

На правах рукописи

Шешкус Александр Владимирович

**Использование преобразования Хафа в качестве слоя
нейронной сети**

Специальность 1.2.2 —
«Математическое моделирование, численные методы и комплексы
программ»

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Москва — 2023

Работа выполнена в Федеральном государственном учреждении “Федеральный исследовательский центр “Информатика и управление” Российской академии наук” в отделе №92.

Научный руководитель: **Николаев Дмитрий Петрович**
кандидат физико-математических наук,
Федеральное государственное бюджетное
учреждение науки “Институт проблем пе-
редачи информации им. А.А. Харкевича
Российской академии наук”,
заместитель директора по научной работе

Официальные оппоненты: **Никоноров Артем Владимирович**
доктор технических наук, Институт систем
обработки изображений Российской академии
наук (ИСОИ РАН) , ведущий научный сотруд-
ник

Шоломов Дмитрий Львович
кандидат технических наук,
АО “Когнитив”,
руководитель отдела машинного обучения

Ведущая организация: Федеральное государственное учреждение
“Федеральный исследовательский центр
Институт прикладной математики им. М. В.
Келдыша Российской академии наук”

Защита состоится «06» июня 2023 г. в 13:00 часов на заседании диссер-
тационного совета 24.1.224.01 на базе Федерального исследовательского
центра «Информатика и управление» Российской академии наук (ФИЦ
ИУ РАН) по адресу: 117312, г. Москва, пр-т 60-летия Октября, 9.

С диссертацией можно ознакомиться в библиотеке ФИЦ ИУ РАН по адре-
су: Москва, ул. Вавилова, д. 40 и на официальном сайте ФИЦ ИУ РАН:
<http://www.frccsc.ru>.

Отзывы на автореферат в двух экземплярах, заверенные печатью учре-
ждения, просьба направлять по адресу: 119333, г. Москва, ул. Вавилова, д.
44, кор. 2, ученому секретарю диссертационного совета 24.1.224.01.

Автореферат разослан «__» 2023 года.
Телефон для справок: +7 (499) 135-51-64.

Ученый секретарь
диссертационного совета 24.1.224.01,
канд. физ.-мат. наук, доцент

И. В. Смирнов

Общая характеристика работы

Актуальность темы. Активное развитие области науки, связанной с нейронными сетями, привело к тому, что все большее количество задач решается с их использованием. При этом от создателей зачастую не требуется глубокого понимания внутренних принципов работы, что в совокупности с возросшими вычислительными мощностями персональных компьютеров и видеокарт сделало нейронные сети инструментом действительно массового пользования. В этих условиях критически важно проводить исследования, направленные на разработку методов создания оптимальных архитектур нейронных сетей и на понимание внутренней работы нейросетей. Первое важно, так как все большее количество алгоритмов, работающих на устройствах с все меньшими вычислительными мощностями и конечным запасом заряда батареи, используют нейронные сети. Второе необходимо для того, чтобы у исследователей и инженеров были механизмы построения архитектур, исходя из специфики задачи, что в свою очередь тоже позволит сократить вычислительные расходы. Рассматриваемые в работе задачи и методы их решения направлены на 1) создание оптимальных архитектур нейронных сетей для решения конкретных задач и 2) построения нейронных сетей с понимаемым принципом работы.

Искусственные нейронные сети появились в конце 1940-х годов. Тогда были разработаны первые модели, включая перцептрон Розенблатта. И хотя сверточные слои использовались и раньше, например в работах ЛеКуна еще в 1998-м году, самый большой за всю историю скачок их развития начался сравнительно недавно, в 2006-м году. Именно тогда начали выходить работы Хинтона, Салахутдинова и Крижевского, посвященные построению многослойных конструкций из сверточных слоев. Эти слои принципиально отличались от полносвязных тем, что могли учитывать лишь информацию в локальной окрестности на картах признаков в противовес полносвязным слоям, которые всегда учитывали всю входную карту признаков. Таким образом, у выходных нейронов сверточных слоев получается ограниченное рецептивное поле, что соответствует специфике задачи обработки изображений. Именно эта особенность подобных слоев и сыграла ключевую роль в их высокой эффективности в задачах обработки изображений.

Возросшие вычислительные мощности позволяют создавать большие нейронные сети, содержащие десятки и даже сотни миллионов обучаемых параметров, которые демонстрируют превосходное качество работы во многих задачах. Однако, в противовес этому, разработка нейросетевых архитектур и, особенно, новых типов слоев, которые позволили бы добиваться того же или сравнимого результата при значительно меньших вычислительных затратах – актуальная проблема. Например, введение слоев пулинга помогло увеличить обобщающую способность сверточных

нейронных сетей путем понижения размера карты признаков, а последующее введение слоев, которые способны тем или иным способом увеличить изображение, породило целое семейство нейросетей типа “автокодировщик”, которые теперь повсеместно используются в задачах обработки, сжатия и улучшения изображений.

Однако, так или иначе, проблема рецептивного поля, т.е. влияния каждого входного нейрона на финальный результат, остается и решать ее “локальными” средствами совсем не просто. Разумеется, если мы знаем о систематических связях нейронов, находящихся далеко друг от друга на изображении, мы можем объединять их в “нелокальные” группы, что и используется в некоторых конкретных задачах.

С другой стороны, в задачах обработки изображений широко используется интегральное преобразование Хафа. Преобразование Хафа - дискретизованный вариант преобразования Радона, где каждое значение на преобразованной плоскости есть сумма значений вдоль соответствующей прямой исходной плоскости. Самое частое использование преобразования Хафа – выделение прямых: ясно, что если на изображении есть одна яркая прямая, то в соответствующей точке Хаф-образа значение будет максимальным. Ввиду важности значения, которое играют прямые на изображениях (дороги, дома, лучи томографа, границы документа, строки и т.п.), роль Хаф-анализа можно сравнить с ролью Фурье анализа в акустике. В то же время проведение Хаф-анализа изображения представляет собой довольно сложную задачу. На изображениях мы имеем дело не с прямыми, а с отрезками различной длины. Эти отрезки почти всегда не совсем прямые, часто зашумлены и/или частично не видны, поэтому Хаф-анализ изобилует алгоритмами эвристического характера. Этому посвящены, в частности, работы Ершова, Карпенко, П. П. Николаева.

Между тем, как раз с теми проблемами, которые затрудняют Хаф-анализ изображения, нейронные сети прекрасно справляются во многих задачах обработки изображений и, более того, существуют систематические методы их решения. Иногда нейронные сети используются непосредственно в Хаф-анализе, но в качестве исключительно обработчиков изображений.

