

Федеральное государственное автономное образовательное учреждение  
высшего образования

«Московский физико-технический институт

(национальный исследовательский университет)»

Физтех-школа прикладной математики и информатики

Кафедра Управляющие и информационные системы

**Направление подготовки:** 03.03.01 Прикладные математика и физика

**Направленность (профиль) подготовки:** Компьютерные технологии  
интеллектуальный анализ данных

**Разработка алгоритма повышения разрешения изображения ( с  
использованием глубоких конволюционных нейронных сетей)**  
(бакалаврская работа)

**Студент:**

Ревар Владимир Игоревич

---

*(подпись студента)*

**Научный руководитель:**

Желтов Сергей Юрьевич

---

*(подпись научного руководителя)*

## СОДЕРЖАНИЕ

<b>ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ . . . . .</b>	<b>5</b>
<b>1 ВВЕДЕНИЕ . . . . .</b>	<b>6</b>
1.1 Актуальность темы . . . . .	6
1.2 ОБЗОР РАБОТ . . . . .	6
1.3 Цели . . . . .	7
1.4 Описание работы . . . . .	8
<b>2 ПОСТАНОВКА ЗАДАЧИ . . . . .</b>	<b>8</b>
<b>3 ПРИНЦИПЫ РАБОТЫ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ . . .</b>	<b>8</b>
3.1 Сверточная нейронная сеть . . . . .	8
3.2 Архитектура и принцип работы . . . . .	9
3.3 Слой свертки . . . . .	10
3.4 Пулинг или слой субдискретизации . . . . .	10
<b>4 ОБРАБОТКА ДАННЫХ . . . . .</b>	<b>11</b>
4.1 Загрузка данных . . . . .	11
4.2 Ухудшение качества изображений . . . . .	11
<b>5 НЕЙРОСЕТЬ ДЛЯ ПОВЫШЕНИЯ КАЧЕСТВА ИЗОБРАЖЕНИЯ</b>	<b>13</b>
5.1 Нейросеть для повышения качества изображения . . . . .	13
5.1.1 Архитектура . . . . .	13
5.2 Принцип работы . . . . .	13
5.2.1 Извлечение и представление патчей . . . . .	13
5.2.2 Нелинейное отображение . . . . .	14
5.2.3 Реконструкция . . . . .	14
5.3 Обучение . . . . .	14
<b>6 НЕЙРОСЕТЬ ДЛЯ ОПРЕДЕЛЕНИЯ КОРАБЛЕЙ НА ИЗОБРАЖЕ-</b>	
<b>НИИ . . . . .</b>	<b>16</b>
6.1 Архитектура . . . . .	16
6.2 Принцип работы . . . . .	16

6.3	Обучение . . . . .	17
<b>7</b>	<b>ОЦЕНКА РЕЗУЛЬТАТОВ . . . . .</b>	<b>19</b>
7.1	Нейросеть для повышения качества изображения . . . . .	19
7.1.1	Нейросеть для определения кораблей на изображении .	20
7.1.2	Объединение двух нейросетей . . . . .	20
<b>8</b>	<b>ЗАКЛЮЧЕНИЕ . . . . .</b>	<b>23</b>

## АННОТАЦИЯ

В данной работе рассматривается влияние повышения качества изображения на качество обнаружения объектов на изображении. В частности, мы улучшаем качество изображений, полученных с беспилотных летательных аппаратов(БЛА) , и проверяем, можем ли мы идентифицировать корабли с большей точностью, чем на исходном изображении. Используя нейросеть для повышения качества изображения (SRCNN), мы генерируем изображения, полученные с помощью двухкратного увеличения исходного изображения с повышенным качеством. Затем, используя оригинальные изображения изображения с повышенным качеством, мы тренируем нейросеть для обнаружения объектов. Этот подход позволяет количественно оценить влияние улучшения качества изображения на точность обнаружения объекта. Для тестового набора изображений мы отмечаем, что средняя точность обнаружения кораблей на изображении увеличилась на 8-10%, измеряя точность с помощью площади пересеченных частей ограничивающих прямоугольников в отношении сумм этих площадей(IoU) и порядка 6% для mean average precision(mAP). Так же мы отмечаем, что повышение качества изображения позволяет достичь улучшенных результатов в обнаружении объектов, имеющих маленький размер на изображении.

## ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

MSE — среднеквадратичная ошибка;

PSNR — пиковое отношение сигнала к шуму;

ADAM — adaptive moment estimation;

IoU — Intersection over Union;

Upscale — изображение с увеличенным качеством;

classic — исходное изображение.

GPU - видеокарта

mAP - mean average precision

SSIM - Индекс структурного сходства

# 1 ВВЕДЕНИЕ

Задача сверхразрешения относится к алгоритмам восстановления и улучшения изображения (например, удаление шума и размытия), широко изученным в компьютерном зрении и графике. В обоих сообществах цель состоит в том, чтобы восстановить изображение из вырожденной версии как можно точнее. Качество восстановленного изображения оценивается с помощью количественных метрик, основанных на пикселях, таких как PSNR и SSIM.

## 1.1 Актуальность темы

Развитие технологий машинного зрения и вычислительных средств для их реализации обуславливают их применение в новых областях для решения задач, традиционно решаемых человеком. Одной из таких важных и актуальных задач является разработка алгоритмов повышения разрешения изображений для улучшения качества обнаружения объектов. Методы решения данной задачи востребованы в широком спектре приложений, и, в частности, в спасательных операциях.

В России на сегодняшний день во время спасательных операций основным способом поиска является пешее прочесывание окрестности с применением достаточно примитивных технических средств, таких как сирена или гудящий маяк. Также во время спасательных операций могут использоваться воздушные беспилотные летательные аппараты (БЛА), которые позволяют получать большие объёмы видео информации, просмотр и дешифрирование которой оператором представляет сложную и трудоёмкую задачу. Чтобы найти объект поисковой миссии, например, человека или корабль, попавший в шторм, спасатели должны просмотреть множество изображений. При выполнении такого рода рутинных процедур человек достаточно быстро устаёт и теряет концентрацию внимания, что не может не приводить к ошибкам. Авторы инициативы Machine Learning for Social Good[1] утверждают, что через два часа человек уже неспособен продолжать поиск по фотографиям, а ведь от качества и скорости распознавания зависит здоровье, а иногда и жизнь людей. Данные факторы определяют актуальность рассматриваемой задачи.

## 1.2 ОБЗОР РАБОТ

В работе [2] рассматривается улучшение качества нахождения объекта на изображении, полученного со спутника, с помощью увеличения изображения в 8 раз с последующим улучшением качества. Авторы приводят пример готовых архитектур для улучшения качества изображения и нахождения объекта на изображении. Ключевой особенностью данной работы является сравнение результатов на изображениях различных размеров, от 15см до 4.8м. Авторам удалось достичь прироста в 15% точности.

