



ГОСУДАРСТВЕННАЯ ИНСПЕКЦИЯ
ПО КОНТРОЛЮ ЗА ИСПОЛЬЗОВАНИЕМ
ОБЪЕКТОВ НЕДВИЖИМОСТИ
ГОРОДА МОСКВЫ



МОСГОРБИ

ХАКАТОН



ЛИДЕРЫ ЦИФРОВОЙ
ТРАНСФОРМАЦИИ

Задача

«Автоматизированная классификация
документов ограниченного набора
технических документов Главархива»

К о м а н д а

FROM
STACK
OVER
FLOWS

ПРОБЛЕМАТИКА ОТРАСЛИ



Медленный и сложный
доступ к документам



Обработка и структурирование
документов **вручную**



Колossalные объемы **ресурсов** и
времени тратится на поиск



Высокий процент **ошибок** в
оцифрованных документах



РЕШЕНИЕ



Автоматизированная система
оцифровки для идентификации,
классификации и анализа
скан-копий технических и
контрактных документов.



ЦЕЛЕВОЕ РЕШЕНИЕ



Централизованное веб-приложение на базе сверточных нейронных сетей для оцифровки документов и скан-копий.



ОСОБЕННОСТИ



Автоматизированная идентификация
сущностей с использованием сверточных
нейросетей и компьютерного зрения



Централизованное веб-приложение с
возможностью контекстного поиска по базе
документов с расширенной фильтрацией



Машинное обучение и дообучение
алгоритмов и моделей нейросетей



Пакетная загрузка скан-копий для
последующего распознавания



The screenshot shows the CIFR digital network interface. At the top, there is a search bar with the query "Q 12:372:789:777289" and a "Бросить" (Drop) button. Below the search bar is a navigation menu with items: "Все" (All), "Кадастр" (Cadastral) which is checked, "Тех.паспорт" (Technical Passport), "Постановление" (Decision), "Накладная" (Delivery Note), and "Счета" (Accounts). A date selector labeled "Дата выдачи документа" (Date of document issuance) is shown with a calendar icon. Below the date selector is a monthly calendar for May 2020, with the 21st highlighted as the "Точная дата" (Exact date). To the right of the calendar, there are four preview cards for "Технический паспорт" (Technical Passport) dated March 16, 2020, each with a small thumbnail and the text "«О внесении изменений в План проведения плановых проверок»" (On making changes to the plan for planned inspections). At the bottom right of the interface, there is a small note: "Все права защищены. 2020" (All rights reserved. 2020).

РЕЗУЛЬТАТЫ MVP



Кластеризация

Для решения задачи классификации мы используем кластеризацию, что значительно увеличивает возможности масштабируемости. Данный подход позволяет автоматически делать классы страниц, даже тех существование которых было неизвестно.

Модель алгоритма



Свидетельство об утверждении архитектурно-градостроительного решения

1. Административный округ
2. Район
3. Адрес
4. Наименование объекта
5. Функциональное назначение объекта

Распознано 6 атрибутов документа

ПРАВИТЕЛЬСТВО МОСКОВЫ
КОМИТЕТ ПО АРХИТЕКТУРЕ И ГРАДОСТРОИТЕЛЬСТВУ Г.МОСКОВЫ
125047, Москва, Триумфальная площадь, 1
тел. 250-55-20

Главный архитектор города

Код объекта недвижимости:
Код строительного объекта:
Регистрационный №: 125-3-502

СВИДЕТЕЛЬСТВО ОБ УТВЕРЖДЕНИИ
АРХИТЕКТУРНО-ГРАДОСТРОИТЕЛЬНОГО РЕШЕНИЯ

АДРЕС ОБЪЕКТА:	Город: Москва
Административный округ	Центральный
Район:	Пресненский
Адрес:	улица Малая Бронная
Владение:	37/2
Наименование объекта:	Проект строительства административного здания с подземной автостоянкой
Функциональное назначение объекта:	Офисы, 1 этаж - торговля и бытовое обслуживание
Заказчик:	
Застройщик:	
Разрешение на проектирование:	Постановление Правительства Москвы от 29.08.2000 №48-ПП