Итак, в области обработки и анализа изображений известно множество алгоритмов, плодотворно использующих преобразование Хафа, но без нейронных сетей качество их работы в реальных условиях уже не удовлетворяет ожиданий. Поэтому включение преобразования Хафа в нейронные сети, причем не как поставщика первичных признаков, а в качестве полноценного слоя выглядит очень естественным.

Дополнительно стоит отметить, что использование преобразования Хафа переводит карты признаков в пространство проекций. Задача анализа пространства проекций встает при томографической реконструкции.

Для реконструкции используются свертки и транспонированное преобразование проецирования, отображающее проекции в исходное пространство. Уже предлагаются нейросетевые модели, включающие соответствующие слои. Важно отметить, что здесь же возникает задача обращения подобного преобразования, которая может быть решена с использованием транспонированного преобразования. Например, в работе Адлера и Октема рассматривается глубокая end-to-end модель, которая содержит слои трансформации карт признаков из координатного пространства в пространство признаков и обратно, а Сибен и соавторы для этого использовали целую комбинацию из слоев, решая аналогичную задачу. Таким образом, авторы этих и других работ решают схожую задачу: с использованием Хафовских слоев специального вида они пытаются создать архитектуру нейронной сети, способную решать конкретную задачу.

Однако для того, чтобы систематически сделать преобразование Хафа слоем нейронной сети, необходимо серьезное исследование, связанное с тем, как вести обучение сети при наличии таких слоев, а также с тем, в каких архитектурах использование будет эффективным. Таким образом, исследование, направленное на создание слоя преобразования Хафа и создание нейросетевых архитектур, содержащих такие слои, а так же исследование их свойств является актуальной задачей. Этой теме и посвящена диссертация.

Основные результаты были получены в процессе выполнения работа по следующим научным грантам РФФИ:

1. 15-29-06083 – “Исследования устойчивости методов статистического распознавания к искажениям в неконтролируемых условиях мониторинга аномалий земной поверхности в арктических широтах”
2. 17-29-03161 – “Каскадные алгоритмы локализации и прослеживания протяженных объектов на мобильных устройствах”
3. 17-29-07092 – “Методы машинного обучения для построения оптимальных тепловых карт в пространстве параметров элементов решения для комбинаторных задач”
4. 18-29-26027 – “Использование аппроксимаций операторов прямого и обратного проецирования арифметическими сетями низкой сложности в качестве слоев искусственной нейронной сети”
5. 19-29-09064 – “Исследование нейросетевых методов кооперации локальных и интегральных признаков для получения карт информационной плотности при анализе изображений сцен”
6. 19-29-09092 – “Модели и методы компьютерного зрения в задачах обеспечения информационной безопасности для индустрии 4.0”

Целью данной работы является исследование свойств нейронных сетей, включающих слои прямого и транспонированного преобразований Хафа при решении прикладных задач компьютерного зрения.

Для достижения этой цели поставлены и решены следующие задачи:

1. Исследование трансформации рецептивного поля при добавлении слоя быстрого преобразования Хафа. Это необходимо для мотивированного построения нейросетевых архитектур при решении прикладных задач.
2. Изучение способов и особенностей распространения градиента через слои быстрого преобразования Хафа. Без такого способа обучения нейронной сети в end-to-end режиме невозможно. Оценить вычислительную сложность и практическую применимость.
3. Создание семейства нейросетевых архитектур для решения набора практических задач. Данные архитектуры должны быть объяснимы с точки зрения классического использования интегральных операторов.
4. Исследование поведения сетей предложенных архитектур при решении ими прикладных задач компьютерного зрения: поиск внутренней точки схода, поиск внешней точки схода, детекция выпуклых объектов. Показать, что нейросети разработанных архитектур обеспечивают конкурентное качество при меньшем количестве обучаемых параметров и/или при требовании меньшего количества вычислительных мощностей.

Научная новизна:

1. В работе впервые проведено исследование трансформации рецептивного поля нейронной сети при добавлении в нее слоев прямого и транспонированного быстрого преобразования Хафа. Показано, что для того, чтобы нейронная сеть была способна решать задачи компьютерного зрения, ее рецептивное поле должно обладать достаточным размером.
2. В работе предложена новая архитектура сетей со слоями прямого преобразования Хафа, а также метод ее использования для решения задачи поиска внешней точки схода.
3. В работе впервые предложена архитектура нейронной сети с использованием транспонированного преобразования Хафа для решения задач компьютерного зрения и показано, что такая архитектура способна решать задачи поиска внутренних точек схода и семантической сегментации четырехугольных объектов с превосходным качеством, при этом обладая на два порядка меньшим количеством обучаемых параметров.
4. При применении известной технологии переноса знаний для domain adaptation выявлен новый эффект: перенос знаний успешно происходит, но для высокоуровневых, а не для низкоуровневых признаков.

Практическая значимость работы подтверждается тем, что методы обучения нейронных сетей с использованием исследованных в работе

слоев преобразования Хафа были внедрены в систему обучения нейронных сетей, используемую в компании ООО “Смарт Энджинс Сервис”. Это позволило внедрить предложенные архитектуры в уже имеющиеся 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.

Соответствие диссертации паспорту научной специальности.

В соответствии с формулой специальности 1.2.2 “Математическое моделирование, численные методы и комплексы программ” (технические науки) в работе выполнено исследование слоя преобразования Хафа, приведено доказательство того, что транспонированный оператор для быстрого преобразования Хафа выполняется с той же сложностью, что и прямой, разработаны архитектуры с такими слоями и исследованы их свойства. Работа соответствует следующим пунктам паспорта специальности: п. 2 “Разработка, обоснование и тестирование эффективных вычислительных методов с применением современных компьютерных технологий”, п. 3 “Реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента”, п. 7 “Качественные или аналитические методы исследования математических моделей (технические науки)”.

Методология и методы исследования. В работе используются методы цифровой обработки изображений, методы стохастической непрерывной оптимизации, методы статистического анализа. Все результаты обосновываются формально-дедуктивно или вычислительным экспериментом.

Основные положения, выносимые на защиту:

1. Введение слоев прямого и транспонированного быстрого преобразования Хафа позволяет сделать рецептивным полем выходных нейронов все входное изображение без увеличения числа обучаемых параметров.
2. При помощи представленной архитектуры Хафнет можно построить конечную карту признаков, которая с помощью кусочно-проективных преобразований отображается в бесконечную плоскость, содержащую исходное изображение, что позволяет работать в том числе и с бесконечно удаленными точками.
3. Использование комбинации прямого и транспонированного преобразований Хафа в нейронной сети позволяет построить автокодировщик, способный учитывать интегральные признаки, но не имеющих дополнительных обучаемых параметров.
4. При обучении блоков сверточных слоев, разделенных слоями преобразования Хафа для решения задачи поиска точки схода,

накопление специфичных для конкретных изображений признаков происходит в первом блоке сверточных слоев, в то время как второй и третий блоки можно использовать для решения той же задачи на данных из другого домена.

