

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу **Шешкуса Александра Владимировича**

на тему «Использование преобразования Хафа в качестве слоя нейронной сети», представленную на соискание ученой степени кандидата технических наук по специальности 1.2.2 – «Математическое моделирование, численные методы и комплексы программ»

Актуальность темы исследования. В настоящее время применение искусственных нейронных сетей (ИНС) является магистральным направлением в области искусственного интеллекта и машинного обучения. Несмотря на тот факт, что ИНС были придуманы, исследованы и применены для анализа изображений ещё в прошлом веке, в настоящее время решение и внедрение задач, связанных с применением ИНС, переживает определенный бум. Это связано как с реальными достижениями в области создания архитектур ИНС, а также с совершенствованием методов обучения ИНС, так и с возрастающими возможностями вычислительной техники, позволяющей обучать все более сложные ИНС. Диссертационная работа А. В. Шешкуса нацелена на совершенствование ИНС, используемых в задачах компьютерного зрения. Разработанные математически модели и программные средства призваны улучшить решение нескольких задач компьютерного зрения в области анализа изображений документов, таких как поиск внутренних и внешних точек схода, а также оконтуривание произвольно ориентированных объектов. В работе А. В. Шешкуса показано, как высокие требования, предъявляемые к точности характеристик ИНС обеспечиваются в разработанных архитектурах при меньшем количестве обучаемых параметров и/или при требовании меньшего количества вычислительных мощностей по сравнению с известными архитектурами.

Рассмотрение содержания диссертационной работы. Диссертационная работа А. В. Шешкуса состоит из введения, четырех глав, заключения, списка литературы, списка сокращений и условных обозначений, списков рисунков и таблиц и приложений. Объем диссертационной работы составляет 123 страницы, в том числе 103 страницы основного текста, 9 таблиц и 38

иллюстраций. Список использованных источников из 95 наименований занимает 10 страниц, а приложения занимают 3 страницы.

Во введении определена проблематика и обоснована актуальность темы исследования, поставлена цель исследования, определены объект и предмет исследования, выбраны методы исследования, сформулирована научная новизна и практическая ценность результатов. Приведены основные положения, выносимые на защиту. Перечислены научные гранты, в процессе выполнения которых были получены научные результаты, а также приведены сведения о внедрении результатов работы и их апробации.

В первой главе вводятся основные понятия машинного обучения, используемые в исследовании, такие как интегральный оператор, карта признаков. Описывается преобразование Радона, преобразование Хафа (ПХ) и быстрое преобразование Хафа (БПХ). Приводятся основные методы создания ИНС, предназначенных для анализа изображений. В выводах отмечается, что большинство используемых методов основаны на одном интегральном преобразовании, а предлагаемый подход, базирующийся на ПХ, должен быть адекватен алгоритмам, использующим несколько интегральных преобразований. Цель работы детализируется поставленными задачами.

Во второй главе в разделе 2.1 вводится понятие рецептивного поля, исследуется его форма и размер в зависимости от количества применённых ПХ. Указывается, что такие формы рецептивных полей ИНС, использующих ПХ, как пучок прямых или набор полос, обеспечивают новые свойства ИНС, отличных от классических. В разделе 2.2 рассматривается прямое БПХ, приводятся формулы для преобразования прямолинейных сегментов в точки, в том числе для нормированных изображений. В разделе 2.3 показано, как с помощью диадического паттерна организовывать повторные вычисления, что определяет уменьшение сложности алгоритма БПХ. Рассмотрены квадратные изображения с размером 2^p и указана возможность аппроксимации изложенных алгоритмов для иных изображений. В последующих разделах рассмотрена возможность вычисления прямого и транспонированного БПХ с одной реализацией и с одной асимптотической точностью. Для этого приводится исследование, завершающееся доказательством теоремы.

Экспериментально показывается, что описанный Хаф-образ изображения пригоден для использования в качестве входа ИНС для анализа изображений.

