

Оглавление

Введение	3
1 Анализ предметной области	4
1.1 Задача распознавания образов	4
1.2 Возможные сложности	5
2 Предобработка исходных изображений	6
3 Традиционные методы компьютерного зрения	7
3.1 Метод опорных векторов	7
3.2 Метод k-ближайших соседей	8
3.3 Вывод	9
4 Нейросетевой подход	10
4.1 Полносвязные нейросети	11
4.2 Свёрточная нейронная сеть	11
4.3 Архитектуры нейросетей	13
4.3.1 YOLO	14
4.3.2 Капсульная нейросеть	14
Заключение	16
Список используемых источников	17

Введение

Распознавание изображений представляет собой один из важнейших компонентов систем управления и обработки информации, автоматизированных систем и систем принятия решений. Задачи, связанные с классификацией и идентификацией предметов, явлений и сигналов, характеризующихся конечным набором некоторых свойств и признаков, возникают в таких сферах как робототехника, информационный поиск, мониторинг и анализ визуальных данных, исследования искусственного интеллекта. В настоящий момент в производстве широко используются системы распознавания рукописного текста, автомобильных номеров, отпечатков пальцев или человеческих лиц, находящие применение в интерфейсах программных продуктов, системах безопасности и идентификации личности, а также в других прикладных целях.

1 Анализ предметной области

1.1 Задача распознавания образов

Распознавание изображений - научное направление, связанное с разработкой принципов и построением систем, предназначенных для определения принадлежности данного объекта к одному из заранее выделенных классов объектов.

Распознавание изображений — это метод компьютерного зрения для идентификации объектов на изображениях или видео. Распознавание изображений является основным результатом алгоритмов глубокого и машинного обучения. При просмотре фотографий или видео, человек может легко распознать людей, предметы, сцены и визуальные детали. Цель состоит в обучении компьютера делать то, что естественно для людей: достичь уровня понимания того, что содержит изображение.

Алгоритм распознавания изображений (также известный как классификатор изображений) принимает изображение (или фрагмент изображения) в качестве входных данных и выводит то, что содержит изображение. Другими словами, вывод — это метка класса (например, «кошка», «собака», «таблица» и т.д.).

Существует несколько специализированных задач, основанных на распознавании, например:

- Поиск изображений по содержанию: нахождение всех изображений в большом наборе изображений, которые имеют определённое содержание. Содержание может быть определено различными путями, например в терминах схожести с конкретным изображением (найдите мне все изображения похожие на данное изображение), или в терминах высокоуровневых критериев поиска, вводимых как текстовые данные (найдите мне все изображения, на которых изображено много домов, которые сделаны зимой и на которых нет машин).
- Оценка положения: определение положения или ориентации определённого объекта относительно камеры. Примером применения этой техники может быть содействие руке робота в извлечении объектов с ленты конвейера на линии сборки.

- Оптическое распознавание знаков: распознавание символов на изображениях печатного или рукописного текста, обычно для перевода в текстовый формат, наиболее удобный для редактирования или индексации (например, ASCII).

1.2 Возможные сложности

Классическая задача в распознавании изображений - определение содержат ли видеоданные некоторый характерный объект, особенность или активность. Эта задача может быть достоверно и легко решена человеком, но до сих пор не решена удовлетворительно в компьютерном зрении в общем случае: случайные объекты в случайных ситуациях.

Кроме того есть некоторые проблемы с распознаванием изображений:

- Изменение точки зрения.
- Изменение освещения.
- Деформация.
- Изображения частично скрыто.
- Изображения совпадает с фоном.

2 Предобработка исходных изображений

- Часто входное изображение предварительно обрабатывается для нормализации эффектов контрастности и яркости. Очень распространенный этап предварительной обработки — вычесть среднее значение интенсивности изображения и разделить его на стандартное отклонение. Иногда гамма-коррекция дает немного лучшие результаты. При работе с цветными изображениями преобразование цветового пространства (например, цветовое пространство RGB в LAB) может помочь получить лучшие результаты.

В рамках предварительной обработки входное изображение или фрагмент изображения также обрезаются и изменяются до фиксированного размера. Это важно, потому что следующий шаг, извлечение признаков, выполняется на изображении фиксированного размера.

- Входное изображение содержит слишком много дополнительной информации, которая не нужна для классификации. Следовательно, первым шагом в классификации изображений является упрощение изображения путем извлечения важной информации, содержащейся в изображении, и исключения остальной части. Этот шаг называется извлечением признаков. В традиционных подходах к компьютерному зрению разработка этих функций имеет решающее значение для производительности алгоритма.

