

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МО ЭВМ

ОТЧЕТ

По научно-исследовательской работе

Тема: Применение сверточных нейронных сетей для распознавания
объекта в видеопотоке

Студентка	_____	Ханукашвили В.Д.
Руководитель	_____	Чернокульский В.В.
	_____	Кринкин К.В.

Санкт-Петербург
2018

ЗАДАНИЕ

Студентка Ханукашвили В.Д.

Группа 3303

Тема работы: Применение сверточных нейронных сетей для распознавания объекта в видеопотоке

Задание на НИР:

Провести обзор предметной области, рассмотреть существующие решения, описать собственный вариант решения

Сроки выполнения НИР: 01.10.2018 – 20.12.2018

Дата сдачи отчета: 19.12.2018

Дата защиты отчета: 20.12.2018

Студентка

Ханукашвили В.Д.

Руководитель

Чернокульский В.В.

Кринкин К.В.

АННОТАЦИЯ

Для многих областей жизни и деятельности человека, задача распознавания объектов в видеопотоке является очень актуальной в настоящее время. В данной работе приведен ряд определений, специфичный для данной области, представлены существующие методы решения задачи распознавания объектов в видеопотоке. Описан математический аппарат, который необходим для решения поставленной задачи.

SUMMARY

The problem of recognition of objects in the video library is very relevant at the moment for many areas of human life and activity. This paper presents a number of definitions specific to this field, presents the existing methods for solving the problem of recognizing objects in a video stream. The mathematical apparatus, which is necessary for solving the problem, is described.

СОДЕРЖАНИЕ

Введение	5
1. Обзор предметной области	7
1.1. Обзор методов распознавания объектов	7
1.1.1. Детерминированные методы	7
1.1.2. Статистические методы	8
1.1.3. Нейросетевые методы	9
2. Выбор метода решения	9
2.1. Описание математического аппарата	9
2.1.1. Структура сверточной нейронной сети	10
2.1.2. Топология сверточной нейронной сети	12
2.1.3. Метод обучения сверточной нейронной сети	12
2.1.4. Функция активации	14
2.2. Обзор архитектур сверточных сетей	14
2.2.1. AlexNet	14
2.2.2. VGG Net	14
2.2.3. ZF Net	14
2.2.4. GoogLeNet	15
2.2.5. VoxNet	15
ЗАКЛЮЧЕНИЕ	17
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	18

Введение

Одним из основных направлений машинного обучения является исследование, разработка и реализация методов для решения задач распознавания объектов в режиме реального времени в видеопотоке. Данная задача является важной частью автоматизированных систем, систем управлений и обработки информации и систем принятия решений. Классификация явлений, сигналов и предметов, которые определяются набором признаков и свойств, является задачей, которая находит свое применение во многих отраслях таких как, диагностическая медицина, робототехника, исследование искусственного интеллекта, автоматизация различных систем, информационный поиск и многих других. Обработка видеопотока и классификация объектов находит применение в системах безопасности и видеонаблюдения, в управлении и контроле доступа, в системах виртуальной реальности и поиска по изображению. Системы по распознаванию автомобильных номеров, идентификации и маркировки изделий, человеческих лиц и отпечатков пальцев, рукописных текстов на данный момент широко используются в различных сферах современной жизни.

Несмотря на внушительный прогресс достигнутый в последнее время в сфере классификации образов современные исследования подчеркивают тот факт, что существующие методы и алгоритмы по распознаванию образов до сих пор еще отстают от модели биологической зрительной системы, потому что они не способны функционировать на неограниченном множестве классов распознавания, не имеют устойчивости к вариативности объектов в пределах категорий и их инвариантным преобразованиям. Исходя из этого актуальной проблемой на данном этапе развития систем по распознаванию образов остается классификация объектов на изображении под действием преобразований, при которых значительно изменяется форма объекта на изображении, но при этом принадлежность объекта к классу остается неизменной. В теории классификации образов данная проблема имеет название проблема инверсии.

