

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ ФЕДЕРАЛЬНОЕ
ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ

Факультет физико-математических и естественных наук
Математический институт им. С.М. Никольского

"Допустить к защите"

Директор Математического института
им. С.М. Никольского, д.ф.-м.н.

_____ А. Б. Муравник

«__» _____ 2022г.

**Выпускная квалификационная работа магистра
на тему**

**"Сегментация изображений с использованием нейросетевой архитектуры
Transformer"**

Направление 01.04.01 – "Математика"

Выполнил:

Студент группы НМТмд-01-20 Студ.
билет No: 1032202184

_____ Е. Ю. Котляров

«__» _____ 2022г.

Руководитель:

Доцент, к.ф.-м.н.

_____ Я. М. Карандашев

«__» _____ 2022г.

Москва

2022 г.

Оглавление	
Введение	4
Обзор литературы	5
Методы	7
1. Постановка задачи	8
2. Данные	9
2.1 Описание данных	9
2.2 Методы предобработки	10
3. Описание архитектуры Transformer	19
3.1 Описание архитектуры Self attention	19
3.2 Архитектура transformer для анализа текста	19
3.3 Архитектура transformer для анализа изображений	21
4. Обогащение данных с использование архитектуры GAN	23
4.1 Описание архитектуры нейронной сети GAN	23
4.2 Архитектура GAN генерирующая изображения с инородными объектами	24
5. Классификация изображения разбитого на части	27
6. Сегментация изображения разбитого на части	29
Результаты	31
Вывод	32

Abstract

This paper deals with the problem of finding foreign objects on X-ray images obtained from personal inspection scanners. The available data have a number of problems, which are described and solved in the text. To solve the segmentation problem, a new approach is used where we first classify parts of the images, finding foreign objects, and then segment those parts. The Transformer neural network architecture is used as the classifier. This method of solving the problem allows to carry out segmentation without compressing the image and without increasing the size of the model or the amount of memory required.

Аннотация

В данной работе рассматривается задача нахождения инородных объектов на рентгеновских снимках, полученных со сканеров персонального досмотра. Имеющиеся данные имеют ряд проблем, которые описываются и решаются в тексте. Для решения задачи сегментации применяется новый подход, когда сначала мы классифицируем части изображений, находя инородные объекты, а затем сегментируем эти части. В качестве классификатора используется архитектура нейронной сети Transformer. Данный метод решения задачи позволяет проводить сегментацию не сжимая изображение, при этом не увеличивая размер модели или объём требуемо памяти.

Введение

Нейронные сети находят всё большее применение в различных задачах нашей жизни. Уже сейчас они позволяют ускорить поиск информации, находить людей по фотографии, детектировать дорожные знаки и прочее. Это в значительной сфере ускоряет рабочий процесс, по этой причине различные сферы бизнеса стараются внедрить в работу решения на основе нейронных сетей.

Данная работа посвящена созданию автоматизированного решения для поиска запрещённых объектов на снимках со сканеров персонального досмотра. Это позволит в значительной степени ускорить время проверки людей на пунктах досмотра, а также улучшит качество проверки, так как решение на основе нейронных сетей позволит находить даже очень мелкие объекты.

Обзор литературы

Сначала рассмотрим развитие архитектуры внимания(attention), и применение её в нейронных сетях трансформер, затем покажем развитие нейронных сетей для задачи сегментации рентгеновских изображений.

Подход вычисления внимания впервые был применён к нейронным сетям RNN для задачи анализа текста в 2014 году [1], что позволило улучшить качество модели перевода текста. Это послужило отправной точкой к исследованию архитектур нейронных сетей на основе attention. Следующим важным событием в развитии нейронных сетей послужила статья [2] 2017-го года, в которой был предложен новый подход внимания – само-внимание (self-attention), который является одним из слоёв новой архитектуры - transformer. В этом подходе вычисляется взаимосвязь слов в тексте друг с другом, что приводит к высоким результатам во всех задачах анализа текста. В 2020 был предложен подход анализа изображений на основе архитектуры transformer. Изображение разбивается на квадратики 16x16, и используются в качестве слов.

Сегментация объектов на рентгеновских снимках является весьма распространённой задачей. Сначала эта задача решалась с помощью классических методов обработки изображений [3, 4]. Но со временем эта задача стала решаться преимущественно с помощью свёрточных нейронных сетей. В работе [5] произведена модификация архитектуры SegNet [6], упрощающая оригинальную сеть и позволяющая проводить обучение на небольшом наборе данных. Дальнейшее развитие сети SegNet - архитектура XNet [7], специализирующаяся на сегментации рентгеновских изображений, а именно на разделении мягкой ткани и костей. Задача сегментации получила особое распространение в сфере медицины. Огромное количество работ было посвящено сегментации клеточных структур, например [8, 9]. Среди множества подобных работ выделяется сеть U-Net [8], подходящая для более широкого класса задач. Эта сеть входит в число базовых архитектур для сегментации. В ней используются только свёрточные слои, что позволяет

подавать на вход изображения произвольного размера и получать маску с классами на выходе.

Методы

В данной работе используются различные подходы к улучшению изображения: ThreshTrunc, EqualizeHist, EqualizeAdapthist.

Используются различные архитектуры нейронных сетей: GAN, Transformer, Unet.

Все вычисления проводились с помощью языка программирования python версии 3.7. Для методов анализа изображений использовалась библиотека skimage. Для работы с нейронными сетями задействована библиотека torch.

1. Постановка задачи

Требуется создать автоматизированное решение поиска запрещённых(инородных) объектов на изображениях, полученных из сканеров персонального досмотра. В качестве данных предоставлено 4 набора данных, полученных с разных сканеров персонального досмотра. В данных существует две проблемы:

- Распределение интенсивности пикселей в снимках, полученных с разных наборов данных отличаются
- Визуально невозможно разобрать что находится на снимке

Рассмотренные проблемы с данными решаются с помощью алгоритмов предобработки рентгеновских изображений.

Следующей задачей является увеличение вариативности различных запрещённых объектов, представленных на снимках. Так как заранее неизвестно что может быть у человека в качестве запрещённого объекта, а набор данных ограничен несколькими тысячами снимков, то требуется разработать алгоритм, который сможет генерировать и накладывать запрещённые изображения на снимки.

Финальной задачей является поиск всех объектов, не принадлежащих телу человека. Предложен новый подход для решения этой задачи.

2. Данные

2.1 Описание данных

Данные представляют из себя изображения в формате .tiff. Все снимки одноканальные, то есть в них содержится только значение интенсивности пикселя. Значения интенсивности варьируются от 0 до 2^{16} . Размер изображений различный, но составляет в среднем 1700 на 600 пикселей. Всего предоставлено 1643 изображения. Все предоставленные снимки сделаны со спины человека, и в полный рост. На 1343 снимках есть те или иные объекты, не принадлежащие телу человеку, например застёжки, пуговицы, ножи, проволока, пистолеты. Оставшиеся изображения не содержат никаких посторонних объектов.

Все снимки предоставлены без разметки, поэтому потребовалось воспользоваться ресурсом для разметки изображений – [яндекс.толока](#)[10]. Данный ресурс позволяет быстро разметить большой объём данных.

По результатам разметки получены булевы маски, где значение 1 соответствует инородному объекту, значение 0 – отсутствию инородного объекта. Всего было получено около 18000 запрещённых объектов. Пример изображения и маски к нему представлены на рис. 1.

