающихся направлений анализа данных. Однако, несмотря на высокую эффективность современных подходов обработки изображений общего плана, существует значительное количество областей, в которых современные решения, предназначенные для широкого класса задач, не обеспечивают необходимую эффективность. В связи с этим можно отметить анализ изображений, содержащих семантически значимую текстовую информацию наряду с семантически значимыми элементами фона и визуальными объектами переднего плана. Дополнительно сложностью анализа такого класса изображений часто оказываются сложные условия съемки – вариативные углы съемки, непостоянные условия освещения, подвижность устройства захвата кадра. Подобные обстоятельства создают на изображении блики, затенения, шумы, существенно затрудняющие анализ полученных изображений. Таким образом, **актуальность** данной области исследований не вызывает сомнений и подтверждена большим числом публикаций в высокорейтинговых российских и международных научных изданиях.

Диссертационная работа Самарина Алексея Владимировича посвящена решению задач анализа изображений в рамках приведенного выше контекста. В работе рассмотрены две частные задачи классификации изображений с визуальными элементами текста на сцене в сложных условиях съемки. Первая задача связана с условиями, в которых отдельный семантический анализ области с текстом невозможен или нецелесообразен. Вторая задача связана с противоположной ситуацией, когда отдельный семантический анализ области сцены с присутствием графических элементов текста часто носит определяющее значения для вынесения вердикта о принадлежности входного изображения к тому или иному классу. Автор предлагает разработанные им глубокие нейросетевые архитектуры для эффективного решения этих задач.

Работа имеет следующую **структуру**. Диссертация состоит из введения, 3 глав, заключения, списка литературы и приложения, где прилагаются акты о внедрении. Диссертация включает 92 страницы в версии на русском языке. Список литературы содержит 108 наименований. Приведу краткое содержание работы. *Первая глава* посвящена обзору предметной области, содержит описание современных базовых нейросетевых архитектур, нейросетевых классификаторов, а также введение в область оптического распознавания текста. *Вторая глава* посвящена описанию первой задачи, решаемой автором, а именно, распознаванию изображений без получения явного текстового изображения посредством создания нейросетевой структуры. Работа предложенной структуры демонстрируется на имеющихся данных, результаты приводятся в конце этой главы. *Третья глава* посвящена решению второй задачи, описанной выше, которая заключается в распознавании изображений с явным использованием текстовой информации. В этой главе, также как и во второй, дается подробное описание

предложенной нейросетевой архитектуры, включающей визуальные дескрипторы. Глава завершается описанием экспериментов и их сравнительным анализом. В заключении даются выводы по проделанной работе и подробно перечисляются полученные в работе результаты.

**Теоретическая ценность** представленной диссертационной работы заключается в предложенных автором новых комбинированных нейросетевых архитектурах, предоставляющих универсальный подход к задачам выделенных классов, а также эффективных методах извлечения специальных дескрипторов и предварительной обработки информации. Среди преимуществ разработанных методов по сравнению с существующими аналогами особенно выделяется низкая вычислительная сложность процедуры извлечения дескрипторов и универсальность предложенных способов предварительной обработки изображений.

**Практическая ценность** диссертационной работы А. В. Самарина состоит в программной реализации предложенных архитектур, а также их адаптации для двух конкретных представителей выделенных классов задач. В частности, первая архитектура позволила разработать подсистему предварительной массовой верификации изображений удостоверений личности при восстановлении доступа к социальной сети ВКонтакте (к работе приложены акты о внедрении).

**Научная новизна** диссертационной работы заключается в оригинальном решении задачи классификации с ограничениями на качество съемки, а также способами учета текстовой информации, имеющейся на изображении. Автором предложена нейросетевая архитектура, новизна которой заключается в совмещении нескольких дескрипторов изображений. Также автором предложен специальный метод предобработки изображений, основанный на модифицированной идее автокодировщика. Также в работе предложена нейросетевая архитектура с комбинаторной схемой, использующая дескрипторы текстовых областей изображений.

**Замечания к диссертационной работе.** Работа производит положительное впечатление, она имеет четкую структуру, все постановки задач и предложенные автором решения аккуратно описаны. Результаты работы предложенных технических решений продемонстрированы на реальных данных, проведен сравнительный анализ структур, показано преимущество использования тех или иных решений. В работе имеется достаточное число графического материала, что существенно улучшает описание поставленных задач и предложенных решений. Однако работа содержит ряд недостатков, сформулирую их в виде списка замечаний и вопросов.

1. На рис. 1.6 изображен график функции Leaky\_ReLU с параметром  $\alpha$ , тогда как в определении этой функции несколькими строками выше используется другой параметр  $\beta$ . Как мне кажется, это опечатка.
2. В начале второй главы приводится постановка задачи, где в качестве документа (для распознавания) используется паспорт. Можно ли использовать предложенную автором архитектуру для распознавания других документов, нужно ли будет что-то менять в предложенной архитектуре? Таким образом, возникает вопрос об области применения предложенной архитектуры.
3. На стр. 40 написано, что выбраны архитектуры с 16 и 19 слоями соответственно. Чем обусловлено именно такое количество слоев?
4. Непонятно, что такое  $t$  в последней выделенной формуле на стр. 56, и какие значения он принимает.

5. В работе иногда опускается описание используемых в формулах параметров. Например, в первом абзаце на стр. 62 не поясняется, что такое  $W$  и  $H$ . Хотя далее мной были найдены эти обозначения на стр. 63, где дается их описание.
6. Ниже на этой же стр. 62 говорится, что  $c$  определено константой, равной 0,1. Почему выбрано именно это значение?
7. На стр. 65 описывается алгоритм построения дескриптора, в том числе стратегия перемещения агентов, и используется термин «приоритетное направление», но нет описания того, что это означает.
8. Не считаю корректным использование фразы «все более оптимальные стратегии перемещения агентов» в данном контексте на стр. 73.
9. В работе имеется большое число опечаток, которые я не буду приводить в отзыве, но которые автор легко может найти, внимательно прочитав текст.

Однако эти замечания не являются определяющими и не влияют на представленные результаты и общее положительное впечатление от диссертационной работы. Следует отметить, что результаты диссертационной работы А.В. Самарина были доложены на многих российских и международных конференциях, опубликованы в журналах и трудах конференций, внедрены в деятельность компании ООО «В контакте».

**Заключение.** Диссертация Самарина Алексея Владимировича на тему: «Комбинированные нейросетевые модели для классификации специфичных изображений» соответствует основным требованиям, установленным Приказом от 19.11.2021 №11181/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», соискатель Самарин Алексей Владимирович заслуживает присуждения ученой 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 установлены.

Председатель диссертационного совета,  
доктор физико-математических наук, доцент,  
профессор Кафедры математической теории игр  
и статистических решений,  
Санкт-Петербургский государственный университет



Е.М. Парилина  
06.02.2024