В статье [3] авторы приводят новую структуру для обучения глубокой нейронной сети, в которой подсеть для повышения качества изображения явно включает потерю в качестве обнаружения посредством компромисса с традиционной потерей обнаружения.

В статье [4] авторы описывают метод повышения качества изображения с использованием глубоких сверточных нейросетей. Отличительной особенностью этого метода является оптимизация всех слоев нейронной сети одновременно, в отличие от классических методов, где каждый слой обновляется отдельно. Данный метод показывает наилучшие результаты на данный момент.

В статье [5] говорится о методе обнаружения объектов с использованием быстрой региональной сверточной сети. Авторы приводят результаты своей работы, которые показывают, что данный метод позволяет ускорить обучение в 9 раз, а также достичь лучшего качества.

В работе [6] авторы проводят исследование эффективности архитектур нейронных сетей и разработки оптимизаторов для глубоких нейронных сетей. Кроме того, в этой статье говорится о текущих проблемах в задаче повышения качества изображения.

В статье [7] затрагивается проблема восстановления тонкой структуры, например, граница объекта, на изображении. В статье описан классический метод восстановления качества изображения с использованием среднеквадратической ошибки, что позволяет достичь хорошего восстановления больших объектов, но не способна восстанавливать объекты тонкой структуры. Авторы предлагают новый метод решения данной проблемы с помощью модернизации функции потерь во время обучения. Также авторы предлагают свой вариант глубокой нейронной сети для восстановления изображения.

В работах [8], [9], [10] рассказывается об основах глубокого обучения, нейронных сетях и методах оптимизации, а именно об оптимизаторе Adam, который используется в большинстве архитектур нейронных сетей. В этих работах показаны сравнения результатов различных оптимизаторов на одинаковых задачах.

В работе [11] авторы представляют алгоритм, способный распознать локальные особенности на изображениях с низким разрешением, а затем соответствующим образом увеличивающий их разрешение, который имеет лучшие результаты относительно классических алгоритмов задачи повышения разрешения.

### 1.3 Цели

В наше время, с появлением мощных GPU и инновационных идей, для спасательных операций начинают использовать сверточные нейронные сети. Проблема такого подхода заключается в том, что из-за различной высоты снимков, объекты, которые нужно найти, могут иметь очень маленький размер, в следствие чего, могут быть распознаны неверно. Целью данной работы является проверка влияния повышения качества изображения на определение объектов.

## 1.4 Описание работы

В данной работе будут построены нейросети для распознавания объектов и для повышения качества изображений, и проведены исследования по улучшению результатов нахождения объектов на изображении с помощью задачи повышения качества изображения. В качестве данных будут использованы снимки кораблей с дронов, которые были взяты с [14]. В данной работе будут рассмотрены статьи, связанные с задачей повышения качества изображения с использованием глубоких нейронных сетей. Будут рассмотрены только научные статьи, которые были опубликованы на достоверных источниках, или, на которые ссылались другие статьи.

## 2 ПОСТАНОВКА ЗАДАЧИ

Рассматривается задача нахождения объекта на изображении. Имеются объекты малого размера, нахождение которых представляет множество трудностей. Требуется проверить, как повышение качества изображения улучшит результат работы нейросети для нахождения объектов на изображении.

## 3 ПРИНЦИПЫ РАБОТЫ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ

### 3.1 Сверточная нейронная сеть

Свёрточная нейронная сеть — специальная архитектура искусственных нейронных сетей, разработанная в 1988 и нацеленная на эффективное распознавание образов, входит в состав технологий глубокого обучения. Использует некоторые особенности зрительной коры, в которой были открыты так называемые простые клетки, реагирующие на прямые линии под разными углами, и сложные клетки, реакция которых связана с активацией определённого набора простых клеток. Таким образом, идея свёрточных нейронных сетей заключается в чередовании свёрточных слоёв и субдискретизирующих слоёв. Структура сети — однонаправленная (без обратных связей), принципиально многослойная. Для обучения используются стандартные методы, чаще всего метод обратного распространения ошибки. Функция активации нейронов (передаточная функция) — любая, по выбору исследователя.

Название архитектура сети получила из-за наличия операции свёртки, суть которой в том, что каждый фрагмент изображения умножается на матрицу (ядро) свёртки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения.



## 3.2 Архитектура и принцип работы

В обычном перцептроне, который представляет собой полносвязную нейронную сеть, каждый нейрон связан со всеми нейронами предыдущего слоя, причём каждая связь имеет свой персональный весовой коэффициент. В свёрточной нейронной сети в операции свёртки используется лишь ограниченная матрица весов небольшого размера, которую «двигают» по всему обрабатываемому слою (в самом начале — непосредственно по входному изображению), формируя после каждого сдвига сигнал активации для нейрона следующего слоя с аналогичной позицией. То есть для различных нейронов выходного слоя используются одна и та же матрица весов, которую также называют ядром свёртки. Её интерпретируют как графическое кодирование какого-либо признака, например, наличие наклонной линии под определённым углом. Тогда следующий слой, получившийся в результате операции свёртки такой матрицей весов, показывает наличие данного признака в обрабатываемом слое и её координаты, формируя так называемую карту признаков. Естественно, в свёрточной нейронной сети набор весов не один, а целая гамма, кодирующая элементы изображения, например линии и дуги под разными углами. При этом такие ядра свёртки не закладываются исследователем заранее, а формируются самостоятельно путём обучения сети классическим методом обратного распространения ошибки. Проход каждым набором весов формирует свой собственный экземпляр карты признаков, делая нейронную сеть многоканальной, а именно, много независимых карт признаков на одном слое. Также следует отметить, что при переборе слоя матрицей весов её передвигают обычно не на полный шаг, равный размеру этой матрицы, а на небольшое расстояние. Так, например, при размерности матрицы весов  $5 \times 5$  её сдвигают на один или два пикселя вместо пяти, чтобы не «перешагнуть» искомый признак.

Операция субдискретизации, выполняет уменьшение размерности сформированных карт признаков. В данной архитектуре сети считается, что информация о факте наличия искомого признака важнее точного знания его координат, поэтому из нескольких соседних нейронов карты признаков выбирается максимальный и принимается за один нейрон уплотнённой карты признаков меньшей размерности. За счёт данной операции, помимо ускорения дальнейших вычислений, сеть становится более инвариантной к масштабу входного изображения.

