

Отзыв

члена диссертационного совета Льва Владимировича Уткина на диссертацию

Алексея Владимировича Самарина

**«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**

Распознавание образов и классификация изображений является традиционной задачей искусственного интеллекта и компьютерного зрения, имея большое количество различных постановок. На сегодняшний день перспективным подходом к решению этой задачи являются глубокие свёрточные нейронные сети, которые позволили достичь существенного прогресса в классификации изображений общего плана. На сегодняшний день автоматическая классификация таких изображений оказывается более качественной, чем обычная, «человеческая». Вместе с тем имеется большое количество частных задач из разных областей науки и практики (медицины, биоинформатики, безопасности, документооборота, компьютерной криминалистики и т.д.), связанных с распознаванием и классификацией специфичных изображений и имеющих различные требования по эффективности. Всё это делает дальнейшие исследования в этой области актуальными и перспективными.

Данная диссертация посвящена классификации изображений с помощью комбинированных нейросетевых архитектур на основе глубоких свёрточных нейронных сетей. Рассматривались изображения двух видов. Первый вид

характеризуется наличием в изображении фрагмента с фотографией, которая должна быть распознана, изображение разбито на области, взаимное расположение которых влияет на классификацию, изображение имеет хорошо выраженные контуры, на изображении могут присутствовать дефекты, условия съёмки изображения могут значительно варьироваться, допустимы пользовательские изменения (например, скрытие на документе, удостоверяющем личность, персональной информации, помимо ФИО). Второй вид изображений характеризуется дополнительно наличием текстовой информации, влияющей на классификацию, но имеющей широко варьируемые атрибуты, значительно препятствующие оптическому распознаванию текста. Соответственно, для изображений первого типа, в целях выполнения классификации, не требовалось распознавание текста, а для изображений второго вида – требовалось.

В работе были получены следующие основные результаты.

1. Комбинированная нейросетевая архитектура VCA (VGG Combined with Autoencoders) для классификации изображений, не выполняющая явное распознавание текстовой информации. Предложенная архитектура была использована для решения задачи определения пригодности изображения документа, подтверждающего личности, при условии варьирования условия съёмки.
2. Комбинированная нейросетевая архитектура CCIT (Combined Classifier of Images with Text) для классификации изображений с явным распознаванием текстовой информации. При этом был разработан ряд дополнительных дескрипторов для извлечения из изображения деталей визуального оформления текста, что существенно улучшило качество классификации. Предложенные подходы использованы при решении задачи классификации фотографии фасадов коммерческих зданий по типу оказываемых услуг.
3. Экспериментальное исследование качества классификации разработанных архитектур по сравнению с современными методами классификации изображений общего плана в контексте рассматриваемых задач.

Теоретическая ценность работы заключается в создании двух нейросетевых архитектур, ориентированных на разные виды классифицируемых изображений, метод предварительной обработки изображений, полученных в нестандартных условиях (включая различные типы корректирующих преобразований), метод построения дескрипторов текстовых областей изображения путем параллельного обхода этих областей несколькими агентами с представлением трёх разных алгоритмов, имеющих линейную сложность относительно размеров изображений.

Практическая ценность работы заключается в программной реализации предложенных архитектур и методов, а также в проведении экспериментов с этими реализациями для сравнения с существующими аналогами. Эксперименты наглядно демонстрируют преимущество предложенных архитектур.

В работе имеется ряд **недостатков**.

1. Хотелось бы отметить, что в тексте работы имеется некоторый дисбаланс в ключе общее/конкретное. Например, в главе 1 подробно описываются азы нейросетей – нейроны, слои, функции активации и пр. И в то же время, порой, совсем не описываются важнейшие детали исследования, например, выбранные для обучения архитектуры VCA классы документов (стр. 47) – указано лишь их общее количество (двенадцать штук).
2. Автор порой весьма непоследователен в изложении имеющихся в его области результатов и обосновании выбранных им методов. Так в разделе 1.2 он начинает с описания архитектуры VGG, дальше критикует эту архитектуру, после этого описывает более современные – Inception, ResNet, MobileNets и др. – и не делает никаких заключений. Потом в разделе 2.2 он снова возвращается к этим архитектурам с частичным повторением их описания и приводит данные эксперимента, из которых следует, что архитектура Inception является лучшей (таблица 1). После этого идут весьма туманные слова о большом количестве ложноположительных срабатываний всех претендентов, кроме VGG и в итоговой архитектуре применяется именно VGG. Было бы целесообразно не возвращаться к одним и тем же технологиям в тексте

дважды и ставить именно такие эксперименты, которые демонстрировали бы финальный выбор автора.

3. В работе содержится ряд опечаток, например, слитно написано слово *рамках* (стр. 10), в ряде других мест вместо буквы *й* используется буква *и*, например: *перспективы дальнейшей работы* (стр. 11); слово *Длина* в середине предложения написано почему-то с большой буквы (стр. 6 3).

Тем не менее указанные замечания не влияют на положительную оценку работы. Таким образом, 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,
профессор, доктор технических наук, профессор
Высшей школы технологий искусственного интеллекта
Санкт-Петербургского политехнического университета Петра
Великого (СПбПУ)
Лев Владимирович Уткин

