

МИНОБРНАУКИ РОССИИ
ФГБОУ ВО «СГУ ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра теоритических основ компьютерной безопасности и криптографии

СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ НЕЙРОННЫМИ СЕТЯМИ
КУРСОВАЯ РАБОТА

студента 2 курса 231 группы
направления 100501 — Компьютерная безопасность
факультета КНиИТ
Окунькова Сергея Викторовича

Научный руководитель
Доцент

И. И. Слеповичев

Заведующий кафедрой
доцент, к.ф.-м.н.

М. Б. Абросимов

Саратов 2021

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Основные понятия и определения	5
2 Подходы к задаче сегментации изображения.....	11
3 Алгоритмы машинного обучения для решения задачи сегментации изображения.....	13
3.1 FCN (fully convolutional network)	13
3.2 U-Net	14
3.3 Mask RCNN	15
3.4 FRRN	16
3.5 PSPNet	17
3.6 RefineNet	18
3.7 GCN	18
ЗАКЛЮЧЕНИЕ	20
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	21

ВВЕДЕНИЕ

В наше время компьютеры стали очень важной частью нашей жизни. Существует огромный пласт задач, решение которых мы уже не представляем без его использования. Поэтому не удивительно, что с каждым годом все больше и больше набирают популярность такие сферы, как нейронные сети, машинное и глубокое обучение. Данные направления не просто способны решать большинство важных задач, но и делать это эффективно с помощью обучаемых систем, а не программируемых.

Поэтому в данный момент одним из главных направлений в развитии IT-технологий является тема искусственного интеллекта и внедрение его во все сферы человеческой жизни. В некоторых областях искусственный интеллект уже превосходит человеческие возможности, так, например, уже существуют модели искусственного интеллекта, которые способны переиграть даже самых умелых шахматистов.

Нейронные сети и машинное обучение уже давно используются для решения многих бытовых задач: распознавание и классификация объектов на изображении, классификация текста, распознавание речи, выдача рекомендаций, основанных на наших предпочтениях, перевод текста с картинки и многих других. Одной из таких задач, а именно задаче сегментации изображения и посвящена данная работа.

Сегментацией изображения называют выделение однородных по некоторому критерию областей на изображении. Таким образом, главной задачей сегментации, можно назвать разбиение изображения по заданному критерию на множества областей, которые у наблюдателя будут ассоциироваться с объектами на нашем исходном изображении или их частями, в результате которого на выходе получаются карты областей (сегментов). Иначе говоря, результатом сегментации является отображение множества точек изображения в множество итогового числа разбиений на изображении.

На практике чаще всего сегментацию используют для анализа медицинских изображений, например, для обнаружения различных опухолей и составлений рекомендаций по их лечению, выделения объектов на спутниковых снимках, распознавания лиц или отпечатков пальцев, распознавания дорожных знаков, для распараллеливания информационных потоков при передаче изображений высокого разрешения, а так же в сфере машинного зрения.

Задача сегментации не новая для человечества, потому ее исследованию было посвящено огромное количество научных работ. Благодаря чему сегодня существует много способов ее решения. В данной работе будут рассмотрены некоторые методы и алгоритмы машинного обучения для решения данной задачи, а также будут подробно описаны принципы построения и обучения искусственных нейронных сетей.

1 Основные понятия и определения

Искусственный интеллект (ИИ, AI, Artificial intelligence) - это наука, описывающая создание машины или программы, способной имитировать человеческое поведение для выполнения определенной задачи и обучаться за счет полученной в результате работы информации.

Другими словами ИИ можно описать, как технологию, которая способна мыслить как человек и выполнять функции человека с большей точностью. На практике же достаточно сложно создать идеальную модель ИИ, способного грамотно выполнять заданные ей функции. Поэтому для решения большинства задач в этой сфере используют машинное обучение.

Машинное обучение (Machine learning, ML) - это метод анализа данных, который автоматизирует построение аналитической модели.[1] Данная область ИИ воплощает в себе идею того, что машина способна обрабатывать полученные данные, анализировать их и на основе анализа обучаться с минимальным вмешательством человека.

Однако данная область ИИ все же требует помощи в самом процессе обучения со стороны человека, а именно загрузку корректных данных, рассматривающих все возможные случаи, на которых машина будет обучаться. Иначе говоря, несмотря на то, что данная система обучения машины позволяет решить огромное количество задач, она не способна сама генерировать эти задачи.

Глубокое обучение (глубинное обучение, Deep learning, DL) - это тип машинного обучения, который обучает компьютер выполнять задачи человека.[1]

Данная система обучения вместо того, чтобы ждать тесты от человека, для дальнейшей их обработки по формулам, заданным изначально, сама устанавливает начальные параметры и учит машину обучаться самостоятельно.