Рассмотрим типовую структуру свёрточной нейронной сети более подробно. Сеть состоит из большого количества слоёв. После начального слоя, то есть входного изображения, сигнал проходит серию свёрточных слоёв, в которых чередуется собственно свёртка и субдискретизация. Чередование слоёв позволяет составлять карты признаков, на каждом следующем слое карта уменьшается в размере, но увеличивается количество каналов. На практике это означает способность распознавания сложных иерархий признаков. Обычно после прохождения нескольких слоёв карта признаков вырождается в вектор или даже скаляр, но таких карт признаков становятся сотни. На выходе свёрточных слоёв сети дополнительно устанавливают несколько слоёв полносвязной нейронной

сети, на вход которому подаются конечные карты признаков.

### 3.3 Слой свертки

Слой свёртки — это основной блок свёрточной нейронной сети. Слой свёртки включает в себя для каждого канала свой фильтр, ядро свёртки которого обрабатывает предыдущий слой по фрагментам, суммируя результаты поэлементного произведения для каждого фрагмента. Весовые коэффициенты ядра свёртки неизвестны и устанавливаются в процессе обучения.

Особенностью свёрточного слоя является сравнительно небольшое количество параметров, устанавливаемое при обучении. Так например, если исходное изображение имеет размерность  $100 \times 100$  пикселей по трём каналам, что значит 30000 входных нейронов, а свёрточный слой использует фильтры с ядром  $3 \times 3$  пикселя с выходом на 6 каналов, тогда в процессе обучения определяется только 9 весов ядра, однако по всем сочетаниям каналов, то есть  $9 \times 3 \times 6 = 162$ , в таком случае данный слой требует нахождения только 162 параметров, что существенно меньше количества искомых параметров полносвязной нейронной сети.

### 3.4 Пулинг или слой субдискретизации

Слой пулинга (иначе подвыборки, субдискретизации) представляет собой нелинейное уплотнение карты признаков, при этом группа пикселей (обычно размера  $2 \times 2$ ) уплотняется до одного пикселя, проходя нелинейное преобразование. Наиболее употребительна при этом функция максимума. Преобразования затрагивают непересекающиеся прямоугольники или квадраты, каждый из которых ужимается в один пиксель, при этом выбирается пиксель, имеющий максимальное значение. Операция пулинга позволяет существенно уменьшить пространственный объём изображения. Пулинг интерпретируется так: если на предыдущей операции свёртки уже были выявлены некоторые признаки, то для дальнейшей обработки настолько подробное изображение уже не нужно, и оно уплотняется до менее подробного. К тому же фильтрация уже ненужных деталей помогает не переобучаться. Слой пулинга, как правило, вставляется после слоя свёртки перед слоем следующей свёртки.

Кроме пулинга с функцией максимума можно использовать и другие функции — например, среднего значения или L2-нормирования. Однако практика показала преимущества именно пулинга с функцией максимума, который включается в типовые системы.

В целях более агрессивного уменьшения размера получаемых представлений, всё чаще находят распространение идеи использования меньших фильтров или полный отказ от слоёв пулинга.

## 4 ОБРАБОТКА ДАННЫХ

### 4.1 Загрузка данных

В качестве обучающей выборки был взят набор изображений кораблей, доступный на [14]. После проверки разметки, было обнаружено, что данный набор изображений несбалансирован. Из 200 тысяч фотографий около 150 тысяч не содержат кораблей, то есть являются снимками моря с воздуха. Чтобы нейросеть не переобучилась на изображениях без кораблей, были отобраны все фотографии, содержащие корабли и 300 изображений без них. Все изображения имеют разметку. Разметка представляет из себя вектор для каждого изображения, содержащий 4 числа, показывающие местоположение центра корабля и размеры прямоугольника, ограничивающего этот корабль. Ниже приведен пример того, как должны распознаваться корабли.

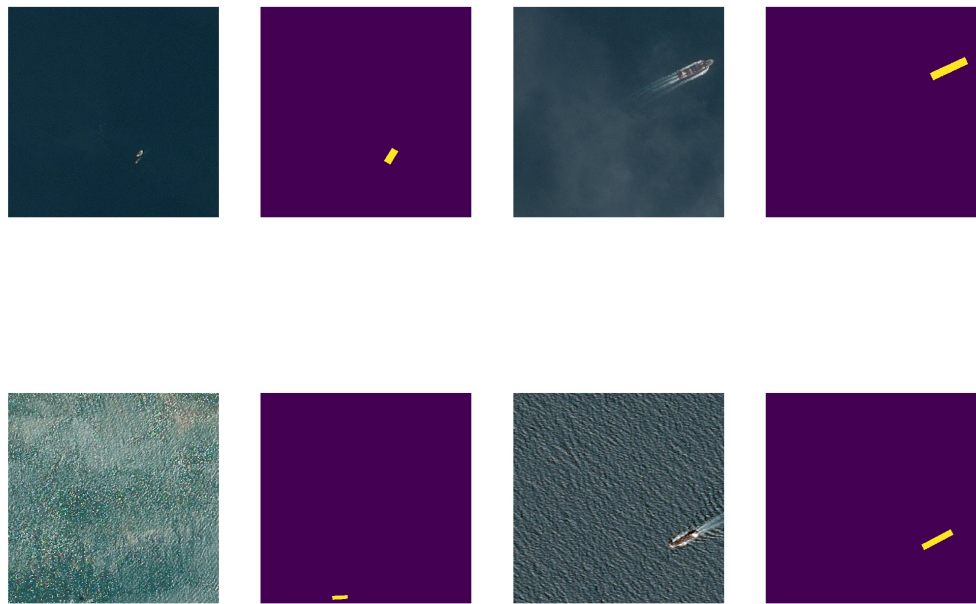


Рис. 1

### 4.2 Ухудшение качества изображений

Для обучения нейросети на повышение качества изображения нужно создать фотографии, имеющие размытое качество, которое может быть получено из фотографии путем сжатия в два раза, а затем возвращением к исходному размеру с помощью бикубической интерполяции. Идея бикубической интерполяции заключается в том, что

используя 4 соседних пикселя, то есть квадрат  $2 \times 2$ , мы получаем 16 соседних пикселей, аппроксимируя их, используя исходные 4 пикселя. Так как изображение получается с помощью аппроксимации, качество обязано быть хуже, так как теряется исходная информация о пикселях.