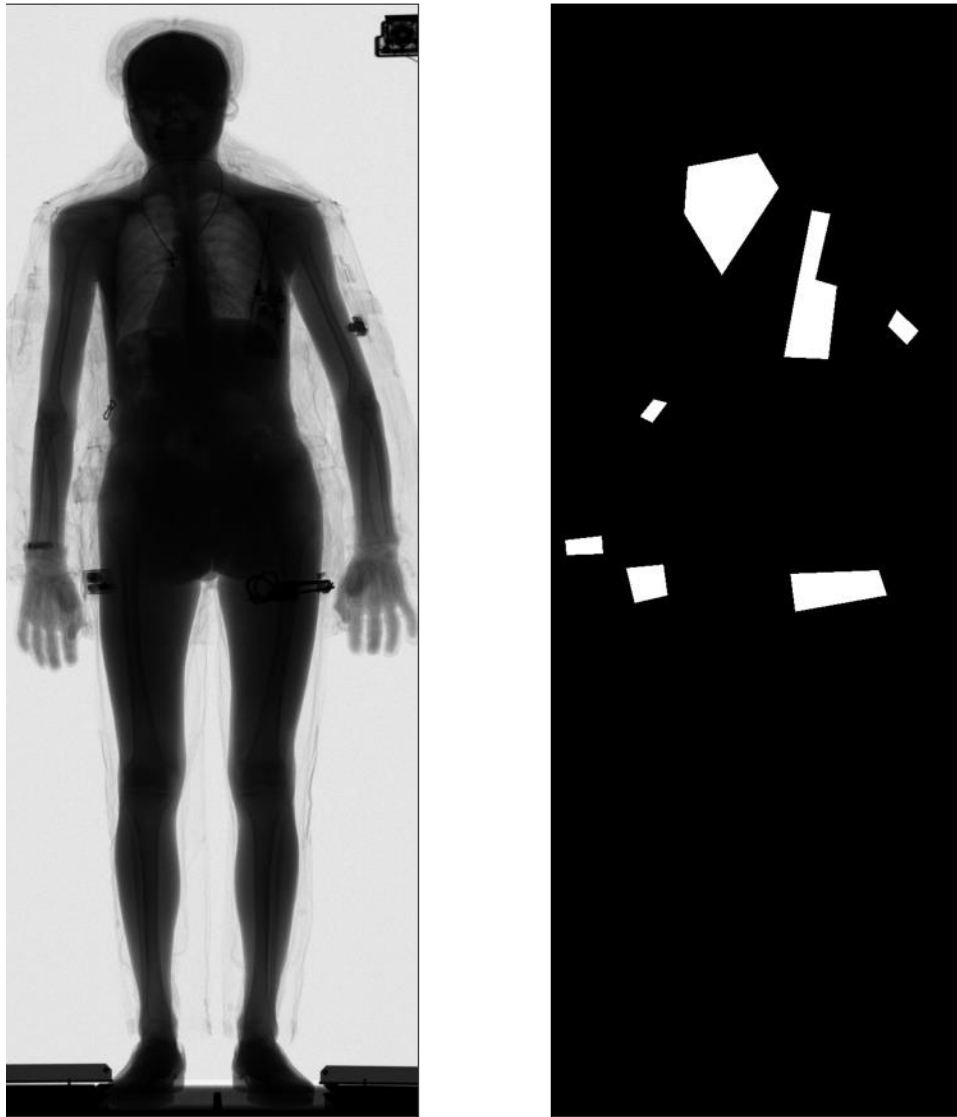


Рисунок 1. Пример изображения и маски к нему

2.2 Методы предобработки

Как можно видеть на рис. 1 значение интенсивности пикселей оригинальных изображений достаточно низкое и очень сложно анализировать изображение визуально. Для решения этой проблемы предлагается последовательное использование нескольких операций по улучшению интенсивности изображений.

Для начала требуется ввести понятие гистограммы распределения интенсивности пикселей. Посчитаем количество пикселей для каждого уникального значения интенсивности пикселей. Отсортируем значения пикселей от 0 до максимального значения, и построим гистограмму распределения. На рис. 2 можно видеть, что наибольшее количество пикселей

сосредоточено возле значения интенсивности 2246 (светлый фон) и возле значения интенсивности равным 7 (тёмные объекты).

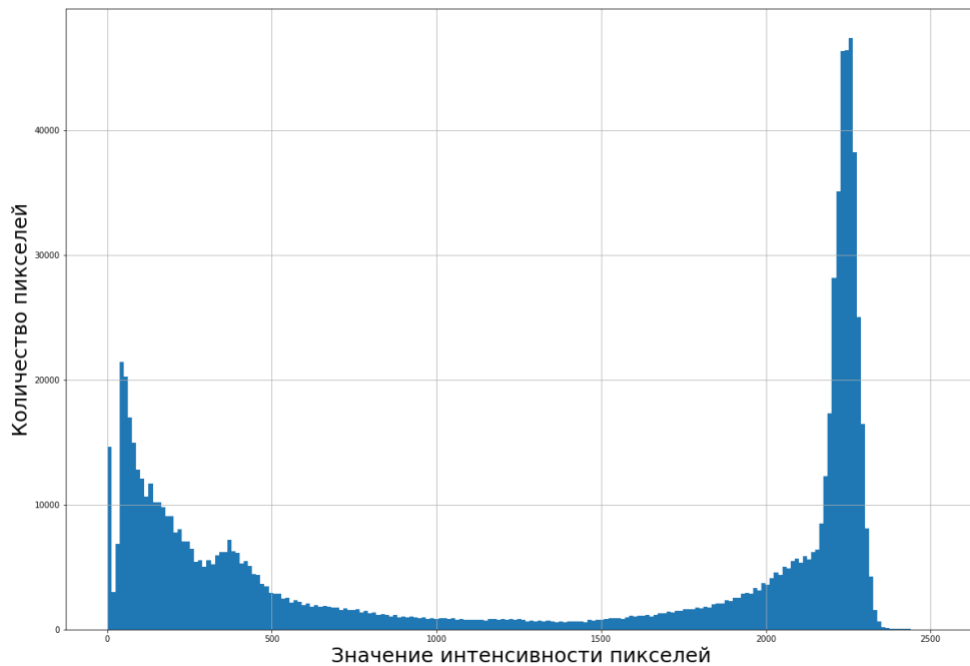


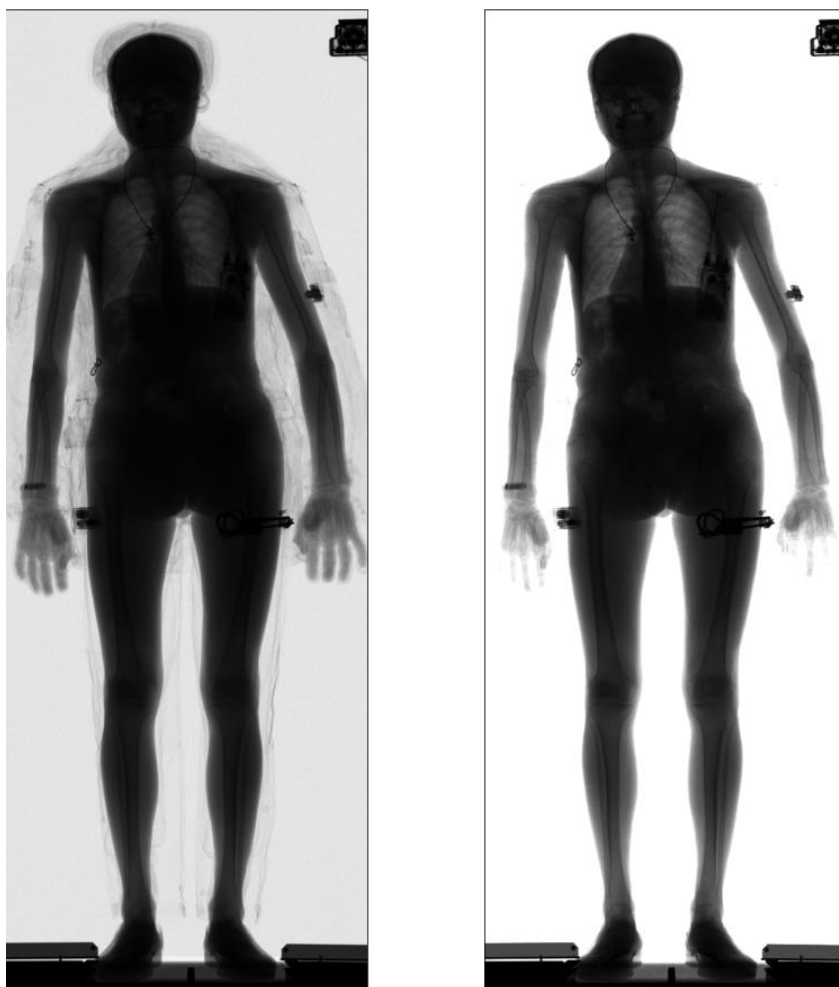
Рисунок 2. Гистограмма распределения интенсивности пикселей