Таким образом, если представить AI, ML и DL в виде трех множеств, то мы получим картину, изображенную на рисунке 1:

$$DL \subset ML \subset AI$$

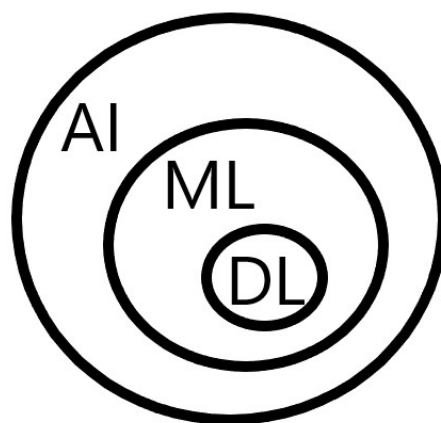


Рисунок 1 – Представление AI, ML и DL в виде множеств

Компьютерное зрение - это одна из областей ИИ, которая отвечает за распознавания компьютером визуального мира: фотографий и видео. Оно подразумевает под собой методы, позволяющие обучить машину анализировать объекты и извлекать из этого какую-то информацию.

К задачам компьютерного зрения относятся:

1. Классификация объектов на изображение
2. Сегментация изображения
3. Обнаружение объектов на изображение

Для решение задач в сфере машинного и глубокого обучения часто используют искусственные нейронные сети.

Искусственные нейронные сети (ИНС, artificial neural networks, ANN) - математическая модель и ее программное воплощение, созданное на основе модели биологических нейронных сетей (сетей нервных клеток организма).

ANN состоит из искусственных нейронов (ИН, artificial neuron, AN), являющихся программной реализацией нейронов живых организмов и представляющих из себя нелинейную функцию, называемую передаточной функцией, от массива входных данных ($[x_1, x_2, \dots, x_n]$). Основная функция ИН - принятие входного сигнала с последующей его обработкой и передачей результата на другие нейроны.

Из всего выше описанного очевидно, что некоторые AN связаны между собой. Как и в биологии данную связь называют синоптической связью.

Синапсом в искусственных нейронных сетях называют связь между нейронами. По нему передаются выходные связи с одного нейрона на другой. У любого синапса имеется такой параметр, как весовой коэффициент w_i , который показывает текущее сосотяние нейрона, а более сложные синапсы также могут

иметь память. Состояние нейрона в определенный момент времени вычисляется как взвешенная сумма его входов:

$$s = \sum_{i=1}^n x_i w_i,$$

где n - число входов нейрона, x_i - i -й входной сигнал, w_i - вес i -го синапса.

В большинстве случаев связь между нейронами представляют в виде матрицы W , которую соответственно называют матрицей веса.

Помимо синапсов важную роль в связи нейронов между собой играют аксоны. **Аксон** - это выходная связь нейрона, с помощью которой выходной сигнал нейрона поступает на синапсы других нейронов. Его значение будет равно:

$$Y = f(s),$$

где s - это состояние нейрона в данный момент, а f - это передаточная функция, в качестве которой чаще всего используются сигмоид, линейный порог, гиперболический тангенс или жесткую пороговую функцию. Данная функция одинакова для нейронов одного слоя, но для нейронов разных слоев она могут выбираться разные функции.

Проанализировав все выше описанное можно представить, как в общем выглядит нейрон:

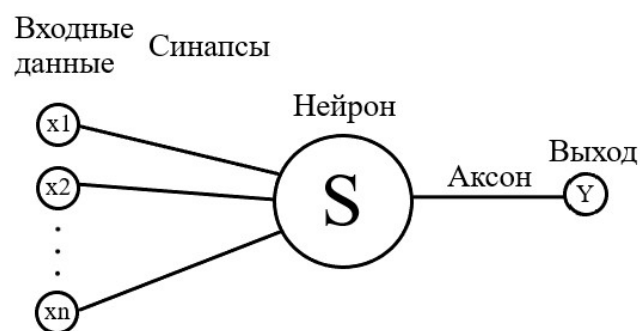


Рисунок 2 – Общий вид искусственного нейрона

Уже на данном этапе достаточно информации, чтобы сделать вывод о том, как схематически можно изобразить искусственную нейронную сеть. Пример нейронной сети изображен на рисунке 3.

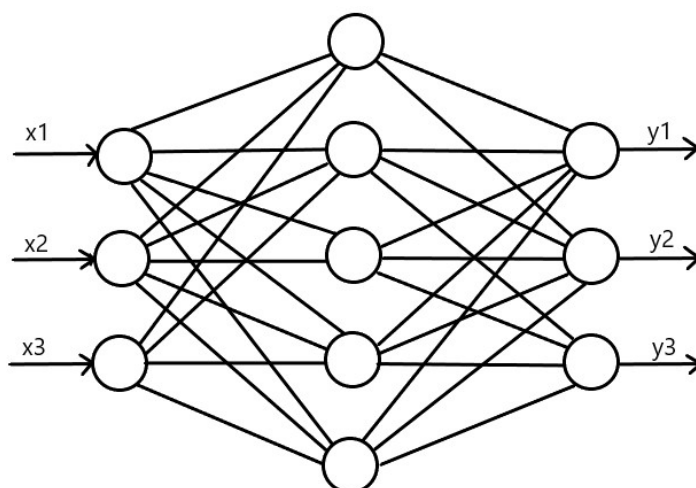


Рисунок 3 – Пример искусственной нейронной сети

Обучение искусственной нейронной сети

Как говорилось выше, для решения задачи не достаточно просто написать рабочую нейросеть, ее еще нужно и обучить. Под процессом обучения понимается выбор таких параметров, при которых нейросеть решает поставленную задачу с большей эффективностью. Схема обучения изображена на рисунке 5. Математически его можно описать следующим образом. В процессе работы нейросеть, как было выяснено выше, формирует выходной сигнал $Y = G(X)$, являющийся реализацией какой-то функции. И пусть ответом на поставленную задачу будет функция $Y = F(X)$, заданная с помощью параметра входных-выходных данных так, что

