ИСПОЛЬЗОВАНИЕ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ RESNET ДЛЯ КЛАССИФИКАЦИИ СТРУКТУРИРОВАННЫХ ДОКУМЕНТОВ

## Постановка задачи

**Введение**  
Организация "Такси Город Барнаул", где я работаю, столкнулась с серьезной проблемой в области документооборота. Документы, связанные с различными аспектами деятельности компании, включая управление автопарком и водителями, хранятся в различных местах и в разрозненном виде. Это затрудняет их поиск и обработку, увеличивает риск ошибок, утечек данных, потери документов, а также существенно повышает затраты на их хранение и поддержание в актуальном состоянии. В связи с этим возникла необходимость в разработке системы классификации и систематизации документации, которая позволит упорядочить и оптимизировать весь процесс документооборота.

**Цель**  
Основная цель данного проекта заключается в разработке и внедрении эффективной системы классификации и систематизации документов, связанной с деятельностью компании, что позволит улучшить управление документами, их хранение и доступ к ним. В конечном итоге это должно повысить безопасность данных, уменьшить риск потерь, а также сократить время и ресурсы, затрачиваемые на обработку документации.

## ****Анализ существующих решений для классификации структурированных документов****

В рамках работы был проведен анализ решений, разработанных российскими компаниями в области автоматизации обработки документов. Основное внимание уделялось трем продуктам: Smart Engines [3], Dbrain [4] и Биорг.Паспорт [5]. Эти системы представляют собой разработки, которые могут не только классифицировать документы, но и извлекать из них структурированную информацию с высоким уровнем точности. Каждое из решений поддерживает интеграцию с внешними приложениями через API, что упрощает их внедрение в существующие корпоративные системы. Также продукты соответствуют ключевым требованиям федерального законодательства Российской Федерации по защите персональных данных.

**Smart Engines**

**Особенности:**

* Специализируется на автоматическом распознавании без опции верификации данных.
* Имеет высокую точность распознавания печатных текстов (93–94%), но заметное снижение эффективности при работе с рукописными текстами (около 60%).
* Отсутствуют облачные сервисы и верификации данных для увеличения безопасности данных.
* Предполагает взаимодействие через API.
* Ориентируются на большие объемы документов, с малым количеством документов компании не интересны

**Ограничения:**

* Неудобство для пользователей при повторном вводе нераспознанных данных.
* Отсутствие поддержки для облачных решений, это ограничивает масштабируемость и доступность.

**Dbrain**

**Особенности:**

* Предлагается как облачный сервис, есть возможность установки ПО непосредственно у заказчика на сервер.
* Автоматическое распознавание с декларируемой точностью 85%.
* Верификация данных через интеграцию с Яндекс.Толока, позволяющая использовать как внешних, так и внутренних операторов.
* Система оплаты постраничная, что может быть выгодно при малых объемах данных.

**Ограничения:**

* Проблемы с качеством распознавания на некачественных изображениях.
* Не обеспечивает поддержку установки и настройки системы, ограничиваясь API.
* Отсутствие необходимых сертификатов безопасности согласно 152-ФЗ Российской Федерации о защите персональных данных.

**Биорг.Паспорт**

**Особенности:**

* Работает как в формате облачного сервиса, так и "коробочного" решения в контуре заказчика.
* Опциональная верификация данных с быстрым подключением.
* Высокая скорость и точность обработки данных, подтвержденная реальными тестами (скорее всего достигается путём ручной верификации на стороне разработчика).
* Обеспечение высоких стандартов безопасности, включая страхование профессиональной ответственности и наличие всех необходимых лицензий согласно законодательству РФ.

**Ограничения:**

* Скорость обработки уступает некоторым конкурентам, хотя и предлагает лучшее качество.
* Возможно, потребуется дополнительное время и ресурсы для полной интеграции и оптимизации процессов.

Тем не менее, данные продукты обладают рядом ограничений. Прежде всего, они не покрывают весь спектр документов, необходимых для комплексной автоматизации документооборота в "Такси Город". Кроме того, стоимость использования данных решений формируется индивидуально и доступна только по предварительному запросу, что затрудняет планирование бюджета и оценку экономической эффективности внедрения данных технологий на предприятии.

