Исследование сверточных нейронных сетей для обнаружения объектов на аэрокосмических снимках

В.О. Скрипачев, М.В. Гуйда, Н.В. Гуйда, А.О. Жуков

Аннотация. **В статье рассматриваются актуальные алгоритмы для решения задач распознавания объектов на изображениях, их основные особенности и преимущества. Проведен краткий анализ существующих моделей работы с изображениями на основе сверточных нейронных сетей. Проведен краткий обзор особенностей архитектур сверточных нейронных сетей, количественные показатели оценки качества их функционирования и типы решаемых задач, рассмотрены основные особенности работы с изображениями и основные возникающие сложности, выделены особенности обработки аэрокосмических снимков. Осуществлена постановка задачи распознавания объектов на аэрокосмических снимках посредством адаптации существующих актуальных алгоритмов и их сочетаний. Показаны основные проблемы обработки аэрокосмических снимков и подходы к их решению, применение сложившихся методов распознавания объектов на обычных снимках к проблематике распознавания объектов на аэрокосмических снимках. Проведен анализ различных архитектур нейронных сетей в призме решения задач распознавания объектов на аэрокосмических снимках. Сделаны выводы относительно наиболее удачных сочетаний различных алгоритмов в структуре нейронных сетей при распознавании объектов на аэрокосмических снимках. Определены основные факторы, затрудняющие распознавание объектов на аэрокосмических снимках и направления работы для снижения их влияния на точность работы нейронных сетей при распознавании объектов на аэрокосмических снимках.**

Ключевые слова: **сверточные нейронные сети (CNN), субдискретизация, сеть предсказания регионов (Region proposal network (RPN)), свертка, SRCNN, U-NET, ResNet, обработка аэрофотоснимков, ориентированные ограничивающие рамки (oriented bounding box (OBB)), детектирование по характерным точкам (Landmark detection), метод субдискретизации пространственных пирамид (SPP-net), гистограмма ориентированных градиентов (HOG), R-FCN, Faster R-CNN, YOLO, SSD.**

Статья получена 24 апреля 2022г.

В. О. Скрипачев, к.т.н., начальник отдела, Российский технологический университет (РТУ МИРЭА), Москва, Россия (e-mail: skripachev@mirea.ru).

М. В. Гуйда, к.т.н., старший научный сотрудник, Московский государственный университет имени М.В.Ломоносова (МГУ им. М.В. Ломоносова), Москва, Россия (e-mail: guida.mv21@physics.msu.ru).

Н. В. Гуйда, инженер 1 категории, Особое конструкторское бюро Московского энергетического университета (ОКБ МЭИ), Москва, Россия (e-mail: gujda.nv@okbmei.ru).

А. О. Жуков, д.т.н., профессор, ведущий научный сотрудник, Институт астрономии РАН, заместитель директора по научной работе ФГБНУ «Аналитический центр», Россия (e-mail: aozhukov@mail.ru).

I. Введение

Обнаружение объектов на изображениях является ключевым компонентом многих моделей глубокого обучения и претерпело ряд революционных преобразований в последние годы. Алгоритмы обнаружения объектов используются в таких областях, как автономное вождение, камеры безопасности, робототехника, и почти во всех приложениях, которые включают визуализацию, включая медицину, а также в новых направлениях, таких как магазин Amazon Go без продавца и кассы.

На протяжении многих лет основная проблема заключалась в том, что многие приложения требуют обнаружения объектов в реальном масштабе времени.

Сегодня эта проблема решена, существует целый ряд алгоритмов, способных осуществлять обнаружение объектов в реальном масштабе времени.

Несмотря на впечатляющие достижения, одним из самых больших барьеров в обнаружении объектов является то, что модели очень большие и вычислительно тяжелые. Их запуск занимает много времени и вычислительной мощности. Задачам моделирования обнаружения объектов с глубоким обучением свойственен дуализм. Во-первых, поскольку размер и количество объектов могут варьироваться, сеть должна быть в состоянии справиться с этой изменчивостью. Во-вторых, количество возможных комбинаций для границ объекта огромно, и эти сети, как правило, требуют вычислений. Это привело к поиску компромиссных решений, чтобы вычислить выход в реальном времени.

Отдельной актуальной задачей обнаружения объектов на изображениях является обнаружение объектов на аэрокосмических снимках. Данная задача является чрезвычайно актуальной для алгоритмов обнаружения объектов, и хотя она не во всех приложениях требует работы в реальном масштабе времени, необходимо создание новых алгоритмов, адаптированных под особенности поставленной задачи.

II. Алгоритмы обнаружения объектов

Для создания алгоритмов обнаружения объектов необходимо рассмотреть суть задачи при решении

проблемы, а также существующий сегодня задел в решении данной проблемы.

Алгоритмы обнаружения объектов на современном этапе способны решать следующие задачи:

- классификация объектов;
- локализация объектов;
- детектирование объектов;
- сегментация.

При классификации объектов входными данными являются изображения, выходными - класс объекта, который представлен на изображении. Количество классов определяется при проектировании алгоритма.

Под локализацией объектов понимается определение области расположения объекта на изображении. Как правило объект выделяется на изображении прямоугольником.

Под детектированием понимается совокупность операций локализации и классификации, выполненные последовательно. Если на изображении обнаружено несколько объектов, то каждый из них классифицируется отдельно. Существует два основных подхода к детектированию:

-детектирование с использованием ограничивающих рамок (Bounding box detection);

- детектирования по характерным точкам (Landmark detection).

Детектирование с использованием ограничивающих рамок основано на выделении части изображения, в которой находится объект при помощи некоторого прямоугольника имеющего координаты центра, высоту и ширину (Рисунок 1).