$$Y^k = F(X^k), \text{ где } k = 1, \dots, N$$

Обучением же будет состоять из генерации функции $G(X)$, близкой к $F(X)$. Оно будет состоять из множества итераций, на каждой из которых функция $G(X)$ будет все ближе и ближе к функции $F(X)$, а для определения степени их близости друг к другу используют некоторую функцию $D_F(G)$, которую называют функцией ошибок или целевой функцией. Таким образом обучение ИНС будет проходить до того момента, пока функция $D_F(G)$ не примет минимальное свое значение.

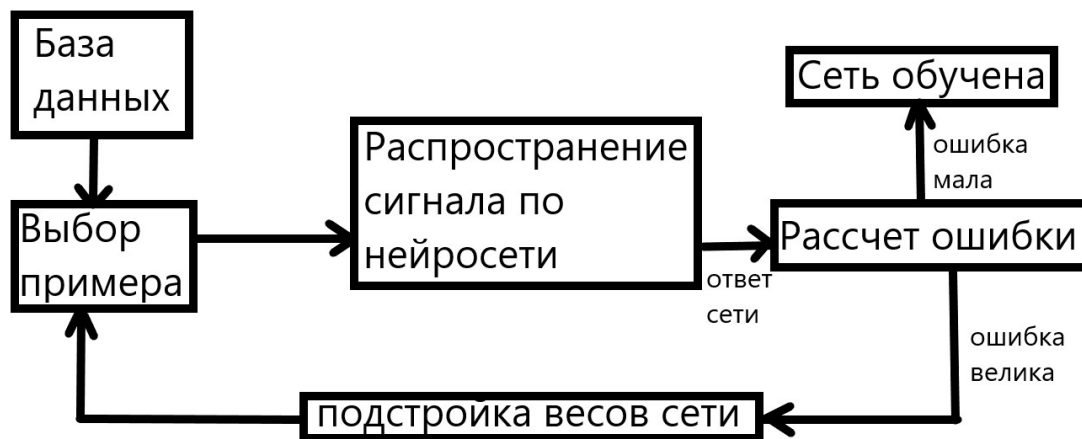


Рисунок 4 – Процесс обучения нейронной сети

Существует две главные стратегии обучения:

- с учителем или по другому контролируемое обучение (Supervised Learning)
- нейросети на вход подаются данные и она обучается путем обработки этих данных
- неконтролируемое обучение (Unsupervised Learning) - нейросеть обучается в соответствии с некоторым правилом, при этом данные для обучения не требуются

Для обучения нейронной сети используют следующие типы алгоритмов:

1. алгоритмы локальной оптимизации с вычислением частных производных первого порядка
2. алгоритмы локальной оптимизации с вычислением частных производных первого и второго порядка
3. стохастические алгоритмы оптимизации
4. алгоритмы глобальной оптимизации (задачи глобальной оптимизации решаются с помощью перебора значений переменных, от которых зависит целевая функция)

Сверточная нейронная сеть

Наилучшие результаты в задачах распознавания объектов на изображениях показала **сверточная нейронная сеть (Convolutional Neural Network, CNN, СНС)**, которую называют логическим продолжением идей когнитрона и некогогнитрона. Главной причиной успеха данной архитектуры является учет двумерной топологии изображения. СНН работает по принципу масштабирования (свертки) изображения.

Свертка - это операция над матрицами A и B, размерностью $(n_x * n_y)$ и $(m_x * m_y)$ соответственно, результатом которой является матрица $C = A * B$, размер которой $(n_x - m_x + 1 * n_y - m_y + 1)$, а элементы рассчитываются по формуле:

$$C_{i,j} = \sum_{u=0}^{m_x-1} \sum_{v=0}^{m_y-1} A_{i+u,j+v} B_{u,v}.$$

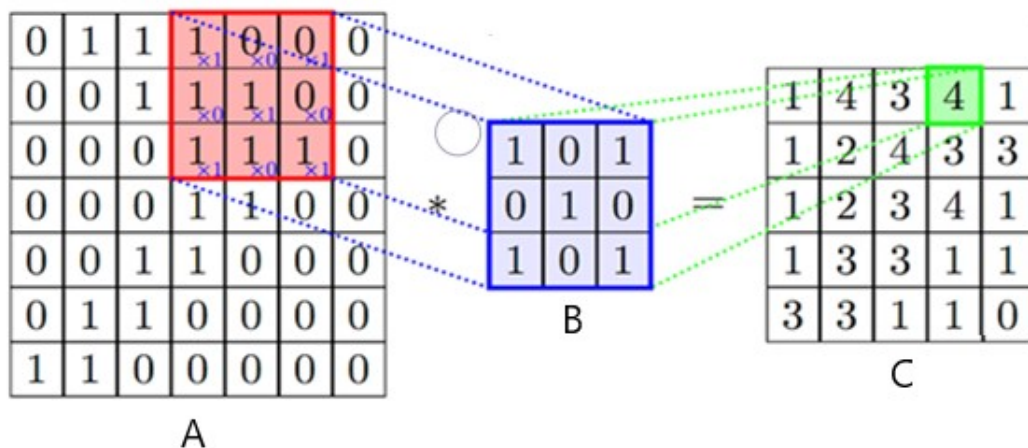


Рисунок 5 – Свертка

Архитектура данного типа ИНС строится на чередовании слоев свертки и подвыборки, которые состоят из карт признаков. Они в свою очередь отвечают за поиск определенных признаков. К примеру одна карта ищет предметы красного цвета, вторая - синего.

2 Подходы к задаче сегментации изображения

В данный момент выделяют два основных альтернативных друг другу подхода, которые используют алгоритмы сегментации изображений:

1. Выделения границ областей;
2. Нарращивания точек области.

Первый подход основан на том, что переход от одного объекта к другому на изображении происходит не плавно. Иначе говоря, данный подход представляет из себя определение пикселей, которые находятся на границе двух областей изображения с дальнейшим выделением контуров(точек, между которыми происходит изменение яркости). Его очевидным плюсом является то, что по мимо самих границ областей он позволяет идентифицировать и сами области. Ниже приведен пример сегментации с помощью данного подхода.



Рисунок 6 – Сегментация изображения с помощью выделения границ областей

Второй подход наиболее распространенный и представляет из себя выделение точек на изображение, которые обладают одинаковыми свойствами, а после объединение их в область, которой присваивается смысловая метка. Именно данный подход преимущественно используется в алгоритмах найросетей для решения задачи сегментации.

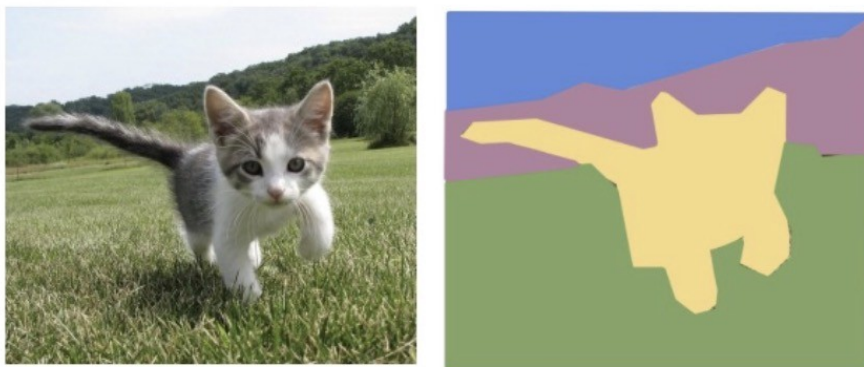


Рисунок 7 – Сегментация изображения с помощью наращивания точек области

3 Алгоритмы машинного обучения для решения задачи сегментации изображения

Для решение задачи сегментации изображения в основном выделяют следующие виды алгоритмов:

1. Пороговые алгоритмы;
2. Алгоритм наращивания областей;
3. Граничные алгоритмы;
4. Алгоритмы сегментации на основе кластеризации.

3.1 FCN (fully convolutional network)

Один из самых простых и наиболее распространенных алгоритмов искусственных нейронных сетей для сегментации изображения, который позволяет работать с изображениями любого размера.

FCN представляет из себя полносверточную НС, которая за некоторое количество шагов выдает карту классов изображения. При этом на каждом шаге происходит свертка или транспонированная свертка изображения. На вход такая НС получает само изображение размерностью ($W \times H$) и множество классов (C), к которым могут принадлежать пиксели с изображения. Первая часть слоев такой нейронной сети является реализацией уменьшения размерности (кодировки) с целью сбора контекстной информации, а вторая часть - увеличения размерности (декодировки) с целью восстановления пространственной информации, это связано с тем, что свертка матриц больших размерностей с большими количествами каналов является очень затратным процессом. В итоге размер изображения на входе равен размеру изображения на выходе. Конечная классификация выбирается поиском максимума по классам из многомерного массива, называемого тензором, который имеет размерность $C \times W \times H$. Для обучения такой ИНС используют функцию кросс-энтропии в качестве функции потерь для метода обратного распространения ошибок, имеющую вид:

$$H(p, q) = - \sum_x p(x) \log(q(x)),$$

где $p(x)$ - искомая вероятность, $q(x)$ - действительная вероятность.