## Описание процесса решения

## Выбор нейронной сети для классификации структурированных документов

Для классификации типов документов на основе их изображений можно использовать несколько типов нейронных сетей, каждая из которых предлагает преимущества для обработки и распознавания структурных особенностей документов. Сверточные нейронные сети (CNN) идеально подходят для распознавания визуальных паттернов и структур. Они могут обучаться на различных типах документов и научиться распознавать общую структуру и расположение элементов, характерных для каждого типа. Это включает в себя распознавание распределения текста, наличие или отсутствие таблиц, типичные заголовки или форматы.

Основные нейронные сети, используемые для классификации по изображению:

**ResNet (Residual Networks)**: Сверточная нейронная сеть, известная своими остаточными блоками, которые позволяют обучать очень глубокие сети без потери эффективности. Часто используется для сложных задач классификации изображений, включая документы, благодаря своей способности избегать проблемы исчезающего градиента.

**Inception Network (GoogLeNet)**: Сеть, состоящая из Inception-блоков, которые позволяют сети самостоятельно выбирать размер фильтров для каждого уровня. Подходит для задач, где требуется распознавание на различных масштабах, например, разные размеры шрифтов и элементов документа.

**EfficientNet**: Сеть, которая балансирует глубину, ширину и разрешение изображения для оптимальной производительности и эффективности. Полезна для задач классификации с ограниченными вычислительными ресурсами, обеспечивая высокую точность при меньших затратах.

**Vision Transformers (ViT)**: Сеть, использующая архитектуру трансформера для обработки изображений, разделяя их на патчи и обрабатывая их как последовательность. Подходит для задач, требующих захвата глобального контекста изображения, например, структурные особенности документа.

Сети на основе трансформеров также могут быть адаптированы для работы с изображениями документов, используя техники, подобные тем, что применяются в моделях, обученных на текстах. Такие сети могут эффективно обрабатывать последовательности данных и улавливать более глубокие контекстные связи между элементами документа.

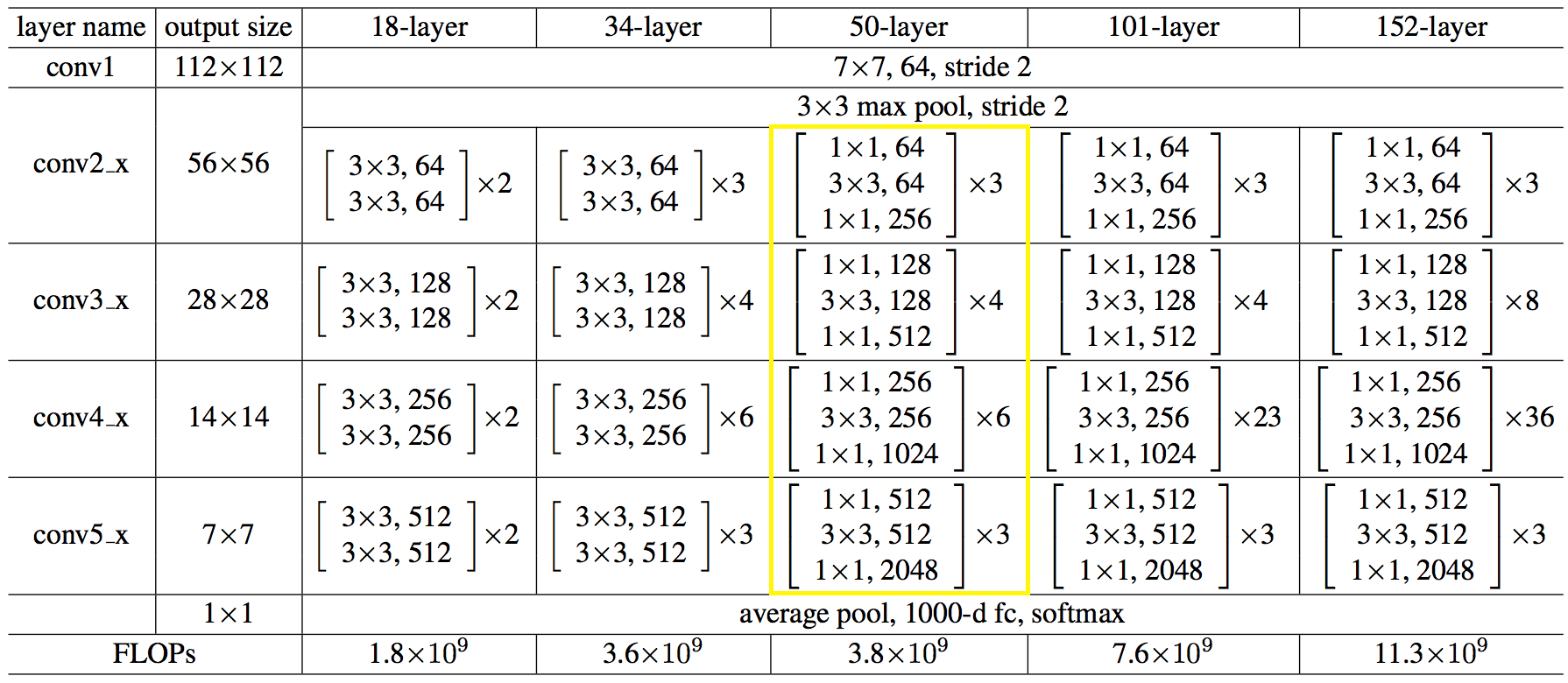
Рассмотрим более подробно ResNet (Residual Neural Network). Эта сверточная нейронная сеть представляет собой одну из самых значимых архитектур в области глубокого обучения и компьютерного зрения. Основное преимущество ResNet заключается в использовании остаточных блоков, которые позволяют передавать входные данные через соединения с пропуском (skip connections) мимо одного или нескольких слоев. Эти соединения помогают преодолеть проблему угасания градиента, которая возникает при увеличении глубины нейросети.