Полученное изображение после предварительной обработки будет использоваться в алгоритмах/методах классификации изображений.

3 Традиционные методы компьютерного зрения

Компьютерное зрение (иначе техническое зрение) — теория и технология создания машин, которые могут производить обнаружение, отслеживание и классификацию объектов.

Цель компьютерного зрения заключается в формировании полезных выводов относительно объектов и сцен реального мира на основе анализа изображений, полученных с помощью датчиков [?].

Принцип работы традиционных методов компьютерного зрения заключается в извлечении векторов объектов из изображений для их дальнейшей классификации.(2)

3.1 Метод опорных векторов

Метод опорных векторов (англ. support vector machine, SVM) — один из наиболее популярных методов обучения, который применяется для решения задач классификации и регрессии. Основная идея метода заключается в построении гиперплоскости, разделяющей объекты выборки оптимальным способом.

Алгоритм работает в предположении, что чем больше расстояние (зазор) между разделяющей гиперплоскостью и объектами разделяемых классов, тем меньше будет средняя ошибка классификатора.

Применение метода опорных векторов в распознавании изображений:

- Рассматриваем изображение как массив пикселей, если размер изображения 200 X 200, то размер массива будет 200 X 200 X 3, где первые 200 — ширина, а вторые 200 — высота, а затем 3 по значению канала RGB. Значения в массиве будут находиться в диапазоне от 0 до 255, что описывает интенсивность пикселя в каждой точке.
- Набор образов изображений загружается для обучения. Сначала преобразует все изображения в опеределённый формат, а затем сглаживает все изображения, в результате чего получается массив n-меры чисел для каждого изображения. Эти массивы используются как 1 точка в

n-мерном пространстве.

- Построить классификатор изображений методом опорных векторов на основе набора изображений для обучения.
- Преимущества метода
 - Хорошо работает с пространством признаков большого размера;
 - Хорошо работает с данными небольшого объема;
 - Метод находит разделяющую полосу максимальной ширины, что позволяет в дальнейшем осуществлять более уверенную классификацию.
- Недостатки метода
 - Долгое время обучения (для больших наборов данных);
 - Неустойчивость к шуму: выбросы в исходных данных становятся опорными объектами-нарушителями и напрямую влияют на построение разделяющей гиперплоскости.

3.2 Метод k-ближайших соседей

Метод k-ближайших соседей (k-nearest neighbors) – это простой алгоритм машинного обучения с учителем, который можно использовать для решения задач классификации и регрессии.

Алгоритм K-NN сохраняет все доступные данные и классифицирует новую точку данных на основе сходства. Это означает, что когда появляются новые данные, их можно легко классифицировать по категории наборов с помощью алгоритма K-NN.

Согласно принципу алгоритма KNN, структура классификатора включает в себя 4 параметра: данные для классификации, набор выборочных данных, набор выборочных меток и значение K. Затем вычислить расстояние между новыми данными и выборочными данными, упорядочить расстояния от наименьшего к наибольшему, возьмите первые K ближайших данных. Наиболее часто встречающаяся метка может быть идентифицирована как новая метка данных путем определения количества вхождений каждого введенного типа данных в K первых точках.

- Преимущества метода
 - Алгоритм прост и легко реализуем.
 - Не чувствителен к выбросам.
 - Нет необходимости строить модель, настраивать несколько параметров или делать дополнительные допущения.
 - Алгоритм универсален. Его можно использовать для обоих типов задач: классификации и регрессии.
- Недостатки метода
 - Алгоритм работает значительно медленнее при увеличении объема выборки, предикторов или независимых переменных.
 - Из аргумента выше следуют большие вычислительные затраты во время выполнения.
 - Всегда нужно определять оптимальное значение k .

3.3 Вывод

Таким образом, традиционный подход, использующий методы машинного обучения, имеет недостатки: потребность в данных для процесса обучения, низкая скорость обучения для больших наборов данных и неустойчивость к шуму, однако рассмотренные методы просты и легко реализуемы. В настоящее время интерес к машинному обучению возрастает, поскольку машинное обучение выполняет вычислительную обработку намного эффективнее. Также оно позволяет быстро создавать модели, позволяющие анализировать данные большего размера и сложности и дающие результаты быстрее и точнее. Благодаря своей эффективности и выдающимся преимуществам машинное обучение станет целенаправленным и увлекательным.