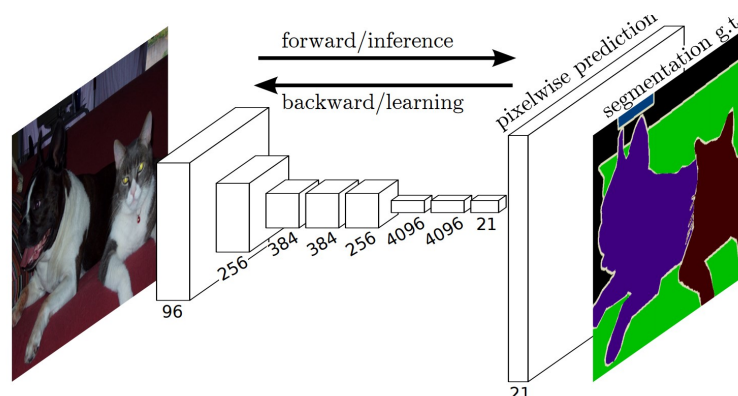


Рисунок 8 – Архитектура FCN

Не смотря на высокую эффективность имеет несколько больших недостатков, главными из которых являются артефакты, расположенные в шахматном порядке, появляющиеся в связи с перекрытием выходов в операции транспонированной свертки, низкая разрешающая способность по краям, которая обусловлена потерей информации во время кодирования, и отсутствие полностью связанных слоев.

3.2 U-Net

На сегодняшний день U-Net считается одним из стандартов искусственных нейронных сетей для выполнения задачи сегментации изображения. ИНС типа U-Net представляют из себя улучшение архитектуры FCN. Свое название данная нейросеть получила за счет своей симметричной формы, отличающейся от других модификаций нейросети FCN. Главным различием между FCN и U-Net является наличие skip связей между выходами блоков уменьшения размерности и входами блоков увеличения размерности, находящихся на одном уровне, у последней, при этом очевидно, что за счет симметричной архитектуры количество блоков данных типов одинаково, что позволяет более точно передать информацию о свойствах каждого пикселя. Для обучения данной ИНС используется та же функция потерь, что и для обучения ИНС FCN.

Архитектуру U-Net можно разделить на 3 части:

1. Путь кодировки
2. Узкое место
3. Путь декодировки

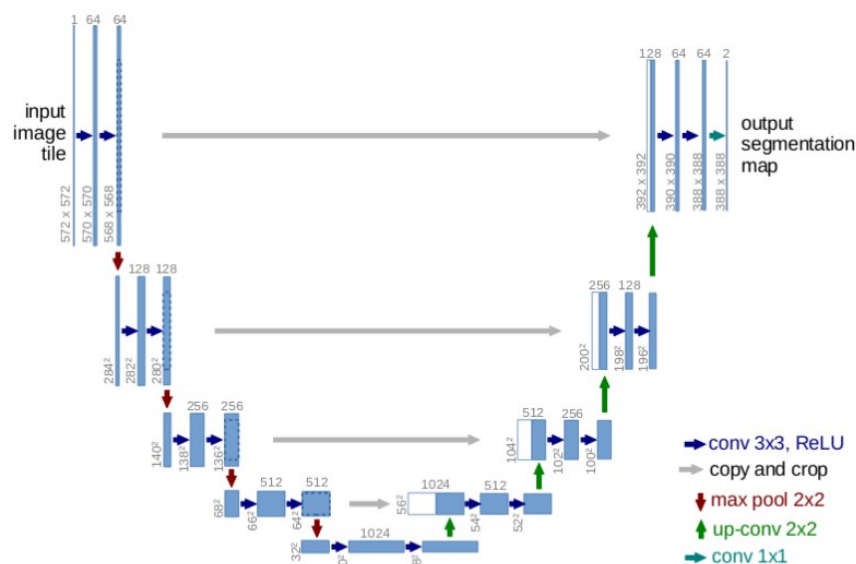


Рисунок 9 – Архитектура U-Net

Основное преимущество алгоритма U-Net является скорость по сравнению с другими алгоритмами сегментации изображения.

3.3 Mask RCNN

Данный алгоритм совмещает в себе два алгоритма: выше описанный FCN и Faster R-CNN, являющийся улучшением алгоритма Fast R-CNN, который в свою очередь является улучшением обычного R-CNN.