Архитектура ResNet включает в себя несколько ключевых элементов:

**Остаточные блоки (Residual Blocks)**: Эти блоки содержат прямые соединения, которые пропускают один или несколько слоев, добавляя входные данные к их выходам. Это упрощает обучение глубоких сетей и улучшает точность моделей.

**Обратное распространение градиента**: Остаточные соединения облегчают процесс обратного распространения градиента, что снижает проблемы затухания или взрыва градиента, часто возникающие в глубоких сетях.

**Масштабируемость**: ResNet может быть адаптирована под различные задачи, увеличивая или уменьшая количество слоев в зависимости от сложности задачи и доступных вычислительных ресурсов.



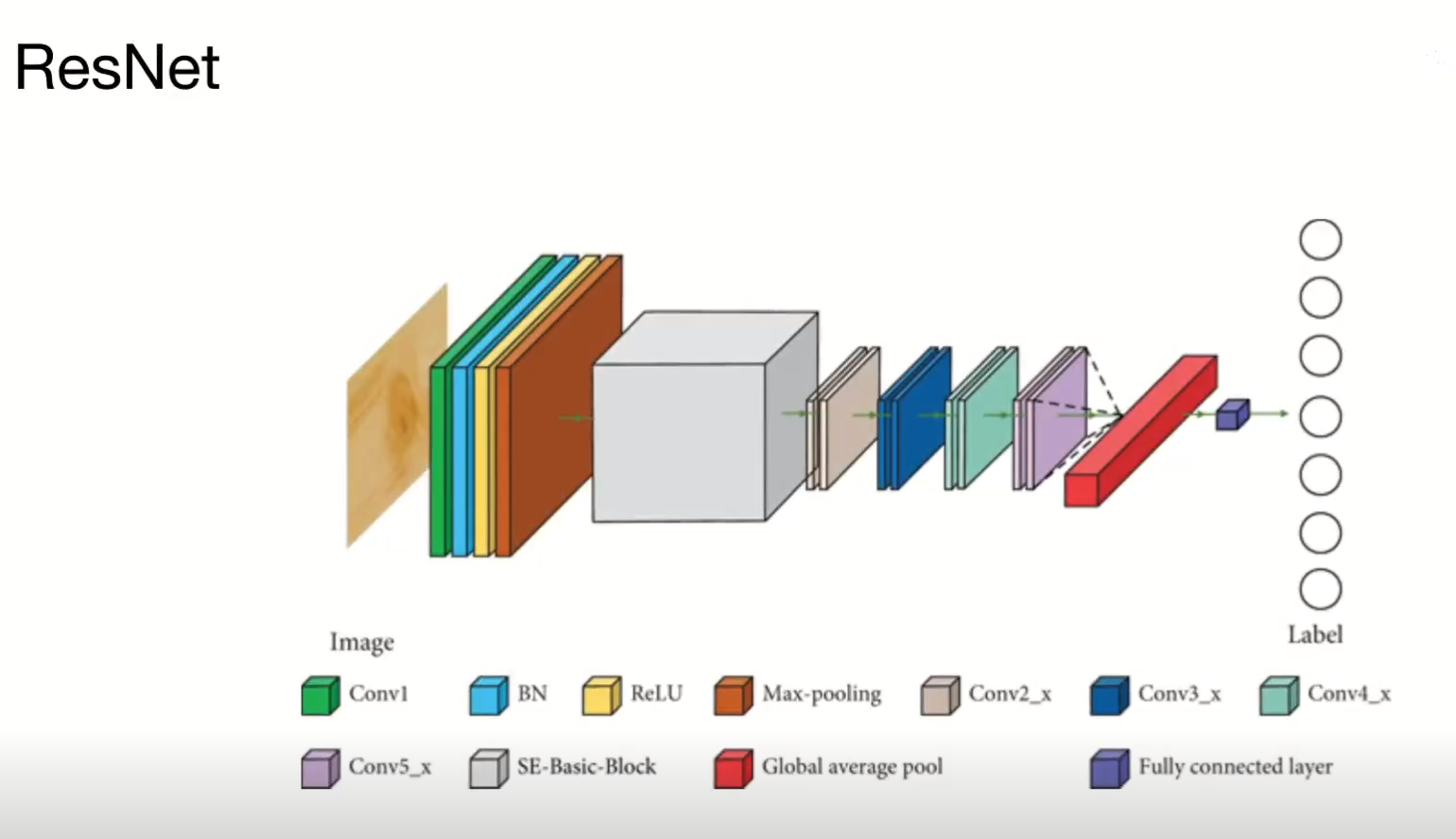
**

Рисунок 1.1. – Схема архитектуры сверточной нейронной сети ResNet   
с различным количеством слоёв

Среди сверточных сетей ResNet выделяется благодаря своим остаточным блокам, которые позволяют обучать глубокие архитектуры без потери эффективности. Версии ResNet, такие как ResNet-18, ResNet-50 или ResNet-152, могут быть настроены для распознавания типа документа, опираясь на обучающие примеры различных документов. Они изучают общую композицию и структуру документов и используют эти знания для точной классификации. ResNet является хорошим выбором из-за способности сети обрабатывать большие и сложные наборы данных документов через остаточные соединения, что обеспечивает масштабируемость и эффективность обучения. Это позволяет эффективно использовать предобученные модели, адаптируя их под новые задачи с минимальным дополнительным обучением. Таким образом, ResNet идеально подходит для классификации типов документов, опираясь на их структурные и форматные особенности, даже без детального анализа содержимого.