Достоверность полученных результатов подтверждается соответствием теоретических и экспериментальных результатов, продемонстрированных в работе, успешной апробацией результатов и внедрением в коммерческие системы распознавания документов.

Апробация работы. Основные результаты работы докладывались и обсуждались на следующих конференциях:

1. Международной конференции “International Conference on Machine Vision” (ICMV) в 2016-м, 2017-м и 2019-м годах.
2. Международной конференции “International Conference on Document Analysis and Recognition” (ICDAR) в 2019-м году.
3. Международной конференции “International Conference on Image Processing” (ICIP) в 2020-м году.

Публикации. Основные результаты по теме диссертации опубликованы в 8 печатных изданиях, 1 работа издана в журнале, рекомендованном ВАК, 7 – в изданиях, индексируемых в Web of Science и Scopus. Также получено свидетельство о регистрации программы для электронных вычислительных машин (ЭВМ) и патент США.

Личный вклад. Все основные результаты, изложенные в данной работе, были получены лично автором. Постановка задач и обсуждение результатов проводились совместно с научным руководителем. В [7] автором представлено доказательство выполнимости транспонированного преобразования Хафа с такой же асимптотической сложностью, как и быстрое прямое преобразование Хафа. В [3, 4] анализируется преобразование Хафа в качестве слоя нейронной сети для решения задач поиска внутренней точки схода и классификации изображений, содержащих линии. В [5] автором предлагается архитектура Хафнет и способ ее применения для решения задачи поиска внешних точек схода на изображениях, содержащих документы. В [2,6,7] автором предлагается архитектура Хафэнкодер, где решаются задачи поиска машиночитаемой зоны, поиска внутренней точки схода и семантической сегментации. В [1] автор исследует проблему переноса знаний между данными из различных доменов для сети архитектуры Хафнет. В этой работе рассматривается два домена (изображения документов и изображения дорог с видео регистраторов) и показывается, что накопление знаний происходит в слоях, идущих после преобразования признаков в пространство (s,t) . В [8] выполнено исследование слоя быстрого преобразования Хафа в задаче классификации изображений и опубликовано без соавторства. В [9,10] приводятся ссылки на свидетельство о регистрации программы для ЭВМ и патент США.

Содержание работы

Во введении обосновывается актуальность диссертационной работы и ее научная новизна, формулируются основные цели и задачи диссертационного исследования и приводятся положения, выносимые на защиту.

В первой главе диссертационной работы проводится анализ развития нейросетевых технологий в параллели с инженерными алгоритмами для решения задач компьютерного зрения. Выделены основные этапы их развития и уделено отдельное внимание наличию качественных скачков: создание первых моделей в 50-х годах, обретение сетями глубины в 80-х, глубокое обучение и сверточные сети в 00-х. Также рассмотрено место алгоритмов машинного обучения в задачах обработки изображений.

Рассмотрены некоторые классические алгоритмы технического зрения с использованием различных интегральных преобразований. Среди них алгоритмы для поиска точек схода, различных объектов и анализа трехмерных сцен. Показано, что такие алгоритмы, кроме ряда преимуществ, таких как простота, вычислительная эффективность, понятность принципа работы и предсказуемость ожидаемого результата, обладают и характерными недостатками. Если рассматривать задачи обработки изображений, то, прежде всего, в эти недостатки входит неустойчивость к различным неидеальностям входных данных. Но изображения, подаваемые на вход таким алгоритмам, часто обладают рядом проблем, среди которых присутствуют искажения геометрии, наличие противоречий в данных, посторонние объекты, шум. Тем самым, классические алгоритмы обработки изображений в современных условиях становятся трудно применимыми.

С другой стороны, показано, что задачи улучшения изображений хорошо решаются нейронными сетями, которые могут, прежде всего, подавлять большинство типов шумов, исправлять геометрию и игнорировать признаки, исходящие от нерелевантных объектов. Такие сети, в основном, основываются на локальных признаках, и поэтому задачи обработки изображений, в которых необходимо учитывать глобальную информацию, в общем виде для них неразрешимы.

В первой главе также рассмотрены некоторые алгоритмы, в которых одновременно используются и нейронные сети, и интегральные операторы. В большинстве случаев это совершенно раздельные части, которые применяются последовательно для построения сложной схемы.

Таким образом получается, что идея совместить интегральный оператор и нейронную сеть в одну сущность кажется логичной и перспективной. В данной диссертационной работе в качестве рассматриваемого интегрального оператора выбрано преобразование Хафа, как одно из самых широко используемых интегральных преобразований при решении задач обработки изображений. Основной целью данной работы является исследование поведения нейронных сетей, включающих в свою архитектуру слой

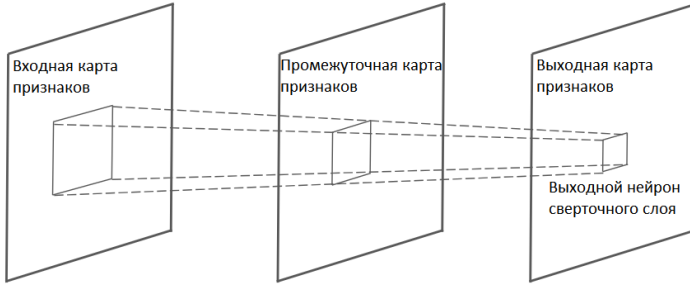


Рис. 1 — Образование рецептивного окна в полносверточной сети.

быстрого преобразования Хафа в явном виде, в практических задачах компьютерного зрения.

В конце первой главы перечислены цели работы и задачи, которые необходимо решить для их достижения.

Во второй главе описаны теоретические основы, на которых базируется проведенное в диссертации исследование. В первую очередь, подробно рассмотрен вопрос трансформации рецептивного поля при применении преобразования Хафа к картам признаков. Как известно, рецептивным полем полносверточной сети является некоторый прямоугольник, размер которого зависит от конфигурации слоев и может быть вычислен по формуле 1.

$$\begin{cases} k_i = k_{i-1} * s_{i-1}/a_{i-1} \\ R_i = R_{i-1} + (f_i - 1) * k_{i-1}, \end{cases} \quad (1)$$

где i — номер слоя, f — размер фильтра, a и s — шаг и апсэмпл коэффициенты, k — кумулятивный коэффициент к i -у слою, а R — размер рецептивного поля к i -у слою. С каждым новым слоем выходные нейроны захватывают прямоугольник все большего размера, что также изображено на рисунке 1.