4 Нейросетевой подход

Нейронная сеть (далее нейросеть) – это математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей – сетей нервных клеток живого организма(4). Сегодня такие сети активно используют в практических целях за счет возможности не только разработки, но и обучения.

Нейросеть для распознавания изображений – это наиболее популярный способ применения нейронной сети.(3)

Три основных компонента нейросети включают в себя(5):

- Входной слой представляет входные данные.
- Выходной слой представляет собой выходные данные нейронной сети.
- Скрытый слой представляет собой промежуточные узлы, которые делят входное пространство на области с границами. Он принимает набор взвешенных входных данных и производит выходные данные с помощью функции активации.

Область применения нейросетей в настоящее время постоянно расширяется, существует множество удачных решений с использованием данного подхода. Столь успешное внедрение нейросетевых решений, прежде всего, обусловлено их преимуществами перед обычными методами(6, 7):

- Существование быстрых алгоритмов обучения, нейронная сеть даже при сотнях входных сигналов и десятках-сотнях тысяч эталонных ситуаций может быть быстро обучена на обычном компьютере;
- Возможность работы при наличии большого числа неинформативных, шумовых входных сигналов;
- Нейронная сеть одновременно может решать несколько задач на едином наборе входных сигналов;

Несмотря на большие возможности, существует ряд недостатков, которые все же ограничивают применение нейросетевых технологий(8):

- Нейронные сети позволяют найти только субъоптимальное решение, и соответственно неприменимы для задач, в которых требуется высокая точность.
- Функционируя по принципу черного ящика, они также неприменимы в случае, когда необходимо объяснить причину принятия решения.

4.1 Полносвязные нейросети

Первая классическая архитектура - полносвязные нейросети. В ней каждый узел (кроме входного и выходного) выступает как входом, так и выходом, образуя скрытый слой нейронов, и каждый нейрон следующего слоя соединён со всеми нейронами предыдущего. Входы подаются с весами, которые в процессе обучения настраиваются и не меняются в последствии. Выходными сигналами сети могут быть все или некоторые выходные сигналы нейронов после нескольких тактов функционирования сети. Все входные сигналы подаются всем нейрона (9). Это хорошая сеть, так как многие задачи классификации могут быть решены. Однако она не применяется на практике из-за 2 проблем

- Много параметров. Например, для сети, у которой на входе картина 100x100, три скрытых слоя по 100 нейронов каждый, и выходом на 10 классов, число параметров будет больше 1000000.

4.2 Свёрточная нейронная сеть

Свёрточная нейронная сеть (СНС) — это тип модели глубокого обучения для обработки данных, имеющих структуру сетки, например изображений, которая вдохновлена организацией зрительной коры животных [13, 14])(10) и предназначена для автоматического и адаптивного изучения пространственных иерархий признаков, начиная с низких уровней.

СНС - это математическая конструкция, которая обычно состоит из трех типов слоев (или строительных блоков): свертки, пулинга (подвыборки) и полносвязных слоев. Первые два слоя, свертки и пулинга, выполняют извлечение признаков, тогда как третий, полносвязный слой, отображает извлеченные признаки в конечный результат.(10)

Принцип работы СНС (11):

- В качестве первого слоя всегда выступает сверточный слой. Вводимое изображение представляет матрицу некоторого размера, например $32 \times 32 \times 3$ с пиксельными значениями. Фильтр представляет собой матрицу (её ещё называют матрицей весов или матрицей параметров) размером, например $5 \times 5 \times 3$ и этот фильтр движется по всей области вводимого изображения. После прохода фильтра по всей области (движение с шагом один) в итоге получается новая матрица размера $28 \times 28 \times 1$ (можно получить и другую размерность, это зависит от применимости граничных условий при движении фильтра по изображению). Если использовать несколько фильтров размерностью $5 \times 5 \times 3$ вместо одного. Тогда выходным значением будет $28 \times 28 \times N$, где N количество фильтров.
- В архитектуре СНС обычно применяется слой пулинга (подвыборки) между последовательности свёрточных слоев. Основная задача состоит в последовательном уменьшении пространственных габаритов (разрешение) изображения с намерением уменьшения количества входных параметров для следующего слоя и, соответственно, вычислительных операций в сети, а также контроля обучаемости. Слои пулинга работают независимо от глубины данных на входе и масштабируют весь объем пространственно. Среди функций можно выделить функцию максимума и функцию среднего, но также есть и другие.
- Последним слоем СНС является полносвязная нейронная сеть. Входными данными для полносвязной нейронной сети являются предыдущие слои и определение свойства, которые больше связаны с определенным классом. Скрытый слой может состоять из нескольких скрытых слоев (обычно два), что позволяет сократить общее количество нейронов в полносвязном слое. Слой полносвязной нейронной сети наблюдают за тем, как высокоуровневые карты свойств сильно связаны с каким-либо классом и содержит конкретные веса, поэтому, когда вычисляются взаимодействие весов с предыдущими слоями, то получаются верные вероятности для разных классов. На выходе получаем N -пространственный вектор, где N соответствует числу классов. Каждое значение в этом N -

пространственном векторе представляет собой вероятность конкретного класса [3](11).