## Постановка задачи и основные этапы решения

В рамках текущей деятельности организации «Такси Город» возникла потребность в систематизации и классификации документов, связанных с управлением транспортными средствами и водителями. Эта задача охватывает спектр документации, включая следующие виды:

1. **Паспорт транспортного средства** – документ, удостоверяющий право собственности на автомобиль.
2. **Свидетельство о регистрации транспортного средства** – официальный документ, подтверждающий факт регистрации автомобиля в государственных органах.
3. **Водительское удостоверение** – разрешение на управление транспортными средствами определенной категории.
4. **Разрешение Минтранса** – специальные документы, выдаваемые Министерством транспорта для осуществления определенных видов деятельности или перевозок.
5. **ИНН физического лица** – идентификационный номер налогоплательщика, который необходим для ведения налогового учета.
6. **Паспорт гражданина РФ** – основной документ, удостоверяющий личность в Российской Федерации.
7. **Справка о наличии или отсутствии судимости** – официальный документ, который может потребоваться для подтверждения законопослушности при трудоустройстве или при выполнении определенных видов деятельности.
8. **Технический осмотр или диагностическая карта ТС** – документы, подтверждающие техническую исправность транспортного средства.
9. **СНИЛС** – страховой номер индивидуального лицевого счета, используемый в системе социального страхования.