Таким образом выходит, что возможности сети ограничены: она никогда не сможет надежно решать задачу, для которой требуется анализировать большую зону входного изображения, чем это позволяет размер рецептивного поля. В некоторых случаях можно решить эту проблему, если добавить слои интегральных операторов. Рассмотрим пример: дан отрезок с двумя точками и необходимо определить, находится ли взятая третья точка внутри них? Если попытаться построить сверточную сеть, способную решать эту проблему, то можно видеть, что для этого ей необходимо рецептивное поле, способное видеть обе ограничивающие точки, иначе решение построить невозможно. Рассмотрим теперь интеграл от такого входа, как показано на рисунке 2.

По каждой точке интеграла можно однозначно принять решение, находится ли она внутри заданных точек или же снаружи и при этом обучаемым слоям уже не придется иметь огромное рецептивное поле. Этот

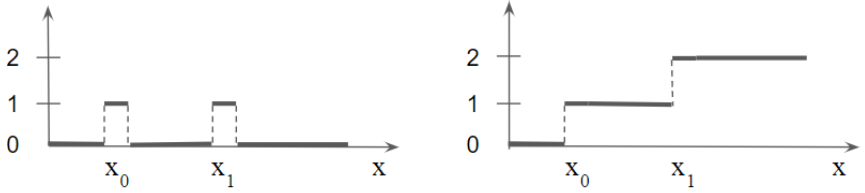


Рис. 2 — Одномерная функция с двумя единичными точками (слева) и ее интеграл (справа).

механизм лежит в основе построения рецептивных полей для нейронных сетей с использованием слоев преобразования Хафа.

Далее во второй главе подробно описано быстрое преобразование Хафа, так как именно такая версия алгоритма выбрана для работы. Важной деталью в этом случае является способ задания и дискретизации прямых, по которым считаются суммы. Алгоритм построен так, что за один раз вычисляет суммы по любой из четырех четвертей: $H_1 : [-45, 0]$, $H_2 : [0, 45]$, $H_3 : [45, 90]$, $H_4 : [90, 135]$ (для квадратного изображения). Кроме этого, вводятся следующие обозначения: H_{12} для половины диапазона, состоящей из преимущественно вертикальных прямых, и H_{34} — для второй половины, состоящей из преимущественно горизонтальных. Для H_{12} , который соответствует преимущественно вертикальным линиям на изображении, можно посчитать координаты результирующей точки (s, t) для прямолинейного сегмента $(x_0, 0) - (x_1, h - 1)$ на входном изображении с помощью уравнения 2:

$$\begin{cases} s = x_0 \\ t = h - (x_1 - x_0) \end{cases} \quad (2)$$

Для H_{34} сегмент прямой задается как $(0, y_0) - (h, y_1)$, а (s, t) могут быть вычислены с помощью уравнения 3:

$$\begin{cases} s = y_0 \\ t = h - (y_0 - y_1) \end{cases} \quad (3)$$

Важным фактом является то, что для построения быстрого преобразования Хафа необходимо использовать специальный вид пикселизации прямой: диадический паттерн. Это необходимо для возможности повторного использования частичных сумм. Для простоты будем рассматривать только случай квадратного изображения со стороной $h = 2^p$. Рассмотрим двоичное представление значений x и t , т.е. набор нулей и единиц такие, что:

$$t = \sum_{r=0}^{p-1} t_r 2^r; x = \sum_{i=0}^{p-1} x_i 2^i \quad (4)$$

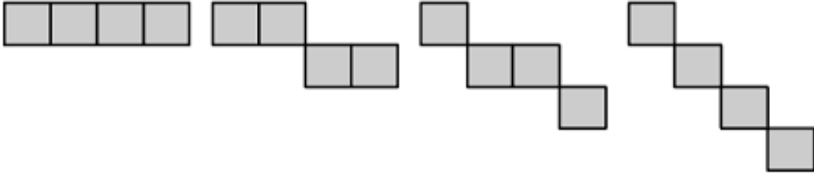


Рис. 3 — Все возможные паттерны для длины 4.

Тогда можно определить матрицу $H(x, t)$, которая бы задавала смещение каждого пикселя в паттерне длины 2^p :

$$H(x, t) = \sum_{r=0}^{p-1} t_r \left[\frac{2^r x}{2^p - 1} \right], \quad (5)$$

где x — номер пикселя в паттерне, а t — смещение последнего пикселя относительно первого.

Примеры паттернов для $p = 2$ изображены на рисунке 3. По этому рисунку можно видеть, что, например, $H(0, 0) = 0$, $H(3, 3) = 3$, $H(1, 2) = 1$ и т.д.

Наконец, во второй главе рассмотрено транспонированное быстрое преобразование Хафа, которое естественным образом возникает при построении слоя нейронной сети. Так как быстрое преобразование Хафа является линейным оператором, то при распространении градиента через него необходимо уметь строить транспонированный оператор. Доказывается, что транспонированное быстрое преобразование Хафа можно вычислять тем же самым быстрым алгоритмом, так как матрица этого оператора становится симметричной относительно главной диагонали, если правильным образом произвести отражения входных и выходных изображений (рисунок 4). Это важное свойство позволяет не только распространять градиент алгоритмом с хорошей асимптотической сложностью, но и строить сети, которые используют транспонированный оператор быстрого преобразования Хафа в явном виде. Далее в работе будет показано в каких случаях это полезно.

Третья глава посвящена описанию разработанных нейросетевых архитектур, использующих слои прямого и транспонированного быстрого преобразования Хафа (БПХ), и исследованию некоторых их свойств и способов применения. Здесь рассматривается два принципиально разных варианта использования таких сетей и обсуждаются особенности, которыми должна обладать архитектура.

Первый случай — это построение архитектуры для решения задачи поиска внешних точек схода. Такая задача возникает, например, при анализе трехмерных сцен или фотографии документов в произвольной ориентации в пространстве. В подавляющем большинстве случаев такие изображения

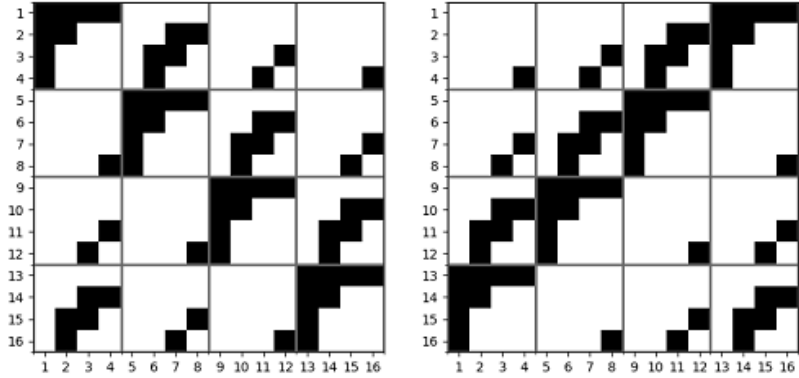


Рис. 4 — Матрица оператора быстрого преобразования Хафа, исходная слева, преобразованная справа.