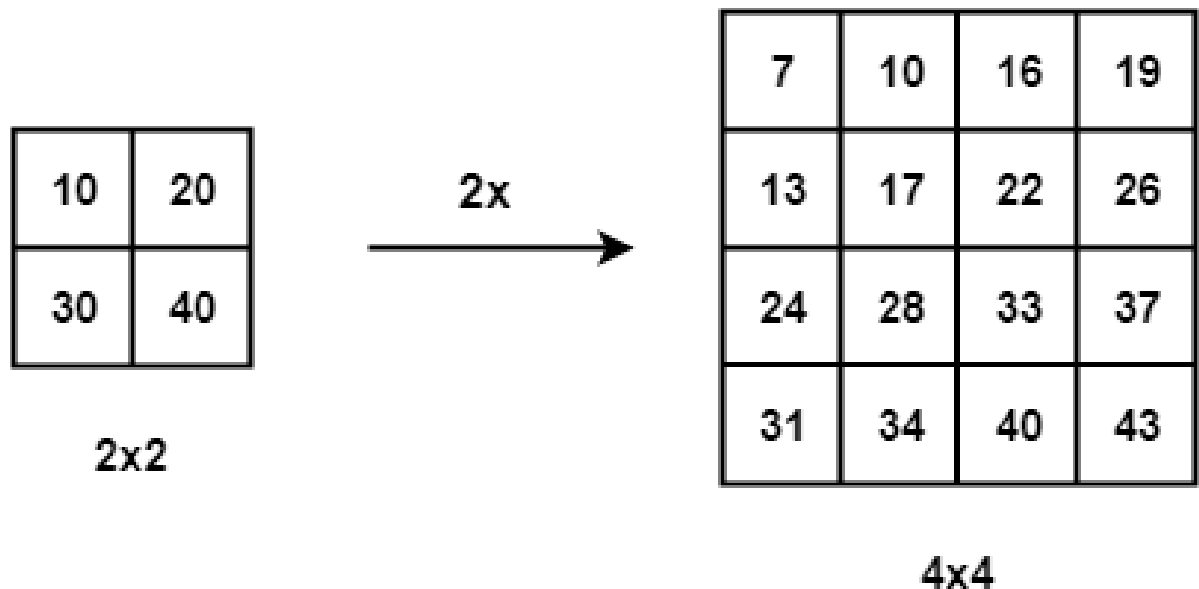


Рис. 2

Далее данный способ ухудшения качества изображения был проверен на ухудшение качества изображения и результат измерялся с помощью среднеквадратичной ошибки для каждого из трех каналов оригинального изображения и его ухудшенной версии. Чтобы проверить, что алгоритм действительно ухудшает качество изображения использовалась средне квадратичная ошибка (MSE). Пусть дана фотография  $I$ , размера  $m \times n$ , и ее шумная аппроксимация  $K$ , тогда среднеквадратичная ошибка может быть рассчитана по формуле:

$$MSE = \frac{1}{mn} * \sum_{i=0}^{m-1} * \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2$$

Средний результат ошибки, вычисленный по всем изображениям, получился следующим:

$$MSE_{mean} = 4.96,$$

что свидетельствует о том, что изображения получаются с ухудшенным качеством.

## 5 НЕЙРОСЕТЬ ДЛЯ ПОВЫШЕНИЯ КАЧЕСТВА ИЗОБРАЖЕНИЯ

### 5.1 Нейросеть для повышения качества изображения

#### 5.1.1 Архитектура

На рисунке 3 показана архитектура нейросети для повышения качества изображения. Для сохранения исходного размера изображения, сверточные слои данной нейросети имеют заранее определенные отступ и размер шага. Данная архитектура позволяет улучшить качество изображения, не поменяв его размер.

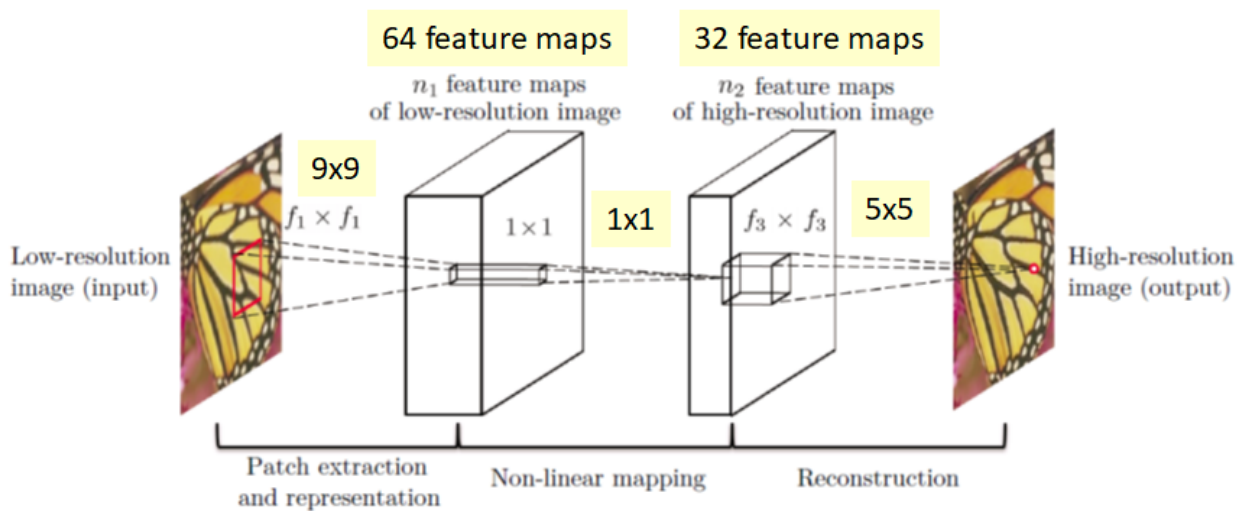


Рис. 3

### 5.2 Принцип работы

В данном параграфе описывается принцип работы нейронной сети для повышения качества изображения. Имея изображение с пониженным качеством, сначала производится увеличение изображения до необходимого размера с использованием бикубической интерполяции. Обозначим увеличенное изображение  $Y$ , исходное изображение  $X$ . Тогда задача формулируется следующим образом: имея изображение  $Y$ , необходимо восстановить изображение  $F(Y)$ , которое максимально близко к изображению  $X$ . Таким образом, необходимо найти отображение  $F$ , которое концептуально состоит из трех операций.

#### 5.2.1 Извлечение и представление патчей

Данная операция извлекает перекрывающиеся патчи из изображения  $Y$  с низким разрешением и представляет каждый патч как вектор высокой размерности. Эти векторы составляют набор карт признаков, число которых равно размерности векторов.

### 5.2.2 Нелинейное отображение

Данная операция нелинейно отображает каждый многомерный вектор на другой многомерный вектор. Каждый картированный вектор концептуально представляет патч высокого разрешения. Эти векторы составляют другой набор карт характеристик.

### 5.2.3 Реконструкция

Данная операция объединяет вышеприведенные патч-представления высокого разрешения для генерации окончательного изображения высокого разрешения. Ожидается, что это изображение будет похоже на исходное изображение  $Y$ .