Многослойные сверточные сети предоставляют частичное решение проблемы инверсии, обеспечивая стойкость только к ограниченному подмножеству преобразований. В тех сферах жизнедеятельности, где для классификации используются объекты, взятые из естественной среды, например, зрительная система в робототехнике, видеонаблюдении, анализе данных с фотографий и тому подобное, данная проблема является наиболее актуальной.

Целью настоящей выпускной квалификационной работы является разработка программы по распознаванию объекта в видеопотоке, используя метод глубокого обучения сверточной нейронной сети. В качестве средства разработки будет использована библиотека TensorFlow.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести обзор предметной области. Сравнить метод, используемый в данной ВКР, с его аналогами, обосновать свой выбор;
2. Построить и описать математическую модель сверточной нейронной сети со всеми выбранными параметрами. А также подробно описать способ решения поставленной задачи;

3. Выбрать стек технологий, который потребуется для реализации программы;
4. На основе построенной модели реализовать программу по распознаванию объектов;
5. Подготовить обучающую выборку;
6. Провести обучения сети;
7. Осуществить тестирование реализованного приложения

Объектом исследования в данной работе является система компьютерного зрения, с помощью которой осуществляется распознавание объектов в видеопотоке

Предметом исследования являются алгоритм и математическая модель для решения задачи распознавания.

1. Обзор предметной области

Видеоаналитика – технология, использующая методы компьютерного зрения для автоматизированного получения различных данных на основании анализа последовательности изображений, поступающих с видеокамер в режиме реального времени или из архивных записей.

Под задачей обнаружения динамических объектов понимается задача обнаружения и выделения изменяющихся участков изображения в последовательности кадров. Соответственно, под обнаружением определённого объекта понимается выбор одного или нескольких обнаруженных динамических объектов, которые имеют некоторые схожие признаки с заданным объектом поиска - эталоном. Признаки выбираются согласно алгоритму.

В рамках выпускной работы необходимо разработать программу, которая в режиме реального времени распознаёт объекты в видеопотоке. В контексте данной работы, в качестве объектов рассматриваются бутылки с напитками (Sprite, Pepsi, Coca-Cola, Bonaqua и т.д.). Эти объекты перемещаются по движущейся ленте, а камера их снимает и передает видеоданные рассматриваемой в данной работе программе.

Ниже приведен обзор разновидностей существующих методов распознавания объектов.

1.1. Обзор методов распознавания объектов

Методы обнаружения и слежения можно разбить на следующие группы:

- Детерминированные методы;
- Статистические методы;
- Нейросетевые методы.

Ниже приведено описание и примеры для каждого из методов.

1.1.1. Детерминированные методы

Детерминированные методы выдают уникальный и предопределённый результат для заданных входных данных. Детерминированные методы рассматривают объект слежения, как объект с неизменяющимися признаками. Эти методы можно разделить на группы:

- Методы поиска оптического потока;
- Методы поиска особенных точек;
- Метод перебора (методы поиска по шаблону).

Методы поиска оптического потока основаны на вычислении разреженного оптического потока. Эти методы строят векторное поле скоростей выделенных точек (пикселей изображения).

Методы поиска особенных точек основаны на вычислении характерных особенностей на изображении и на нахождении соответствия между ними в видеопоследовательности.

Методы перебора не имеют этапа обучения (методы без учителя). Эти методы вычисляют набор признаков на одном заданном изображении с объектом для поиска.

Методы поиска по шаблону имеют сложный этап обнаружения объекта.

Метод гибкого сравнения на графах (Elastic graph matching)

Данный метод относится к методам перебора (детерминированные методы). Суть метода сводится к эластичному сопоставлению графов, описывающих изображения лиц. Лица представлены в виде графов со взвешенными вершинами и ребрами. На этапе распознавания один из графов (эталонный) остается неизменным, в то время как другой деформируется с целью наилучшей подгонки к первому. Результат распознавания системы – эталон с наилучшим значением ценовой функции деформации.