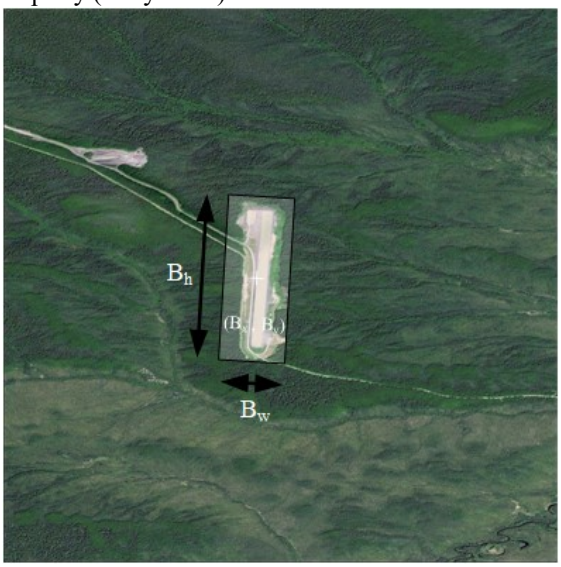

Рисунок 1 - Детектирование с помощью ограничивающих рамок.

Детектирование по характерным точкам основано на выделении формы или характеристик объекта (дороги, здание, корабль). Метод обеспечивает большую детализацию. Для описания объекта используются набор координат - референсных точек (Рисунок 2).

Рисунок 2 - Детектирование по характерным точкам.

Под сегментацией понимается процесс разделения изображения на несколько сегментов путем установления идентичных визуальных характеристик пикселей, относящихся к одному типу объекта на изображении. Результатом работы алгоритма является множество сегментов, покрывающих 100% изображения. Для оценки качества работы алгоритма должны быть введены некоторые количественные оценки качества его работы.

Сегодня, алгоритмы обнаружения объектов оцениваются с помощью двух параметров:

- пересечение над объединением IoU (intersection over union);
- средняя точность AP (average precision).

Суть параметра IoU сводится к следующему. На рисунке 3 показаны результаты определения границ объекта при подготовке обучающих выборок и прогнозирования. Зеленая линия — это правильный результат маркировки при подготовке выборки. Красная линия - результат, прогнозируемый алгоритмом.

Рисунок 3 – Границы объекта.

Значением IoU будет отношение площади области

пересечения представленных на рисунке 4 прямоугольников к площади области их объединения [1]. На современном этапе развития технологий детектирования объектов значение IoU>0,5 считается удовлетворительным (рисунок 4).

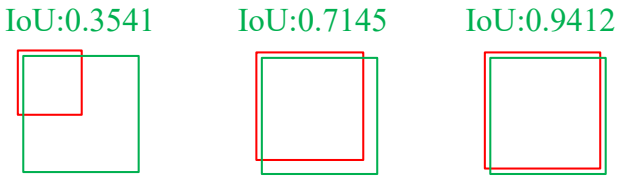

Рисунок 4 – Наглядное изображение значений IoU.

Средняя точность вычисляет среднее значение точности для значения отклика от 0 до 1 [2]. Для определения этого параметра необходимо ввести некоторые дополнительные параметры.

Точность (precision) показывает, на сколько правильны прогнозы, то есть какой прогноз является правильным.

Полнота (recall) показывает, на сколько хорошо определяются все правильные результаты. Эти параметры рассчитываются по следующим формулам:

$$\text{Т} = \frac{\text{ИП}}{\text{ИП+ЛП}}; \quad (1)$$

$$\text{П} = \frac{\text{ИП}}{\text{ИП+ЛО}}. \quad (2)$$

где Т - точность, П - полнота, ИП - истинно положительный, ЛП - ложноположительный, ЛО - ложноотрицательный.

Средняя точность определяется как площадь под кривой точности полноты:

$$AP = \int_0^1 p(r)dr, \quad (3)$$

где - (AP – средняя точность (average precision), p – точность (precision), r – полнота (recall)).

III. - Применение сверточных нейронных сетей для распознавания объектов на аэрокосмических снимках

Сверточные нейронные сети, или CNN, являются специализированным видом нейронных сетей для обработки данных, которые имеют характерную топологию, подобную сетке. Сверточные сети чрезвычайно хорошо зарекомендовали себя в практическом применении. Название "сверточная нейронная сеть" указывает на то, что сеть использует математическую операцию, называемую сверткой. Свертывание является специализированным видом линейной операции. Сверточные сети — это нейронные сети, использующие свертывание вместо общего умножения матрицы по крайней мере в одном из их слоев.

CNN состоит из одного или нескольких сверточных слоев, за которыми следуют один или несколько полносвязных слоев, как в стандартной многослойной нейронной сети. Архитектура CNN разработана для использования 2D-структуры входного изображения. Это достигается использованием локальных связей и привязанных весов, за которыми следует некоторая форма объединения, которая обеспечивает выделение признаков. Такая сеть имеет гораздо меньше параметров, чем полносвязные сети с тем же числом скрытых слоев, что обеспечивает меньшую вычислительную сложность процессе их обучения. В CNN применяется 5 видов слоев:

- входной слой;
- сверточный слой;
- слой функции активации;
- субдискретизационный слой;
- полносвязный слой.

Количество нейронов входного слоя должно соответствовать количеству пикселей в изображении для каждого цвета. Если изображение имеет размер 32x32 пикселя в трех цветах (RGB), то входной слой должен содержать 32x32x3=3072 нейрона.

Сверточный слой вычисляет выходной объем путем поэлементного умножения фильтров и областей изображения. В этой операции важнейшее значение играют применяемые фильтры, которые позволяют выявлять те или иные ключевые характеристики на изображении, например, фильтры границ. Для произвольного количества каналов в изображении, результат работы сверточного слоя описывается следующей формулой:

$$[H]_{i,j,d} = \sum_{a=-\Delta}^{\Delta} \sum_{b=-\Delta}^{\Delta} \sum_{c=1}^{k} [V]_{a,b,c,d} [X]_{i+a,j+b,c} \quad (4)$$

где d индексирует выходные каналы в результирующем изображении H, a (i,j) - координаты каждого пикселя на изображении, k - количество каналов в изображении, (a,b) - координаты элементов фильтра. Примеры результатов применения некоторых фильтров показаны на рисунках 5 - 7.

Рисунок 5 - Пример применения фильтра горизонтальных границ.

Рисунок 6 - Пример применения фильтра вертикальных границ.

Рисунок 7 - Фильтр границ независимо от ориентации.