ThreshTrunc [11] – операция изменения снимка, срезающая чрезмерно высокие значения интенсивности пикселей. В случае предоставленных снимков использовался частный случай этого метода, когда значения интенсивности выше среднего значения заменяются на средние значения интенсивности по всему изображению. Применение этого метода влечёт за собой смещение гистограммы распределения интенсивности пикселей в более светлую часть. К тому же это убирает некоторые шумы, которые возникают из-за особенностей рентгеновских снимков, когда некоторые значения интенсивности пикселей на снимке сильно выше значений интенсивности соседних пикселей. Ниже представлена формула операции:

$$\text{ThreshTrunc}(x, y) = \begin{cases} \text{src}(x, y), & \text{если } \text{src}(x, y) < \text{threshold} \\ \text{threshold}, & \text{если } \text{src}(x, y) \geq \text{threshold} \end{cases} ,$$

где threshold – порог равный среднему значению интенсивности всех пикселей снимка, $\text{src}(x, y)$ - значение интенсивности пикселя с координатами (x, y) .

Применение операции к оригинальному снимку показано на рис. 3. Можно видеть, что весь фон стал светлым и однородным. Исчезли низко контрастные объекты и снимок стал более светлым. На рис. 4 представлена гистограмма распределения пикселей после применения операции *ThreshTrunc*. Как и ожидалось, больше всего пикселей сосредоточены в максимальной точке - предыдущем среднем значении.



*Рисунок 3. Применение операции *ThreshTrunc**

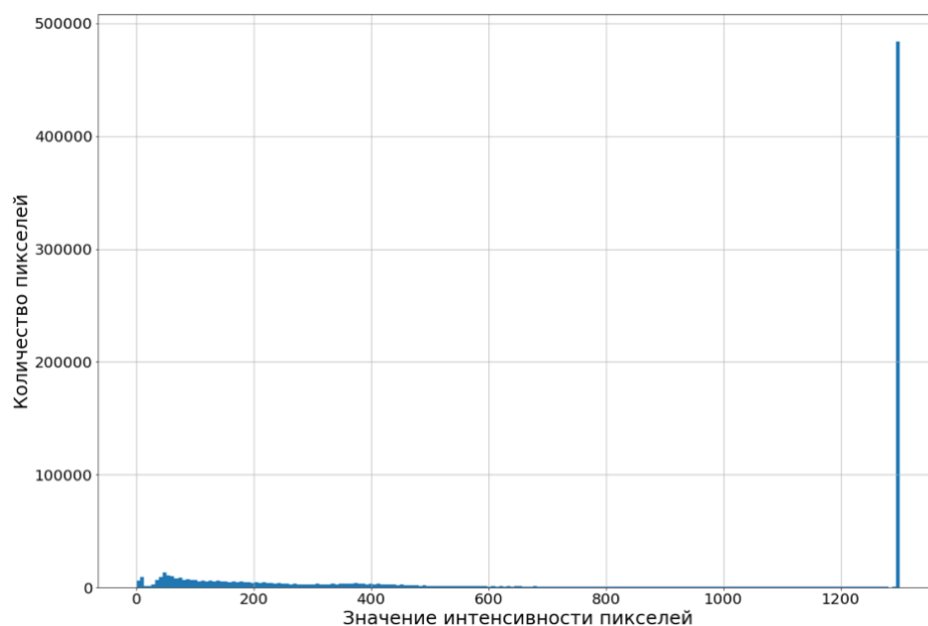


Рисунок 4. Гистограмма распределений после применении операции `ThreshTrunc` к оригинальному изображению

`EqualizeHist` – процедура выравнивания гистограммы изображения [12], позволяющая улучшить общий контраст изображения. Данный метод уменьшает влияние высоких значений интенсивности и увеличивает влияние низких, что способствует улучшению видимости больших низко контрастных объектов, делая более прозрачными «тёмные» участки снимка. Является одним из базовых способов предобработки рентгеновских изображений.

$H(i)$ – гистограмма для значения интенсивности каждого пикселя, то есть по x – значения интенсивности пикселя, по y – количество пикселей с такой интенсивностью. Находим кумулятивное распределение:

$$H'(i) = \sum_{0 \leq j < i} H(j)$$

Затем заменяем значения интенсивности пикселя в изображении на получившиеся значение из распределения:

$$\text{EqualizeHist}(x, y) = H'(\text{src}(x, y)),$$

Которые потом делятся на количество всех пикселей. Поэтому значения получаются от 0 до 1.

Пример преобразования приведён на рис. 5. Можно видеть, что низко контрастные объекты (верхняя одежда) стали хорошо видны, при этом

интенсивность пикселей тела стало заметно выше (светлее), можно разглядеть невидимые ранее объекты (рация, цепочка и др).

Значения интенсивности пикселей в гистограмме распределения пикселей (рис 6) выглядит довольно равномерно, что говорит о том, что она сглаживает все максимумы и минимумы интенсивности, приводя их к среднему значению.