содержат точки схода, которые лежат за их пределами (вплоть до бесконечности). Их нахождение позволяет оценить позиции объектов интереса в кадре и, следовательно, критически важно для дальнейшего анализа изображения. Классический алгоритм поиска таких точек содержит два последовательных применения быстрого преобразования Хафа. На этой идее и базируется первая разработанная архитектура, названная “Хафнет”. Конфигурация слоев этой архитектуры представлена в таблице 1.

Важным свойством такой архитектуры является то, что каждой точке плоскости, содержащей исходное изображение (вплоть до бесконечно удаленных) соответствует некоторая точка выходной карты признаков. Зависимость между координатами точек в исходной плоскости и на выходной карте признаков $H_m(H_n)$ можно вычислить используя следующие формулы:

$$\begin{cases} s' = \frac{3h}{2} + \frac{4hx - w(2y + h)}{2(h - 2y)} \\ \alpha' = \frac{(2y + h)(w + h)}{2y - h} \end{cases} \quad (6)$$

$$\begin{cases} s' = \frac{3w}{2} + \frac{4wy + h(2x - 3w)}{2(2x - w)} \\ \alpha' = \frac{(2x - 3w)(w + h)}{2x - w} \end{cases} \quad (7)$$

Уравнение 6 соответствует преобразованию $H_{34}(H_{12})$, а уравнение 7 — $H_{34}(H_{34})$.

При известных s' и α' можно вычислить координаты точки на исходном изображении (координатная система которого подверглась двойному

Таблица 1 — Архитектура Хафнет

Слои			
#	Тип	Параметры	Функция активации
1	сверт	12 фильтры 5×5 , шаг 1×1 , без расширения	<i>relu</i>
2	сверт	12 фильтры 5×5 , шаг 2×2 , без расширения	<i>relu</i>
3	сверт	12 фильтры 5×5 , шаг 1×1 , без расширения	<i>relu</i>
4	БПХ	H_{12} для вертикальных, H_{34} для горизонтальных	-
5	сверт	12 фильтры 3×9 , шаг 1×1 , без расширения	<i>relu</i>
6	сверт	12 фильтры 3×5 , шаг 1×1 , без расширения	<i>relu</i>
7	сверт	12 фильтры 3×9 , шаг 1×1 , без расширения	<i>relu</i>
8	сверт	12 фильтры 3×5 , шаг 1×1 , без расширения	<i>relu</i>
9	БПХ	H_{34}	-
10	сверт	16 фильтры 5×5 , шаг 3×3 , без расширения	<i>relu</i>
11	сверт	16 фильтры 5×5 , шаг 3×3 , без расширения	<i>relu</i>
12	сверт	1 фильтры 5×5 , шаг 1×1 , без расширения	$1 - rbf$

преобразованию Хафа) с помощью следующих уравнений для $H_{34}(H_{12})$ и $H_{34}(H_{34})$ соответственно:

$$\begin{cases} x = \frac{\alpha'w + (3h - 2s')(w + h)}{2(\alpha' - (w + h))} \\ y = \frac{h(\alpha' + w + h)}{2(\alpha' - (w + h))} \end{cases} \quad (8)$$

$$\begin{cases} x = \frac{w(\alpha' - 3(w + h))}{2(\alpha' - (w + h))} \\ y = \frac{\alpha'h + (3w - 2s')(w + h)}{2(\alpha' - (w + h))} \end{cases} \quad (9)$$

Таким образом, разработана не только архитектура нейронной сети, но и способ ее применения.

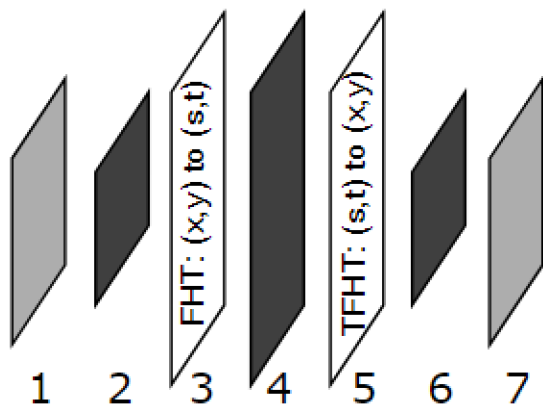


Рис. 5 — Архитектура Хафэнкодера в общих чертах. Блоки 1 и 7 - вход и выход соответственно. Блок 2 - сверточные слои, работающие с исходным изображением. Блоки 3 и 5 слои прямого и транспонированного БПХ. Блок 4 - сверточные слои, работающие в координатах (s, t) . Блок 6 - сверточный блок, работающий в координатах (x, y) , но уже с глобальными признаками.

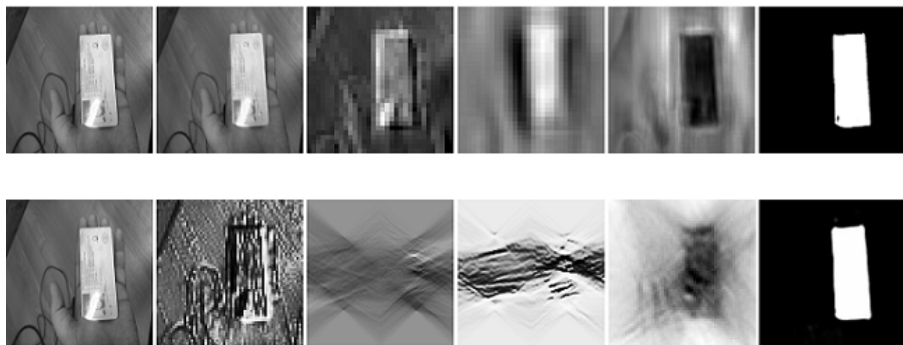


Рис. 6 — Трансформация признаков в Хафэнкодере и в UNet.

Во втором рассмотренном случае точка схода внутренняя. Несмотря на то, что этот случай также возможно разрешить при помощи архитектуры Хафнет, это можно сделать проще, если признаки из пространства (s, t) перевести обратно в (x, y) прямо внутри нейронной сети. Это позволит избежать неочевидных преобразований координат точек и превратит сеть в автоэнкодер. Такая архитектура названа “Хафэнкодер”. Главная ее особенность заключается в том, что второе преобразование Хафа не прямое, а транспонированное. Ключевые блоки такой архитектуры изображены на рисунке 5.