Слой функции активации обеспечивает применение функции активации к выходам слоя свертки.

Субдискретизационный слой периодически вставляется между сверточными слоями сети для уменьшения объема информации, получаемой в результате применения фильтров к изображению. В основном используются три типа слоев субдискретизации:

- по максимальному значению;
- по среднему значению;
- по сумме значений.

Основными параметрами слоя является размер фильтра и шаг (stride). Пример субдискретизации по среднему значению приведен на рисунке 8.

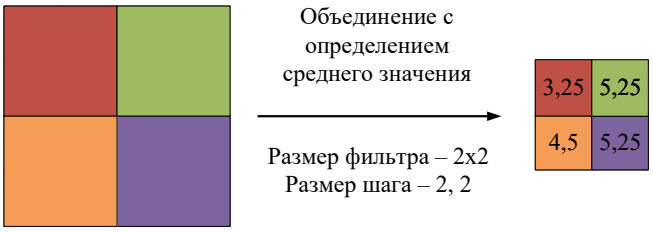

Рисунок 8 - Пример работы субдискретизационного слоя.

Полносвязный слой — это нейронная сеть, которая получает на вход результаты работы предыдущих слоев, и в качестве выходного результата дает вектор, содержащий вероятности отнесения объекта к тому или иному классу. Длина вектора равна количеству классов, а сумма его значений равна 1 (Рисунок 9).

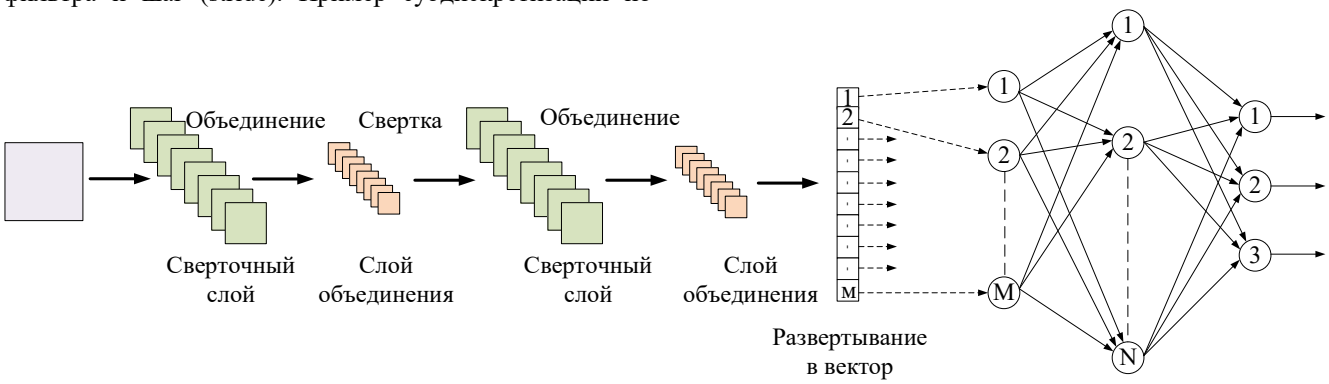

Рисунок 9 - Общая структура сверточной нейронной сети.

Рассмотрим возможность применения CNN в задачах обработки аэрокосмических снимков, получаемых с использованием искусственных спутников Земли. В настоящее время технологии дистанционного зондирования земли ((Earth observation and remote sensing (ДЗЗ)) позволяют вести наблюдение за поверхностью земли с разрешением до 0,5 метра. Несмотря на то, что разработка численных алгоритмов является сложной математической задачей, такие алгоритмы необходимы для обработки получаемых средствами ДЗЗ огромных объемов изображений, целью которой является локализация и классификация представляющих интерес объектов: транспортных средств, судов, построек, различных природных объектов и явлений.

За последнее десятилетие в области обнаружения объектов на изображениях достигнут значительный прогресс, однако это утверждение не относится к аэрофотоснимкам, что объясняется значительными различиями в масштабе и направлением съемки [6].

Основными трудностями при обработке аэрофотоснимков являются:

- произвольная ориентация объектов;
- вариации масштаба;
- неравномерная плотность объектов;
- сложные фоновые условия;
- большое соотношение сторон (large aspect ratios (ARs)).

Для решения проблем произвольной ориентации сторон и их большого соотношения при локализации используются ориентированные ограничивающие рамки (oriented bounding box (OBB)). Они больше подходят для решения задачи локализации на аэрофотоснимках [7, 8], по той причине, что представление объектов в горизонтальных ограничивающих рамках (horizontal bounding box (HBB)), которое используется при обработке обычных изображений, не может обеспечить достаточный уровень локализации ориентированных объектов. Под ориентированными объектами в данном случае понимаются объекты, имеющие соотношение сторон ограничивающей рамки более 1:2.

Использование ориентированных ограничивающих рамок позволяет различать компактно расположенные объекты и извлекать инвариантные к вращению признаки для последующей классификации [9]. Использование OBB фактически вводит новую задачу обнаружения объектов - ориентированное обнаружение объектов. Это недавно сформированное направление исследований и основные усилия в нем направлены на транслирование успешно работающих с HBB

детекторов объектов на базе глубоких нейронных сетей [6]. На сегодняшний день масштабных наборов аннотированных аэрофотоснимков для обучения сетей не так много. Наиболее распространенными наборами данных аэрофотоснимков являются DOTA и HRSC2016.

Еще одной проблемой является то, что дизайн и настройка гиперпараметров детекторов объектов, полученных на обычных изображениях, не подходят для работы с аэрофотоснимками из-за областных различий. Таким образом, при построении детекторов, работающих с аэрофотоснимками, требуется сравнительный анализ базовых алгоритмов, а также проведение абляционного анализа.

Рассмотрим некоторые из наиболее эффективных моделей обработки обычных изображений, существующих на сегодняшний день.

В основе работы региональных сверточных нейронных сетей (R-CNN) лежит получение набора регионов, которые предположительно содержат объекты для классификации, и затем их дальнейшую обработку в сверточной нейронной сети. Представителями подобных сетей является R-CNN и Fast R-CNN, а Faster R-CNN является последней моделью в семействе алгоритмов детекторов объектов.