Преимущества метода:

- Высокая эффективность распознавания

Недостатки метода:

- Высокая вычислительная сложность процедуры распознавания;
- Низкая технологичность при запоминании новых эталонов;
- Линейная зависимость времени работы от размера базы данных лиц.

1.1.2. Статистические методы

В статистических методах предполагается установление связи между отнесением объекта к тому или иному классу (образу) и вероятностью ошибки при решении этой задачи. В ряде случаев это сводится к определению апостериорной вероятности принадлежности объекта образу при условии, что признаки этого объекта приняли значения. Основное преимущество статистических методов распознавания состоит в возможности одновременного учета признаков различной физической природы, так как они характеризуются безразмерными величинами — вероятностями их появления при различных состояниях системы.

Скрытые Марковские модели (СММ)

Одним из статистических методов распознавания лиц являются скрытые Марковские модели (СММ) с дискретным временем. СММ используют статистические свойства сигналов и учитывают непосредственно их пространственные характеристики. Элементами модели являются: множество скрытых состояний, множество наблюдаемых состояний, матрица переходных вероятностей, начальная вероятность состояний. Каждому соответствует своя Марковская модель. При распознавании объекта проверяются сгенерированные для заданной базы объектов Марковские модели и ищется максимальная из наблюдаемых вероятностей того, что последовательность наблюдений для данного объекта сгенерирована соответствующей моделью.

Преимущества метода:

- Возможность одновременного учета признаков различной физической природы

Недостатки метода:

- Необходимо подбирать параметры модели для каждой базы данных;
- СММ не обладает различающей способностью, то есть алгоритм обучения только максимизирует отклик каждого изображения на свою модель, но не минимизирует отклик на другие модели.

1.1.3. Нейросетевые методы

В нейросетевых методах объект в видеопотоке может рассматриваться одновременно с изменяющимися и неизменяющимися признаками. В этих методах неотъемлемой частью является этап обучения нейронной сети. Обучать сеть необходимо под каждый тип задач.

Существуют различные виды нейронных сетей:

Классическая нейронная сеть - Основная идея, лежащая в основе нейронных сетей – это последовательное преобразование сигнала, параллельно работающими элементарными функциональными элементами, нейронами. Основной принцип настройки нейронной сети заключается в применении оптимизационных методов к минимизации среднеквадратичной ошибки, как следствие – склонность к переобучению.

Сверточная нейронная сеть (далее – СНС)

Наилучшие результаты в области распознавания лиц показала сверточная нейронная сеть (СНС - Convolutional Neural Network), которая является логическим развитием идей таких архитектур НС как когнитрона и неокогнитрона. Успех обусловлен возможностью учета двумерной топологии изображения, в отличие от многослойного перцептрона. Отличительными особенностями СНС являются локальные рецепторные поля (обеспечивают локальную двумерную связность нейронов), общие веса и иерархическая организация с пространственным сэмплингом (spatial subsampling). Благодаря этим нововведениям СНС обеспечивает частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям.

Преимущества метода:

- Очень высокая точность распознавания;
- Метод является более устойчивым к различного рода изменениям в изображении;
- Есть возможность распознавать объекты в видеопотоке.

Недостатки метода:

- Добавление нового эталонного объекта в базу данных требует полного переобучения сети на всем имеющемся наборе;
- Трудно формализуемый этап выбора архитектуры сети.

2. Выбор метода решения

2.1. Описание математического аппарата

Сверточная нейронная сеть была впервые предложена Я.Лекуном и Й.Бенгуа. Ученые исследовали зрительную кору головного мозга кошки, содержащую карты местных рецептивных полей, уменьшающиеся по мере детализации объекта, на который смотрит животное. Я.Лекун и Й.Бенгуа пришли к выводу, что математическую модель поведения зрительных рецепторов кошки можно построить, опираясь на следующие этапы обработки изображений:

- Свертка исходного изображения объекта при помощи нескольких небольших фильтров;
- Субдискретизация (объединение) полученных на предыдущем шаге ключевых признаков;

- Повтор предыдущих шагов (свертки, а затем субдискретизации) до тех пор, пока на выходе не получится достаточное количество признаков исходного изображения;
- Использование модели полносвязного слоя для получения решения конкретной задачи.