## 5.3 Обучение

В качестве функции ошибки была взята функция  $PSNR$ .  $PSNR$  можно выразить с помощью средней квадратичной ошибки. Тогда  $PSNR$  может быть записана в виде:

$$PSNR = 10 * \log_{10}(\frac{MAX_I^2}{MSE}),$$

где  $MAX_I$  - максимально возможное значение пикселя, то есть в данном случае, так как пиксель представим в виде 8 бит,  $MAX_I = 255$ .

Так же в качестве оценки результатов был взят индекс структурного сходства (SSIM). SSIM является одним из методов измерения схожести между двумя изображениями. SSIM-индекс это метод полного сопоставления, другими словами, он проводит измерение качества на основе исходного изображения (не сжатого или без искажений). SSIM-индекс является развитием традиционных методов, таких как PSNR (peak signal-to-noise ratio) и метод среднеквадратичной ошибки MSE, которые оказались несовместимы с физиологией человеческого восприятия. Отличительной особенностью метода, помимо упомянутых ранее (MSE и PSNR), является то, что метод учитывает «восприятие ошибки» благодаря учёту структурного изменения информации. Идея заключается в том, что пиксели имеют сильную взаимосвязь, особенно когда они близки пространственно. Данные зависимости несут важную информацию о структуре объектов и о сцене в целом. SSIM метрика рассчитывается для двух изображений одинакового размера по следующей формуле:

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$

где  $\mu_x, \mu_y$  - среднее  $x$  и  $y$ ,  $\sigma_x, \sigma_y$  - дисперсия  $x$  и  $y$ ,  $\sigma_{xy}$  - ковариация  $x$  и  $y$ ,  $c_1 = (k_1L)^2$ ,  $c_2 = (k_2L)^2$  - две константы,  $L$  - динамический диапазон пикселей,  $L = 255$ ,  $k_1 = 0.01$ ,  $k_2 = 0.03$  - две константы.

Для обучения была взята подвыборка из всего набора данных, включающая в себя 20% от общего количества изображений. Были исследованы оптимизаторы  $ADAM$ ,

*Nesterov Accelerated Gradient* и *Adagrad*. Лучшие результаты показал оптимизатор *ADAM*, сочетающий в себе идею накопления движения и идею более слабого обновления весов для типичных признаков. На обучение подавались батчи из 128 фотографий, обучение длилось 300 эпох.

## 6 НЕЙРОСЕТЬ ДЛЯ ОПРЕДЕЛЕНИЯ КОРАБЛЕЙ НА ИЗОБРАЖЕНИИ

### 6.1 Архитектура

В качестве нейросети для нахождения объекта на изображении была взята полносвязная нейронная сеть для сегментации изображений U-NET[15]. Данная нейронная сеть была доработана для задачи нахождения объекта на изображении. На рисунке 5 приведена архитектура нейросети, позволяющая определить корабль на изображении. Так как сеть глубокая, чтобы избежать возникновения затухания градиента были использованы *skip connection*, позволяющие решить эту проблему. На рисунке 4 показано, как работает *skip connection*. Также для детекции кораблей используются свертки  $3 \times 3$ ,

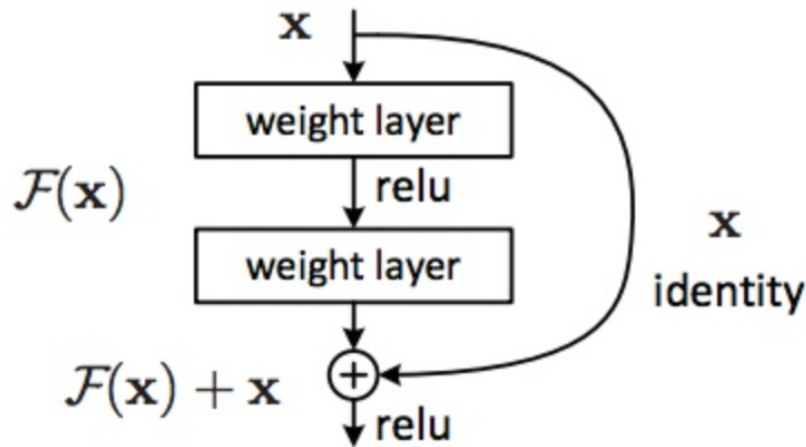


Рис. 4

которые, как и в предыдущей сети, используются с определенными отступами и шагом, чтобы не изменять размер изображения.

### 6.2 Принцип работы

Задача нахождения объекта на изображении сильно связана с задачей классификации объекта на изображении. В задаче классификации результатом работы сети является число, показывающее какому классу принадлежит объект. В задаче нахождения объекта результатом работы сети будет являться вектор, содержащий следующие значения: число, показывающее к какому классу относится объект и дополнительно 4 числа, которыми описывается прямоугольник, в котором находится объект. Обнаружение объекта происходит с помощью скользящего окна. Это окно можно представить как прямоугольник определенного размера, передвигающегося по изображению. Если в этот прямоугольник попал объект, то нейросеть регистрирует этот объект и производит классификацию. Само скользящее окно меняет свои размеры. Во время обучения, зная правильную разметку для исходного изображения, можно получить разметку для