R-CNN принимает на вход изображение и формирует на нём при помощи алгоритма селективного поиска [10] до 2000 регионов различных размеров. Регион — это часть изображения, в котором высока вероятность нахождения целевых объектов. Каждому региону присваивается некоторый класс и ограничивающая рамка. На следующем этапе в R-CNN использует большую CNN для вычисления признаков для каждого, предложенного ранее региона. На заключительном этапе происходит классификация каждого региона методом опорных векторов (support vector machines (SVMs)) и линейной регрессии (Рисунок 10). Недостатки R-CNN известны - эти модели медленные и энергозатратные.

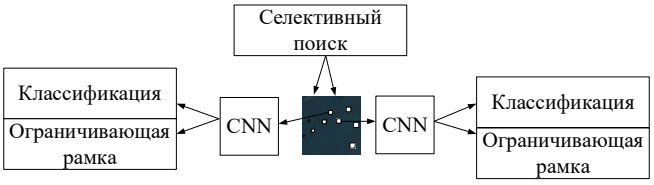

Рисунок 10 - Архитектура R-CNN.

Так как R-CNN обрабатывает каждый регион в CNN независимо, то это значительно замедляет модель. Для решения этой проблемы Fast R-CNN выполняет обработку изображения в CNN один раз на всем изображении.

Fast R-CNN параллельно обрабатывает всё изображение обычной CNN, для извлечения признаков, и алгоритмом селективного поиска, который обеспечивает получение предложений по регионам размещения объектов. Затем полученные признаки и регионы обрабатываются в слое субдискретизации регионов (RoI pooling). В этом слое регион преобразуется из координат изображения в координаты на карте признаков, получая на выходе вектор признаков фиксированной длины. (Рисунок 11).

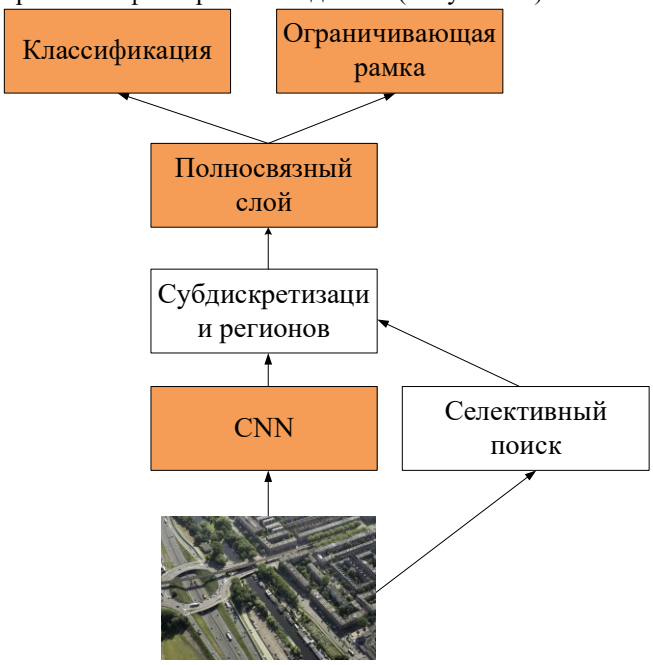

Рисунок 11 - Архитектура Fast R-CNN.

Каждый вектор признаков подается в полносвязные слои (FC), результат работы которых затем выводится в два выходных слоя:

- softmax - для оценки принадлежности объекта к классу;
- регрессии - который уточняет ограничивающие рамки объекта.

Первый слой с помощью функции softmax определяет вероятность отнесения объекта к тому или иному классу с учетом класса фона всего изображения. Второй слой выводит четыре вещественных числа, описывающих положение ограничивающей рамки для каждого объекта.

Таким образом, основными отличиями Fast R-CNN являются:

- при обработке генерируется набор признаков для всего изображения сразу, а не для каждой отдельной ограничивающей рамки, из которого затем при помощи специального слоя выделяются признаки для полученных параллельно регионов;
- не используется метод опорных векторов и линейной регрессии в пользу использования дополнительных слоев полносвязной нейронной сети.

В Faster R-CNN алгоритм селективного поиска заменен на небольшую нейронную сеть (Region proposal network (RPN)) для поиска регионов. Для детектирования используется Fast R-CNN (Рисунок 12).

Рисунок 12 - Архитектура Faster R-CNN.

Anchor boxes — алгоритм нахождения объектов, основанный на предсказании категории объекта и отступа от истинной ограничивающей рамки для большого количества сгенерированных ключевых рамок с последующей их фильтрацией.

RPN получает на вход от CNN признаки, на основе которой формирует набор предложений по регионам размещения объектов с некоторой оценкой. Для снижения количества регионов используется алгоритм не-максимального подавления (NMS), существенно снижающий количество регионов. Полученные данные подаются в алгоритм Fast R-CNN. За счет использования одних и тех же сверточных слоев в обоих сетях, скорость работы значительно увеличивается и модель обнаружения объектов может работать в режиме, близком к режиму реального времени.

Региональная полная сверточная сеть (R-FCN) использует RPN для получения предложений по регионам, но в отличие от семейства R-CNN, полносвязные слои после слоя субдискретизации регионов удаляются. Все основные вычисления производятся до этого слоя. После слоя субдискретизации регионов, все регионы имеют оценки вероятности нахождения в них объектов для дальнейшей выборки по среднему (Рисунок 13). Такой подход значительно сокращает число параметров, и в результате R-FCN быстрее, чем модели R-CNN.