Таким образом, сверточная нейронная сеть благодаря своей архитектуре в основном используется для решения круга сложных графических задач распознавания, а распознавание объекта в видеопотоке, как раз можно отнести к данному типу задач.

Далее приведено описание того, как с помощью выбранного метода будет решаться поставленная задача.

2.1.1. Структура сверточной нейронной сети

Сверточные нейронные сети (СНС) состоят из трех основных видов слоев: сверточный слой, субдискретизирующий (пулинга или подвыборки) слой и выходной слой (полносвязный). Слои СНС расположены друг за другом: сначала сверточный слой, а затем субдискретизирующий, за последним сверточным слоем следует выходной слой. Сверточный и субдискретизирующий слои считаются слоями двумерной размерности, а выходной слой, как правило, представляет собой вектор. В СНС каждый двумерный слой имеет несколько уровней. Каждый уровень представляет собой двумерный массив. Выход каждого уровня в дальнейшем будем называть картой признаков.

В обычном перцептроне, который представляет собой полносвязную нейронную сеть, каждый нейрон связан со всеми нейронами предыдущего слоя, причем каждая связь имеет свой персональный весовой коэффициент. В СНС в операции свёртки используется лишь ограниченная матрица весов небольшого размера, которую «двигают» по всему обрабатываемому слою, формируя после каждого сдвига сигнал активации для нейрона следующего слоя с аналогичной позицией. То есть для различных нейронов выходного слоя используются общие веса — матрица весов, которую также называют набором весов или ядром свёртки. Она построена таким образом, что графически кодирует какой-либо один признак, например, наличие кривой линии, расположенной под определенным наклоном. Тогда следующий слой, получившийся в результате операции свёртки такой матрицей весов, показывает наличие данной наклонной линии в обрабатываемом слое и её координаты, формируя так называемую карту. В свёрточной нейронной сети набор весов не один, а целая гамма, кодирующая всевозможные линии и дуги под разными углами. При этом такие ядра свертки не закладываются исследователем заранее, а формируются самостоятельно путем обучения сети классическим методом распространения ошибки.

Если коротко, что делает СНС — на вход подается изображение, пропускается через серию свёрточных слоев, слоев субдискретизации и полносвязных слоёв, и генерируется вывод. Простым примером может послужить архитектура:

[INPUT— CONV— POOL— FC]

Сложный и основной момент — понимание того, что происходит на каждом из этих слоев.

- В INPUT (входные данные) $[32 \times 32 \times 3]$ содержатся исходная информация об изображении (в данном случае 32 — ширина, 32 — высота, 3 — цветовые каналы R, G, B).
- Слой CONV (слой свёртки) - представляет из себя набор карт (карты признаков), у каждой карты есть синаптическое ядро (фильтр). Количество карт определяется требованиями к задаче, если взять большое количество карт, то повысится качество распознавания, но увеличится вычислительная сложность. Ядро представляет из себя фильтр или окно, которое скользит по всей области предыдущей карты и находит определенные признаки объектов. Ядро представляет собой систему разделяемых весов или синапсов, это одна из главных особенностей сверточной нейросети. В обычной многослойной сети очень много связей между нейронами, то есть синапсов, что весьма замедляет процесс детектирования. В сверточной сети – наоборот, общие веса позволяет сократить число связей и позволить находить один и тот же признак по всей области изображения. Неформально эту операцию можно описать следующим образом — окном размера ядра проходят с заданным шагом (обычно 1) все изображение на каждом шаге поэлементно умножаем содержимое окна на ядро, результат суммируется и записывается в матрицу результата, как на рисунке 1. Также этот слой можно описать формулой 1:

$$x^l = f(x^{l-1} * k^l + b^l) \quad (1)$$

Где x^l – выход слоя l

$f()$ – функция активации;

b^l – коэффициент сдвига слоя l;

* - операция свертки входа x с ядром k.