Цель классификации заключается в обеспечении эффективного управления, хранения и доступа к документам, улучшении организации работы с документацией и оптимизации процессов обработки персональных и корпоративных данных. Это поможет улучшить взаимодействие с государственными органами, ускорить процедуры регистрации и управления автопарком, а также повысить общую безопасность и надежность документооборота.

## Подготовка датасета для классификации сканированных документов

Подготовка датасета включала несколько ключевых этапов:

**Сбор и хранение данных**: Все необходимые данные были собраны и систематизированы в одном месте. Первоначально данные были размещены в разных каталогах на разных сетевых ресурсах, что усложняло доступ к информации. На этом этапе данные были сведены в одно место и разложены по соответствующим каталогам.

**Преобразование форматов данных**: Данные были преобразованы в единый формат с использованием Python-скриптов, что позволило автоматизировать процесс преобразования многостраничных PDF и TIFF файлов в JPG, сохраняя их многостраничность.

**Проверка целостности данных**: после преобразования форматов была проведена тщательная проверка данных на целостность, чтобы удостовериться, что не произошло потерь или искажений информации.

**Визуальная проверка и классификация**: Завершающим этапом была визуальная проверка документов на соответствие определенным классам. Это включало проверку правильности организации и классификации данных после всех преобразований.

**Выявленные проблемы и их решения:** в ходе визуальной проверки были выявлены проблемы с перемешиванием данных, что могло привести к ошибкам при классификации нейронной сетью. Были также обнаружены ошибки в начальной стадии каталогизации и различия в качестве исходных документов. Для решения этих проблем было принято решение использовать методику машинного обучения с применением модифицированной версии ResNet18 бинарного вида.

**Этапы оптимизации датасета:** для обучения модели был выбран тип документа СНИЛС из-за его уникальных характеристик, упрощающих задачу идентификации. Нейронная сеть ResNet18 была дообучена на специализированном малом датасете с использованием техник аугментации данных. Это позволило увеличить эффективность распознавания при ограниченном количестве исходных данных, предотвращая переобучение и повышая обобщающую способность модели.

Полученная модель использовалась для первоначальной фильтрации датасета, классифицируя документы по принципу "принадлежит или не принадлежит" к классу СНИЛС. Это значительно улучшило точность дальнейших классификаций документов. В дальнейшем, рекуррентно дообучая нейросеть ResNet18 для определения одного вида документа и применяя её ко всему датасету, удалось эффективно разложить документы по каталогам.

**Увеличение количества классов документов:** В процессе обработки данных выяснилось, что некоторые документы являются двусторонними, где одна сторона содержит релевантную информацию, а другая нет. Это привело к необходимости дополнительной классификации и увеличению количества типов документов. Количество классов было увеличено до 19, список классов можно увидеть в таблице , что позволило более точно описывать и разделять документы внутри датасета.

*Таблица*   
Объем тренировочного датасета и категории документов

|  |  |  |
| --- | --- | --- |
| **Папка** | **Общий объем Мбайт** | **Файлы JPG** |
| **01 ПТС 1страница** | 2 587.7 | 1775 |
| **01 ПТС 2страница** | 2333.23 | 625 |
| **01 ПТС новые** | 2 258.5 | 806 |
| **02 СТС 1страница** | 1 332.7 | 1862 |
| **02 СТС 2страница** | 414.2 | 573 |
| **03 ОСАГО 1страница** | 6 036.7 | 2660 |
| **03 ОСАГО 2страница** | 1 654.0 | 1418 |
| **04 Разрешение на авто новое** | 129.1 | 259 |
| **04 Разрешение старое** | 679.4 | 729 |
| **05 Техосмотр** | 3 832.2 | 2778 |
| **06 Паспорт 1-я страница** | 2 437.8 | 1659 |
| **06 Паспорт прописка** | 742.6 | 643 |
| **07 ВУ 1 страница** | 1 190.4 | 2569 |
| **07 ВУ 2 страница** | 1 361.3 | 2783 |
| **07 ВУ старые права** | 210.7 | 381 |
| **08 СНИЛС 1 страница** | 812.9 | 1279 |
| **08 СНИЛС 2 страница** | 569.5 | 906 |
| **09 Судимость** | 456.4 | 282 |
| **10 ИНН** | 718.1 | 732 |