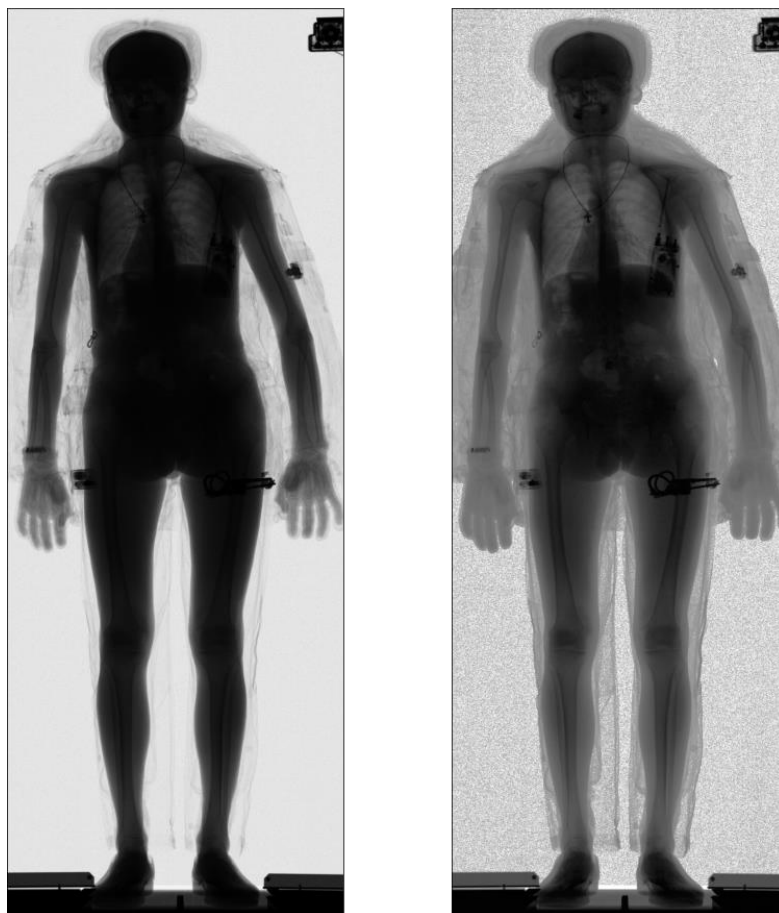


Рисунок 5. Применение преобразования EqualizeHist к оригинальному снимку

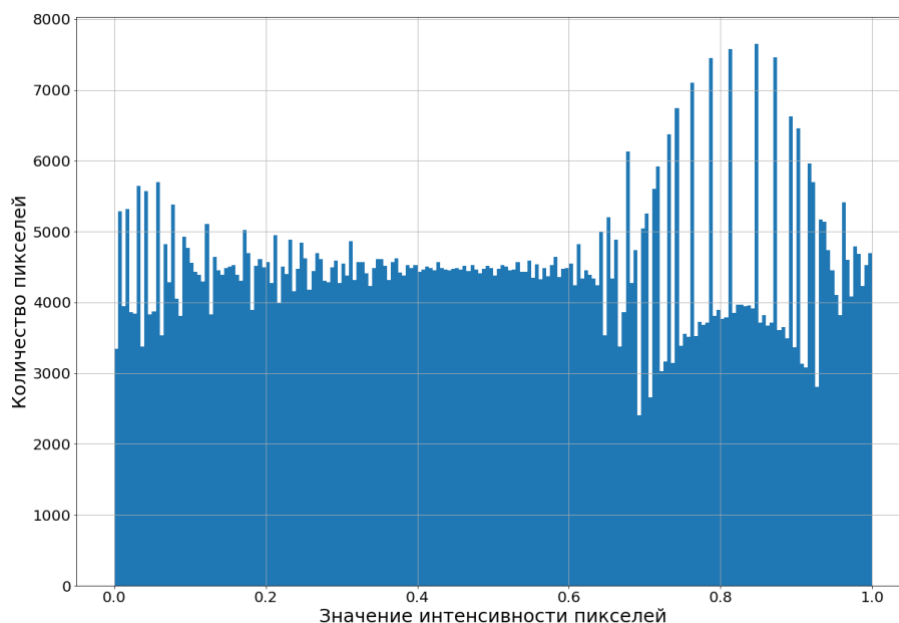


Рисунок 6. Гистограмма распределений после применение операции $EqualizeHist$

$EqualizeAdaphist(CLAHE)$ - адаптивное выравнивание гистограммы. Этот метод базируется на $EqualizeHist$, но в отличии от него выравнивает гистограммы на отдельных участках изображения, что позволяет учитывать локальный контекст интенсивности. Применительно к нашей задаче, он сильно улучшает видимость низко контрастных объектов, которые занимают не очень большое место на снимке.

Результат работы алгоритма представлен на рис. 7. Можно заметить локальное повышение контрастности одежды, при это более тёмные объекты стали светлее. Гистограмма распределения показана на рис. 8, по ней можно судить что значения интенсивности пикселей немного распределились по координате x . Это говорит о сглаживании максимумов и минимумов распределения.

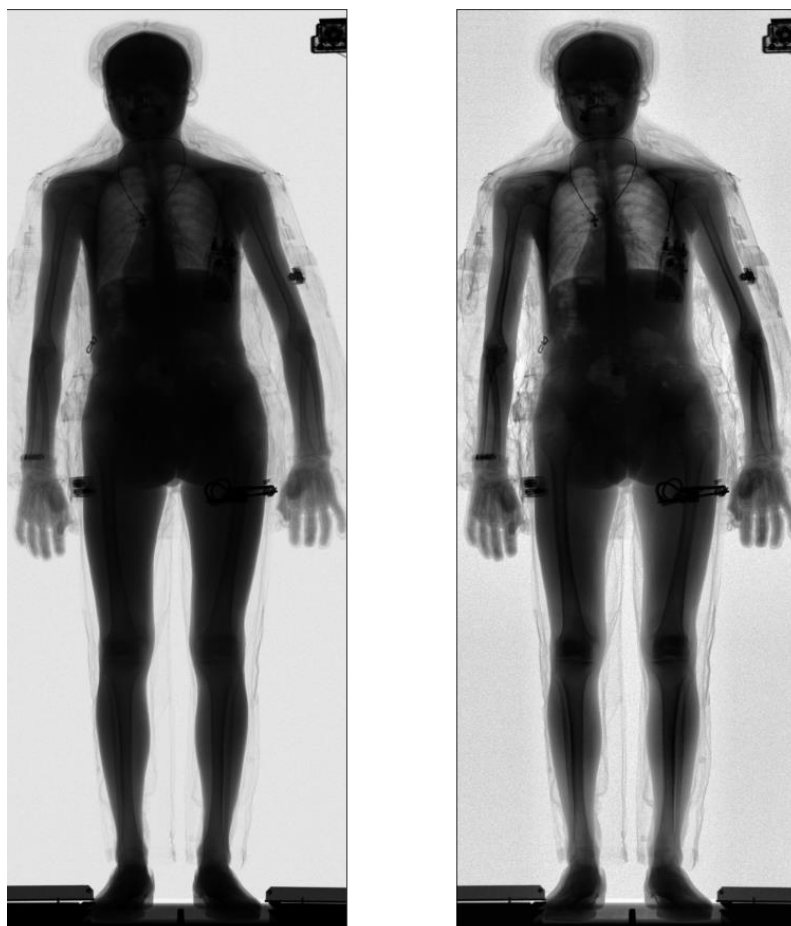


Рисунок 7. Применение операции EqualizeAdapthist к оригинальному изображению

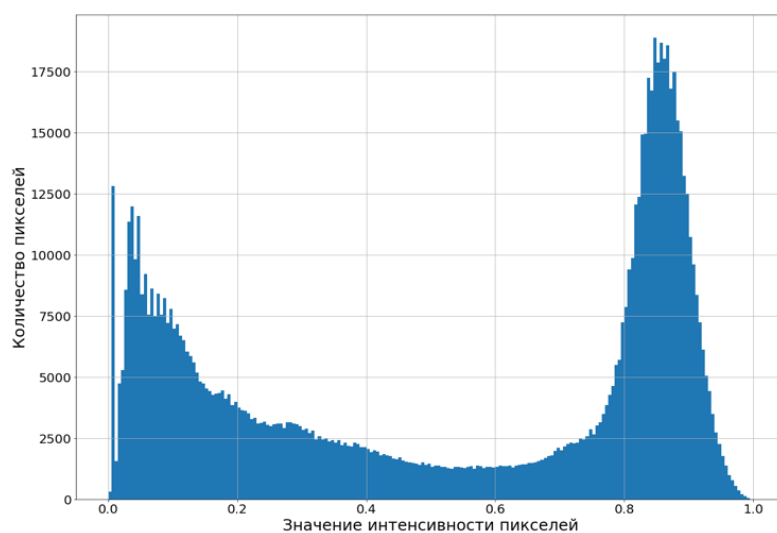


Рисунок 8. Гистограмма распределения интенсивности пикселей после применения операции EqualizeAdapthist

Финальное преобразование проводилось в такой последовательности:

1. ThreshTrunc
2. EqualizeHist
3. ThreshTrunc
4. EqualizeAdapthist
5. ThreshTrunc

Это позволяет в значительной мере нормализовать значения интенсивности пикселей так, что все объекты становятся визуально различимы, и средние значения интенсивности на снимках с разных сканеров персонального досмотра совпадают.

