

ОТЗЫВ

члена диссертационного совета Владимирова Андрея Георгиевича на диссертацию Самарина Алексея Владимировича на тему Комбинированные нейросетевые модели для классификации специфичных изображений, представленную на соискание ученой степени кандидата физико-математических наук по научной специальности 2.3.5 Математическое и программное обеспечение, вычислительных систем, комплексов и компьютерных сетей

Компьютерное зрение занимается анализом медиаданных, полученных с фото- и видеокамер, являясь на сегодняшний день одной из самых успешных областей искусственного интеллекта. Вероятно, самой распространённой и востребованной задачей компьютерного зрения является классификация изображений. Различные решения этой задачи находят применение в самых разных областях, от компьютерной криминалистики до сервисов по массовой идентификации личности и классификации архивных документов при оцифровке.

При этом особый интерес представляют задачи классификации специфичных изображений, в частности, содержащих различные дефекты съёмки. Более того, если такие изображения содержат текст, то возникают также проблемы эффективного распознавания этого текста для успешной классификации. Полное семантическое распознавание текста оказывается дорогой процедурой, что сильно замедляет работу итогового сервиса, часто оказывается трудноосуществимым из-за дефектов изображения и, в то же время, далеко не всегда требуется для успешного решения задачи классификации.

Диссертационная работа Самарина Алексея Владимировича нацелена на решение задачи классификации изображений, содержащих текст и полученных в сложных условиях – с фотокамер различного разрешения, при разном (и часто недостаточном) освещении. Рассматриваются изображения, имеющие блики, различный цветовой баланс, содержащие модификации пользователя и т.д. Автор выделяет два типа задач: (i) «легковесную» классификацию документов, содержащих различные особенности, и (ii) классификацию изображений, включающих текст с различными, заранее неизвестными, визуальными эффектами – экзотическими шрифтами, необычными цветами фона и размером букв, разнообразными стилями оформления и т.д. Для

решения первой задачи автор предлагает не использовать текст для классификации, при решении второй задачи – использовать только визуальные признаки текста.

Теоретическая значимость работы заключается в том, что оба вида задач рассмотрены в общем виде: для каждой из них сформулированы требования и предлагается обобщённая нейросетевая архитектура. Для первой задачи автор разработал методы коррекции цветовой гаммы и устойчивого извлечения контурной информации, ориентированные на использование глубоких нейросетевых кодировщиков. Для второй задачи, на основе нейросетевых подходов извлечения семантической информации из фона, автор предложил методы обработки и анализа изображений, включая специальные дескрипторы фрагментов изображений с текстом. Для алгоритмов построения дескрипторов в работе доказаны сложностные оценки – в отличие от существующих алгоритмов, имеющих квадратичную сложность, предложенные имеют линейную сложность.

Практической значимостью работы является реализация каждой из архитектур, а также их адаптация для двух конкретных задач – первичной верификации фотографий удостоверений личности для системы восстановления доступа к аккаунтам социальной сети и классификация фотографий фасадов коммерческих зданий для нужд прикладного маркетинга. Эффективность разработанных целевых нейросетевых моделей была проверена экспериментами. Линейная сложность алгоритмов вычисления дескрипторов фрагментов изображений с текстом важна с практической точки зрения в виду неуклонного возрастания разрешения сотовых телефонов и фотокамер, что, в свою очередь, влечёт за собой увеличение размеров, а также других численных характеристик изображений.

Представленная работа не свободна от некоторых недостатков.

1. В тексте диссертационной работы нейросетевые архитектуры, предложенные для класса задач, не всегда достаточно точно отделены от конкретных целевых моделей.

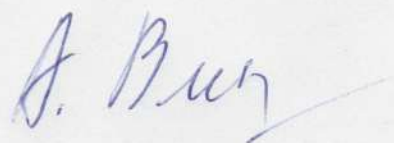
2. В тексте работы следовало бы более полно осветить мотивацию выбора метрик в экспериментах для второй архитектуры. При этом не ясно, почему не использовалась метрика Top1 Accuracy, являющаяся общепринятой для оценки качества классификации изображений общего плана.
3. В работе имеется ряд опечаток – стр. 11 («в приложении представлены акты»), стр. 12 (написано «архитекткры» вместо «архитектуры»), стр. 18 («то есть выходной значение») и некоторые другие.

Однако эти недостатки не носят принципиального характера и не оказывают существенного влияния на качество работы.

Таким образом следует заключить, что работа 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 научных степеней в Санкт-Петербургском государственном университете», а сам 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.

Член диссертационного совета

доктор физико-математических наук,
старший научный сотрудник
Института прикладного анализа и стохастических
процессов им. К. Вейерштрасса
(Германия, г. Берлин),



Владимиров Андрей
Георгиевич

29 января 2024 года