**Обучение и тестирование нейронной сети ResNet:** Для начального тестирования была создана сбалансированная выборка из 1000 документов различных типов. В течение 30 эпох модели ResNet18, ResNet50 и ResNet152 тренировались на этих тестовых данных. После завершения обучения проводилась проверка результатов категоризации на других реальных данных объемом тоже в 1000 изображений, используя ручную проверку классифицированных данных. Полученные результаты моделей представлены в таблице. Из таблицы можно увидеть, что модель ResNet50 продемонстрировала наивысшую точность на реальных данных. Это показало эффективность данной архитектуры для задач классификации документов, подтверждая её преимущество над другими рассмотренными моделями.

**Проверка моделей ResNet на классификацию реальных данных**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Название сети** | **название файла модели** | **для тестовых данных** | | **реальные данные** | | |
| **acc** | **loc** | **файлов** | **ошибок** | **acc** |
| **ResNet18** | model\_STRUCT\_DOCUMENT\_ResNet18\_30\_acc\_94.77\_loss\_0.1772.pth | 94,8 | 0,18 | 1000 | 98 | 90,2 |
| **ResNet50** | model\_STRUCT\_DOCUMENT\_ResNet50\_30\_acc\_95.16\_loss\_0.1699.pth | 95,2 | 0,17 | 1000 | 89 | 91,1 |
| **ResNet152** | model\_STRUCT\_DOCUMENT\_ResNet152\_19\_acc\_95.96\_loss\_0.1391.pth | 96 | 0.14 | 1000 | 99 | 90.1 |

**Обучение нейросети ResNet50 на полном датасете.** После тестирования различных моделей было принято решение использовать модель ResNet50 для обучения на полном датасете. Выбор этой модели был обусловлен её высокой эффективностью и оптимальным балансом между скоростью обучения и точностью результатов. Процесс обучения проводился в течение 40 эпох с использованием CPU, что заняло примерно 10 дней.

В результате обучения нейросеть ResNet50 продемонстрировала точность в 97.80% и низкое значение функции потерь, составляющее всего 0.0738%, это показано на рисунке ниже. Эти показатели подчеркивают высокую эффективность выбранной модели и её способность корректно классифицировать документы с высокой степенью точности.

Итоги работы с ResNet50:

Продолжительность обучения: 10 дней

Количество эпох: 40

Точность нейросети: 97.80%

Значение функции потерь: 0.0738%

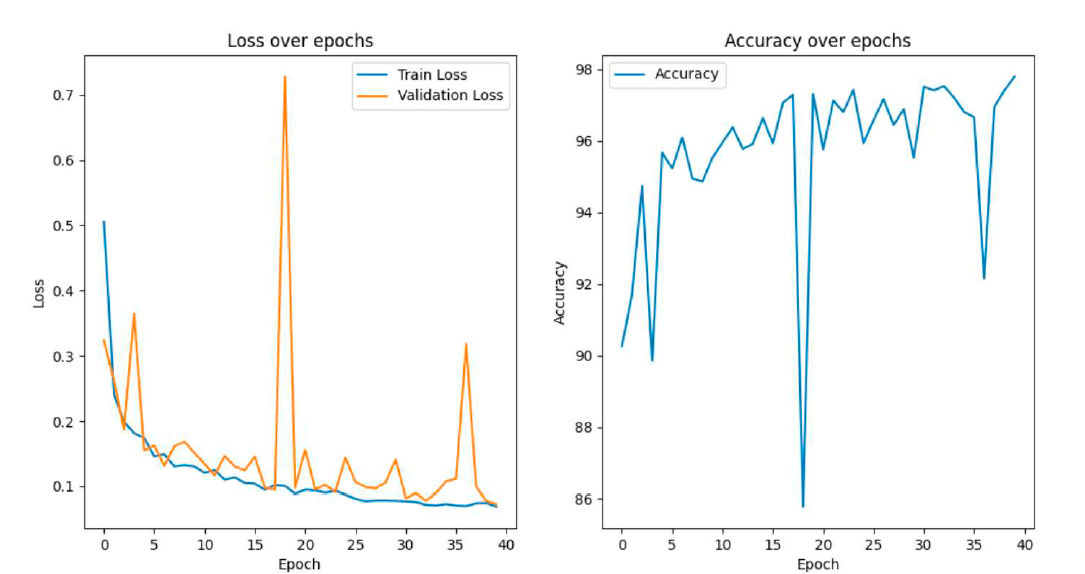
****

Рисунок – Обучение на реальных данных ResNet50

**Итоги работы с ResNet50.** Результаты подтвердили правильность выбора архитектуры ResNet50 для задачи классификации документов. Обученная модель была успешно применена в компании «Такси Город» для автоматизации процесса классификации новых документов и приведения в соответствие существующих данных. Это позволило компании значительно улучшить управление документами и повысить общую эффективность рабочих процессов.

ResNet50 показала высокую эффективность благодаря своему оптимальному балансу между скоростью обучения и точностью результатов. За 40 эпох обучения с использованием СPU, модель достигла точности в 97.80% и низкого значения функции потерь (0.0738%). Эти показатели подтвердили её способность корректно классифицировать документы.

Интеграция модели в корпоративные системы позволила автоматизировать задачи, связанные с обработкой документов, такие как сортировка, классификация и хранение. Это повысило продуктивность и эффективность работы, а также ускорило взаимодействие с государственными органами и партнёрами. Быстрая и точная классификация документов сократила время на регистрацию и управление автопарком, улучшив надёжность и безопасность документооборота.

ResNet50 также сыграла ключевую роль в улучшении качества обслуживания клиентов, позволив быстрее и точнее обрабатывать их запросы, что повысило удовлетворённость клиентов. Точная классификация документов способствовала снижению ошибок в документообороте и улучшила внутренний контроль и управление документацией.

## Заключение

Внедрение системы классификации документов на основе модели ResNet-50 стало важным шагом в цифровой трансформации компании "Такси Город Барнаул". В результате этого проекта компания смогла существенно улучшить качество работы с документацией, снизить затраты на её хранение и обработку, а также повысить безопасность и доступность данных. Применение современных нейронных сетей позволило не только автоматизировать процесс, но и сделать его более надежным и эффективным. В будущем планируется кроме классификации внедрить и распознавание самой структуры в документах, что укорит ввод первичных данных в компании.

## 

## СПИСОК ЛИТЕРАТУРЫ

1. Лапань М., Глубокое обучение с подкреплением. AlphaGo и другие технологии. – СПб.: Питер, 2020. – 496 с.: ил. – (Серия «Для профессионалов»).
2. Николенко С., Кадурин А., Архангельская Е., Глубокое обучение, погружение в мир нейронных сетей. – СПб.: Питер, 2018. – 480с.: ил. – (Серия «Библиотека программиста).
3. Smart Engines. [Электронный ресурс]. Заглавие с экрана. Режим доступа: <https://smartengines.ru> (дата обращения 01.09.2024).
4. Dbrain. [Электронный ресурс]. Заглавие с экрана. Режим доступа: <https://dbrain.io> (дата обращения 01.09.2024).
5. Распознавание паспортов. [Электронный ресурс]. Заглавие с экрана. Режим доступа: <https://beorg.ru/raspoznavanie-pasportov/> (дата обращения 01.09.2024).
6. Residual Neural Network (Res-Net). [Электронный ресурс]. Заглавие с экрана. Режим доступа: <https://en.wikipedia.org/wiki/Residual_neural_network> (дата обращения 01.09.2024).
7. Аксенов С.В., Новосельцев В.Б. Организация и использование нейронных сетей (методы и технологии) / Под общ. ред. В.Б. Новосельцева. – Томск: Изд-во НТЛ, 2006. – 128 с.
8. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с англ. А.А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.:
9. Горбань А.Н. Обучение нейронных сетей. – М.: изд. СССР-США СП "ParaGraph", 1990. – 160 с. (English Translation: AMSE Transaction, Scientific Siberian, A, 1993, Vol. 6. Neurocomputing, PP. 1-134).