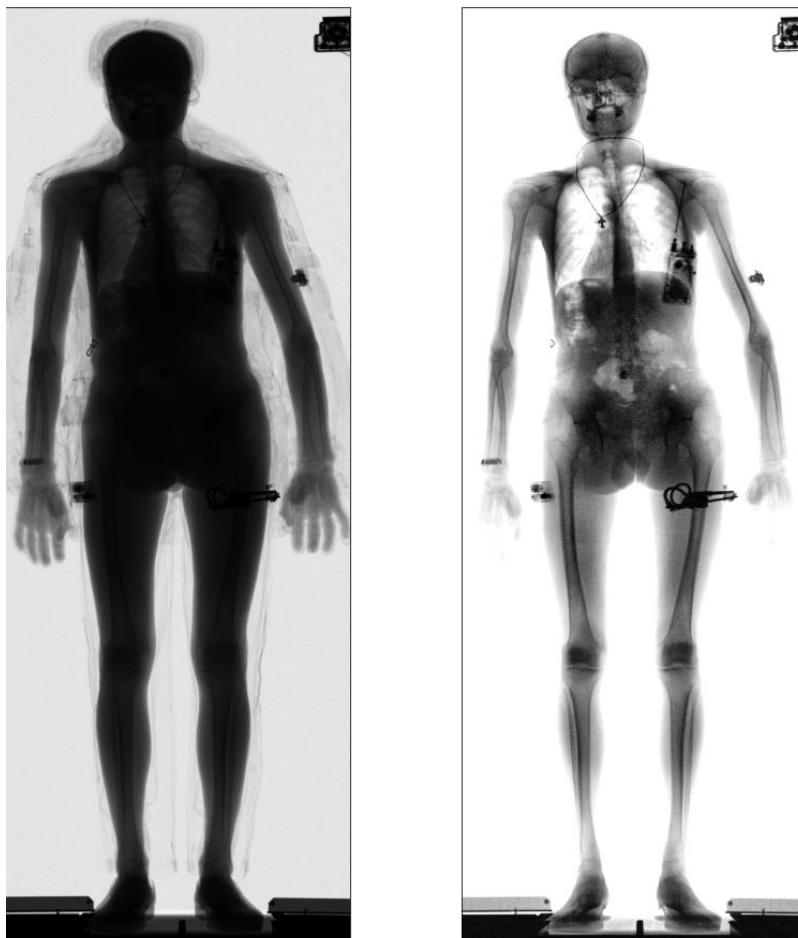


Рисунок 9. Пример изображения после применения последовательности преобразований

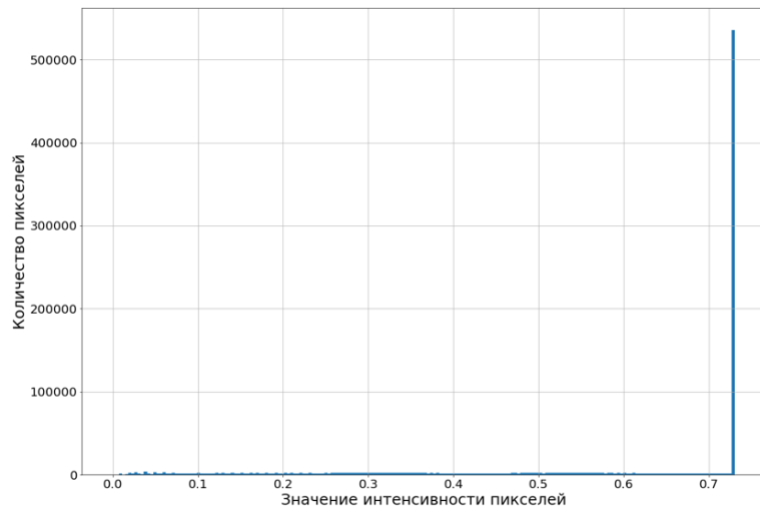


Рисунок 10. Гистограмма распределения пикселей после применения последовательности преобразований

Проведя все преобразования, мы получаем распределение пикселей, представленное на рис. 10. Как следует из этой гистограммы, большинство пикселей являются светлыми (т.е. фоном и участками мягких тканей), а остальные участки снимка распределены по градациям интенсивности более равномерно.

После применения всех преобразований (рис. 9) металлические объекты на снимке становятся визуально различимы. Это позволяет выделять инородные телу объекты, не прибегая к специализированному программному обеспечению для индивидуальной предобработки каждого изображения. В приложении 1 показан процесс улучшения снимка после каждого этапа обработки.

3. Описание архитектуры Transformer

3.1 Описание архитектуры Self attention

Self attention - один из видов механизма внимания, задачей которого является выявление закономерности между входными данными. Данный метод оказался очень эффективным в задаче машинного перевода и позволил отказаться от использования сетей RNN, которые долгое время занимали лидирующие позиции в этом направлении. На данный момент вместо RNN используются полносвязные нейронные сети в комбинации с механизмом Self attention. Данное изменение позволяет значительно ускорить работу алгоритма, так как все вычисления могут выполняться параллельно.

Для описания функции само внимания нужно ввести некоторые обозначения. Дан эмбединг вектора нескольких слов. Для вычисления взаимозависимости этих слов, требуется получить три матрицы: Q, K, V. Они получаются путём применения трёх различных полносвязных однослойных нейронных сетей к каждому из эмбедингов слова. Затем мы производим матричное перемножение вектора Q и транспонированного вектора K, делённое на \sqrt{d} , где d – длина векторов. Деление необходимо для масштабируемости значений. Таким образом мы получаем взаимосвязь слов друг с другом. Применяя функцию softmax построчно, мы нормализуем значения полученной матрицы, так что сумма каждой строки равна 1. Затем мы умножаем получившуюся матрицу на матрицу V. Таким образом мы вычисляем самовнимание.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

3.2 Архитектура transformer для анализа текста

В статье [2] впервые была представлена архитектура трансформатора (рис. 11). Эта архитектура основана на механизме Self-attention. Представленная архитектура хорошо подходит для обработки последовательностей. В настоящее время она почти полностью заменила

рекуррентные нейронные сети, поскольку показывает гораздо лучшие результаты почти во всех задачах, связанных с последовательностями.

Давайте посмотрим, как эта архитектура работает на задаче перевода текста с одного языка на другой. Сеть состоит из двух частей - кодера и декодера. Сначала рассмотрим кодировщик. Слово преобразуется в числовую форму с помощью операции эмбединга [13]. Затем оно передается на слой самовнимания, затем нормализуется и передается в полносвязную нейронную сеть, после чего снова нормализуется. Эти операции можно повторить несколько раз, что позволит увеличить размер сети и получить лучшие результаты. Затем, после всех этих операций, появляются закодированные векторы. Они подаются в декодер вместе со словом на другом языке. После прохождения слоя внимания, нормализации и полносвязной сети модель выдает вектор, длина которого равна длине количества всех слов. Каждое значение вектора ассоциируется со словом. А наибольшее значение вектора - это слово на другом языке.

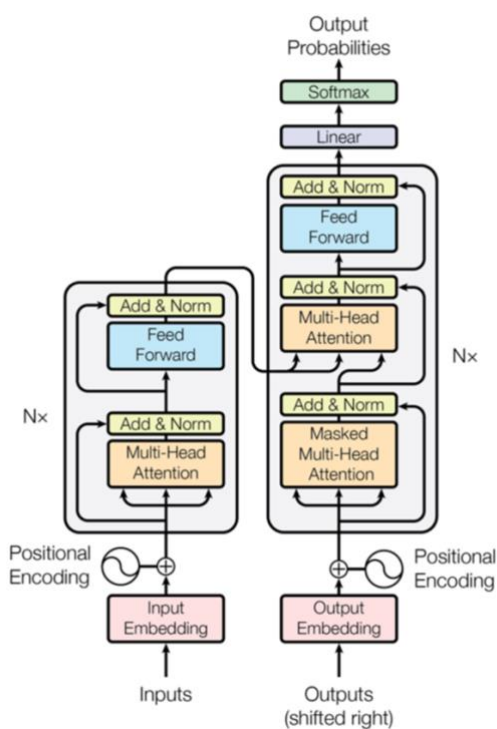


Рисунок 11. Архитектура transformer

В рассматриваемой статье сеть работает над задачей перевода текста. То есть над задачей обработки текста.

3.3 Архитектура transformer для анализа изображений

После успеха архитектуры transformer исследователи начали применять ее для решения других задач. Работа [14] посвящена применению этой технологии для обработки изображений, а именно для решения задачи классификации изображений. Это задача, в которой на вход подается изображение, а модель выдает вероятность того, какие объекты присутствуют на изображении. Авторы описывают новую архитектуру, которая называется Vision Transformer - ViT (рис. 12).

Поскольку нейронная сеть основана на архитектуре трансформатора, сначала необходимо представить изображение в виде последовательности. Для этого авторы предложили оригинальный подход, в котором не используются сверточные нейронные сети. Изображение разбивается на плитки размером 16x16 пикселей. Затем каждая плитка раскладывается в одномерный массив, который уже можно передать в модель. Чтобы определить расположение данного набора пикселей на изображении, исследователи отправили в модель расположение пикселей на изображении вместе с разложенными плитками. После обработки каждой плитки выходные векторы объединяются в массив и умножаются с дополнительными весами, что позволяет объединить информацию от всех плиток для сбора информации о полном изображении. Далее работает полносвязная сеть, на которую подается полученный вектор, и она выдает класс объекта. В этой архитектуре, в отличие от предыдущей статьи, мы не используем декодер, только кодер. Это связано с тем, что нам не нужно подавать несколько видов информации, как при переводе, когда сначала нужно подать слово на одном языке, а затем на другом.