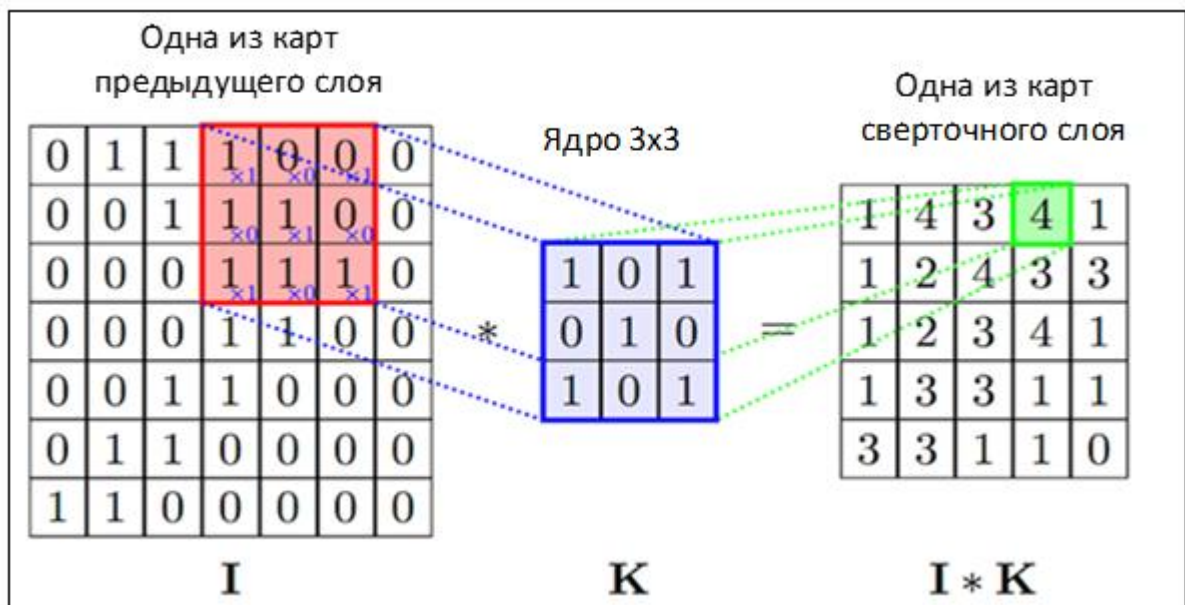


Рисунок 1 — Визуализация работы слоя свертки

- Слой POOL (слой пулинга) выполняет операцию по понижающей дискретизации пространственных размеров (ширина и высота), то есть на этом этапе выполняется нелинейное уплотнение карты признаков. Логика работы такова: если на предыдущей операции свертки уже были выявлены некоторые признаки, то для дальнейшей обработки настолько подробное изображение уже не нужно, и оно уплотняется до менее подробной картинки. Формально слой может быть описан формулой 2:

$$x^l = f(a^l * \text{subsample}(x^{l-1}) + b^l) \quad (2)$$

Где x^l – выход слоя l

$f()$ – функция активации;

a^l, b^l – коэффициенты сдвига слоя l ;

$\text{subsample}()$ - операция локальных максимальных значений входа x с ядром k .

- Слой FC (полносвязный слой) выводит N-мерный вектор (N — число классов) для определения нужного класса. Работа организуется путем обращения к выходу предыдущего слоя (карте признаков) и определения свойств, которые наиболее характерны для определенного класса.

Именно таким образом СНС слой за слоем преобразует исходное изображение, начиная с исходных значений пикселей и заканчивая определением класса, к которому принадлежит изображение в видеопотоке.