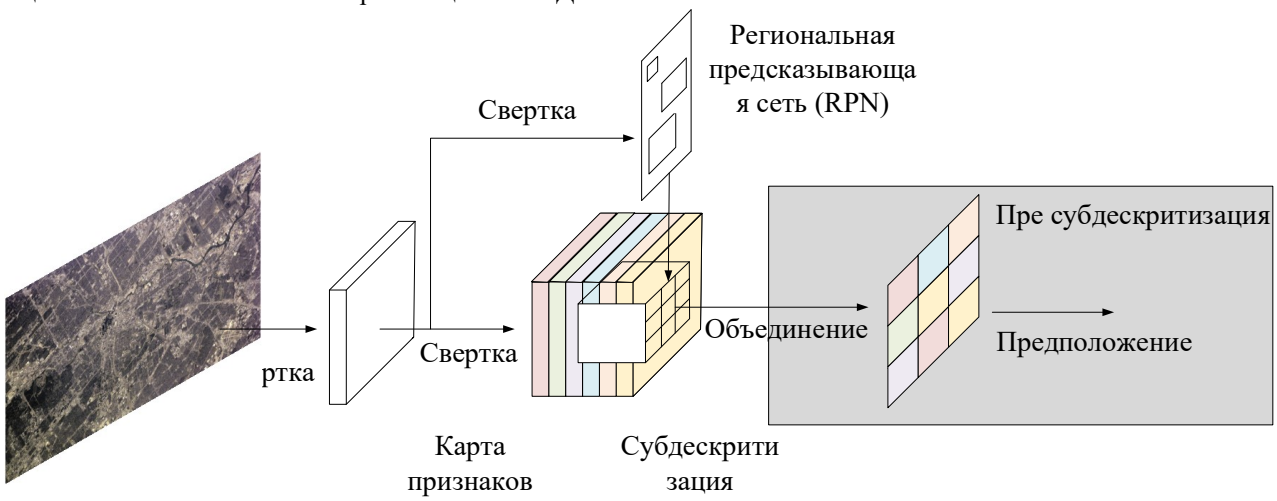

Рисунок 13 - Архитектура R-FCN.

В модели детектор одиночного выстрела (SSD) объекты на изображении обнаруживаются за один прямой проход. SSD предсказывает объекты на изображении, используя однонаправленную свёрточную нейронную сеть, которая определяет фиксированное количество ограничивающих рамок и делает оценку присутствия в них объектов. Особенности сверточных слоев позволяют обнаруживать объекты различных масштабов [4] (Рисунок 14).

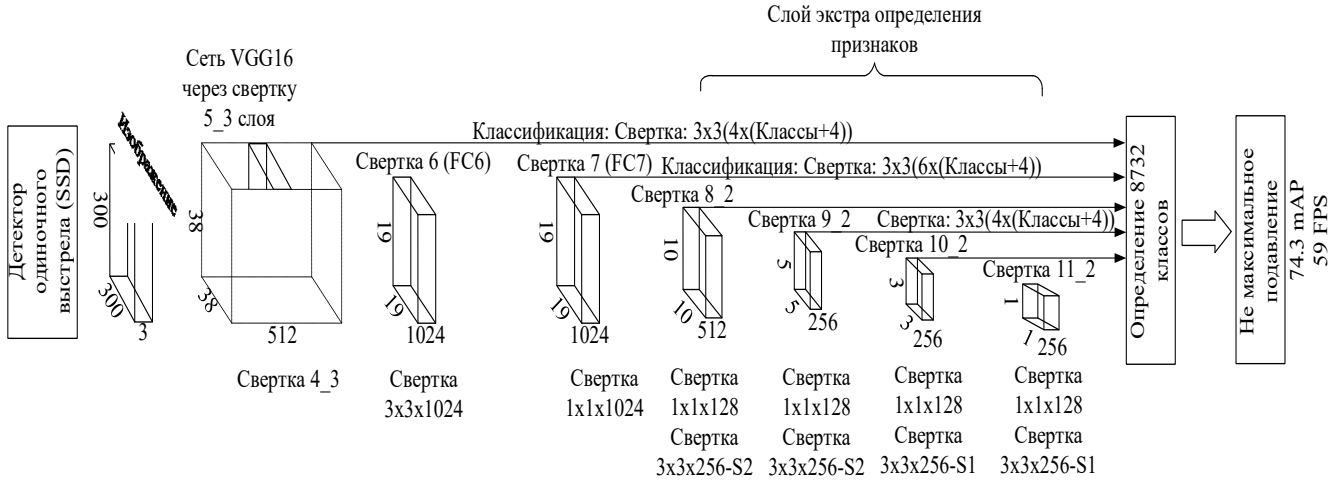

Рисунок 14 - Архитектура сети SSD.

SSD состоит из двух частей:

- базовая CNN для извлечения признаков;
- слои SSD для обнаружения объектов.

Базовая модель представляет собой предварительно обученную сеть для классификации изображений (например, ResNet), из которой удален последний полностью связанный классификационный слой. В результате получается глубокая нейронная сеть, которая может извлечь семантическое значение из входного изображения, сохраняя пространственную структуру изображения. Слои SSD — это просто один или несколько сверточных слоев, добавленных к этой основной модели. Выходы интерпретируются как ограничивающие ячейки и классы объектов в пространственном расположении активации конечных слоев.

SSD в качестве базовой модели для извлечения признаков использует сеть VGG-16. В целом, сеть SSD работает быстрее чем региональные сети, потому что требует только одного прохода.

Модель YOLO (You Only Look Once) использует функции от всего изображения для предсказания ограничивающих рамок, предсказывая их по всем классам одновременно для изображения. Существует несколько реализаций: YOLO, YOLOv2, YOLOv3, YOLO 4 и YOLO 5.

YOLO делит входное изображение на сетку SxS. Если ячейка сетки содержит центр объекта, она отвечает за обнаружение этого объекта. Каждая ячейка сетки предсказывает некоторое количество ограничивающих рамок и их оценки. Оценки показывают, насколько модель уверена в том, что рамка содержит объект, и, кроме того, насколько точна эта рамка.

Каждая рамка содержит предсказания следующих параметров: координаты центра ограничивающей рамки относительно границ ячейки, ширину и высоту рамки относительно всего изображения, а также оценку IoU нахождения объекта в рамке. Каждая ячейка также предсказывает условные вероятности отнесения находящегося в ней объекта к некоторому классу. Эти вероятности относятся к ячейке, содержащей объект. Независимо от числа ассоциированных ограничивающих рамок, для каждой ячейки предсказывается только один набор вероятностей отнесения объекта к классам.