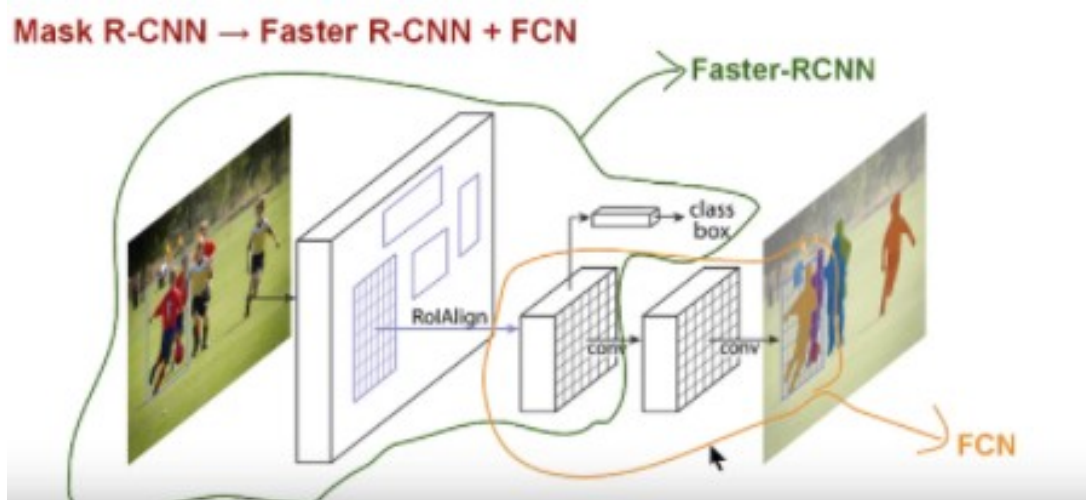


Рисунок 10 – Архитектура Mask RCNN

В основе Mask R-CNN лежит следующая архитектура:

1. Изображение, поданное на вход нейронной сети, обрабатывается первым слоем. По итогу такой обработки получается карта признаков.

2. Полученная карта передается следующему слою, на котором скользящее окно проходит по ней. По итоге данного шага выносятся решение о наличии объекта в регионах.
3. Результат предыдущего этапа передается на вход алгоритма FCN, для определения более точного расположения объекта на изображении.

Функция потерь для данного алгоритма - это сумма потерь полученных во время процесса классификации, создании ограниченной области вокруг объекта и создании самой маски. Т.е. :

$$L = L_{cls} + L_{box} + L_{mask}, \text{ где}$$

$$L_{mask} = \frac{1}{m^2} \sum_{1 \leq i, j \leq m} y_{ij} \log(y_{ij}^k) + (1 - y_{ij}) \log(1 - y_{ij}^k),$$

где k основной класс истинности.

Главным плюсом данного алгоритма является высокая точность вкупе с простым обучением и возможностью дообучить ИНС, но при этом сам процесс обучения является весьма долгим.

3.4 FRRN

FRRN (Остаточные сети с полным разрешением) - алгоритм многомерной двух поточной обработки, один из этих потоков называют остаточным, а другой - поток объединения. Первый из них отвечает за обработку карт объектов вполное разрешение, а второй сокращает их. Другими словами поток объединения работает с семантической информацией высокого уровня, а поток - с информацией пикселей низкого уровня.

При обучении данной ИНС, после каждого максимального пула, выполняется совместная обработка карт характеристик из каждого потока с целью объединения их информации.

Главный минус такого алгоритма является его затратность: обработка высокого разрешения требует значительное количество вычислительных ресурсов.

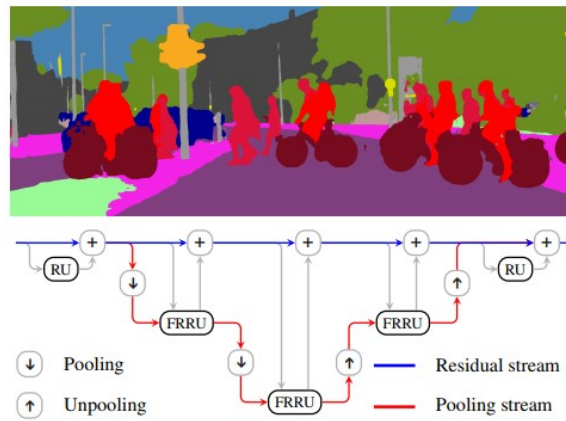


Рисунок 11 – Архитектура FRRN

3.5 PSPNet

Рассмотренный выше алгоритм FRRN является отличным примером масштабной обработки, имеющим свои плюсы и минусы. Алгоритм PSPNet в свое время предложил хороший способ обойти большинство главных минусов FRRN, который заключается в нескольких масштабах объединения.

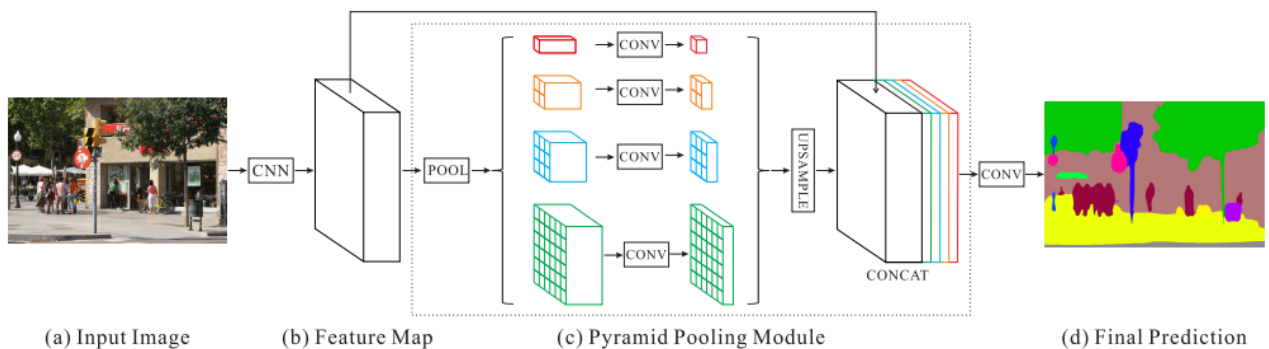


Рисунок 12 – Архитектура FRRN

Реализация данного алгоритма как обычно начинается со стандартного извлечения признаков, а для дальнейшей обработки использует функции третьей понижающей дискретизации. Для получения многомасштабной информации алгоритм применяет четыре различные операции максимального пула к четырем различным размерам изображения, полученным на разных шагах, при этом не обрабатывая сами изображения, а после делает легкую свертку карты признаков каждого из них, объединяя их после и получая по итогу единую карту признаков, которую в свою очередь масштабируют до нужного размера с помощью билинейной интерполяции. Все это позволяет избежать реализации множества сверток, что решает проблему со временем выполнения данного алгоритма и затратностью.