2.1.2. Топология сверточной нейронной сети

Следующим шагом, после описания структуры сверточной нейронной сети, идет выбор топологии сети. Данный выбор делается исходя из следующих пунктов, влияющих на топологию сети:

- Определение задачи, которую решает нейронная сеть (классификация, прогнозирование, модификация);
- Определение ограничений в решаемой задаче (скорость, точность ответа);
- Определение входных (тип: изображение, звук, размер: 100x100, 30x30, формат: RGB, в градациях серого) и выходных данных (количество классов).

2.1.3. Метод обучения сверточной нейронной сети

Существует две парадигмы обучения нейронных сетей – с учителем и без учителя. В первом случае, на входной вектор имеется готовый ответ, во втором случае нейронная сеть самообучается. У каждого вида обучения есть своя ниша задач и по большому счету они не пересекаются. На данный момент придумано и запатентовано большое количество архитектур нейронных сетей и методов их обучения. В данной работе будет реализовано обучение СНС с учителем. Методов обучения СНС с учителем также существует несколько:

- Метод коррекции ошибки

- Метод обратного распространения ошибки
- Метод коррекции ошибок без квантования
- Метод коррекции ошибок с квантованием

Ниже приведено более детальное описание каждого из методов.

Метод коррекции ошибок

Метод коррекции ошибки — метод обучения перцептрона, предложенный Фрэнком Розенблаттом. Представляет собой такой метод обучения, при котором вес связи не изменяется до тех пор, пока текущая реакция перцептрона остается правильной. При появлении неправильной реакции вес изменяется на единицу, а знак (+/-) определяется противоположным от знака ошибки.

Метод обратного распространения ошибки

Метод обратного распространения ошибки (англ. backpropagation)— метод обучения многослойного перцептрона. Впервые метод был описан в 1974 г. А.И. Галушкиным. Это итеративный градиентный алгоритм, который используется с целью минимизации ошибки работы многослойного перцептрона и получения желаемого выхода. Основная идея этого метода состоит в распространении сигналов ошибки от выходов сети к её входам, в направлении, обратном прямому распространению сигналов в обычном режиме работы. Барцев и Охонин предложили сразу общий метод («принцип двойственности»), приложимый к более широкому классу систем, включая системы с запаздыванием, распределённые системы, и т.п.

Метод коррекции ошибки без квантования

Представляет собой такой метод обучения, при котором вес связи не изменяется до тех пор, пока текущая реакция перцептрона остается правильной. При появлении неправильной реакции вес изменяется на единицу, а знак (+/-) определяется противоположным от знака ошибки.

Если реакция на стимул S_i правильная, то никакого подкрепления не вводится, но при появлении ошибок к весу каждого активного А-элемента прибавляется некоторая величина $\delta = \rho_i \cdot \Delta x_i$, где Δx_i — число единиц подкрепления, выбирается так, чтобы величина сигнала превышала некоторого порога ρ_i , а , при этом S_i^+ — стимул, принадлежащий положительному классу, а S_i^- — стимул, принадлежащий отрицательному классу.

Метод коррекции ошибки без квантования

Отличается от метода коррекции ошибок без квантования только тем, что Δx_i , то есть равно одной единице подкрепления. Этот метод и метод коррекции ошибок без квантования являются одинаковыми по скорости достижения решения в общем случае, и более эффективными по сравнению с методами коррекции ошибок со случайным наком или случайными возмущениями.

2.1.4. Функция активации

Одним из этапов разработки нейронной сети является выбор функции активации нейронов. Вид функции активации во многом определяет функциональные возможности нейронной сети и метод обучения этой сети. Варианты функций перечислены ниже:

- Единичный скачок или жесткая пороговая функция
- Гиперболический тангенс
- Сигмоидальная функция или сигмоид
- Линейный порог или гистерезис

2.2. Обзор архитектур сверточных сетей

В данном подразделе будут описаны варианты архитектур свёрточных нейронных сетей.