В третьей главе рассматриваются две предложенные соискателем архитектуры ИНС: Хафнет и Хафэнкодер. Обе архитектуры используют слои прямого и транспонированного БПХ. Архитектура Хафнет предназначена для поиска внешних точек схода в изображениях трехмерных сцен или документов. Архитектура Хафнет реализуется с помощью двух последовательных применений БПХ. Свойством архитектуры Хафнет является возможность установления соответствия каждой точке плоскости, содержащей исходное изображение, некоторой точке карты признаков. Это преобразование плоскости и обратное преобразование выражаются аналитически. Архитектура Хафэнкодер предназначена для поиска внутренних точек схода и использует возможность пересчёта признаков из пространства образов Хафа в координаты изображения. Особенность Хафэнкодера состоит в том, что используется прямое и транспонированное ПХ. Отличием архитектуры Хафэнкодер от других сверточных ИНС типа автоэнкодер является видоизменить свое рецептивное поле без изменения количества обучающих параметров и учитывать интегральные признаки на изображении. В главе была рассмотрена вычислительная сложность реализации слоев Хафэнкодера и показано, что добавление слоя быстрого преобразования Хафа вносит небольшой вклад в сложность ИНС. Соискателем был проведен эксперимент, показавший, что можно в 2 раза уменьшить ошибку детектирования объектов одной природы при сохранении знаний, накопленных в последних слоях при обучении на выборке из объектов другой природы.

В четвертой главе описаны прикладные задачи, в решении которых используются полученные автором результаты. В разделе 4.2.1. описана задача (поиск внешних горизонтальной и вертикальной точек схода на изображениях оцифрованных камерами мобильных устройств документов), решенная с помощью архитектуры Хафнет. Для обучения и тестирования ИНС были использованы два опубликованных набора изображений: MIDV-500 и ICDAR 2011. Для обоих наборов с помощью найденных точек схода проводилось выправление формы документа. Качества оценивалось, в частности, с помощью последующего распознавания текстового содержимого.

Было показано улучшение почти в 20% по сравнению с выправлением известными методами. В разделе 4.2.2 рассмотрена задача семантической сегментации изображений документов. На опубликованных датасетах MIDV-500 и MIDV-2019 были протестированы сеть с архитектурой Хафнет, сеть с архитектурой Хафэнкодер и ИНС UNet. Было показано, что сеть с архитектурой Хафэнкодер уменьшает количество ошибок сети UNet на 20-30%, при этом количество обучаемых параметров сети Хафэнкодер было в 132 раза меньше количества обучаемых параметров сети UNet. В разделе 4.2.3 описаны результаты, полученные при решении задачи поиска машиночитаемой зоны. Для экспериментов использовался синтетический набор данных и сеть Хафэнкодер. Было достигнуто высокое качество сегментации по метрике mIOU, равный 98,1%.

Научная новизна и практическая значимость. Работа А. В. Шешкуса сочетает в себе результаты и подходы математического моделирования, вычислительной геометрии и современные инструменты машинного обучения. Основным элементом научной новизны является исследование трансформации рецептивного поля нейронной сети при добавлении в нее слоев прямого и транспонированного быстрого преобразования Хафа. Предложены новые архитектур нейронных сетей, слои которых базируются на прямом и транспонированном преобразовании Хафа. Показано, что при переносе знаний для доменной адаптации успешно происходит для высокоуровневых, а не низкоуровневых признаков. Разработан комплекс программ, который был внедрен в систему обучения нейронных сетей ООО «Смарт Энджинс Сервис». Это позволило внедрить предложенные архитектуры в решения для обработки изображений и распознавания документов, которые интегрированы в информационную инфраструктуру ряда коммерческих организаций, а также в ряд информационных решений государственных структур Российской Федерации.

Апробация результатов диссертации

Результаты диссертационной работы были доложены на 3 международных конференциях, опубликовано 8 работ, из них 7 – в изданиях, индексируемых в Web of Science и Scopus. Основные положения диссертации докладывались на научных семинарах ФГУ ФИЦ «Информатика и

управление» РАН и ФГБУН ИППИ РАН им. А. А. Харкевича. Имеются 3 акта о внедрении результатов диссертации.