3.6 RefineNet

Еще одним улучшением алгоритма FRRN можно назвать алгоритм RefineNet. Он так же работает с несколькими картами признаков объектов разномасштабных вариаций изображения, но делает это по принципу снизу вверх. Его главным плюсом является то, что он обрабатывает данные карты как по отдельности не зависимо друг от друга, так и вместе, благодаря чему нужно обрабатывать и комбинировать вариации с низкими и высокими разрешениями.

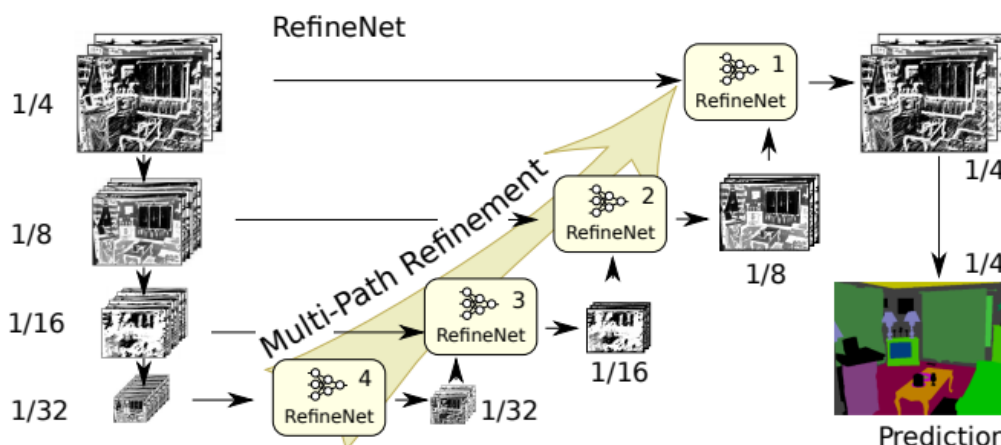


Рисунок 13 – Архитектура RefineNet

3.7 GCN

В нескольких предыдущих алгоритмах был рассмотрен такой процесс, как многомасштабная обработка. Однако в вышеописанных алгоритмах он выполнялся отдельно, а результаты работы алгоритма впоследствии объединялись. Данный алгоритм решает этот вопрос, позволяя получить многомасштабную информацию за один раз. Это достигается за счет использования в данном алгоритме больших одномерных ядер вместо квадратных. Их главными плюсами являются возможность эффективного масштабирования и возможность обработки более одной шкалы за раз.

В остальном GCN работает по тому же принципу, что и другие алгоритмы его типа, но из-за эффективности одномерной свертки ему не нужно впоследствии масштабировать обработанные разномасштабные вариации изображения, а достаточно просто выполнить обработку полноразмерного изображения. Это позволяет учитывать постоянное уточнение сегментации по мере увеличения масштаба, а не сужение, которое может возникать в результате масштабирования.