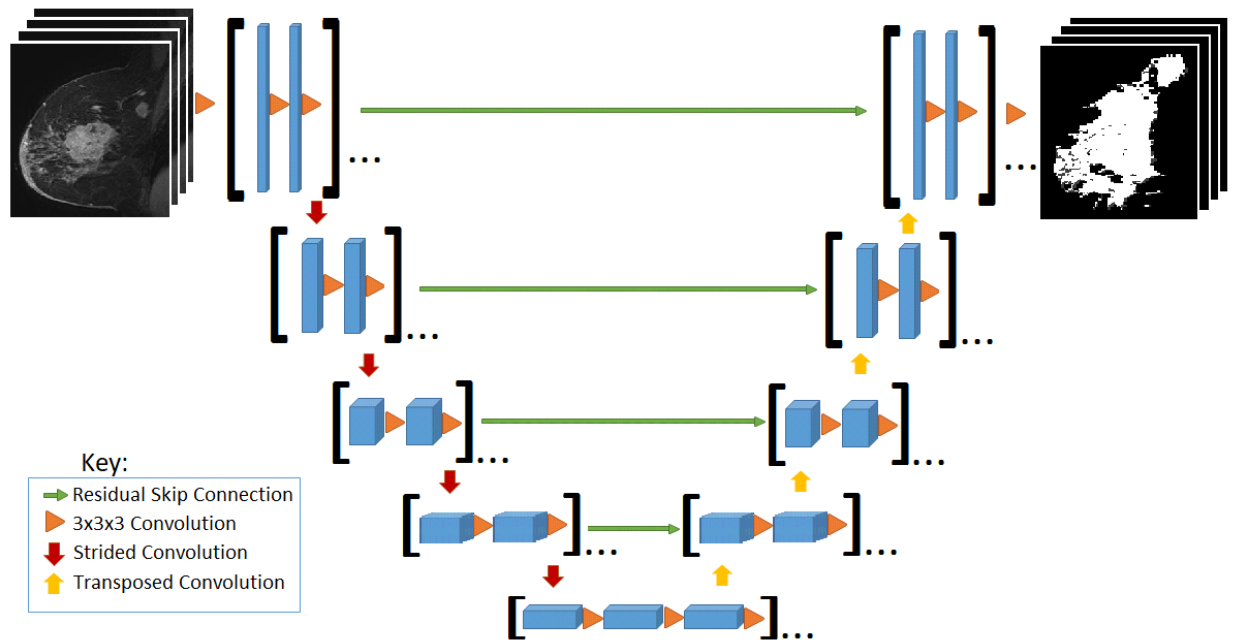


Рис. 5

каждого окна, которое используется скользящим окном, в следствие чего, нейросеть способна научиться определять объект на изображении.

### 6.3 Обучение

Определение корабля на изображение описывается ограничивающим параллелепипедом, в котором находится искомый объект. В качестве функций ошибки были взяты  $IoU$  - отношении площади пересечения двух ограничивающих параллелепипедов к площади их объединения и  $back IoU$  - для определения фона.  $IoU$  для двух множеств  $A$  и  $B$  можно записать в виде:

$$IoU(A, B) = \frac{A \cap B}{A \cup B}$$

Так как по определению  $IoU \leq 1$ , то в качестве функции ошибки в данном случае была взята  $IoU_{loss} = 1 - IoU$ , чтобы ее можно было минимизировать. На рисунке 6 показано соответствие правильности определения объекта и значение  $IoU$ .

Так же для оценки качества работы нейросети использовалась функция mAP. Идея mAP заключается в том, что выбрав некоторый порог IoU, сеть определяет только те объекты, IoU которых выше этого порога. mAP считается для 11 порогов IoU, а именно от 0.5 до 1 с шагом 0.05. Для каждого из порога подсчитывается точность определения объектов, а затем усредняется по всем порогам.

Из рисунка 6 можно сделать вывод, что при  $IoU \geq 0.7$  объект определяется достаточно хорошо.

В качестве оптимизатора снова был использован  $ADAM$ , обучение проходило всем на-



Рис. 6

боре изображений, на вход подавались батчи из 20 изображений, обучение длилось порядка 15 эпох.

## 7 ОЦЕНКА РЕЗУЛЬТАТОВ

### 7.1 Нейросеть для повышения качества изображения

Результаты работы сети были проверены на тестовой выборке, которая не задействовалась в обучении. Сеть показывает лучшие результаты по сравнению с обычной бикубической интерполяцией. На рисунке 7 приведены примеры повения качества изображения с помощью бикубической интерполяции и с использованием нейросети.

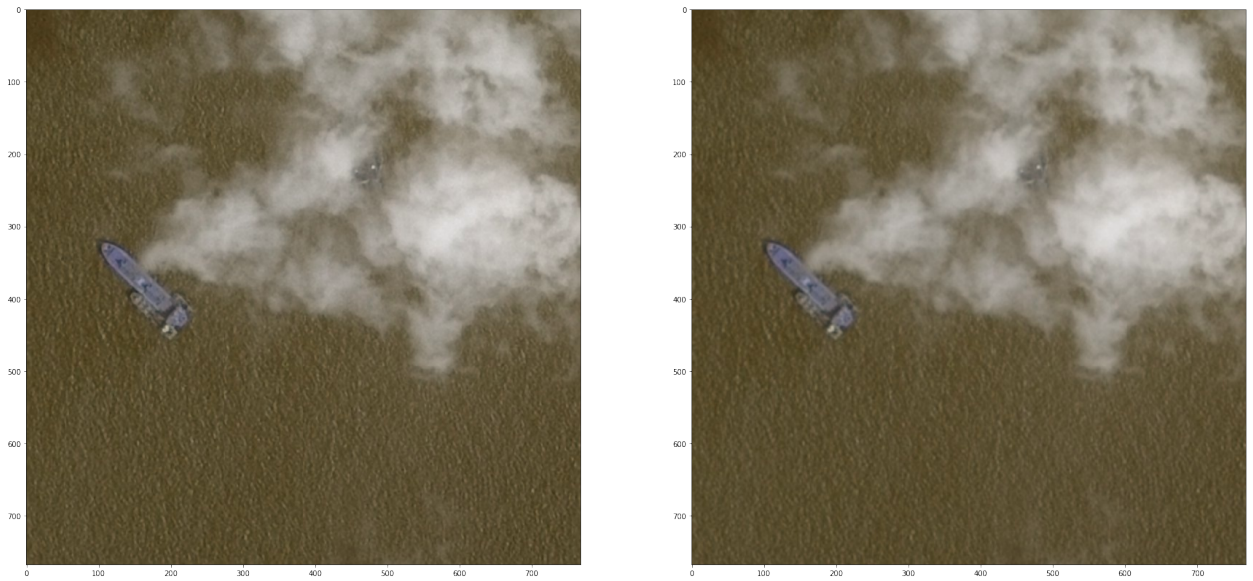


Рис. 7

Сравнив  $PSNR$  для данных двух фотографий с их оригинальной, были получены следующие результаты:

$$PSNR_{bicubic} = 34,574 \quad PSNR_{nn} = 41,193,$$

$$SSIM_{bicubic} = 36,213 \quad SSIM_{nn} = 43,586$$

что свидетельствует о том, что повышение качества изображения с помощью нейросети дает лучшие результаты, чем классические методы. Средние результаты данной сети на всем тестовом наборе данных оказались следующими:

$$PSNR_{bicubic} = 35,082 \quad PSNR_{nn} = 42,711,$$

$$SSIM_{bicubic} = 37,002 \quad SSIM_{nn} = 44,489$$

### 7.1.1 Нейросеть для определения кораблей на изображении

Результаты работы нейросети для определения кораблей на изображении оценивались с помощью IoU и mAP. Средние результаты по всей тестовой выборке оказались следующими:

$$IoU = 0,713 \quad Back\_IoU = 0,989 \quad mAP = 0,692$$

Результаты работы нейросети на изображениях из тестового набора, восстановленных с помощью бикубической интерполяции, оказались следующими::

$$IoU = 0,663 \quad Back\_IoU = 0,991 \quad mAP = 0,618$$

Так как корабль занимает очень малую часть на изображение, метрика *BackIoU* достигает таких высоких значений. На рисунке 8,9 приведены примеры исходного изображения, размеченного изображения, и то, как нейросеть определяет корабль.