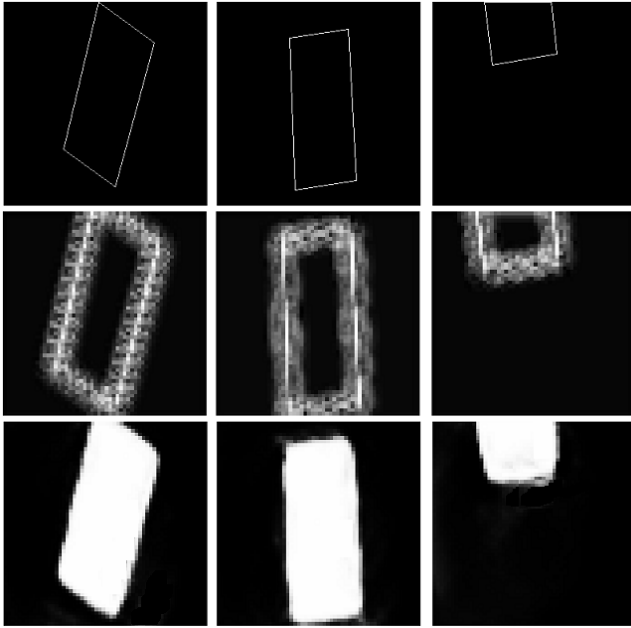


Рис. 7 — Примеры контуров (сверху), ответы полносверточной сети (посередине) и ответы Хафэнкодера (снизу).

В отличие от стандартных полносверточных автоэнкодеров, такая сеть работает по другому принципу. На рисунке 6 изображены сравнительные формы карты признаков для Хафэнкодера и для Unet. Можно видеть, что в обычной сети карты признаков постепенно трансформируются из исходного изображения в результирующее, когда в Хафэнкодере есть два места, в которых происходят преобразования системы координат из (x,y) в (s,t) и обратно. Это позволяет нейронной сети видоизменить свое рецептивное поле без изменения количества обучающих параметров и учитывать интегральные признаки на изображении.

Рассмотрен пример, демонстрирующий пользу от этого. Аналогично примеру с отрезком и двумя точками, рассмотренному во второй главе, здесь решается постановочная задача: дан четырехугольный контур на изображении. Необходимо построить такую нейронную сеть, которая была бы способна его закрасить (т.е. классифицировать точки на внутренние и внешние). Примеры входных данных изображены на рисунке 7 в верхней строке.

Для решения этой задачи протестируем две архитектуры, различающиеся наличием слоев прямого и транспонированного преобразования Хафа. Выход обычной сети можно видеть на рисунке 7 в средней строке. Такая сеть в силу ограниченного рецептивного поля не способна решить данную задачу. Выход Хафэнкодера показан на рисунке 7 в нижней строке.

Здесь видно, что использование интегральных признаков позволяет при такой же конфигурации сверточных слоев (т.е. их параметрах и количестве), успешно классифицировать точки на внутренние и внешние.

Отдельно рассмотрена вычислительная сложность предлагаемых слоев в сравнении с другими слоями нейронной сети. Количество требуемых операций - $O(cs^2 \log s)$, где c – количество каналов изображения, s – линейный размер изображения. Для сверточного слоя требуется порядка $O(cs^2 f^2 m)$ операций, где f – линейный размер фильтра, m – количество фильтров. Соотношение между вычислительными сложностями сверточного и БПХ слоев равно $\frac{f^2 m}{\log s}$. Следовательно, добавление слоя быстрого преобразования Хафа вносит небольшой вклад в сложность сети, который тем меньше, чем сложнее сверточная часть нейронной сети.

Дополнительно показано, что признаки в пространстве (s, t) имеют общий вид для входных изображений различной природы, что позволяет расширить обучающие данные для конкретной задачи данными из других областей. Также это позволяет использовать перенос знаний, накопленных в последних слоях (в отличие от стандартного переноса, где, как правило, участвуют ранние слои), с одной задачи на другую. Проведен подтверждающий это эксперимент с использованием двух существенно различающихся наборов данных: изображения документов и изображения с видеорегистратора автомобиля. Этот эксперимент показал, что при дообучении первых слоев нейронной сети, обученной на документах, на небольшом количестве изображений дорог и при сохранении остальных параметров без изменений удастся достигнуть примерно в 2 раза меньшей ошибки в сравнении с обучением такой же сети исключительно на этих же изображениях дорог в задаче поиска внешней точки схода.

Четвертая глава диссертации посвящена практическим применениям разработанных архитектур и описанию проведенного внедрения.

Первая важная задача, которая решается с помощью архитектуры Хафнет – поиск внешних точек схода на изображениях документов. Такая задача естественным образом возникает, например, при оцифровке данных с документов, так как определенные позиция и ориентация позволяют найти нужный регион изображения, привести документ к прямоугольной форме и сохранить или распознать его, так как в таком виде это делать проще и надежнее. Для того, чтоб найти ориентацию объекта можно либо найти его края и сам четырехугольник, либо же найти точки схода. Первое не всегда возможно, так как края могут быть заслонены какими-либо объектами на изображении или же вообще выходить за его пределы. Точки схода в таком случае можно все равно найти, так как они образованы строками текста, началами выравненных текстовых полей, даже элементами букв. Таким образом, изображение документа содержит две точки схода (вертикальная и горизонтальная). Геометрическое место таких точек

Таблица 2 — Результаты на наборе данных MIDV-500.

Данные	Обуч.	Обуч. испр.	Тест.	Тест. испр.
$d_1(^{\circ})$	2.56	1.63	2.70	1.65
$d_2(^{\circ})$	5.76	0.93	4.77	0.91

Таблица 3 — Результаты на наборе данных ICDAR 2011.

Данные	Исходные	Радон-1	Радон-2	Предл. метод
Распознано (%)	31.3	49.6	50.1	59.7

в обычном случае лежит далеко за пределами картинки вплоть до бесконечности. Чем меньше прямоугольник документа подвержен проективным искажениям, тем дальше эти точки находятся.

Эксперименты проводились на двух наборах данных: MIDV-500 и ICDAR 2011. В первом наборе данных находятся изображения муляжей различных документов, удостоверяющих личность, расположенные на сложных фонах и подверженные проективным и другим, специфичным для съемки с мобильной камеры, искажениям. Второй набор данных — данные с соревнования “ICDAR 2011 Page Dewarping Contest”. Здесь расположены сканы страниц формата A4, заполненные текстом и другими элементами (картинки, таблицы).