МАСШТАБИРОВАНИЕ



Масштабируемость за счет кластеризации - автоматическое добавление новых классов



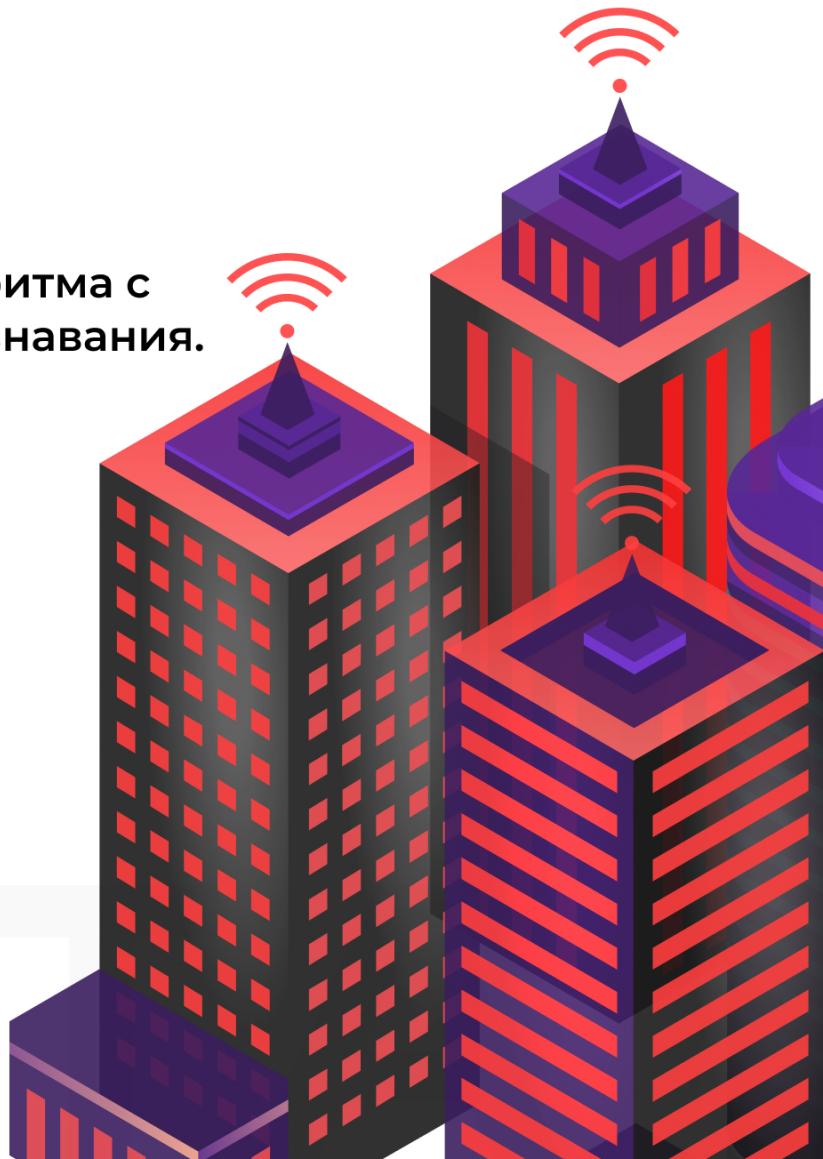
Аналитика и отчет-протоколы о работе алгоритма с расчетом применяемых мер качества распознавания.
BI система построения отчетов.



REST API



Обучение аналитической системы оператором.



ЭКОНОМИЧЕСКИЙ ЭФФЕКТ



Внедрением централизованного веб-приложения на базе сверточных нейронных сетей порядок упрощается и ускоряется работа с различного рода документацией и ее оцифровкой.

ГлавАрхив

Разбор одного комплекта документов у специалиста занимает от **120 до 180 мин.**



Обработка одного документа занимает **3 секунды**. А обработка **100 000** документов займет всего **3.5 дня**.

Скорость оцифровки

В 2400 раз быстрее



Экономия человекочасов

99.95%

РАЗВИТИЕ



12:22 ↗



Внедрение новых алгоритмов классификации и структуризации баз данных для повышения эффективности работы с документооборотом.



Разработка мобильного приложения для возможности мобильного сканирования документов и корпоративного мессенджера внутри организации для быстрой координации работ сотрудников и отделов между собой.



Подключение картографических сервисов к базе данных для интеллектуальной классификации документов.



Введите кадастровый номер

0 1500705100750027

Настройки поиска

Что будем искать?

Документы · 2

Кадастр

Постановления

Дата выдачи документа

Необязательно, но влияет на результат

Применить

СТЕК ТЕХНОЛОГИЙ



HTML



CSS



TensorFlow



Tesseract OCR



ХАКАТОН
ЛИДЕРЫ ЦИФРОВОЙ
ТРАНСФОРМАЦИИ



Дамиан
Баженов

Backend, Rest API
[@arny_meier](https://twitter.com/arny_meier)



Михаил
Дмитриев

Data Science
[@Potap_potap](https://twitter.com/Potap_potap)



Вадим
Бештень

Computer Vision
[@rbv_008](https://twitter.com/rbv_008)



Николай
Шаповалов

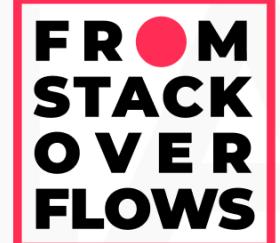
Frontend
[@Sarkozi](https://twitter.com/Sarkozi)



Дмитрий
Ефремов

UI/UIX, Design
[@Brandmaker1](https://twitter.com/Brandmaker1)

КОМАНДА



ХАКАТОН



ЛИДЕРЫ ЦИФРОВОЙ
ТРАНСФОРМАЦИИ



ГОСУДАРСТВЕННАЯ ИНСПЕКЦИЯ
ПО КОНТРОЛЮ ЗА ИСПОЛЬЗОВАНИЕМ
ОБЪЕКТОВ НЕДВИЖИМОСТИ
ГОРОДА МОСКВЫ



МОСГОРБТИ

FROM
STACK
OVER
FLOWS

СПАСИБО ЗА
ВНИМАНИЕ