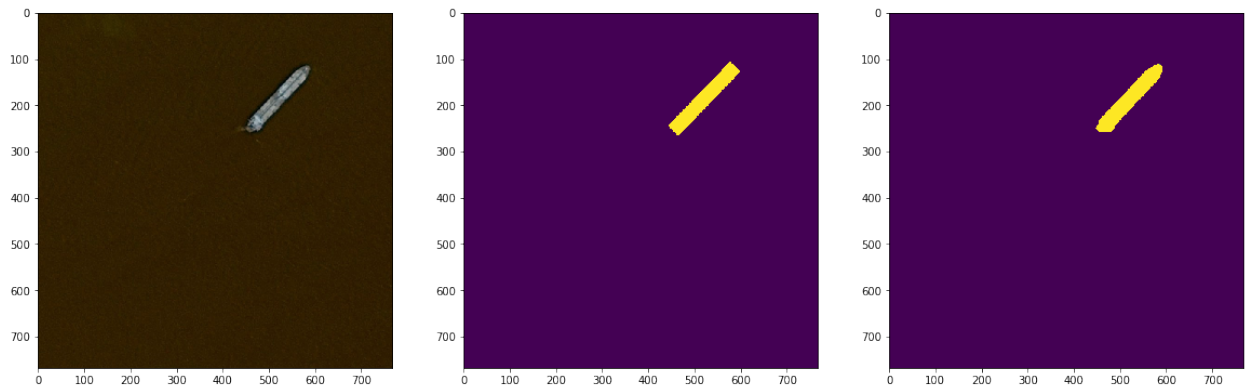


Рис. 8

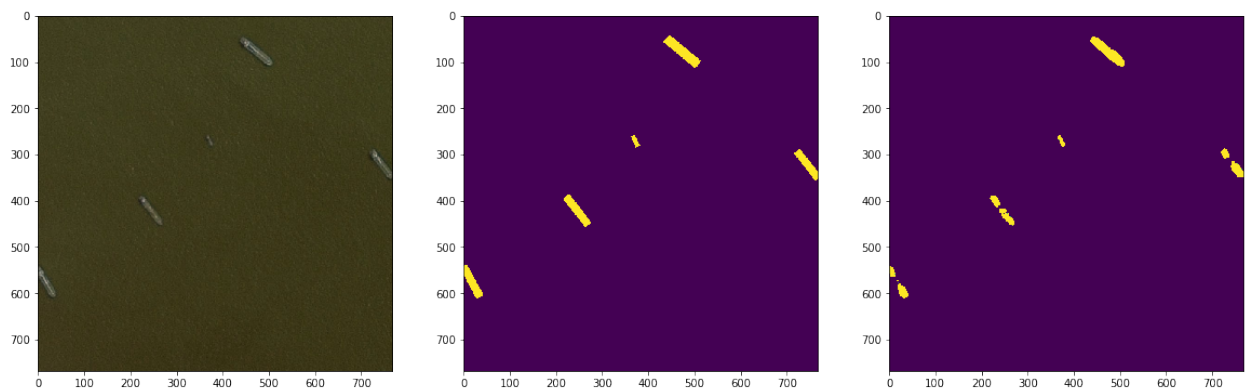


Рис. 9

### 7.1.2 Объединение двух нейросетей

Для оценки влияния повышения качества изображения для задачи обнаружения объекта, мы провели следующие сравнения. Исходная фотография делилась на 4

равных части, каждая из которых приводилась к исходному размеру с помощью нейросети, повышающей качество. На этих фотографиях был На рисунке 10,11 приведены исходное изображение, результат работы нейросети, определяющая корабли на изображении, и результат этой же сети на тех участках увеличенного изображения, где находится корабль

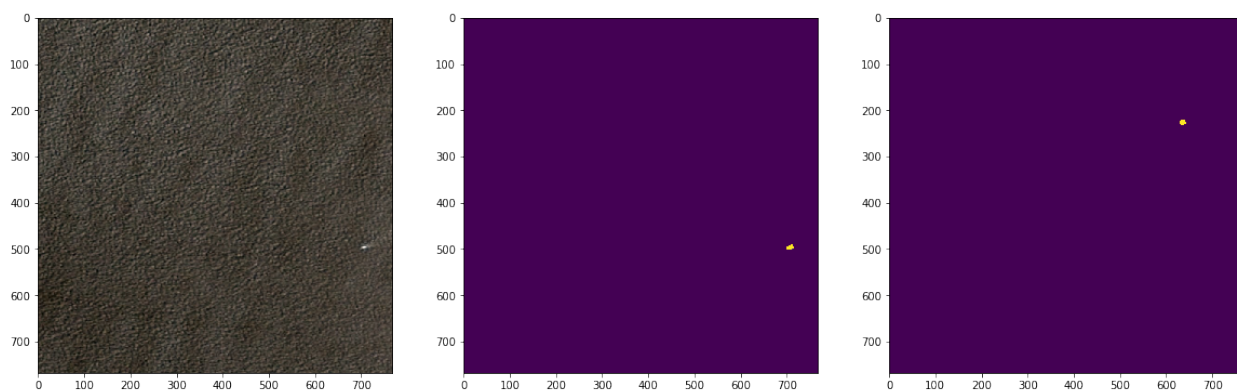


Рис. 10

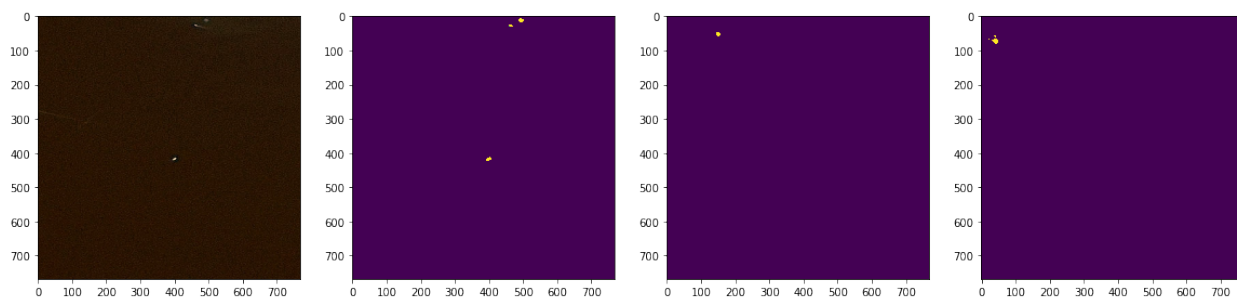


Рис. 11

Как можно заметить из рисунка 10, сеть определяет корабль более точно. Разница количества выделенных областей в пикселях равна 144. На рисунке 11 сеть не смогла определить один из кораблей, но увеличилось качество определение другого корабля. Измерив разницу количества пикселей на рисунке 10, результат равняется 83, то есть определяется большая часть корабля, но ухудшается количество верно определенных кораблей. Этот результат может быть объяснен тем, что один из кораблей имеет слишком маленький размер, и при увеличении качества этого изображения он становится неразличим с фоном. Данная проблема может быть решена путем более долгого обучения сети, повышающего качество изображения, или, путем увеличения числа слоев в этой же сети.