2.2.1. AlexNet

В 2012 году на конкурсе ILSVRC по классификации изображений впервые победила нейронная сеть — AlexNet. В AlexNet были собраны новейшие на тот момент техники для улучшения работы сети. Обучение AlexNet из-за количества параметров сети происходило на двух GPU, что позволило сократить время обучения в сравнении с обучением на CPU. Также оказалось, что использование функции активации ReLU (Rectified Linear Unit) вместо более традиционных функций сигмоиды и гиперболического тангенса позволило снизить количество эпох обучения в шесть раз. Формула ReLU следующая:

$$y(x) = \max(0, x) \quad (3)$$

ReLU позволяет побороть проблему затухания градиентов, свойственную другим функциям активации. Помимо прочего, в AlexNet была применена техника отсева (Dropout) [Hinton et al., 2012]. Она заключается в случайном отключении каждого нейрона на заданном слое с вероятностью p на каждой эпохе.

2.2.2. VGG Net

VGG Net — модель свёрточной нейронной сети, предложенная в 2014. В данной сети отказались от использования фильтров размером больше, чем 3×3 . Авторы показали, что слой с фильтром 7×7 эквивалентен трём слоям с фильтрами 3×3 , причём в последнем случае используется на 55 % меньше параметров. Аналогично слой с фильтром 5×5 эквивалентен двум слоям с фильтром 3×3 , которые экономят 22 % параметров сети.

2.2.3. ZF Net

ZF Net — победитель ILSVRC 2013. Основным достижением данной архитектуры является создание техники визуализации фильтров — сети развёртки (deconvolutional network), состоящей из операций, в каком-то смысле обратных операциям сети. В итоге сеть развёртки отображает скрытый слой сети на оригинальное изображение.

Чтобы изучить поведение фильтра на определённом изображении с помощью обученной нейронной сети, необходимо сначала осуществить вывод сетью, после чего в слое изучаемого фильтра обнулить все веса, кроме весов самого фильтра, и затем подать полученную активацию на

слоя сети развёртки. В сети развёртки последовательно применяются операции Unpooling, ReLU и фильтрации. Unpooling частично восстанавливает вход соответствующего слоя субдискретизации, запоминая координаты, которые выбрал слой субдискретизации. ReLU — обычный слой, применяющий функцию ReLU. Слой фильтрации выполняет операцию свёртки с весами соответствующего слоя свёртки, но веса каждого фильтра «перевернуты» вертикально и горизонтально.

2.2.4. GoogLeNet

GoogLeNet — победитель ILSVRC 2014, также известный как Inception-v1. Создатели этой сети исходили из факта, что после каждого слоя сети необходимо сделать выбор — будет ли следующий слой свёрткой с фильтром 3x3, 5x5, 1x1 или же слоем субдискретизации. Каждый из таких слоёв полезен — фильтр 1x1 выявляет корреляцию между каналами, в то время как фильтры большего размера реагируют на более глобальные признаки, а слой субдискретизации позволяет уменьшить размерность без больших потерь информации. Вместо того чтобы выбирать, какой именно слой должен быть следующим, предлагается использовать все слои сразу, параллельно друг другу, а затем объединить полученные результаты в один. Чтобы избежать роста числа параметров, перед каждым слоем свёртки используется свёртка 1x1, которая уменьшает число карт признаков. Такой блок слоёв называли модулем Inception. Также в GoogLeNet отказались от использования полносвязного слоя в конце сети, используя вместо него слой Average Pooling, благодаря чему резко уменьшилось число параметров в сети. Таким образом, GoogLeNet, состоящая из более чем ста базовых слоёв, имеет почти в 12 раз меньше параметров, чем AlexNet.

2.2.5. VoxNet

VoxNet — это архитектура трехмерной сверточной нейронной сети (3DCNN), которая используется для распознавания трехмерных объектов в реальном времени (видеоанализ). В качестве третьего параметра выступает время.

Исходные данные для алгоритма, представляют собой сегмент облака точек, который может быть получен различными методами сегментации или при помощи алгоритма «скользящего трехмерного окна». Сегмент, как правило, определяется пересечением облака точек с отграничивающим параллелепипедом и может включать в себя фоновые шумы. Задача заключается в определении принадлежности объекта для данного сегмента к определенному классу. Система решения данной задачи состоит из двух компонентов: «объемной сетки», которая представляет оценку пространственного наполнения, и 3DCNN, которая классифицирует объекты, непосредственно используя объемную сетку. На рисунке 2 представлена архитектура VoxNet.