Для обучения такой сети требуется много изображений. Но результаты побили рекорд во многих задачах классификации.

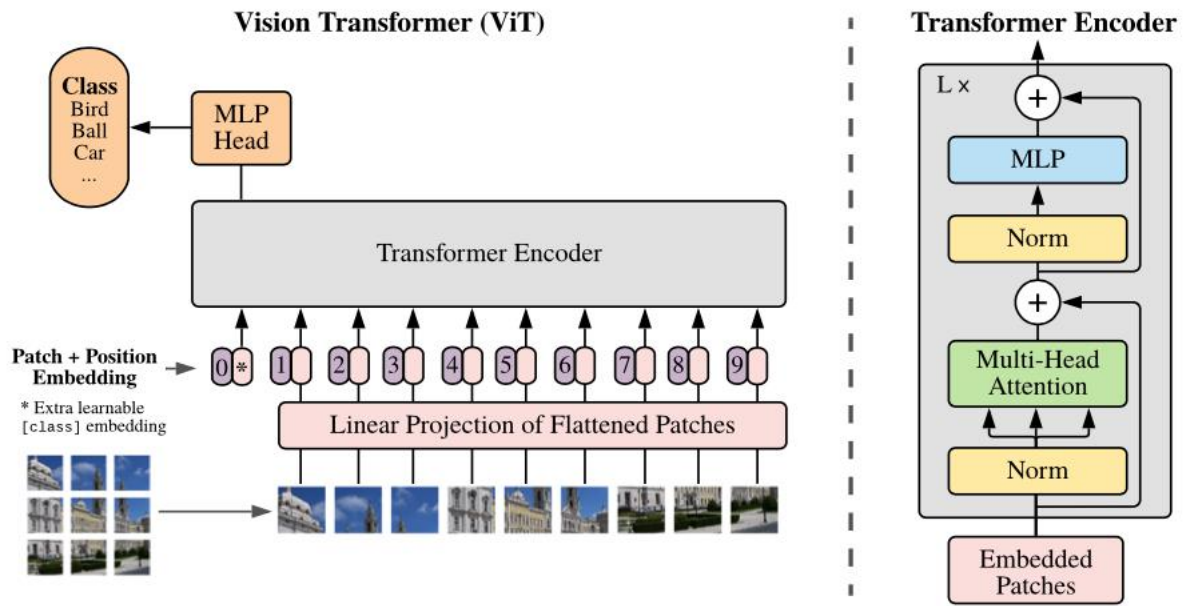


Рисунок 12. Архитектура ViT (трансформер для изображений)

4. Обогащение данных с использование архитектуры GAN

Одна из проблем данных – малое количество данных. Мощные модели будут быстро переобучаться, то есть запоминать выборку. Для преодоления это проблемы предлагается новый подход – генерировать дополнительные запрещённые объекты в случайных местах. Для этого применены генеративные нейронные сети.

4.1 Описание архитектуры нейронной сети GAN

Классический GAN [15] представляет из себя две нейронные сети, одна из которых называется генератор, другая называется дискриминатор.

Генератор – нейронная сеть, генерирующая необходимый объект, в нашем случае это изображение. В классическом варианте в генератор подаётся случайный вектор, из которого он должен сгенерировать реалистичное изображение.

Дискриминатор – нейронная сеть, которая учится понимать, какие изображения сгенерировал генератор, а какие являются настоящими.

Обе сети обучаются одновременно, конкурируя друг с другом, то есть генератор учится генерировать всё более правдоподобные изображения, а дискриминатор учится отличать их от настоящих. Сеть считается обученной, когда вероятность предсказания дискриминатора равняется $\frac{1}{2}$, что говорит о том, что генератор обучился, и дискриминатор не может отличить сгенерированное изображение и настоящее. Для практического эксперимента использовался модифицированный GAN - WGAN-GP [16] показывающий более стабильные результаты при обучении.

4.2 Архитектура GAN генерирующая изображения с инородными объектами

Пусть дано изображение – X , возьмём Ω – часть изображения X размера $n \times n \times 1$. Наша задача – найти такую функцию, которая отображала бы $G: \Omega \rightarrow Z$, где Z – часть изображения с нарисованным инородным объектом. Для обучения GAN введём дискриминатор, который умеет отличать изображение на котором изображён запрещённый объект от изображения без запрещённого объекта. $D_a: K \rightarrow \{0, 1\}$; $\Omega, Z \subset K$. Где 0 – на изображении нет инородного объекта (Ω), и 1 – в изображение присутствует инородный объект (Z).

Положим X_Z – изображение с заменённой на Z часть. Наложим дополнительное ограничение, введём функцию, которая умеет отличать изображение без изменений от изображения с наложенной частью $D_i: H \rightarrow \{0, 1\}$; $X, X_Z \subset H$. Где 0 – на изображении нет изменений (X), и 1 – в изображение присутствует вставлена часть (X_Z). Z в X должно не нарушать общего вида изображения.

Таким образом архитектура GAN состоит из трёх частей, а именно: одного генератора и двух дискриминаторов.

Опишем параметры обучения более подробно. Из изображения берётся прямоугольник размера 100×100 , вырезанный в случайно области. К этой вырезанной части добавляется канал со случайными значениями, сгенерированными из нормального распределения, это необходимо для генерации различных инородных объектов. Затем получившейся тензор подаётся в нейронную сеть, которая является генератором состоящем из энкодера и декодера. Энкодер отображает изображение в вектор, затем декодер восстанавливает изображение. На выходе генератора получается изображение такого же размера, на котором изображено реалистично сгенерированная вырезанная часть изображения и правдоподобный запрещённый объект.

Так же вторым каналом генератор генерирует маску для инородного объекта. Один дискриминатор учится понимать реалистично ли сгенерирован

инородный объект и его маска. Второй дискриминатор получает на вход всё изображение с вставленной сгенерированной частью и смотрит, реалистично ли выглядит весь снимок. Таким образом генератор учится генерировать правдоподобные части изображения, но с помещённой на них инородными объектами, а также маски для этих объектов.

На рис. 13 показан пример сгенерированной части изображения (обведено чёрной рамкой) и маску для сгенерированного участка. Часть ноги восстановлена достаточно хорошо, при этом добавились инородные объекты, при этом маска соответствует сгенерированным инородным объектам. В приложении 2 больше примеров генерации.

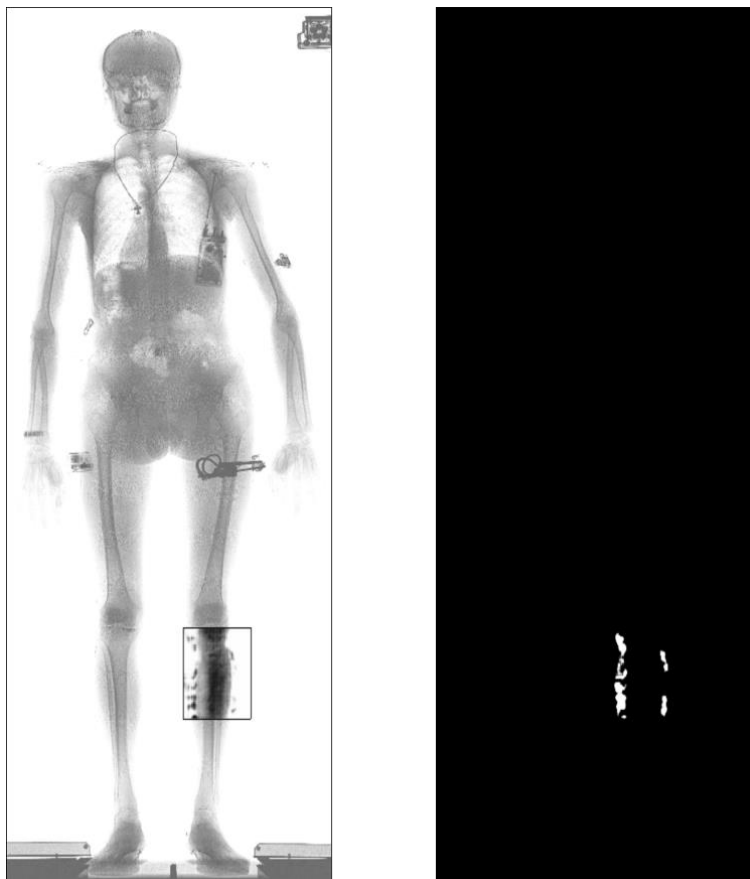


Рисунок 13. Пример сгенерированной части изображения

В качестве дискриминатора используется нейронная сеть unet [8] глубины 2 и 64 каналами свёртки. Алгоритм обучался с помощью алгоритма оптимизации adam[17] с шагом обучения 0.0001.