Затем, вероятность отнесения к классу и предсказания ограничивающей рамки умножаются. Это дает оценку доверия по каждому классу для каждой ограничивающей рамки. Эти оценки кодируют как вероятность появления этого класса в рамке, так и то, насколько хорошо предсказанная рамка подходит к объекту.

Гистограмма ориентированных градиентов (HOG) предполагает работу дескрипторов признаков, получающих изображение и вычисляющих векторы признаков. Эти признаки действуют как своего рода числовой "отпечаток пальца", который может быть использован для отличия одного признака от другого. Алгоритм HOG подсчитывает появление градиентной ориентации в локализованных частях изображения. Он делит изображение на небольшие связные области, называемые ячейками, а для пикселей внутри каждой ячейки алгоритм HOG вычисляет градиент изображения вдоль оси x и оси y.

Эти градиентные векторы отображаются в значения цвета от 0 до 255, пиксели с отрицательными изменениями - чёрные, пиксели с большими положительными изменениями - чёрные, а пиксели без изменений серые. Используя эти два значения цвета, конечный градиент вычисляется путем сложения векторов. HOG обычно используется в сочетании с алгоритмами классификации, такими как метод опорных векторов, для выполнения обнаружения объектов.

Метод субдискретизации пространственных пирамид (SPP-net) используется для снижения потерь признаков объектов связанных с оптимизацией размеров изображений для обработки в CNN. CNN требует входное изображение фиксированного размера, которое должно быть ограничено соотношением сторон и ориентацией. При использовании произвольных изображенией, соответствие достигается обрезанием или деформацией исходного изображения. Однако, обрезание может привести к потери значимой части изображения, а деформация - к геометрическому искажению. Зафиксированный на стадии проектирования сети статический размер изображения плохо работает с объектами различных масштабов. В CNN фиксированного размера требуют только полносвязные слои, сверточные и субдискретизационные слои могут работать с изображениями произвольного размера.

Такие входные векторы могут быть получены с помощью подхода Bag-of-Words (BoW), который объединяет признаки вместе. Метод субдискретизации пространственных пирамид улучшает подход BoW, он сохраняет пространственную информацию путем объединения признаков в локальные пространственные области. Эти области имеют размеры, пропорциональные размеру изображения, что означает, что количество областей остается неизменным независимо от размера изображения. Чтобы использовать любую глубокую нейронную сеть с изображениями произвольного размера, необходимо просто заменить последний слой субдискретизации слоем пространственных пирамид. Это позволяет использовать не только произвольные размеры, но и произвольную ориентацию.

SPP-net вычисляет карты признаков со всего изображения один раз, а затем объединяет их в произвольные области для создания представлений фиксированной длины для детектора. Это позволяет избежать многократных вычислений признаков в CNN. SPP-net быстрее, чем методы R-CNN, при достижении большей точности.

Согласно теореме об универсальной аппроксимации, при достаточной емкости, сеть прямого распространения с одним слоем достаточна для представления любой функции. Однако слой может быть массивным, и сеть будет склонна к переобучению данных. Поэтому в научно-исследовательском сообществе существует общая тенденция, согласно которой сетевая архитектура должна быть более

глубокой, в связи с чем на практике применяются глубокие остаточные нейронные сети для распознавания изображений.

Начиная с AlexNet, современная архитектура CNN становится всё глубже и глубже. В то время как AlexNet имел только 5 сверточных слоев, сеть VGG и GoogleNet имели 19 и 22 слоя соответственно [11].

Однако увеличение глубины сети не работает простым сложением слоев вместе. Глубокие сети трудно обучать из-за проблемы исчезновения градиента. По мере углубления производительность такой сети перестает расти или даже начинает быстро деградировать. Эта проблема была решена в ResNet.

Основная идея заключается в ведении так называемого “короткого пути”, который пропускает один или несколько слоев (Рисунок 15).

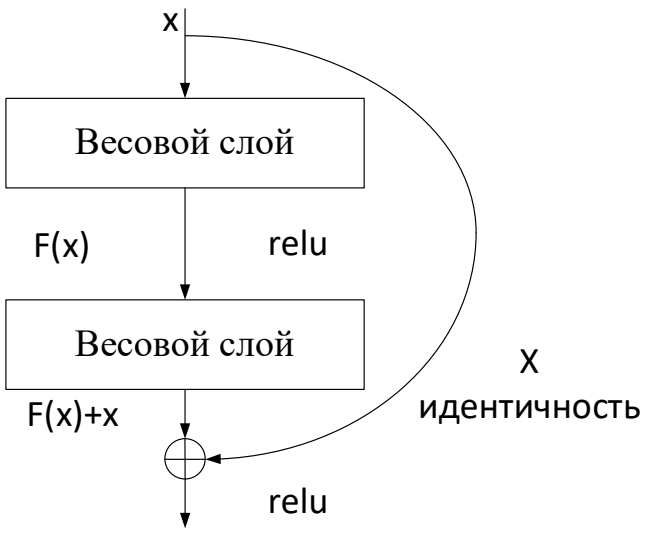

Рисунок 15 - Основная идея ResNet.

Включение таких “коротких путей” в архитектуру сети позволяет формировать сети, глубиной более 1000 слоев.

U-NET представляет собой полносвязную сверточную сеть, включающую сверточную и разверточную части. На каждом шаге количество каналов признаков удваивается. Сверточная часть представляет собой обычную сверточную сеть, содержащую слои свертки, активации и субдискретизации.

Каждый шаг разверточной части содержит слой, обратный слою субдискретизации, который расширяет карту признаков, в сочетании со сверточным слоем, который уменьшает количество каналов признаков. Затем выполняется конкатенация с соответствующим образом обрезанной картой признаков из сверточной части и два последовательных слоя свертки и активации. На последнем слое свертка используется для приведения каждого вектора признаков до требуемого количества классов. Структура сети показана на рисунке 16.