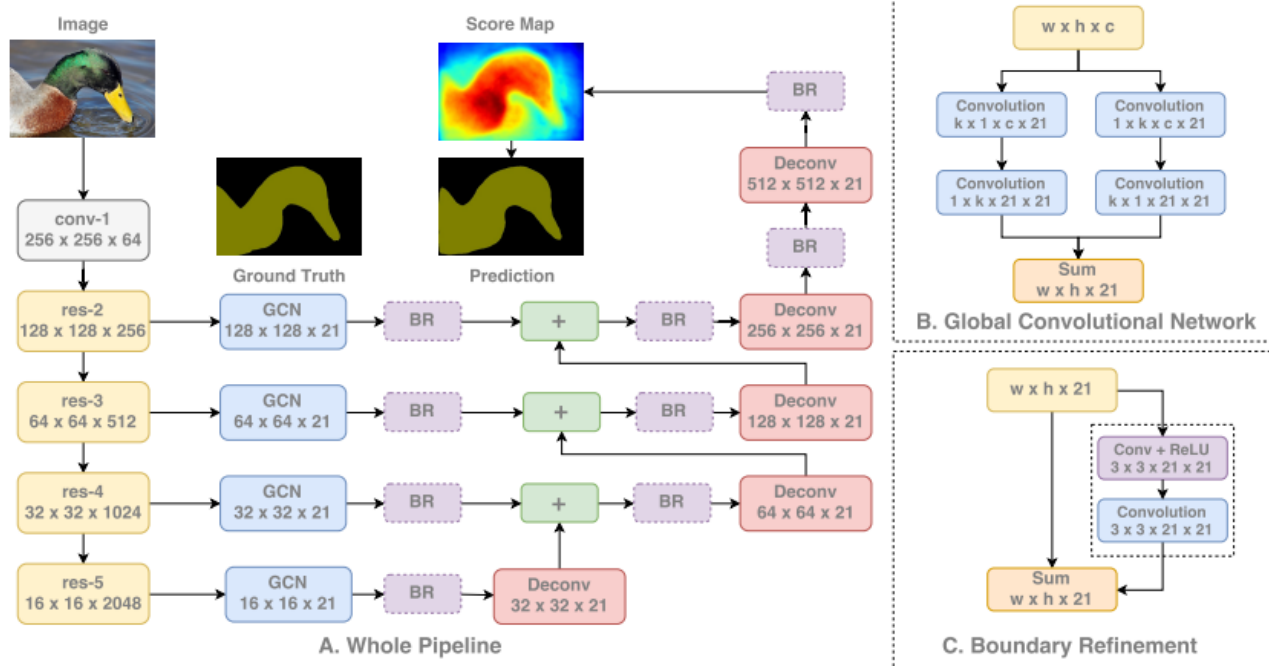


Рисунок 14 – Архитектура GCN

ЗАКЛЮЧЕНИЕ

В данной работе были рассмотрены следующие темы:

1. Основная теоритическая информация о задачах искусственного интеллекта, машинного и глубокого обучения, компьютерного зрения.
2. Основы функционирования нейронных сетей и их обучения.
3. Сверточные нейронные сети и процесс свертки.
4. Краткая теории о задаче сегментации изображения и подходах ее решения
5. Алгоритмы ИНС для сегментации
6. Применение сегментации в жизни

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Пять технологий искусственного интеллекта, о которых вам нужно знать [Электронный ресурс] – URL: https://www.sas.com/ru_ru/insights/articles/analytics/five-ai-technologies.html (дата обращения 27.04.2021) - Загл. с экрана. Яз. рус.
- 2 Нейронные сети для начинающих. Часть 1 [Электронный ресурс] – URL: <https://habr.com/ru/post/312450/> (дата обращения 27.04.2021) - Загл. с экрана. Яз. рус.
- 3 Нейронные сети — математический аппарат [Электронный ресурс] – URL: <https://basegroup.ru/community/articles/math> (дата обращения 27.04.2021) - Загл. с экрана. Яз. рус.
- 4 Нейронные сети, или как обучить искусственный интеллект [Электронный ресурс] – URL: <http://internetinside.ru/neyronnye-seti-ili-kak-obuchit-iskuss/> (дата обращения 28.04.2021) - Загл. с экрана. Яз. рус.
- 5 Как работает нейронная сеть: алгоритмы, обучение, функции активации и потери [Электронный ресурс] – URL: <https://neurohive.io/ru/osnovy-data-science/osnovy-nejronnyh-setej-algoritmy-obuchenie-funkcii-aktivacii-i-poteri/> (дата обращения 29.04.2021) - Загл. с экрана. Яз. рус.
- 6 Сверточная нейронная сеть, часть 1: структура, топология, функции активации и обучающее множество [Электронный ресурс] – URL: <https://habr.com/ru/post/348000/> (дата обращения 29.04.2021) - Загл. с экрана. Яз. рус.
- 7 Сегментация изображения (Image segmentation) [Электронный ресурс] – URL: <https://api-2d3d-cad.com/segment/> (дата обращения 29.04.2021) - Загл. с экрана. Яз. рус.
- 8 Семантическая сегментация: краткое руководство [Электронный ресурс] – URL: <https://neurohive.io/ru/osnovy-data-science/semantic-segmentation/> (дата обращения 29.04.2021) - Загл. с экрана. Яз. рус.
- 9 U-Net: нейросеть для сегментации изображений [Электронный ресурс] – URL: <https://neurohive.io/ru/vidy-nejrosetej/u-net-image-segmentation/> (дата обращения 01.05.2021) - Загл. с экрана. Яз. рус.

- 10 Семантическая сегментация - популярные архитектуры [Электронный ресурс] – URL: <https://www.machinelearningmastery.ru/semantic-segmentation-popular-architectures-dff0a75f39d0/> (дата обращения 02.05.2021) - Загл. с экрана. Яз. рус.
- 11 3 МЕТОДА ДЕТЕКТИРОВАНИЯ ОБЪЕКТОВ С DEEP LEARNING: R-CNN, FAST R-CNN И FASTER R-CNN [Электронный ресурс] – URL: <https://python-school.ru/r-cnn-methods-for-object-detection/> (дата обращения 03.05.2021) - Загл. с экрана. Яз. рус.
- 12 Семантическая сегментация с глубоким обучением [Электронный ресурс] – URL: <https://www.machinelearningmastery.ru/semantic-segmentation-with-deep-learning-a-guide-and-code-e52fc8958823/> (дата обращения 02.05.2021) - Загл. с экрана. Яз. рус.