В качестве дискриминаторов использовалось две нейронные сети ResNet18[18], оптимизатором также послужил алгоритм adam с шагом обучения 0.0001.

Обучение проводилось на видеокарте Nvidia GTX 1080ti с 11 гигабайтами памяти. Размер батча равняется 10. Прохождение одной эпохи занимало 20 минут.

5. Классификация изображения разбитого на части

Так как в нашей задаче не важен глобальный контекст изображения, нужно просто понимать есть ли в определённом месте инородный объект или нет, то задачу можно переформулировать следующим образом: нужно классифицировать есть ли в выделенной части изображения запрещённый объект или нет. Для этого введём функцию $C: x_i \rightarrow [0, 1]$, где $x_i \subset X$, $x_i \cup x_j = \emptyset \forall i \neq j$, $i, j \in [1, N]$, функция C является обучаемой нейронной сетью-классификатором.

Таким образом мы приходим к новому подходу решения задачи выделения запрещённых объектов на теле человека. Разбиваем изображение на равные плитки, затем каждую плиточку независимо классифицируем на присутствие инородного объекта, таким образом получаем карту запрещённых объектов на изображении. В качестве классификатора используется нейронная сеть ViT, с оптимизатором adam и с шагом обучения 0.001. Обучение проводилось на видеокарте Nvidia GTX 1080ti с 11 гигабайтами памяти. Размер батча равняется 2. Прохождение одной эпохи занимало 10 минут. К последнему слою применялась функция softmax, обучение проводилось с помощью функции потерь – бинарная кросс энтропия. Преимущество такого метода – не нужно сжимать изображение, что позволяет сохранять детали снимка, что очень важно для поиска мелких объектов, к тому же такой подход позволит оператору смотреть только на важные участки, и полностью игнорировать не важные части изображения.

На рис. 14 приведён пример классификации участков размера 100x100 изображения со сгенерированной генератором частью. Модель выделила все ячейки, в которых присутствуют инородные объекты. В дальнейших экспериментах размер квадратиков может быть уменьшен для большей детализации, но это потребует дополнительной информации об их окружении.

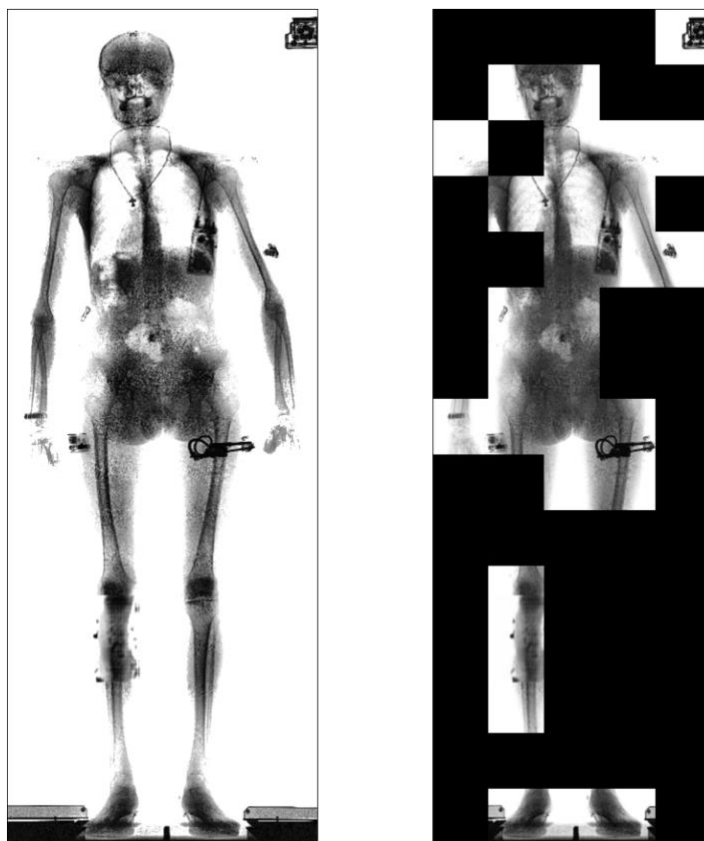


Рисунок 14. Пример классификации отдельных частей изображения

6. Сегментация изображения разбитого на части

На предыдущем шаге мы получили маску запрещённых объектов для рентгеновского снимка человека. Дальнейшим шагом является непосредственное нахождение инородных объектов на теле. Так как мы заранее (с помощью классификатора) знаем в какой части изображения находятся такие объекты, мы можем подавать в сегментирующую нейронную сеть только их, выкинув ненужные нам части. Это приведёт к более выстрой сегментации, за счёт того, что нейронной сети нужно провести меньше вычислений. К тому же считается, что задача классификации является более простой по отношению к задаче сегментации, что предположительно позволяет классифицирующей нейронной сети замечать более мелкие объекты, которые сегментирующая нейронная сеть затем выделит, потому что точно знает, что на этом месте что-то есть.

В качестве сегментирующей сети используется Unet [8] глубиной сети 5 и 64 каналами свёртки. Для обучения применялся алгоритм adam со скоростью обучения равной 0.001 и размером батча 2. Обучение проводилось на видеокарте Nvidia GTX 1080ti с 11 гигабайтами памяти. Размер батча равняется 3. Прохождение одной эпохи занимало 3 минуты. Применялась лосс функция dice.

Как можно видеть на рис. 15, модель хорошо выделяет инородные объекты. В некоторых местах заметны полосы на месте соединения частей изображения. Но несмотря на это геометрический вид объектов сохраняется. Так же видно, что модель успешно выделила все мелкие объекты, которые могли быть стёрты при сильном сжатии изображения.

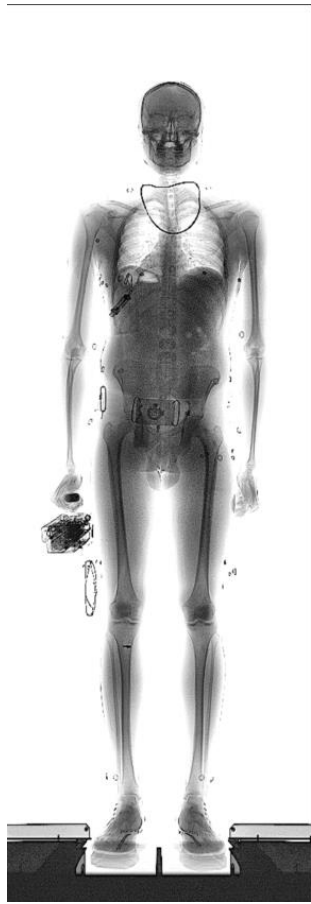


Рисунок 15. Сегментация изображения по частям

Результаты

Описаны алгоритмы преобразования рентгеновского изображения и разработана последовательность применения рассмотренных алгоритмов обработки рентгеновских изображений, полученных со сканеров персонального досмотра, которая переводит гистограмму распределения интенсивности пикселей в более равномерную форму, в которой все объекты на снимки различимы глазу.

Рассмотрена проблема нехватки данных, а именно вариации различных запрещённых объектов. Данная проблема была решена с помощью нейронной сети архитектуры GAN, которая позволяет генерировать правдоподобную часть тела человека, помещая на неё инородный объект.

Так же разработан новый подход к сегментации рентгеновских снимков со сканеров персонального досмотра, в которых не важен общий контекст. Суть метода заключается в разбиении изображения на независимые части, которые классифицируются на наличие запрещённого объекта. По сути, строится карта инородных объектов на теле человека. Затем сегментирующая сеть сегментирует части изображения, в которых классификатор нашёл инородных объект. Данный подход позволяет находить объекты не сжимая изображения.