Оценка качества работы проводилась по двум метрикам:

$$d_1 = \frac{1}{4N} \sum_{j=0}^{N-1} \sum_{i=0}^3 |90^{\circ} - \alpha_i^j| \quad (10)$$

$$d_2 = \frac{1}{2N} \sum_{j=0}^{N-1} |\alpha_v^j| + |\alpha_h^j| \quad (11)$$

Расстояние d_1 показывает среднее отклонение углов исправленного документа от 90° независимо от его ориентации, как видно из формулы 10. Расстояние d_2 , наоборот, показывает насколько хорошо была исправлена ориентация документа независимо от степени его прямоугольности, как можно видеть из формулы 11. Чем меньше эти значения, тем точнее была проведена корректировка формы и ориентации. В таблицах 2 и 3 представлены результаты и для исходных, и для исправленных изображений документов, чтоб можно было оценить эффект от исправления. Аналогичные результаты для обучающего набора данных здесь приведены для демонстрации того, что переобучения не произошло: качество работы сети примерно одинаковое. Дополнительно, в таблице 3 приведены соответствующие замеры для двух инженерных алгоритмов, базирующихся на преобразовании Радона.

Другой рассмотренной задачей является сегментация изображений документов. На основании эксперимента, проведенного в третьей главе, который показал способность нейросетей архитектуры Хафнет классифицировать точки на внутренние и внешние, была решена реальная задача.

Для тестирования было обучено три нейронных сети: Хафэнкодер (обозначен “HE”), полносверточная сеть с конфигурацией сверток как у Хафэнкодера, но без слоев прямого и транспонированного БПХ (обозначен “HE*”) и стандартная UNet сеть. Все три нейронных сети обучались в одинаковых условиях на одинаковых данных: изображения 30 типов документов из MIDV-500 и 1 типа документов из MIDV-2019. При обучении для всех трех использовались одни и те же раздутья на лету и каждая сеть училась по 200 эпох. Дополнительно, данный эксперимент был проведен по пять раз для каждой сети и взято среднее качество для большей достоверности.

Оценка результата проводилась с помощью оценки среднего отношения пересечения найденных и идеальных зон документа к их объединению, MIoU (mean Intersection over Union), вычисляемого по формуле 12.

$$MIoU = \frac{1}{N} \sum_{i=0}^{N-1} \frac{A_i \cap G_i}{A_i \cup G_i}, \quad (12)$$

В формуле 12 A_i – найденная зона документа, G_i – истинная зона документа, N – общее количество изображений.

Таблица 4 — Сравнительные результаты, MIoU

Сеть	Кол. пар-ов	Качество MIDV-500	Качество MIDV-2019
UNet	1.96×10^6	0.94	0.86
HE*	1.48×10^4	0.91	0.74
HE	1.48×10^4	0.96	0.89

Сравнительные характеристики с точки зрения количества обучаемых параметров и полученного качества приведены в таблице 4. Из результатов можно видеть, что при в 132 раза меньшем количестве обучаемых параметров Хафэнкодер все равно показывает большее качество, чем стандартная полносверточная архитектура. Дополнительно можно видеть, что полносверточная сеть, равная по количеству обучаемых параметров Хафэнкодеру показывает, наоборот, худший результат. Далее в главе отмечается, что аналогичным методом можно решать похожие задачи. Для примера рассматривается задача поиска машиночитаемой зоны на изображении документа и оценивается качество ее распознавания при поиске различными методами. Проведенные эксперименты показали, что качество распознавания при поиске Хафэнкодером достигает 98.1% против лучшего опубликованного результата, равного 97.8%.



Рис. 8 — Примеры добавленных в изображения искажений, призванных сделать невозможным локальный поиск точки схода.

Наконец, третьей важной рассмотренной задачей является поиск внутренней точки схода на изображениях дорог. Ключевым элементом эксперимента сравнения сети Хафэнкодер и стандартной полносверточной сети является попытка решить задачу на заведомо усложненных данных. Для этого без переучивания нейронной сети был модифицирован тестовый набор данных следующим образом: прямоугольные зоны разного размера вокруг искомой точки схода были размыты гауссовым сглаживанием так, чтобы их невозможно было различить исходя из локальной окрестности изображения. Примеры таких модифицированных изображений показаны на рисунке 8.

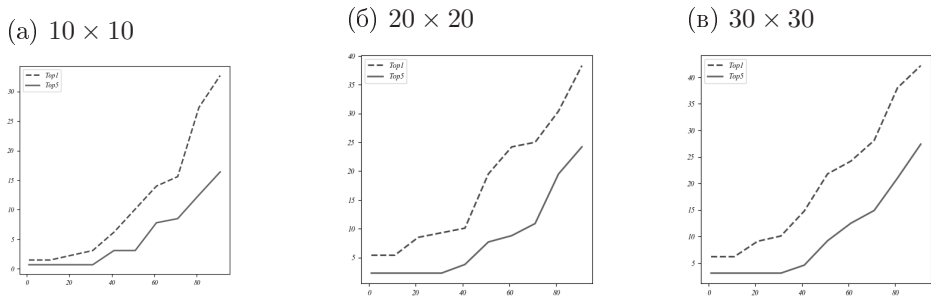


Рис. 9 — Деградация качества при порче изображения для решеток различного размера.

Результаты этого эксперимента частично представлены на рисунке 9. На графиках показана зависимость ошибки от размера размытой области вокруг точки схода для различных размеров решетки. Легко можно видеть, что предложенная нейронная сеть, вообще говоря, не требует наличия области изображения, на котором находится непосредственно сама точка схода. Это демонстрирует одно из главных свойств этой сети: способность

в каждой локальной точке принимать решение, исходя из глобальных признаков, без использования дополнительных обучаемых параметров.

В завершении четвертой главы диссертации описаны детали внедрения слоев прямого и транспонированного быстрого преобразования Хафа в системы обучения и распознавания компании “Смарт Энджинс Сервис”. Кроме распознавания документов, эти системы способны проверять некоторые признаки документов на наличие подделок, а предоставляемый программный интерфейс позволяет строить на их базе многофункциональные решения, включающие в себя, например, сверку фотографии в документе с лицом перед камерой.