- В итоге СНС в каждом слое преобразования трансформирует данное изображение. Преобразование начинается с первоначальных значений исходного изображения и заканчивается определением класса изображения.

Основными достоинствами данной сети являются:

- В сверточном слое данной сети происходят преобразования изображений, который использует ядра. Его наличие значительно уменьшает время и объем вычислительных ресурсов на обучение(12).
- Использование ядра свертки способствует обобщению полученной информации. Восприятие входного изображения по областям позволяет учесть все его свойства, что увеличивает качество распознавания изображений в несколько раз.
- Частичная неизменность к масштабу за счет сжатия изображения.

К сверточной нейронной сети относятся:

- Большое количество параметров (количество слоёв, размер ядра свёртки для каждого из слоёв, количество ядер для каждого из слоёв, шаг сдвига и т.д.). Каждый из параметров существенно влияет на результат работы нейронной сети, поэтому для каждой новой задачи они подбираются эмпирически(12).
- Большое количество обучающего материала.

4.3 Архитектуры нейросетей

Алгоритмы глубоких нейросетей сегодня обрели большую популярность, которая во многом обеспечивается продуманностью архитектур. Некоторые популярные сегодня архитектуры нейронных сетей: Lenet5, YOLO, ResNet, капсульная нейросеть,

4.3.1 YOLO

YOLO — это передовая сеть обнаружения объектов, разработанная Джозефом Редмоном. Главное, что отличает его от других популярных архитектур, — скорость. Модель YOLO действительно быстрая, намного быстрее, чем другие модели. Это означает, что мы можем распознавать объекты в режиме реального времени.

В алгоритме YOLO изображение разделяется на ячейки с использованием сетки. Для каждой ячейки сетки оценивается вероятность присутствия объекта вообще, затем строятся несколько наиболее вероятных положений объекта в виде прямоугольников с центром в данной ячейке, после чего для каждого полученного прямоугольника выполняется оценка вероятностей наличия в нем объектов каждого рассматриваемого класса(13). На следующем шаге выполняется фильтрация прямоугольников по вероятности нахождения в них объектов. И, наконец, дается класс объекта внутри прямоугольника.

4.3.2 Капсульная нейросеть

Современные виды нейросетей имеют ряд ограничений. Так, нейронная сеть может распознать изображение чашки кофе, но не увидит чашку, перевернутую вверх дном. Точнее — не распознает. Данное ограничение можно изменить с помощью нового подхода «капсульная нейросеть», разработанного доктором Джоффри Хинтона с соавторами из Google Brain. Капсульные нейросети не обладают недостатком современных нейронных, которым для обучения требуется огромное число изображений с примерами. Капсулы — небольшие группы виртуальных нейронов — служат для того, чтобы отслеживать различные части предмета, например, нос или ухо кота, и их относительное положение в пространстве. Сеть таких капсул позволит понять, когда на изображении действительно что-то новое, а когда — то же самое, просто под другим углом [2](14).

Вывод

Нейронные сети — это метод глубокого обучения, использующий входные изображения, которые могут быть необработанными. Этот подход обладает

многими преимуществами, такими как высокая скорость обучения, защита от помех и вход, который решает несколько проблем. Однако никакая причина не может быть названа, чтобы определить результат.

При нейросетевом подходе использование только необработанных изображений является большим преимуществом по сравнению с традиционным подходом.

Заключение

В результате выполнения работы были проведены анализ предметной области и постановка проблемы распознавания изображений и некоторых возникающих проблем, а также показаны предобработки входных изображений.

Были исследованы некоторые методы решения задачи распознавания изображений с традиционным подходом и нейросетевым подходом и указаны их преимущества и недостатки. При нейросетевом подходе использование только необработанных изображений является большим преимуществом по сравнению с традиционным подходом.

В настоящее время технология распознавания изображений имеет большой потенциал и широко используется во многих отраслях.

Список используемых источников