Вывод

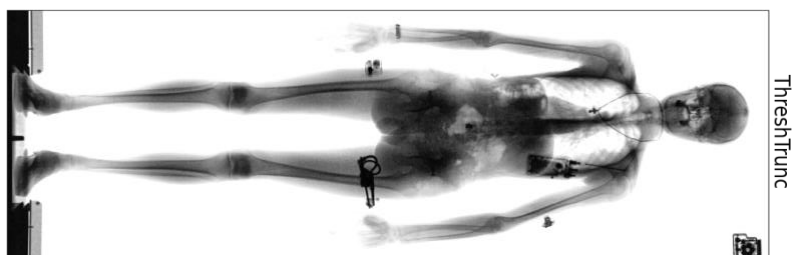
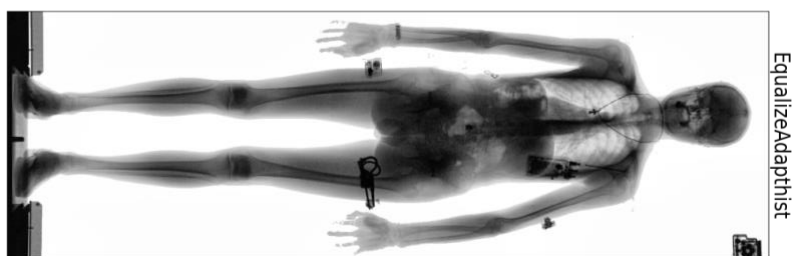
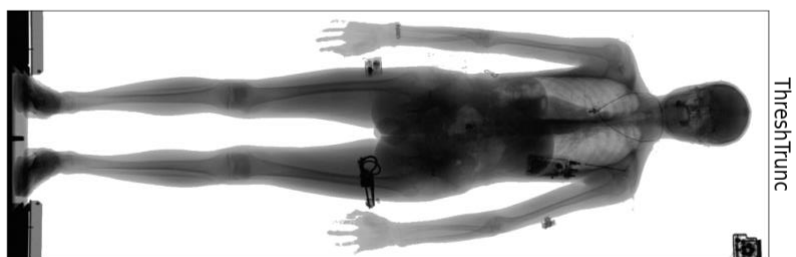
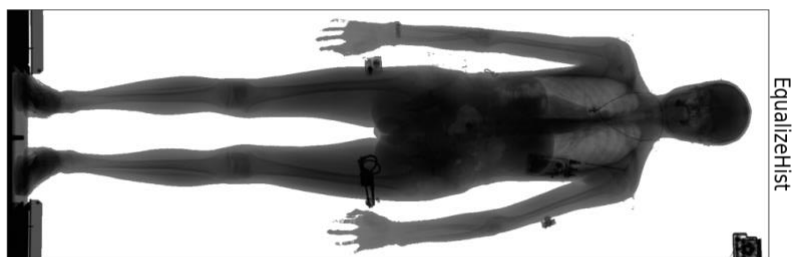
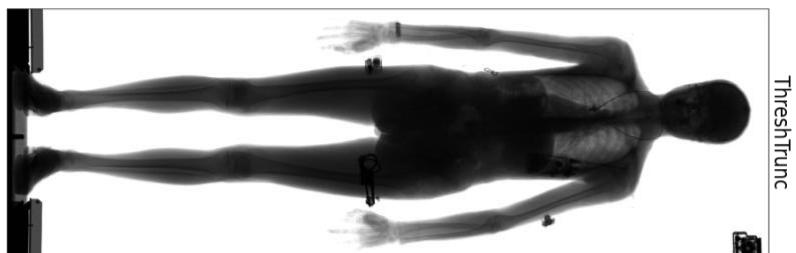
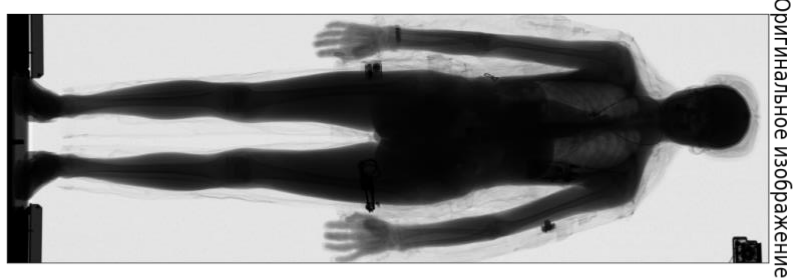
С помощью комбинации нескольких различных алгоритмов предобработки изображений удалось изменить гистограмму распределения интенсивности пикселей таким образом, чтобы все объекты были визуально различимы, при этом общая контекстная информация не теряется. С помощью комбинации генеративных сетей удалось построить алгоритм, который может генерировать дополнительные инородные объекты на изображении, не меняя общей картинки человека. Предложен новый подход по поиску запрещённых объектов, когда каждый участок изображения независимо классифицируется на наличие запрещённого объекта, и выводится карта запрещённых объектов. На основе классифицированных частей обучена сегментирующая сеть, выделяющая запрещённые объекты.

Источники

1. Bahdanau, Dzmitry and Cho, Kyunghyun and Bengio, Yoshua, Neural Machine Translation by Jointly Learning to Align and Translate, 2014, arXiv
2. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In NIPS, 2017.
3. Sharma, N. and Aggarwal, L. M., “Automated medical image segmentation techniques,” *Journal of Medical Physics / Association of Medical Physicists of India* 35(1), 3–14 (2010).
4. Mansoor, A., Bagci, U., Foster, B., Xu, Z., Papadakis, G. Z., Folio, L. R., Udupa, J. K., and Mollura, D. J., “Segmentation and image analysis of abnormal lungs at ct: Current approaches, challenges, and future trends,” *Radiographics* 35(4), 1056–1076 (2015).
5. Badrinarayanan, V., Handa, A., and Cipolla, R., “Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labeling,” arXiv preprint arXiv:1505.07293 (2015).
6. Badrinarayanan, V., Kendall, A., and Cipolla, R., “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(12), 2481–2495 (2017).
7. . XNet: A convolutional neural network (CNN) implementation for medical X-Ray image segmentation suitable for small datasets
8. 6. Ronneberger, O., Fischer, P., and Brox, T., “U-net: Convolutional networks for biomedical image segmentation,” in [Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015], 234–241, Springer International Publishing, Cham (2015).
9. 7. Ciresan, D., Giusti, A., Gambardella, L. M., and Schmidhuber, J., “Deep neural networks segment neuronal membranes in electron microscopy images,” in [Advances in Neural Information Processing Systems 25],

- Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., eds., 2843–2851, Curran Associates, Inc. (2012).
10. Яндекс Толока, <https://toloka.ai/>
 11. https://docs.opencv.org/3.4/db/d8e/tutorial_threshold.html
 12. https://docs.opencv.org/3.4/d6/dc7/group_imgproc_hist.html
 13. Mikolov, Tomas; et al. (2013). "Efficient Estimation of Word Representations in Vector Space". arXiv:1301.3781 [cs.CL].
 14. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. «An image is worth 16x16 words: Transformers for image recognition at scale». In ICLR, 2021.
 15. Goodfellow, Ian J. and Pouget-Abadie, Jean and Mirza, Mehdi and Xu, Bing and Warde-Farley, David and Ozair, Sherjil and Courville, Aaron and Bengio, Yoshua, «Generative Adversarial Networks», arXiv, 2014.
 16. Gulrajani, Ishaan and Ahmed, Faruk and Arjovsky, Martin and Dumoulin, Vincent and Courville, Aaron, «Improved Training of Wasserstein GANs», arXiv, 2017.
 17. Kingma, Diederik P. and Ba, Jimmy, «Adam: A Method for Stochastic Optimization», arXiv, 2014.
 18. He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian, «Deep Residual Learning for Image Recognition», arXiv, 2015.

Приложение 1. Последовательное применение последовательностей обработки.



Приложение 2. Сгенерированные участки изображений.