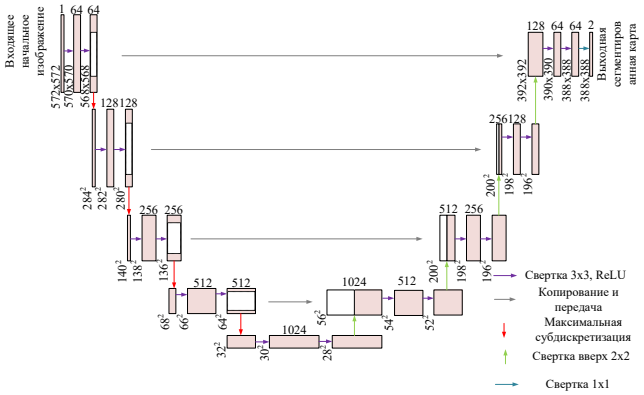

Рисунок 16 - Архитектура U-NET.

Архитектура состоит из большого числа различных операций, проиллюстрированных стрелками в диаграмме архитектуры. Входное изображение подается в сеть, а затем данные распространяются в сеть по всем возможным путям, и в конце появляется готовая сегментированная карта. Каждый розовый прямоугольник соответствует многоканальной карте объектов. Верхний номер розового прямоугольника обозначает каналы, а нижний - размер. Большинство операций являются свертками, за которыми следует нелинейная функция активации.

Основными преимуществами U-NET являются:

- малый объем данных, необходимых для обучения;
- выделение объектов - четко выделяются границы объектов на изображениях со слабой контрастностью.

U-Net хорошо зарекомендовала себя в таких случаях, как сегментация нейронных структур в электронной микроскопии, особенно объектов с нечеткими границами и низкой контрастностью. Достигнутые результаты намного лучше, чем у других сверточных сетей для сегментации изображения. Эта архитектура также хороша для сегментации клеток, которые имели сильные вариации формы, слабые внешние границы и похожие структуры.

Сверточная нейронная сеть супер разрешения (SRCNN) реализует увеличение разрешения входного изображения. Сеть SRCNN имеет следующие особенности [11]:

- SRCNN полностью сверточная, что обеспечивает скорость ее работы;
- обучение проводится по критерию качества работы фильтров, а не достигнутой точности. Такой подход позволяет увеличивать масштаб изображения.
- SRCNN не требует дообучения при эксплуатации. После достижения необходимых параметров фильтров при обучении, сеть может выполнять простой прямой переход (simple forward pass) для получения изображения с высоким разрешением. Для получения результата не требуется оптимизация функции потерь для каждого изображения;
- SRCNN имеет сквозную (end-to-end) архитектуру. На вход сети подается изображение и на выходе получается изображение с высоким разрешением, без каких-либо промежуточных шагов (Рисунок 17).

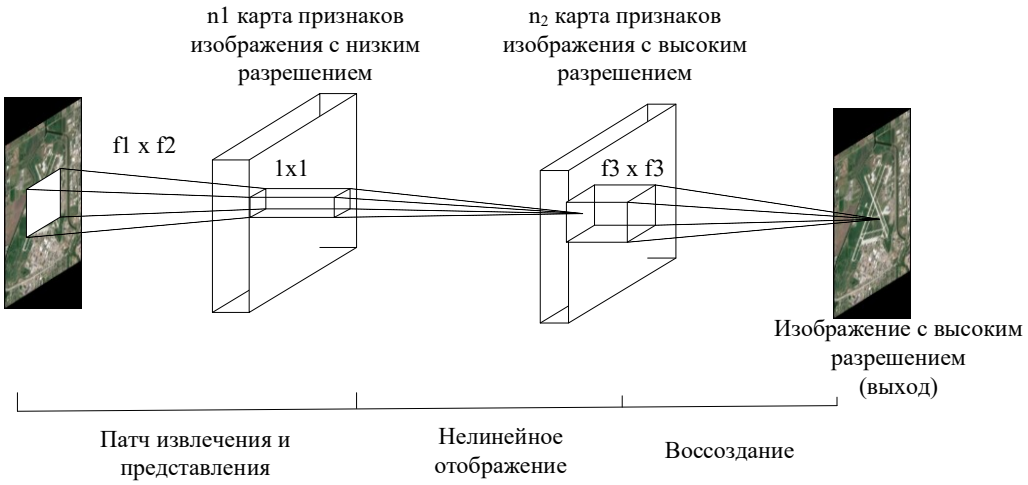

Рисунок 17 - Сверточная нейронная сеть с супер разрешением (SRCNN).

IV. Заключение

В статье приведен краткий обзор особенностей архитектур сверточных нейронных сетей, количественные показатели оценки качества их функционирования и типы решаемых задач, а также рассмотрены основные особенности работы с изображениями и основные возникающие сложности, выделены особенности обработки аэрокосмических снимков. Кроме того, проведен краткий анализ существующих моделей работы с изображениями на основе сверточных нейронных сетей. Рассмотренные в статье вопросы позволяют сделать следующие выводы.

1. Проблема распознавания объектов на аэрокосмических снимках является сложной и на сегодняшний день ее разработка только начинается. Основными сложностями являются:

- произвольная ориентация объектов на аэрокосмических снимках;
- вариации масштаба аэрокосмических снимков;
- неравномерная плотность объектов;
- сложные фоновые условия;
- большое соотношение сторон.

2. Алгоритмы для решения задач по распознаванию объектов на обычных изображениях достаточно хорошо разработаны и применяются в самых различных сферах науки и техники. Для успешного решения задачи обработки аэрокосмических снимках необходимо адаптировать уже известные алгоритмы и их сочетания, для достижения достаточных количественных показателей качества работы сети.