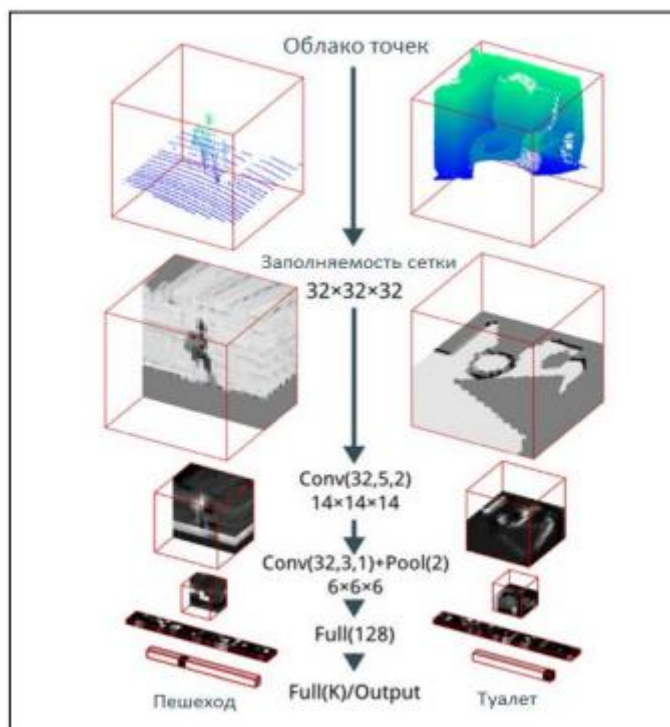


Рисунок 2 – Архитектура VoxNet

ЗАКЛЮЧЕНИЕ

В данной работе были перечислены области, в которых используется распознавание объектов в режиме реального времени. Описаны современные методы распознавания объектов в видеопотоке, а также для каждого из методов приведены примеры. У каждого описанного примера выделены его преимущества и недостатки, исходя из которых можно сделать вывод, что для реализации поставленной задачи, оптимальным решением является использование сверточной нейронной сети.

Также в настоящей работе был описан математический аппарат, который позволит решить задачу распознавания объекта в видеопотоке. Также были перечислены и описаны основные параметры, специфичные для данного метода, которые в дальнейшем предстоит выбрать. И рассмотрены основные архитектуры сверточных нейронных сетей, среди которых можно выделить архитектуру VoxNet. Так в ВКР будет использована именно данная архитектура, так как она позволяет идентифицировать объекты в видеопотоке в режиме реального времени.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Анштедт, Т. Видеоаналитика: мифы и реальность. Издательство: Security Focus, 2012 г. 176 с.
2. Roth, P.M. Survey of Appearance-Based Methods for Object Recognition. Издательство: Graz University of Technology, Austria, January, 2008г. – 68 с.
3. Ле Мань Ха. Свёрточная нейронная сеть для решения задачи классификации. Издательства: ТРУДЫ МФТИ, 2016г.
4. Е.Ю. Минаев. Трекинг объектов в видеопотоке на основе сверточных нейронных сетей и фрактального анализа. Издательство: ИТНТ, 2018 г.
5. Goshin Ye.V. Segmentation of stereo images with the use of the 3D Hough transform. Издательство: CEUR Workshop Proceedings, 2016 г. - 1638 с.
6. Tombari F. Object recognition in 3D scenes with occlusions and clutter by Hough voting. Издательство: Fourth Pacific-Rim Symposium on Image and Video Technology, 2010 г. – 7 с.
7. Свёрточные сети (convolutional neural networks, CNNs) .
URL:http://www.machinelearning.ru/wiki/images/1/13/Dl17_Osokin_Convnets.pdf
(дата обращения: 04.12.2018).