

ОПИСАНИЕ ПРЕДПОЛАГАЕМОГО СПОСОБА РЕШЕНИЯ

Целью настоящей выпускной квалификационной работы является разработка программы по распознаванию объекта в видеопотоке, используя метод глубокого обучения сверточной нейронной сети.

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

Сегодня искусственная нейронная сеть широко используется для решения различных задач в реальной жизни. Такие проблемы, как распознавание объектов в режиме реального времени, теперь решаются с использованием нейронной сети с высокой точностью. В настоящей выпускной работе, для решения поставленной задачи, будет использоваться одна специальная архитектура нейронной сети – свёрточная нейронная сеть.

1. Введение

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. 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 свертки 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 свертки. 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 свертки такой матрицей весов, показывает наличие данной наклонной линии в обрабатываемом слое и её координаты, формируя так называемую карту. В сверточной 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 сверточных слоев, слоев субдискретизации и полносвязных слоев, 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.

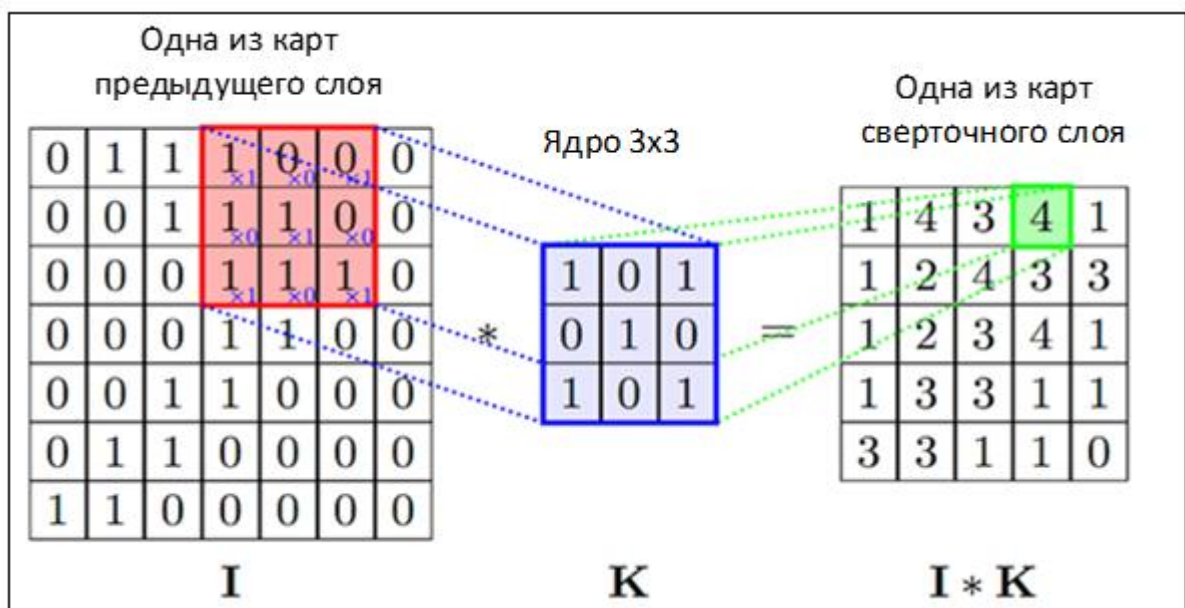


Рисунок 1 — Визуализация работы слоя свертки

- 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:

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

Где x^l – выход слоя l

$f()$ – функция активации;

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

$subsample()$ - 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.

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).

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

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

5. Используемые технологии

Реализация вышеописанной задачи будет осуществляться на языке программирования Python. Предварительно предполагается, что для работы со сверточной нейронной сетью будет использоваться библиотека Theano.

6. Вывод

Таким образом, в данной работе был описан выбранный метод решения задачи распознавания объекта в видеопотоке. Также перечислены и описаны основные параметры, которые в дальнейшем предстоит выбрать, специфичные для данного метода.