Измерив метрику  $IoU$  на рисунках 10 и 11 для определения кораблей на исходном изображении и изображениях, полученным с помощью увеличения качества самого изображения, были получены следующие результаты:

Рисунок 10:

$$IoU_{original} = 0,83 \quad IoU_{upscale} = 0,88$$

Рисунок 11:

$$IoU_{original} = 0,74 \quad IoU_{upscale} = 0,77$$

Кроме того, было проведено сравнение качества определения корабля на исходном изображении и том же изображении, которому было повышено качество.

На рисунке 12 показаны: исходное изображение, результат работы сети на этом изображении и результат работы сети на этом же изображении, но с увеличенным качеством.

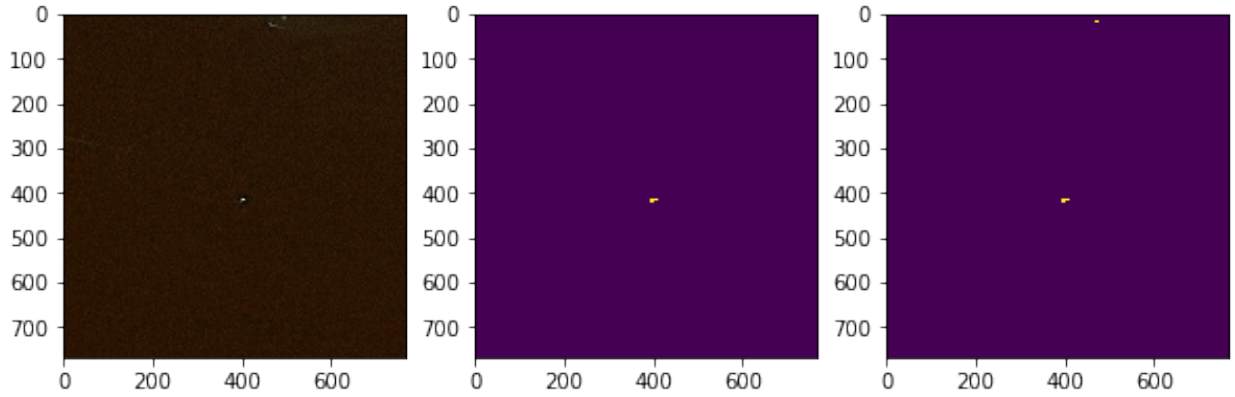


Рис. 12

Как видно из рисунка 12, сеть смогла определить еще один корабль, разница количества выделенных областей в пикселях равна 91.

$IoU$  для рисунка 11:

$$IoU_{classic} = 0,73 \quad IoU_{upscale} = 0,81$$

Средние результаты на всем тестовом наборе оказались следующими:

С разбиением исходного изображения на несколько более мелких с их последующим увеличением и улучшением качества:

$$IoU = 0,782 \quad mAP = 0,731$$

С помощью улучшения качества исходного изображения без разбиения на более мелкие:

$$IoU = 0,748 \quad mAP = 0,709$$

Таким образом, увеличение качества изображения позволяет достичь более точного определения кораблей на изображении.

## 8 ЗАКЛЮЧЕНИЕ

В данной работе было проведено исследование повышения качества определения заданного объекта на изображении с помощью повышения качества изображения. Были исследованы результаты работы нейросети для обнаружения объектов на изображении более, чем 50000 изображений, полученных с беспилотных летательных аппаратов. Чтобы проверить влияние повышения качества изображения на результаты работы, мы сгенерировали изображения худшего качества с помощью сжатия изображения с последующей бикубической интерполяцией. Тесты, проведенные на ухудшенных и оригинальных изображениях показывают, что точность определения объектов падает 7%. Применение методов суперразрешения в качестве предварительной обработки обеспечивает улучшение качества обнаружения объекта (Рис 12). Методы повышения качества изображения позволяют достичь прироста в точности порядка 4% при обработке исходного изображения, и порядка 10% при разбиении исходного изображения на более маленькие изображения с их последующим увеличением. В целом, учитывая относительную простоту применения методов повышения качества изображения, общее улучшение, наблюдаемое в этом исследовании, заслуживает внимания и предполагает, что повышение качества изображения может быть ценным этапом предварительной обработки для будущих приложений обнаружения объектов с помощью изображений, полученных с беспилотных летательных аппаратов.

## Список литературы

- [1] Open Data Science Проект Lascmus: как компьютерное зрение помогает спасать потерявшихся людей -2020
- [2] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang  
Image Super-Resolution Using Deep Convolutional Networks  
– 2015,  
– arxiv.org
- [3] Sharif Elfouly  
Introduction: Fast R-CNN (Object Detection)  
– 2019  
– towardsdatascience
- [4] Jacob Shermeyer and Adam Van Etten The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery -2019 -arxiv.org
- [5] Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita Task-Driven Super Resolution: Object Detection in Low-resolution Images –2018 –arxiv.org
- [6] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, Qingmin Liao Deep Learning for Single Image Super-Resolution: A Brief Review  
– 2019  
– arxiv.org
- [7] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi  
Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network  
– 2017  
– arxiv.org
- [8] Sung Cheol Park ; Min Kyu Park ; Moon Gi Kang Super-resolution image reconstruction: a technical overview – 2003 – ieeexplore.ieee.org
- [9] Diederik P. Kingma, Jimmy Ba ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION – 2017 – arxiv.org
- [10] Yann LeCun, Yoshua Bengio, Geoffrey Hinton Deep Learning – 2015 – nature.com
- [11] S. Baker ; T. Kanade Limits on super-resolution and how to break them – 2002 – ieeexplore.ieee.org



- [12] [wikipedia.org/wiki/Сверточная\\_нейронная\\_сеть](https://wikipedia.org/wiki/Сверточная_нейронная_сеть)  
–Wikipedia
- [13] Визильтер Ю.В., Желтов С.Ю., Бондаренко А.В., Ососков М.В., Моржин А.В.  
Обработка и анализ изображений в задачах машинного зрения –2010
- [14] URL: <https://www.kaggle.com/c/airbus-ship-detection/data>
- [15] Olaf Ronneberger, Philipp Fischer, Thomas Brox U-Net: Convolutional Networks for  
Biomedical Image Segmentation -2015 -arxiv.org