Таблица 1 - Характерные особенности нейронных сетей

Название	Особенность архитектуры	Основной эффект
R-FCN	RPN используется для получения предложений по регионам, но в отличие от семейства Faster R-CNN, полносвязные слои после слоя субдискретизации регионов отсутствуют	Модель может работать в режиме времени, близком к реальному. Количество параметров снижено относительно Faster R-CNN
Faster R-CNN	В качестве алгоритма селективного поиска используется RPN для поиска регионов	Модель может работать в режиме времени, близком к реальному
YOLO	Использует функции от всего изображения для предсказания ограничивающих рамок	Более быстрый чем, SSD метод, однако с несколько меньшей точностью
SSD	Предсказывает объекты на изображении, используя однонаправленную свёрточную нейронную сеть, которая определяет	Все вычисления осуществляются за один прямой проход, что обеспечивает работу в реальном масштабе времени

	фиксированное количество ограничивающих рамок и делает оценку присутствия в них объектов	
Гистограмма ориентированных градиентов (HOG)	Алгоритм делит изображение на небольшие связные области, называемые ячейками, а для пикселей внутри каждой ячейки алгоритм HOG вычисляет градиент изображения вдоль оси x и оси y	Обеспечивает эффективную локализацию объектов
Метод субдискретизации пространственных пирамид (SPP-net)	Вычисляет карты признаков со всего изображения один раз, а затем объединяет их в произвольные области для создания представлений фиксированной длины для детектора	Обеспечивает снижение потерь признаков объектов связанных с оптимизацией размеров изображений для обработки в CNN
ResNet	Наличие в структуре сети “короткого пути”, который пропускает один или несколько слоев	Позволяет формировать сети, глубиной более 1000 слоев.
U-NET	Включает сверточную и разверточную части	Обеспечивает четкое выделение границ объектов даже при слабой контрастности
SRCNN	Полностью сверточная сеть, обучение которой проводится по критерию качества работы фильтров	Обеспечивает увеличение разрешения входного изображения

Работа выполнена при финансовой поддержке Фонда содействия инновациям (ФСИ) в рамках Договора 94С2/МОЛ/73887.

Библиография

[1] Adrian Rosebrock. Intersection over Union (IoU) for object detection. https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection.

[2] Jonathan Hui. mAP (mean Average Precision) for Object Detection https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173.

[3] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. https://arxiv.org/pdf/1506.01497.pdf.

[4] Wei Liu1, Dragomir Anguelov2, Dumitru Erhan3, Christian Szegedy3, Scott Reed4, Cheng-Yang Fu1, Alexander C. Berg1. SSD: Single Shot MultiBox Detector. https://arxiv.org/pdf/1512.02325.pdf

[5] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. https://arxiv.org/pdf/1506.02640.pdf.

[6] Jian Ding, Nan Xue, Gui-Song Xia, Xiang Bai, Wen Yang, Michael Ying Yang, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, Liangpei Zhang. “Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges”. https://arxiv.org/pdf/2102.12219.pdf.

[7] Z. Liu, H. Wang, L. Weng, and Y. Yang, “Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds,” IEEE Geosci. Remote Sensing Lett., vol. 13, no. 8, pp. 1074–1078, 2016.

[8] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, “Dota: A large-scale dataset for object detection in aerial images,” in CVPR, 2018.

[9] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, “Learning roi transformer for oriented object detection in aerial images”. https://arxiv.org/pdf/1812.00155.pdf.

[10] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers2, and A.W.M. Smeulders. Selective Search for Object Recognition. http://www.huppelen.nl/publications/selectiveSearchDraft.pdf.

[11] Vincent Feng. An Overview of ResNet and its Variants. https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035.

[12] Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Computer Vision–ECCV 2014. Springer (2014) 184–199.

Investigation of convolutional neural networks for object detection in aerospace images

Vladimir Skripachev, Mikhail Guida, Nikolay Guida, Alexander Zhukov

Abstract. **The article discusses current algorithms for solving problems of object recognition in images, their main features and advantages. A brief analysis of existing models of working with images based on convolutional neural networks is carried out. A brief overview of the features of convolutional neural network architectures, quantitative indicators for assessing the quality of their functioning and the types of tasks to be solved, the main features of working with images and the main emerging difficulties are considered, the features of processing aerospace images are highlighted. The problem of object recognition in aerospace images is formulated by adapting existing relevant algorithms and their combinations. The main problems of processing aerospace images and approaches to their solution, the application of established methods of object recognition in conventional images to the problems of object recognition in aerospace images are shown. The analysis of various neural network architectures in the prism of solving object recognition problems in aerospace images is carried out. Conclusions are drawn regarding the most successful combinations of various algorithms in the structure of neural networks when recognizing objects in aerospace images. The main factors that make it difficult to recognize objects in aerospace images and the directions of work to reduce their impact on the accuracy of neural networks when recognizing objects in aerospace images are determined.**

Keywords: **convolutional neural networks (CNN), subdiscretization, Region prediction network (RPN), convolution, SRCNN, U-NET, ResNet, aerial image processing, oriented bounding box (OBB), Landmark detection, spatial pyramid subdiscretization method (SPP-net), histogram of oriented gradients (HOG), R-FCN, Faster R-CNN, YOLO, SSD..**

REFERENCES

[1] Adrian Rosebrock. Intersection over Union (IoU) for object detection. https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection.

[2] Jonathan Hui. mAP (mean Average Precision) for Object Detection https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173.

[3] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. https://arxiv.org/pdf/1506.01497.pdf.

[4] Wei Liu1, Dragomir Anguelov2, Dumitru Erhan3, Christian Szegedy3, Scott Reed4, Cheng-Yang Fu1, Alexander C. Berg1. SSD: Single Shot MultiBox Detector. https://arxiv.org/pdf/1512.02325.pdf

[5] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. https://arxiv.org/pdf/1506.02640.pdf.

[6] Jian Ding, Nan Xue, Gui-Song Xia, Xiang Bai, Wen Yang, Michael Ying Yang, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, Liangpei Zhang. "Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges". https://arxiv.org/pdf/2102.12219.pdf.

[7] Z. Liu, H. Wang, L. Weng, and Y. Yang, "Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds," IEEE Geosci. Remote Sensing Lett., vol. 13, no. 8, pp. 1074–1078, 2016.

[8] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "Dota: A large-scale dataset for object detection in aerial images," in CVPR, 2018.

[9] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, "Learning roi transformer for oriented object detection in aerial images". https://arxiv.org/pdf/1812.00155.pdf.

[10] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers2, and A.W.M. Smeulders. Selective Search for Object Recognition. http://www.huppelen.nl/publications/selectiveSearchDraft.pdf.

[11] Vincent Feng. An Overview of ResNet and its Variants. https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035.

[12] Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Computer Vision–ECCV 2014. Springer (2014) 184–199.