В **заключении** приведены основные результаты работы, которые заключаются в следующем:

1. Показано, что добавление в сеть двух слоев преобразования Хафа последовательно, равно как и добавление последовательно слоев преобразования Хафа и транспонированного преобразования Хафа в многослойную сверточную сеть делает рецептивным полем каждого выходного нейрона все изображение.
2. Приведено доказательство возможности построения быстрого транспонированного оператора для быстрого преобразования Хафа, выполняемого со сложностью $O(n^2 \log(n))$. Это позволяет эффективно распространять градиент через слой преобразования Хафа при обучении.
3. Предложена архитектура Хафнет, содержащая в себе ветви с последовательно расположенными слоями быстрого преобразования Хафа, разделенными блоками сверточных слоев, а также метод ее применения, позволяющий кодировать на выходе всю плоскость входного изображения, включая бесконечно удаленные точки. Предложенным методом решена задача поиска внешних (включая несобственные) точек схода на изображении. Показано, что построенная end-to-end модель позволяет сократить ошибку на 20% в сравнении с инженерным подходом, использующим аналогичный алгоритм.
4. Предложена архитектура Хафэнкодер, содержащая в себе слои прямого и транспонированного быстрого преобразования Хафа, разделенные блоками сверточных слоев, что позволяет ей быть сетью типа “автокодировщик”. Предложенная архитектура в задаче семантической сегментации продемонстрировала лучшее качество (0.96 против 0.94 MIoU) при примерно в 100 раз меньшем числе обучаемых параметров, в сравнении со стандартной полносверточной сетью. Также решена задача поиска внутренней точки схода, где, в сравнении со стандартной нейронной сетью, ошибка оказалась примерно в 2 раза меньше во всех экспериментах. Демонстрировано, что благодаря использованию интегральных признаков

такая нейронная сеть способна решать задачу поиска точки схода, даже если в локальной окрестности искомой точки нет релевантных признаков.

5. Показано, что перенос обученных весов последних слоев сети Хаф-нет позволяет без потери качества применять ее к данным из другого домена (исследовались домены изображений документов и изображений с видеорегистратора автомобиля). Это демонстрирует, что перенос знаний в сетях таких архитектур возможен, как и в классических сетях, но происходит в глубоких слоях, работающих с признаками в пространстве Хафа, а не в первых слоях, как это происходит в ранее исследованных архитектурах.
6. Создан программный комплекс для обучения и воспроизведений нейронных сетей, включающих в себя эффективные реализации слоев прямого и транспонированного преобразования Хафа. Все практические результаты, приведенные в работе, были получены с его использованием.

Публикации автора по теме диссертации

В изданиях, входящих в международные базы цитирования Scopus и Web of Science

1. A. V. Sheshkus and D. Nikolaev, "Transfer of a high-level knowledge in HoughNet neural network," ICMV 2019, 11433 ed., Wolfgang Osten, Dmitry Nikolaev, Jianhong Zhou, Ed., Bellingham, Washington 98227-0010 USA, Society of Photo-Optical Instrumentation Engineers (SPIE), Jan. 2020, vol. 11433, ISSN 0277-786X, ISBN 978-15-10636-44-6, vol. 11433, pp. 1143322-1-1143322-6, 2020, DOI: 10.1117/12.2559454.

2. A. V. Sheshkus, D. P. Nikolaev and V. L. Arlazarov, "Houghencoder: neural network architecture for document image semantic segmentation," IEEE ICIP 2020, Washington, DC, United States, IEEE Computer Society, 2020, 704 pp., ISSN 1522-4880, ISBN 978-17-28163-96-3, pp. 1946-1950, 2020, DOI: 10.1109/ICIP40778.2020.9191182.

3. A. Sheshkus, A. Ingacheva and D. Nikolaev, "Vanishing Points Detection Using Combination of Fast Hough Transform and Deep Learning," ICMV 2017, 10696 ed., Antanas Verikas, Petia Radeva, Dmitry Nikolaev, Jianhong Zhou, Ed., Bellingham, Washington 98227-0010 USA, Society of Photo-Optical Instrumentation Engineers (SPIE), Apr. 2018, vol. 10696, 758 pp., ISBN 978-15-10619-41-8, vol. 10696, pp. 106960H1-106960H8, 2018, DOI: 10.1117/12.2310170.

4. A. Sheshkus, E. Limonova, D. Nikolaev and V. Krivtsov, "Combining Convolutional Neural Networks and Hough Transform for

Classification of Images Containing Lines,” ICMV 2016, 10341 ed., Antanas Verikas, Petia Radeva, Dmitry P. Nikolaev, Wei Zhang, Jianhong Zhou, Ed., Bellingham, Washington 98227-0010 USA, Society of Photo-Optical Instrumentation Engineers (SPIE), July 2017, vol. 10341, ISSN 0277-786X, ISBN 978-15-10611-31-3, vol. 10341, pp. 103411C1-103411C5, 2017, DOI: 10.1117/12.2268717.

5. A. Sheshkus, A. Ingacheva, V. Arlazarov and D. Nikolaev, “HoughNet: neural network architecture for vanishing points detection,” ICDAR 2019, Manhattan, New York, U.S., The Institute of Electrical and Electronics Engineers (IEEE), Feb. 2020, ISSN 2379-2140, ISBN 978-17-28130-14-9, 8978201, pp. 844-849, 2020, DOI: 10.1109/ICDAR.2019.00140.

6. S. Пыухин, A. Sheshkus, V. Arlazarov and D. Nikolaev, “Hough Encoder for Machine Readable Zone Localization,” Pattern Recognit. Image Anal., vol. 32, no 4, pp. 793-802, 2022, DOI: 10.1134/S1054661822040150.

7. A. Sheshkus, A. Chirvonaya, D. Matveev, D. Nikolaev and V. L. Arlazarov, “Vanishing Point Detection with Direct and Transposed Fast Hough Transform inside the neural network,” Computer Optics, vol. 44, no 5, pp. 737-745, 2020, DOI: 10.18287/2412-6179-CO-676.

В изданиях из списка ВАК РФ

8. Шешкус А. В. Использование сверточных нейронных сетей в комбинации с преобразованием Хафа для классификации изображений с прямыми линиями // Труды ИСА РАН. — 2017. — Т. 67. — № 1. — С. 83-88.

Патенты и зарегистрированные программы для ЭВМ

9. Шешкус А. В., Чернышова Ю. С., Гайер А. В., Лынченко А. Е., Николаев Д. П. Автоматическая система генерации данных и обучения искусственных нейронных сетей “Smart NNCreator” № 2018615794 // Роспатент.

10. Alexander Vladimirovich SHESHKUS и др. — Artificial Intelligence Using Convolutional Neural Network With Hough Transform. — Pub. No.: US 2022/0122267 A1; Patent No.: 17 / 237539. — 2022.

Шешкус Александр Владимирович

Использование преобразования Хафа в качестве слоя нейронной сети

Автореф. дис. на соискание ученой степени канд. техн. наук

Подписано в печать _____._____._____. Заказ № _____

Формат 60×90/16. Усл. печ. л. 1. Тираж 100 экз.

Типография _____