Замечания по диссертационной работе

1) Одной из заявленных задач исследования является повышение эффективности нейросетей посредством уменьшения числа операций, при этом сравнение времени работы нейросетевых моделей не приводится.

2) В разделе 4.2.1 приводится сравнение предложенного решения с известными ранее алгоритмами, использующими преобразование Радона, но не представлено их описание.

3) Предложенные архитектуры нейронных сетей со слоями Хафа безусловно интересны, но возникает несколько вопросов, пояснения на которые отсутствуют в тексте, а именно:

- в архитектуре сверточной сети используются свертки размера 5×5 , как это влияет на устойчивость обучения?
- хафэнкодер сравнивается с сетью U-Net, не ясно, как на это сравнение влияют кросс-связи U-Net.

4) В разделе 3.6 описывается возможность переноса знаний в архитектурах со слоями Хафа, однако не вполне ясно позиционирование предлагаемого подхода по отношению к устоявшейся технологии transfer learning и доменной адаптации.

5) В тексте присутствуют технические погрешности, в частности: не все формулы в диссертации полностью описаны: для формулы 2.1 не объяснены величины k_i , R_i , s_i , a_i и f_i ; не все понятия вводятся последовательно, например, функция softsign используется начиная со страницы 65, а введена со страницы 105.

Указанные замечания носят по большей части технический характер и не влияют на общую положительную оценку работы.

Заключение

Диссертация А. В. Шешкуса «Использование преобразования Хафа в качестве слоя нейронной сети» является законченной научно-исследовательской работой, в которой содержится решение научной задачи разработки новых архитектур искусственных нейронных сетей для анализа изображений, имеющей значение для развития компьютерного зрения. Совокупность полученных лично диссертантом результатов можно квалифицировать как вклад в решение задачи анализа изображений с помощью искусственных нейронных сетей.

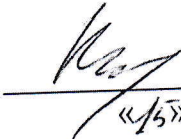
Диссертация написана автором самостоятельно, обладает внутренним единством, содержит новые научные результаты и положения, выдвигаемые для публичной защиты, и свидетельствует о личном вкладе автора диссертации в науку.

Предложенные автором решения аргументированы и оценены по сравнению с другими известными решениями.

Автореферат корректно отражает основное содержание диссертации.

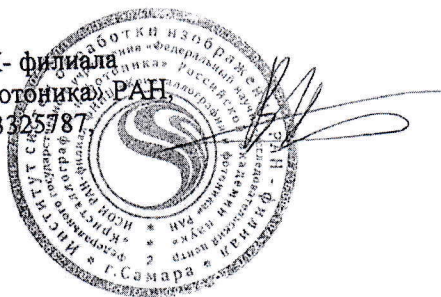
Диссертационная работа отвечает критериям «Положения о присуждении ученых степеней», а ее автор – Шешкус Александр Владимирович – заслуживает присуждения ему ученой степени кандидата технических наук по специальности 1.2.2 – Математическое моделирование, численные методы и комплексы программ.

Официальный оппонент,
заведующий лабораторией, ведущий
научный сотрудник
института систему обработки
изображений РАН (ИСОИ РАН) –
филиала федерального
государственного учреждения
«Федеральный научно-
исследовательский центр
«Кристаллография и фотоника»
Российской академии наук»,
доктор технических наук

 /Никоноров А.В.
«15» мая 2023 г.

Подпись Никонорова Артема Владимировича заверяю

Ученый секретарь ИСОИ РАН- филиала
ФНИЦ «Кристаллография и фотоника» РАН,
д.ф.-м.н., профессор, +7(846) 3325787,
kotlyar@ipsiras.ru



Котляр В.В.

Сведения об организации:

Институт систем обработки изображений РАН (ИСОИ РАН) – филиал
федерального государственного учреждения «Федеральный научно-
исследовательский центр «Кристаллография и фотоника» Российской
академии наук»

Адрес: 443001, ул. Молодогвардейская, 151

ipsi@ipsiras.ru

(846)332-57-83